Społeczeństwo 5.0 : Cyberrozwiązania dla technologii zorientowanych na człowieka

Analiza stochastyczna macierzy Hankela i Toeplitza do zastosowań medycznych i technicznych

Przedstawiono wyniki nowoczesnych metod analizy stochastycznej widm osobliwych macierzy specjalnych w ramach osobliwej reprezentacji operatorów macierzowych (algorytm SVD). Wykorzystując metodę przybliżonego rozwiązania równania Fredholma pierwszego rodzaju, przeprowadzono rekonstrukcję rozkładów widm osobliwych dla następujących macierzy: (1) macierzy granicznej Hankla o złym warunkowaniu; (2) macierzy Toeplitza; (3) specjalnej macierzy dwudiagonalnej. Macierze Toeplitza i Hankla są dość szeroko prezentowane w stosowanych i teoretycznych gałęziach matematyki (analiza funkcjonalna, teoria procesów losowych), zastosowaniach technicznych, biomedycynie (matematyczna symulacja stanów funkcjonalnych układów i narządów człowieka za pomocą równań różniczkowych, badanie szeregów czasowych w diagnostyce z wykorzystaniem metody osobliwej analizy widmowej). Nowe podejście, mające na celu znalezienie praw rozkładu wartości osobliwych w metodzie zaburzeń, jest niezbędne do kontroli i przewidywania stabilności rozwiązań układów równań algebraicznych, do których sprowadzane są równania różniczkowe, a także do badania możliwości regularyzacji wektorów stochastycznych. Wykazano, że do poprawnego rozwiązania problemów odwrotnych w zastosowaniach medycznych i technicznych, w tym w systemach cyberfizycznych (CPS), można zastosować metodę regularyzacji EC. Wyniki badań mogą być przydatne do tworzenia modeli matematycznych struktur, narządów i układów ciała człowieka, które odzwierciedlają ich stochastyczne i chaotyczne właściwości, w tym dla nowoczesnych medycznych systemów ekspertowych opartych na diagnostyce zmiennych i procesów losowych.

Słowa kluczowe: Macierze Hankela i Toeplitza · Analiza stochastyczna · Identyfikacja · Wartości osobliwe · Pseudoinwersja macierzy · Diagnostyka · Systemy ekspertowe

1 Wstęp

Tu przedstawiono wyniki nowoczesnych metod analizy stochastycznej widm osobliwych macierzy specjalnych, jako część metody osobliwej reprezentacji operatorów macierzowych (algorytm SVD). Metoda przybliżonego rozwiązania równania Fredholma pierwszego rodzaju została wykorzystana do identyfikacji rozkładów widm osobliwych macierzy: granicznej macierzy Hankla o złym warunkowaniu; macierzy Toeplitza; specjalnej macierzy dwudiagonalnej. Macierze Toeplitza i Hankla są dość szeroko reprezentowane w stosowanych i teoretycznych działach matematyki (analiza funkcjonalna, teoria procesów losowych), technicznych zastosowaniach prowadzących do problemów brzegowych dla równań różniczkowych cząstkowych (aero- i hydrodynamika, fizyka cieplna) .Modele matematyczne są również wykorzystywane w postaci problemów brzegowych, układów równań różniczkowych, optymalizacji stochastycznej i problemów odwrotnych w pracach z zakresu biomedycyny S.I. Kabanichina, członka korespondenta Rosyjskiej Akademii Nauk (ICM i MG SB RAS), G.A. Bocharov, Profesor (ICM RAS im. G.I. Marchuka). Macierze Hankla i Toeplitza mogą być wykorzystywane w biomedycynie do matematycznej symulacji stanów funkcjonalnych układów i narządów człowieka, do rozwiązywania problemów odwrotnych z równaniami różniczkowymi w badaniu rozprzestrzeniania się chorób zakaźnych, przy analizie szeregów czasowych w diagnostyce w ramach systemów ekspertowych, w tym CPS, z wykorzystaniem metody osobliwej analizy widmowej. Nowe podejście autorów, mające na celu znalezienie praw rozkładu wartości osobliwych w metodzie perturbacyjnej, jest niezbędne do kontroli i przewidywania stabilności

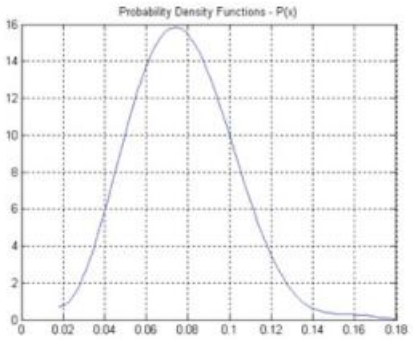
rozwiązań układów równań algebraicznych aproksymujących równania różniczkowe w metodach numerycznych, a także do badania możliwości regularyzacji wektorów stochastycznych. Zidentyfikowane gęstości rozkładu małych wartości osobliwych służą do kontroli i przewidywania stabilności normalnych pseudorozwiązań algebraicznych układów równań w warunkach zaburzeń macierzy i prawej strony układów. W rozdziale procedura została zaimplementowana na dwóch zasadniczo różnych poziomach zaburzeń elementów diagonalnych w macierzach (prawo normalne zniekształcenia elementów). Analiza rozwiązań problemu odwrotnego (źle uwarunkowane SLAE z macierzą kwadratową wyższego rzędu) została przeprowadzona pod kątem odporności na błędy o różnym natężeniu. Oczekuje się, że wyniki przeprowadzonych badań będą mogły zostać wykorzystane do tworzenia modeli matematycznych struktur, narządów i układów ciała człowieka, które wykazują własności stochastyczne i chaotyczne, również dla nowoczesnych systemów medycyny eksperckiej opartych na diagnostyce procesów losowych.

2 Metoda rozwiązania i wyniki eksperymentów obliczeniowych

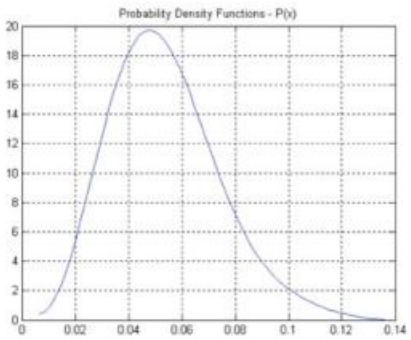
Po raz pierwszy zbadano wpływ pojawienia się polimodalnych (w tym niegaussowskich) praw rozkładu wartości osobliwych dla źle uwarunkowanej macierzy w działaniu algorytmu SVD w warunkach zmienności sztucznych lub eksperymentalnych danych problemu. Niniejsza sekcja przedstawia wyniki rozwoju nowego podejścia na przykładzie macierzy wspomnianych trzech typów. Macierz Hilberta ósmego rzędu jest tu rozpatrywana jako źle uwarunkowana macierz Hankela. Niski rząd tej macierzy testowej pozwala nam na zwartą prezentację wyników identyfikacji bez utraty ogólności. Identyfikacja gęstości rozkładu wartości osobliwych dla tej macierzy przy zaburzeniach głównych elementów przekątnych jest przeprowadzana przy dwóch różnych poziomach szumu gaussowskiego (środowiskiem obliczeniowym jest pakiet MATLAB). W rzeczywistości do testów stochastycznych używany jest dobrze znany algorytm Monte Carlo. Rysunek przedstawia macierz Hankela (Hilberta) (8 × 8) w celach ilustracyjnych.

	1	2	3	4	5	6	7	8
1	1	0.5	0.33333	0.25	0.2	0.16667	0.14286	0.125
2	0.5	0.33333	0.25	0.2	0.16667	0.14286	0.125	0.11111
3	0.33333	0.25	0.2	0.16667	0.14286	0.125	0.11111	0.1
4	0.25	0.2	0.16667	0.14286	0.125	0.11111	0.1	0.090909
5	0.2	0.16667	0.14286	0.125	0.11111	0.1	0.090909	0.083333
6	0.16667	0.14286	0.125	0.11111	0.1	0.090909	0.083333	0.076923
7	0.14286	0.125	0.11111	0.1	0.090909	0.083333	0.076923	0.071429
8	0.125	0.11111	0.1	0.090909	0.083333	0.076923	0.071429	0.066667

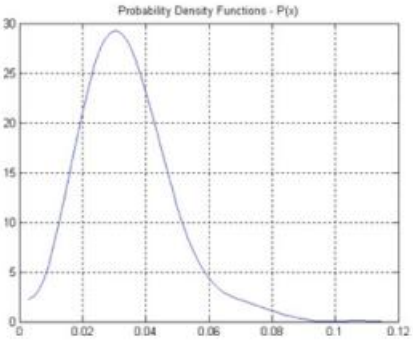
Jest to macierz symetryczna z malejącymi elementami diagonalnymi. Na każdym poziomie wybiera się 1000 cykli zaburzeń. Osiem rozkładów identyfikuje się na podstawie liczby wartości osobliwych. Rysunek 2 przedstawia rozkłady wartości osobliwych dla macierzy H przy zaburzeniu Gaussa na przekątnej według wzoru (1) hii + 0,1*randn.



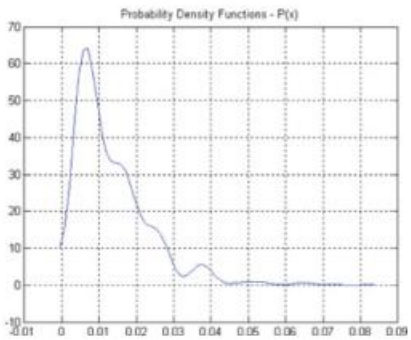
S5 NOPT = 4 t_e = 0.97 s
k = 2.032



S6 NOPT = 5 t_e = 1.00 s
k = 2.002



S7 NOPT = 6 t_e = 1.03 s
k = 1.961

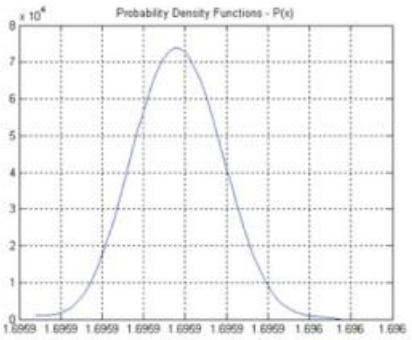


S8 NOPT = 18 t_e = 1.11 s
k = 1.621

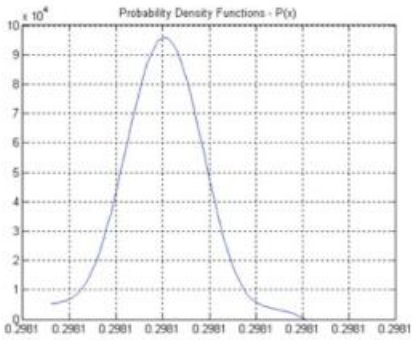
Liczba warunkowa początkowej macierzy Hilberta wynosi cond(H) = 1,53e + 10; rząd wynosi 8. Widmo osobliwe po 1000 zaburzeniach jest określone przez wektor: {1,6556; 0,26975; 0,1226; 0,093268; 0,082738; 0,07313; 0,042492; 0,018888}. Liczba warunkowa macierzy zaburzonej wynosi cond(HV) = 87,6.

	1	2	3	4	5	6	7	8
1	0.88735	0.5	0.33333	0.25	0.2	0.16667	0.14286	0.125
2	0.5	0.4239	0.25	0.2	0.16667	0.14286	0.125	0.11111
3	0.33333	0.25	0.16766	0.16667	0.14286	0.125	0.11111	0.1
4	0.25	0.2	0.16667	0.22597	0.125	0.11111	0.1	0.090909
5	0.2	0.16667	0.14286	0.125	0.0099511	0.1	0.090909	0.083333
6	0.16667	0.14286	0.125	0.11111	0.1	0.018934	0.083333	0.076923
7	0.14286	0.125	0.11111	0.1	0.090909	0.083333	0.0024089	0.071429
8	0.125	0.11111	0.1	0.090909	0.083333	0.076923	0.071429	0.2045

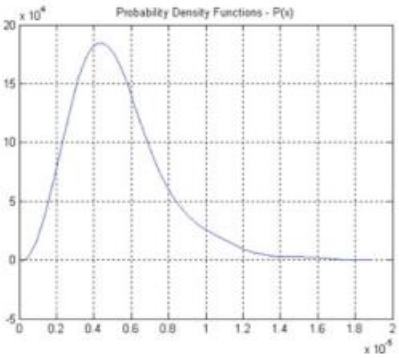
Analiza zidentyfikowanych gęstości pokazuje, że transformacja prawa rozkładu normalnego charakterystyczna dla pierwszych wartości osobliwych staje się widoczna począwszy od szóstej wartości osobliwej. Szósta i siódma wartość mają rozkłady typu Pearsona, a najmniejsza (ósma) charakteryzuje się nieznacznie wyrażoną gęstością polimodalną. Ogólnie gęstość ta przypomina zniekształcony rozkład gamma. Wniosek wstępny jest następujący: po pierwsze, efekt pojawienia się polimodalnych, niegaussowskich praw rozkładu małych wartości osobliwych jest potwierdzony dla źle uwarunkowanej macierzy Hankela; po drugie, zaburzenie diagonalne o określonej intensywności może być użyte do regularyzacji stochastycznej przy rozwiązywaniu SLAE. Przy identyfikacji przy niskim poziomie zaburzenia: (2) h_{ii} + 0,00001*randn bardziej wyraźny efekt polimodalny znajduje się w rozkładzie minimalnej wartości osobliwej .



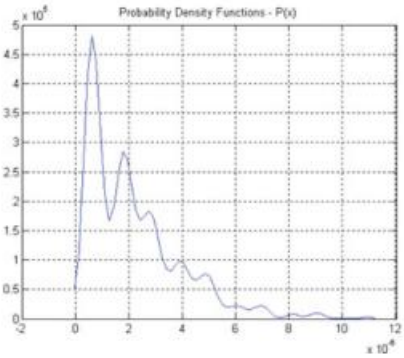
S1 NOPT = 5 elapsed_time = 0.89 s
k = 0.993



S2 NOPT = 5 elapsed_time = 0.89 s
k = 1.226



S7 NOPT = 7 elapsed_time = 0.98 s
k = 1.893



S8 NOPT = 22 elapsed_time = 1.41 s
k = 1.545

Widmo osobliwe po 1000. iteracji (zaburzeniu) jest określone przez wektor: {1,6959; 0,29812; 0,026217; 0,0014732; 6,0361e-005; 1,2659e-005; 7,1063e-006; 2.3334e-006}. Liczba warunkowa macierzy zaburzonej wynosi cond(HV) = 7,3e + 05. Zatem mało intensywne zaburzenia źle uwarunkowanej macierzy Hankela nie mogą zapewnić regularyzacji SLAE w zakresie parametrów Lagrange'a typowych dla klasycznych problemów regularyzacji A.N. Tichonowa; jest to określone przez osobliwą strukturę widmową macierzy Hilberta .

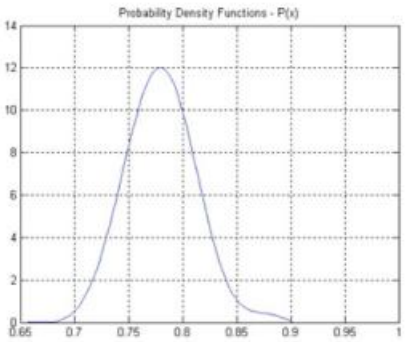
3 Identyfikacja gęstości rozkładu wartości osobliwych dla macierzy Toeplitza

Kwadratowa macierz ósmego rzędu jest uważana za macierz Toeplitza. Macierz Toeplitza (8 × 8) jest pokazana na rysunku 5.

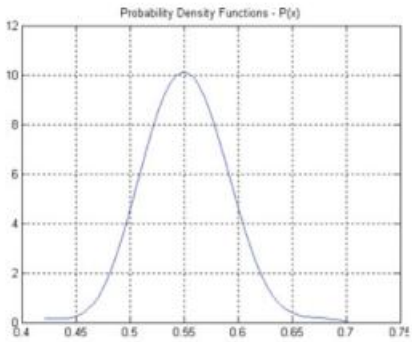
	1	2	3	4	5	6	7	8
1	1	2	3	4	5	6	7	8
2	1.5	1	2	3	4	5	6	7
3	2	1.5	1	2	3	4	5	6
4	2.5	2	1.5	1	2	3	4	5
5	3	2.5	2	1.5	1	2	3	4
6	3.5	3	2.5	2	1.5	1	2	3
7	4	3.5	3	2.5	2	1.5	1	2
8	4.5	4	3.5	3	2.5	2	1.5	1

Jest to macierz asymetryczna o stałych elementach diagonalnych (na głównej i pobocznej przekątnej). Podobnie jak poprzednio, identyfikacja gęstości rozkładu wartości osobliwych dla tej macierzy jest przeprowadzana metodą zaburzeń głównego elementu diagonalnego przy dwóch różnych poziomach szumu gaussowskiego. Liczba zaburzeń na każdym poziomie wynosi również

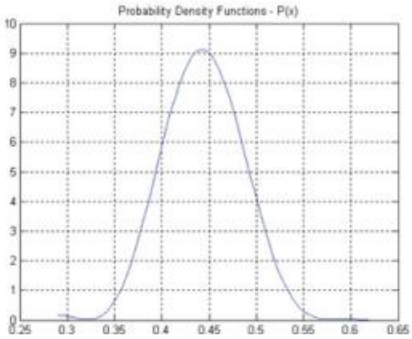
1000. Kody poleceń do ustawiania macierzy (MATLAB) są następujące: c = (1:0,5:4,5); r = (1:8); T = toeplitz(c, r). Liczba warunkowa początkowej macierzy Toeplitza wynosi cond (T) = 65,4; rząd wynosi 8. Widmo osobliwe to wektor: {25,508; 8,5779; 3,0117; 1,2053; 0,76973; 0,54161; 0,44128; 0,38979}, co pokazuje, że macierz jest dość dobrze uwarunkowana. Początkowo identyfikację rozkładu wartości osobliwych dla macierzy T przeprowadzono przy zaburzeniu Gaussa na przekątnej wzorem tii + 0,1*randn. Analiza zidentyfikowanych gęstości sugeruje, że widmo osobliwe macierzy Toeplitza praktycznie nie wykazuje transformacji prawa rozkładu normalnego.



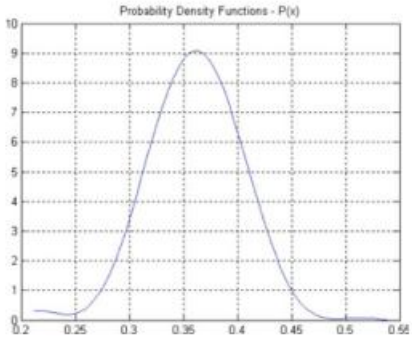
S5 NOPT = 6 elapsed_time = 0.89 s
k = 2.051



S6 NOPT = 5 elapsed_time = 0.92 s
k = 2.057

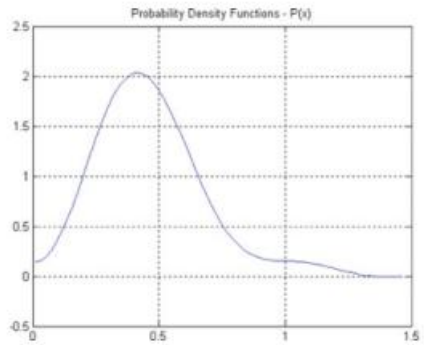


S7 NOPT = 5 elapsed_time = 0.92 s
k = 2.052

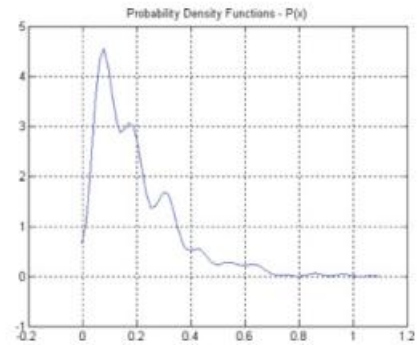


S8 NOPT = 5 elapsed_time = 0.87 s
k = 2.044

Jeśli poziom zaburzenia zostanie zwiększony o rząd wielkości: tii + randn (randn to program generatora liczb losowych Gaussa), efekt polimodalny ponownie występuje w rozkładzie minimalnej wartości osobliwej .



S7 NOPT = 5 elapsed_time = 1.03 s
k = 1.939



S8 NOPT =20 elapsed_time = 1.08 s
k = 1.585

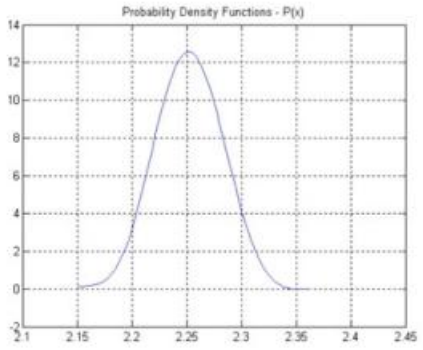
Zatem tendencja w rozkładzie małych wartości osobliwych dla obu macierzy w metodzie zaburzeń jest taka sama i ma charakter polimodalny. Liczba manifestowanych modów zależy od poziomu szumu.

4 Analiza źle uwarunkowanych rozwiązań SLAE pod kątem odporności na błędy o zmiennej intensywności

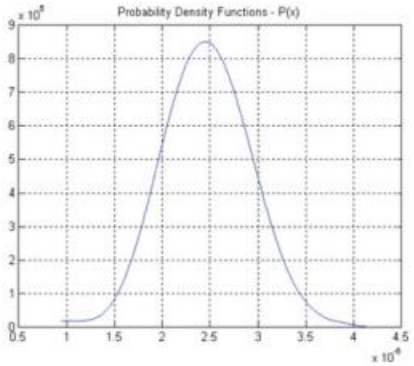
Zajmijmy się najpierw źle uwarunkowaną macierzą ogólną. Rozważmy jako przykład macierz G (30 × 30) będącą sumą dwóch macierzy: dwudiagonalnej, skrajnie źle uwarunkowanej macierzy D2 oraz stochastycznej macierzy B z elementami rozłożonymi zgodnie z prawem normalnym. Macierz D2 pochodzi z przykładu testowego z monografii .

	1	2	3	4	5	6	7	8	9	10	11	12
1	1.4	3.6667	0	0	0	0	0	0	0	0	0	0
2	0	1.4	3.6667	0	0	0	0	0	0	0	0	0
3	0	0	1.4	3.6667	0	0	0	0	0	0	0	0
4	0	0	0	1.4	3.6667	0	0	0	0	0	0	0
5	0	0	0	0	1.4	3.6667	0	0	0	0	0	0
6	0	0	0	0	0	1.4	3.6667	0	0	0	0	0
7	0	0	0	0	0	0	1.4	3.6667	0	0	0	0
8	0	0	0	0	0	0	0	1.4	3.6667	0	0	0
9	0	0	0	0	0	0	0	0	1.4	3.6667	0	0
10	0	0	0	0	0	0	0	0	0	1.4	3.6667	0
11	0	0	0	0	0	0	0	0	0	0	1.4	3.6667
12	0	0	0	0	0	0	0	0	0	0	0	1.4

Liczba uwarunkowań tej macierzy wynosi 5,66e + 12. Druga macierz jest przypisana operatorem B = 0,00001*randn(N); N = 30. Liczba uwarunkowań macierzy G (całkowicie wypełnionej elementami) wynosi 2,03e + 6; rząd = 30. Widmo osobliwe zawiera trzydzieści składowych: {5,0449; 5,0177; 4,9799; 4,9315; ..., 2,3176; 2,2796; 2,4839e-006}. Macierz G ma płynnie malejące widmo osobliwe (wszystkie składowe z wyjątkiem ostatniej są większe od jedności), ale jest wciąż stosunkowo słabo uwarunkowana ze względu na małą wartość 30. wartości osobliwej. Rysunek 9 przedstawia rozkłady małych wartości osobliwych macierzy G przy zaburzeniu Gaussa głównej przekątnej według wzoru: gii + 0,1*randn.

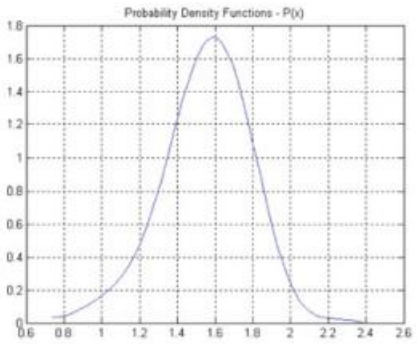


S29 NOPT = 5 elapsed_time = 0.91 s
k = 2.058

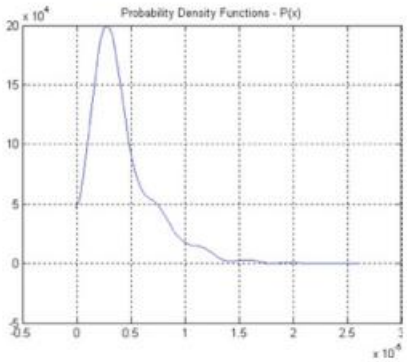


S30 NOPT = 5 elapsed_time = 0.91 s
k = 2.061

Podobnie jak w przypadku macierzy Toeplitza, całkowite widmo osobliwe macierzy G charakteryzuje się brakiem transformacji prawa rozkładu normalnego. Nawet najmniejsza wartość osobliwa ma gęstość rozkładu normalnego. Jeśli poziom zaburzenia zostanie zwiększony o rząd wielkości, ujawni się efekt emergencji polimodalności ,

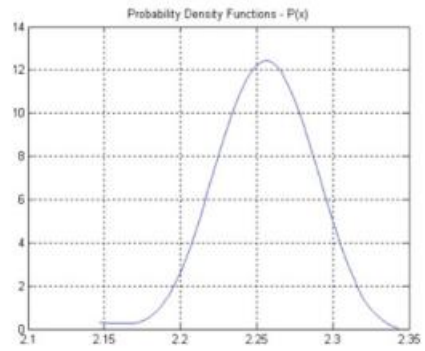


S29 NOPT = 6 elapsed_time = 0.89 s
k = 2.039

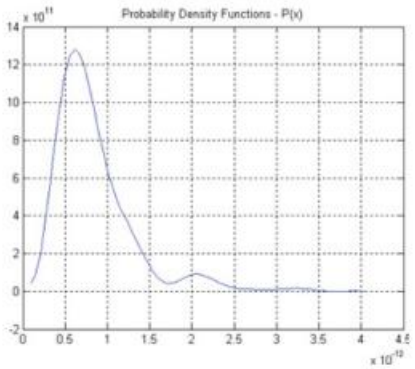


S30 NOPT = 13 elapsed_time = 1.25 s
k = 1.691

który pojawia się przy wystarczająco wysokich natężeniach zaburzeń. Oba poziomy zaburzenia diagonalnego macierzy G nie dają podstaw do stochastycznej regularyzacji diagonalnej podczas rozwiązywania równania SLAE, ponieważ liczba warunkowa macierzy zaburzonej nie maleje (w tysięcznej iteracji cond(G) = 8,6e + 05). Zauważymy przy tym następującą cechę: macierz D2 uzyskuje w cyklu perturbacyjnym słabą bimodalność 30. wartości osobliwej (rys. 11).



S29 NOPT = 4 elapsed_time = 0.84 s
k = 2.051



S30 NOPT = 12 elapsed_time = 1.12 s
k = 1.594

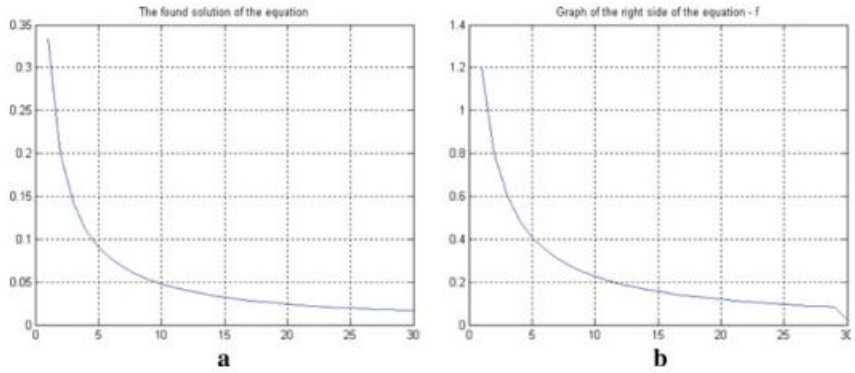
Wprowadzenie losowej macierzy B (sumowanie macierzy B i skrajnie źle uwarunkowanej macierzy D2) pomaga zmniejszyć liczbę warunkową (w klasycznym przypadku regularyzacji poprzez uspójnienie „dyspersji perturbacyjnej" z błędami operatora i prawą stroną równania SLAE). Zatem stabilność (lub niestabilność) rozwiązań i możliwość regularyzacji stochastycznej będą determinowane nie tylko przez elementy diagonalne, ale także przez peryferie źle uwarunkowanej macierzy. Należy zauważyć, że zaburzenie pierwszej niezerowej przekątnej pomocniczej D2 nie powoduje efektu polimodalności małych liczb osobliwych. Analizujemy dalej rozwiązania problemu odwrotnego ze specjalną macierzą o stosunkowo wysokim rzędzie pod kątem odporności na błędy o zmiennej intensywności. Początkowa macierz dwuprzekątna DM = D2 jest podana powyżej i ma rząd (30 × 30). Składowe wektora po prawej stronie są określone za pomocą wzorów z przypadku testowego :

$$f_k = \frac{152k + 118}{15(2k+1)(2k+3)} \quad (1 \le k \le M-1),\ f_M = \frac{7}{5(2M+1)} \quad M = 30.$$

Dokładnym rozwiązaniem układu dla dowolnego M jest wektor ze składowymi

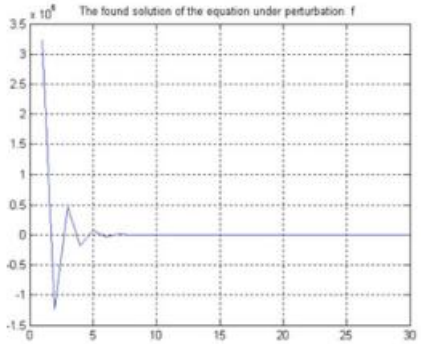
$$x_k = \frac{1}{2k+1} \quad (1 \le k \le M)$$

Numeryczne rozwiązanie układu równań liniowych metodą pseudoinwersji Moore'a–Penrose'a (algorytm SVD pakietu MATLAB) oraz prawa strona układu są pokazane na rysunku 12.



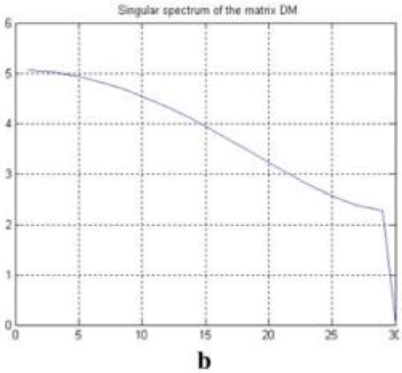
a b

Analiza rozwiązań SLAE pokazuje, że przy minimalnym zaburzeniu składowych wektora fk najsilniejsze zniekształcenie wektora x występuje w kolejnych rozwiązaniach .



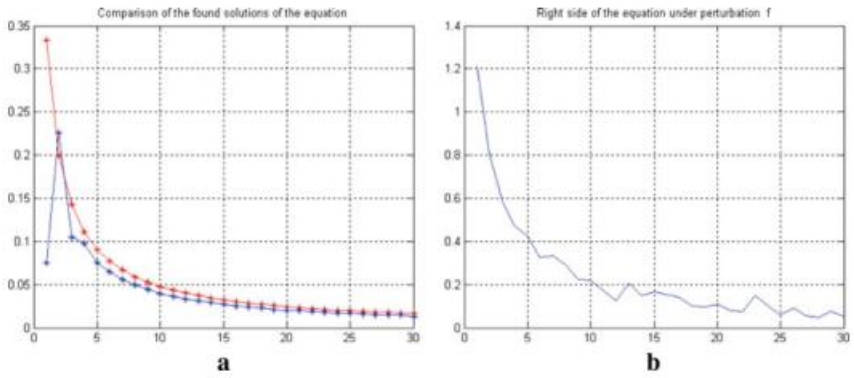
a

Szpice osiągają tam poziom 1,e+06 ÷ 1,e+08 jednostek. Widmo osobliwe macierzy DM jest pokazane na rysunku.

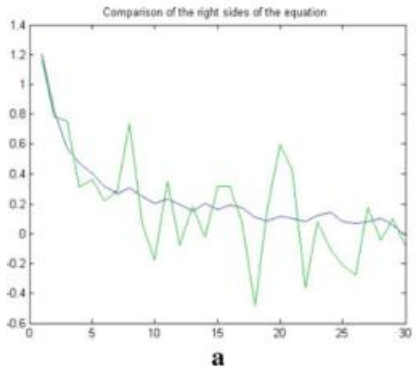


b

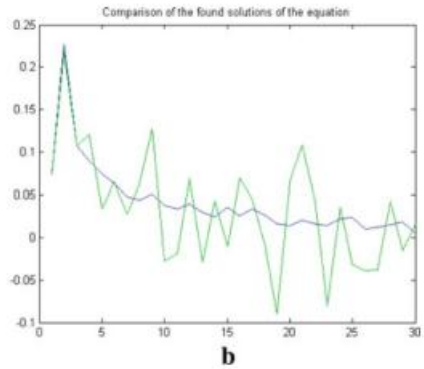
Widmo jest płynnie malejące; wszystkie wartości osobliwe z wyjątkiem jednej (wartość minimalna, równa 8,94e-013) są większe od jedności. Liczba warunkowa układu wynosi 5,66e + 012. Zatem istnieje błędne postawienie problemu pseudoinwersji w układach ze źle uwarunkowanymi macierzami. Aby znaleźć odporne na błędy dane początkowe rozwiązań, można wykorzystać znane algorytmy. demonstrujemy efektywność metody regularyzacji EC badanej i modyfikowanej. Rysunek a przedstawia porównanie rozwiązań SLAE z DM obiema metodami bez zakłócania prawej strony układu.



Rozwiązania z regularyzacją EC są stabilne (w punkcie początkowym występuje tylko jeden skok) i praktycznie nie zmieniają się od poziomu 3.e-10*randn zaburzenia fk do poziomu równego 0,03*randn (wykres przedstawia niebieska linia). Pewien błąd w punkcie początkowym rozwiązania w metodzie regularyzacji EC jest swego rodzaju „płatnością" za niewykorzystanie a priori informacji o rozwiązaniu i może być w inny sposób zredukowany. W metodzie regularyzacji EC, przy dalszym wzroście zaburzeń do wartości 0,3*randn ,



zachodzą pewne spójne zmiany fk i rozwiązań x (rys. 15b).



Zatem proponowana metoda poprawnie reaguje na zmiany fk. Przy wspomnianych poziomach szumu metoda pseudoinwersji Moore'a–Penrose'a jest całkowicie niestabilna i wymaga procedury regularyzacji

5 Wyniki i wnioski

Metodę przybliżonego rozwiązania równania Fredholma pierwszego rodzaju zastosowano do identyfikacji rozkładów widm osobliwych następujących macierzy: macierzy granicznej Hankela

o złym warunkowaniu, macierzy Toeplitza i specjalnej macierzy dwudiagonalnej. Opracowane podejścia są niezbędne do kontroli i przewidywania stabilności rozwiązań układów równań algebraicznych i różniczkowych, a także do badania możliwości regularyzacji wektorów stochastycznych. Wykazano, że metoda pseudoinwersji macierzy oparta na rozkładzie na wartości osobliwe bez regularyzacji prowadzi do dużych zniekształceń pseudorozwiązania normalnego nie tylko w przypadku szybkiego zaniku wartości osobliwych do zera, ale także w przypadku płynnie malejącego widma z dostępną co najmniej jedną małą wartością osobliwą. Metoda regularyzacji EC może być wykorzystana do poprawnego i stabilnego rozwiązania wspomnianych problemów odwrotnych.

Analiza danych klinicznych u kobiet w ciąży z nadciśnieniem tętniczym w oparciu o nowe podejście stochastyczne

Wielu światowych badaczy skupiło się w ostatnich latach na badaniu ryzyka związanego z rozwojem patologicznej ciąży u kobiet z nadciśnieniem tętniczym. Patogeneza powikłań ciążowych w przypadku zaburzeń nadciśnieniowych jest złożona i wiąże się głównie z dysfunkcją śródbłonka oraz zmniejszoną elastycznością ściany naczyń. Niemniej jednak, wysoka wartość prognostyczna oceny elastyczności naczyń została obecnie udowodniona w kontekście rozwoju incydentów sercowo-naczyniowych u pacjentek z nadciśnieniem tętniczym w populacji ogólnej. Aby wcześnie przewidzieć ryzyko powikłań ciążowych, obiecujące jest badanie stanu elastyczności ścian naczyń w oparciu o analizę stochastyczną, co pozwoli na optymalizację postępowania w ciąży i poprawę wyników okołoporodowych u kobiet w ciąży z nadciśnieniem tętniczym. Podstawą podejścia do probabilistycznej (stochastycznej) analizy wyników klinicznych u kobiet w ciąży będą metody prawidłowej identyfikacji praw rozkładu wartości losowych w biomedycynie. Opierają się one na rozwiązaniach problemów odwrotnych źle postawionych, opracowanych przez szkołę akademika A.N. Tichonowa. Aby przeanalizować dane z 24-godzinnego monitorowania ciśnienia tętniczego (BP) w trakcie nadzoru nad kobietami w ciąży, zidentyfikowano prawa rozkładu (gęstości rozkładu prawdopodobieństwa), które między innymi wykazują pewne cechy polimodalne odzwierciedlające aktualny stan pacjentek. Na podstawie analizy stochastycznej zidentyfikowano dwa wskaźniki markerowe (czas przejścia fali odbitej RWTT oraz szacowaną prędkość fali tętna w aorcie PWVao).

Słowa kluczowe: Ciąża · Nadciśnienie tętnicze · Dysfunkcja śródbłonka · Elastyczność ścian naczyń · Prognozowanie ryzyka · Analiza stochastyczna · Błędnie postawione problemy · Polimodalne gęstości rozkładu

1 Wstęp

Nadciśnienie tętnicze (NT) pozostaje obecnie jednym z najpilniejszych problemów medycyny klinicznej ze względu na wysokie ryzyko rozwoju powikłań sercowo-naczyniowych (udar mózgu, zawał mięśnia sercowego, niewydolność serca i nerek) oraz śmiertelność. Ważnym trendem w kardiologii jest poszukiwanie wczesnych markerów uszkodzenia narządów docelowych, takich jak naczynia krwionośne, serce, nerki i mózg u pacjentów cierpiących na nadciśnienie tętnicze Nadciśnienie ciążowe stanowi interdyscyplinarny problem kardiologii i położnictwa, ponieważ prowadzi do niekorzystnego przebiegu ciąży i patologii okołoporodowej. NT to stan, w którym u kobiet w ciąży odnotowuje się ciśnienie skurczowe (SBP) ≥140 mm Hg i/lub ciśnienie rozkurczowe (DBP) ≥90 mm Hg. Nadciśnienie tętnicze wikła 10–20% ciąż na świecie, a w Rosji 5–30%. Należy zauważyć, że w ciągu ostatnich dwóch dekad częstość występowania nadciśnienia ciążowego wzrosła o 40–50%. Wzrost częstości występowania nadciśnienia tętniczego występuje głównie z

powodu jego przewlekłych postaci (choroby nadciśnieniowej) związanych z zachorowalnością na otyłość u kobiet, cukrzycę, starszą średnią wieku pierwiastków, a także powszechnym stosowaniem technik wspomaganego rozrodu . Choroba nadciśnieniowa (HD) zwiększa ryzyko wystąpienia poważnych zespołów położniczych (stan przedrzucawkowy, niewydolność łożyska, zahamowanie wzrostu płodu, poród przedwczesny, odklejenie się prawidłowo położonego łożyska), porodu operacyjnego przez cesarskie cięcie, patologii okołoporodowej, konieczności intensywnej opieki nad noworodkami, śmiertelności matek i niemowląt . Patogeneza powikłań ciąży u kobiet cierpiących na nadciśnienie tętnicze jest złożona, wciąż niedostatecznie zbadana i nie do końca poznana. Jest ona w dużej mierze związana z uogólnioną dysregulacją napięcia naczyniowego, dysfunkcją śródbłonka i zaburzeniami łożyska. Zmiany strukturalne w ścianach naczyń krwionośnych wiążą się ze zmniejszeniem ich elastyczności (usztywnienie, zwiększona sztywność), zwężeniem światła tętnicy oraz upośledzeniem przepływu krwi i tlenu do łożyska i płodu. W literaturze można znaleźć pewne dowody na to, że dysfunkcja śródbłonka rozwija się jeszcze przed wystąpieniem jakichkolwiek objawów klinicznych ciąży i powikłań płodowych . W związku z tym znacznie wzrosła uwaga krajowych i zagranicznych naukowców na ocenę sztywności ścian naczyń, aby móc wcześnie przewidzieć ryzyko rozwoju patologii ciąży w przypadku nadciśnienia tętniczego. Na podstawie badań światowych i rosyjskich przeprowadzonych w ciągu ostatniej dekady wykazano, że jednym z dowodów uszkodzenia narządów docelowych jest nadciśnieniowa przebudowa naczyń spowodowana podwyższonym ciśnieniem krwi. Do tej pory udowodniono, że wyższa sztywność ścian tętnic wiąże się ze zwiększonym ryzykiem zgonu z przyczyn sercowo-naczyniowych u pacjentów z HD w populacji ogólnej .Sztywność ścian tętnic wpływa na zdolność dużych naczyń do wygładzania pulsacji wywołanej rzutem serca i przekształcania ciśnienia uderzeniowego w aorcie wstępującej w stabilny przepływ krwi w naczyniach obwodowych. Zwiększona sztywność ścian tętnic nasila w przypadku HD funkcję tłumienia tętnic, przyspieszenie fal odbitych, wzrost ciśnienia skurczowego i tętna, spadek ciśnienia rozkurczowego oraz ryzyko niedokrwienia mięśnia sercowego. Według badania HARVEST (Hypertension and Ambulatory Recording VEnetia Study) sztywność ścian naczyń może być wyższa nawet we wczesnym stadium choroby; pozwala to na lepszą stratyfikację ryzyka incydentów sercowo-naczyniowych u pacjentów HD w populacji ogólnej.

2 Aktualne Technologie Diagnostyki Funkcjonalnej Układu Sercowo-Naczyniowego

Ciśnienie tętnicze (BP) jest dynamicznym parametrem fizjologicznym. 24-godzinne wahania ciśnienia tętniczego stanowią sumę reakcji na zewnętrzne bodźce presyjne, wahania spontaniczne i regulacyjne. Wysoce informatywną metodą ambulatoryjnej identyfikacji ciśnienia tętniczego jest całodobowy monitoring ciśnienia tętniczego (BPM) z wykorzystaniem 24-godzinnego, programowego systemu monitorowania BPLab (Petr Telegin LLC, Niżny Nowogród). BPM to metoda tonometrii aplanacyjnej. System BPLab jest przeznaczony do automatycznego, nieinwazyjnego pomiaru ciśnienia tętniczego, tętna, a także do oceny wszelkich uzupełniających parametrów charakteryzujących aktywność fizyczną i hemodynamikę.

Zasada techniki BPM jest następująca:

1. Ciśnienie tętnicze pacjenta, żyjącego w naturalnych warunkach i mogącego się swobodnie poruszać, jest okresowo mierzone za pomocą przenośnego monitora noszonego na ciele. Pomiary te są wykonywane zarówno w dzień, jak i w nocy. Czas trwania monitoringu powinien wynosić co najmniej 24 godziny.

2. Po zakończeniu monitorowania analizuje się tabelę uzyskanych danych pomiaru ciśnienia krwi w celu oceny wskaźników 24-godzinnego profilu ciśnienia krwi.

Takie monitory noszone na ciele zapewniają pomiar ciśnienia krwi metodami pośrednimi, opartymi na analizie zależności parametrów sygnału biologicznego od wartości nadciśnienia w mankiecie założonym na ramię pacjenta. Stosowane metody pomiaru ciśnienia krwi to oscylometryczna i osłuchowa. Monitory te rejestrują sygnały podstawowe dla każdego pomiaru, pozycję ciała i aktywność pacjenta w okresach odpowiadających wykonanym pomiarom ciśnienia krwi. Oprogramowanie BPLab jest przeznaczone do monitorowania ciśnienia krwi i fali tętna; jest dostępne w wersji Standard i Advanced. Do diagnozy HD wykorzystywana jest wersja Standard oprogramowania BPLab, która umożliwia analizę 24-godzinnego profilu ciśnienia krwi obwodowej wraz z obliczeniem wskaźników statystycznych (wartości średnie, zmienność, wskaźniki obciążenia ciśnieniowego SBP, DBP, średnie ciśnienie krwi, tętno w ciągu doby, w dzień, w nocy oraz w określonym przedziale czasowym). Do stratyfikacji ryzyka uszkodzenia narządów docelowych u pacjentów z nadciśnieniem tętniczym zastosowano innowacyjną, zaawansowaną wersję oprogramowania BPLab z technologią Vasotens 24, która umożliwia ocenę codziennych trendów parametrów hemodynamicznych (sztywności tętnic i wyników pomiaru ciśnienia centralnego w aorcie). „Złotym standardem" pomiaru sztywności aorty jest ocena prędkości fali tętna w tętnicy szyjnej i udowej (PWV) za pomocą techniki podwójnego przetwornika fali tętna. Współczynnik ten jest bezpośrednio skorelowany ze sztywnością ściany naczynia i odwrotnie proporcjonalny do jej elastyczności. Niektóre niedawne badania pokazują, że usztywnienie aorty oceniane na podstawie stopnia wzrostu PWV jest niezależnym predyktorem ryzyka powikłań sercowo-naczyniowych u pacjentów poddawanych hemodializie. Jednakże, jak wykazano w tradycyjny pomiar PWV (dwupunktowy) ma pewne wady. W szczególności procedura badawcza jest czasochłonna i wymaga odpowiedniego doświadczenia zawodowego. Ponadto elastyczność tętnic, a tym samym mierzona PWV, zależą zasadniczo od wartości ciśnienia tętniczego krwi w momencie pomiaru. Zatem chwilowy pomiar PWV jest ważny tylko dla ciśnienia, przy którym dokonano pomiaru, co ogranicza jego zastosowanie przy porównywaniu pacjentów i ocenie dynamiki zmian specyficznych dla danego pacjenta. Parametr ten nie jest zidentyfikowany w oprogramowaniu BPLab. Aby uprościć procedurę pomiaru, niektórzy autorzy zaproponowali alternatywne wskaźniki sztywności, które dość dobrze korelują z tradycyjnie mierzoną PWV, ale są łatwiejsze w użyciu. Do wskaźników tych należą :

– czas przejścia fali odbitej (RWTT);

– prędkość fali tętna w aorcie, określona przez czas propagacji fali odbitej, PWVao. Ponadto istnieją wskaźniki, które mają niezależne znaczenie, ale również zasadniczo zależą od sztywności głównych tętnic. Do wskaźników tych należą:

– wskaźnik wzmocnienia (AIx), wskaźnik charakteryzujący przede wszystkim intensywność fali odbitej i jej wpływ na wzrost ciśnienia tętniczego;

– maksymalne tempo narastania ciśnienia krwi (dP/dt)max, wskaźnik charakteryzujący przede wszystkim kurczliwość mięśnia sercowego.

Do oceny parametrów hemodynamicznych mierzy się również centralne ciśnienie aortalne (CAP), którego poziom jest regulowany przez właściwości sprężyste dużych naczyń, elementy strukturalne i funkcjonalne tętnic średniej wielkości oraz naczyń oporowych i jest najbardziej informatywnym wskaźnikiem w porównaniu z obwodowym ciśnieniem krwi, który identyfikuje stan ściany tętnicy. Europejscy eksperci uważają sztywność ścian naczyń i CBP za niezależne markery ryzyka sercowo-naczyniowego u pacjentów z HD w populacji ogólnej. Nadciśnienie tętnicze u kobiet w ciąży jest wskazaniem do wykonania BPM w celu rozpoznania nadciśnienia tętniczego, jego nasilenia oraz diagnostyki różnicowej zaburzeń nadciśnieniowych. Jednak

doświadczenie w stosowaniu BPM do oceny sztywności ścian tętnic i ryzyka powikłań ciążowych wykazało, że wyniki badań kobiet w ciąży cierpiących na nadciśnienie tętnicze są mniej informatywne i sprzeczne. Większość badań elastyczności naczyń przeprowadzono u kobiet w ciąży w trakcie rozwoju stanu przedrzucawkowego (trymestr II–III). Niektóre badania wykazały zwiększoną PWV i AIx w porównaniu do kobiet w ciąży z prawidłowym ciśnieniem tętniczym [10], inne odnotowały zwiększoną PWV przy niezmienionym indeksie powiększenia , a w niektórych badaniach parametry te nie różniły się od fizjologicznej ciąży . W literaturze można znaleźć również informacje, że u kobiet z klinicznymi dowodami stanu przedrzucawkowego zidentyfikowano jedynie zwiększony indeks powiększenia, podczas gdy inne parametry sztywności nie były badane Ponadto, w przypadku stanu przedrzucawkowego rozwiniętego w ciąży donoszonej wykazano wzrost PWV i AIx, przy czym wyższe poziomy tych parametrów utrzymywały się również w okresie poporodowym, co może świadczyć o trwałych zmianach w ścianie naczyń. Ci sami autorzy zbadali kobiety z historią stanu przedrzucawkowego. Uzyskano wyniki wysokiego wskaźnika wzmocnienia, prędkości fali tętna, czasu propagacji fali odbitej i wskaźnika sztywności tętnic w drugim trymestrze ciąży, co może świadczyć o sztywniejszej budowie naczyń w porównaniu z kobietami w ciąży bez stanu przedrzucawkowego w wywiadzie . W badaniach hemodynamiki tętna przeprowadzonych w II–III trymestrze ciąży zauważono, że u kobiet z nadciśnieniem tętniczym centralne skurczowe ciśnienie krwi, centralne rozkurczowe ciśnienie krwi, centralne średnie ciśnienie hemodynamiczne i centralne ciśnienie tętna są wyższe niż w fizjologicznej ciąży. Kobiety w ciąży z nadciśnieniem tętniczym powikłanym stanem przedrzucawkowym miały najwyższe poziomy centralnego skurczowego ciśnienia krwi i centralnego rozkurczowego ciśnienia krwi . Istnieją również informacje o pacjentkach z nadciśnieniem tętniczym, u których centralne skurczowe ciśnienie krwi i centralne tętno były niższe przez całą ciążę niż wartości obwodowego skurczowego ciśnienia krwi i obwodowego tętna, co może wskazywać na zmniejszoną elastyczność głównych naczyń w przypadku chorób sercowo-naczyniowych . Badania przeprowadzone w drugim trymestrze ciąży wykazują, że u kobiet z HD wskaźnik wzmocnienia, wskaźnik sztywności tętnic i maksymalna szybkość wzrostu ciśnienia krwi są wyższe w przypadku klinicznych objawów stanu przedrzucawkowego, podczas gdy czas propagacji fali odbitej i wzmocnienia ciśnienia tętna nie różnią się od wartości u kobiet normotensyjnych . W pojedynczych badaniach, wartość prognostyczna CBP i wpływ różnicy między centralnym i obwodowym ciśnieniem tętniczym na przebieg i wynik ciąży były badane podczas badania kobiet w drugiej połowie ciąży. Tak więc, podczas wykonywania BPM w drugim trymestrze ciąży predyktorami rozwoju zahamowania wzrostu płodu u kobiet w ciąży z nadciśnieniem są zmniejszona elastyczność ściany naczyniowej i zmniejszona różnica między CBP a obwodowym ciśnieniem krwi. Różnica między centralnym i obwodowym ciśnieniem skurczowym wynosi 8,3 ± 0,9 mm Hg.; różnica między centralnym i obwodowym ciśnieniem tętna równa się 9,8 ± 1,0 mm Hg; poziom centralnego ciśnienia skurczowego ≥118 mm Hg; obwodowego ciśnienia rozkurczowego ≥82 mm Hg; obwodowego średniego ciśnienia hemodynamicznego ≥99 mm Hg mają znaczenie prognostyczne dla ryzyka zahamowania wzrostu płodu. Bezpośredni związek powikłań okołoporodowych (przedwczesny poród, przedwczesne odpłynięcie płynu owodniowego, długi okres bezwodny, anomalie porodu) ze wzrostem centralnego ciśnienia skurczowego w trymestrach II–III ≥118 mm Hg został zidentyfikowany w badaniach tych samych autorów. Zatem do tej pory w większości badań analiza fali tętna była przeprowadzana w drugiej połowie ciąży, badano właściwości sprężyste naczyń krwionośnych w trakcie klinicznych objawów stanu przedrzucawkowego, a znaczenie prognostyczne pojedynczych parametrów sztywności naczyń oceniano w drugiej połowie ciąży. Istotne jest zbadanie prognostycznego znaczenia parametrów sztywności ścian tętnic w pierwszym trymestrze ciąży w celu stratyfikacji ryzyka powikłań ciążowych u kobiet w ciąży z HD. Wczesne

przewidywanie głównych zespołów położniczych powinno zoptymalizować prowadzenie ciąży, poprawić wyniki okołoporodowe oraz zmniejszyć śmiertelność matek i niemowląt u kobiet z HD.

3. Nowe podejście do przetwarzania wyników badań klinicznych u kobiet w ciąży

Celem badania jest opracowanie wskaźników markerowych sztywności ścian naczyń w celu przewidywania głównych zespołów położniczych u kobiet w ciąży z nadciśnieniem tętniczym w oparciu o analizę stochastyczną.

Materiały i metody

Przeprowadzono badanie 19 kobiet w ciąży z nadciśnieniem tętniczym I stopnia, nadciśnieniem tętniczym I stopnia, ryzykiem 1–2. Wszystkie te kobiety zostały przebadane zgodnie z rozporządzeniem Ministerstwa Zdrowia Federacji Rosyjskiej N 1130n z dnia 20 października 2020 r. w sprawie zatwierdzenia procedury udzielania świadczeń medycznych w zakresie położnictwa i ginekologii. U wszystkich kobiet w ciąży w wieku ciążowym 10–13, 18–21 i 30–34 tygodni wykonano 24-godzinny pomiar ciśnienia tętniczego i ocenę sztywności tętnic przy użyciu systemu monitorowania ciśnienia tętniczego BPLab 24-h z technologią Vasotens24. Podstawą podejścia do probabilistycznej (stochastycznej) analizy danych klinicznych kobiet w ciąży są metody poprawnej identyfikacji praw rozkładu zmiennych losowych w biomedycynie, opublikowane po raz pierwszy w czasopiśmie European Journal w 2015 roku .. Opierają się one na rozwiązaniach problemów odwrotnych źle postawionych opracowanych przez akademika A.N. Tichonowa ze Szkoły Matematycznej . Podstawą teoretyczną metody identyfikacji gęstości rozkładu zmiennych losowych jest przybliżone rozwiązanie równania Fredholma pierwszego rodzaju oraz zasada minimalizacji ryzyka strukturalnego. Algorytmiczna podstawa metody polega na rozwiązywaniu równoważnych układów równań liniowych w celu znalezienia współczynników rozwinięcia o pożądanej gęstości względem wybranego układu funkcji bazowych. Przy założeniu, że kształt gęstości rozkładu jest gładki, prawa rozkładu są przywracane na zbiorze funkcji trygonometrycznych z ograniczeniem liczby wyrazów rozwinięcia N w zależności od objętości L obserwowanych danych poprzez minimalizację ryzyka gwarantowanego. Podstawą podejścia do probabilistycznej (stochastycznej) analizy danych klinicznych kobiet w ciąży są metody poprawnej identyfikacji praw rozkładu zmiennych losowych w biomedycynie, opublikowane po raz pierwszy w czasopiśmie European Journal w 2015 roku. Opierają się one na rozwiązaniach problemów odwrotnych źle postawionych, opracowanych przez akademika A.N. Tichonowa ze Szkoły Matematycznej [20]. Podstawą teoretyczną metody identyfikacji gęstości rozkładu zmiennych losowych jest przybliżone rozwiązanie równania Fredholma pierwszego rodzaju oraz zasada minimalizacji ryzyka strukturalnego. Algorytmiczna podstawa metody polega na rozwiązywaniu równoważnych układów równań liniowych w celu znalezienia współczynników rozwinięcia o pożądanej gęstości względem wybranego układu funkcji bazowych.

Stochastyczna diagnostyka okołoporodowa

Zakładając, że kształt gęstości rozkładu jest gładki, prawa rozkładu są odtwarzane na zbiorze funkcji trygonometrycznych z ograniczeniem liczby wyrazów rozwinięcia N w zależności od objętości L danych obserwowanych poprzez minimalizację ryzyka gwarantowanego. Wyniki badań opartych na proponowanej metodzie identyfikacji były już wcześniej wykorzystywane w zastosowaniach medycznych: w gastroenterologii i immunologii klinicznej. Przykładowo, fraktalny model sygnału EGEG (sygnału elektrogastroenterograficznego) jest rozważany w publikacjach jako ułamkowy ruch Browna w kategoriach rozkładów polimodalnych. W przeprowadzonych badaniach rozwiązywane są problemy identyfikacji praw rozkładu dla lokalnych przekrojów sygnału EGEG (realizacji selektywnych). Zidentyfikowane charakterystyki polimodalne zawierają

informacje o dynamice stanów przewodu pokarmowego (GIT) pacjentów, umożliwiają identyfikację cech klasyfikacyjnych i obliczenie klasycznych parametrów statystycznych rozkładów. W porównaniu ze znanymi metodami diagnostyki i predykcji, proponowane algorytmy pozwalają na identyfikację ewolucji gęstości rozkładu procesów losowych niemal w czasie rzeczywistym, dostarczają cennych informacji do weryfikacji rozwiązań stochastycznych i umożliwiają efektywne wykorzystanie metod analizy entropii. Wysoka wrażliwość algorytmów entropii na zmiany parametrów danych biomedycznych sprawia, że proponowane podejście jest pożądane dla efektywnego doboru leków i wysokiej jakości terapii indywidualnej. Opracowane podejścia pozwoliły na skuteczne rozwiązanie problemów medycyny praktycznej związanych z monitorowaniem stanu pacjentek w okresie okołoporodowym (stochastyczna diagnostyka okołoporodowa). Na początku 2021 roku naukowcy z Państwowego Uniwersytetu Technicznego im. R.E. Aleksiejewa w Niżnym Nowogrodzie (NSTU) i Nadwołżańskiego Uniwersytetu Medycznego (PRMU) przeprowadzili następujące wspólne prace:

1. Do poprawnej analizy danych uzyskanych podczas 24-godzinnego monitorowania ciśnienia tętniczego w trakcie nadzoru nad kobietami w ciąży wykorzystano złożone metody analizy stochastycznej.

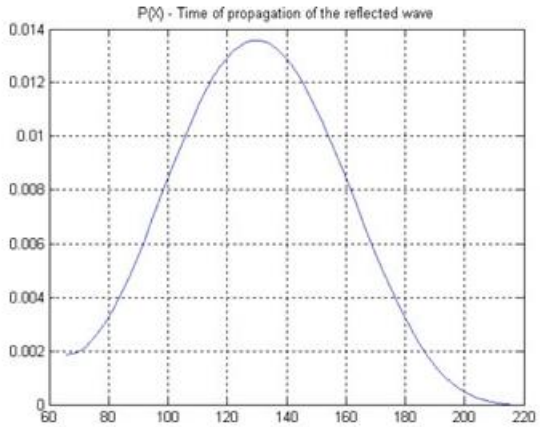
2. Przetwarzanie danych medycznych uzyskanych w 2020 roku poprzez monitorowanie układu sercowo-naczyniowego w grupie kobiet ciężarnych pod nadzorem przeprowadzono zgodnie z ośmioma istotnymi wskaźnikami .

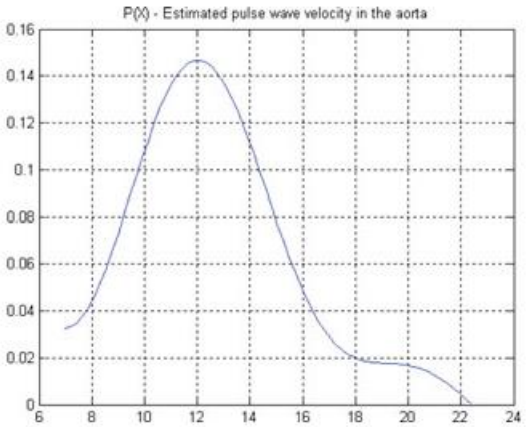
3. Zidentyfikowano prawa rozkładu (gęstości rozkładu prawdopodobieństwa), które między innymi charakteryzowały się cechami polimodalnymi odzwierciedlającymi aktualny stan pacjentek (ponad 150 cech).

4. Na podstawie analizy stochastycznej zidentyfikowano dwa wskaźniki markerowe (czas przejścia fali odbitej – RWTT i szacowaną prędkość fali tętna w aorcie – PWVao), które charakteryzowały sztywność układu naczyniowego i były istotne dla przewidywania przebiegu ciąży i pomyślnego porodu. Różnice różnicowalne dla grupy pacjentek z normą i patologią wynoszą 15–20% i są wiarygodnie identyfikowane proponowaną metodą.

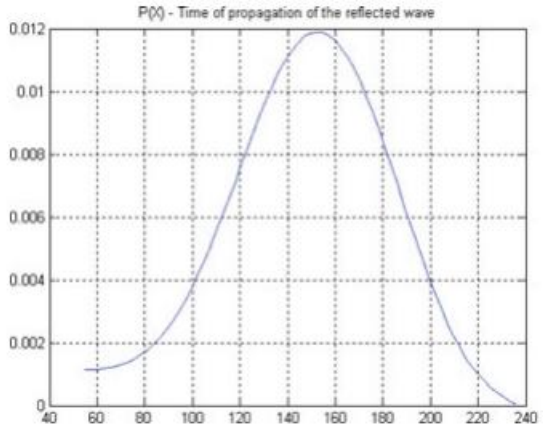
5. Zidentyfikowane cechy markerowe potwierdzono badaniami w ramach koncepcji fraktalnej metodą Hursta i metodą analizy widmowej Fouriera.

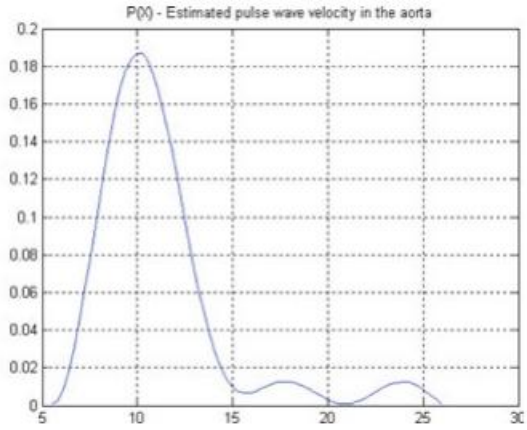
Rysunek przedstawia dla ilustracji kilka faktycznie „uśrednionych" zidentyfikowanych wskaźników markerowych kobiet, które urodziły z powodzeniem (gęstość rozkładu), które dość dobrze korelują z indywidualnymi cechami kobiet w ciąży.



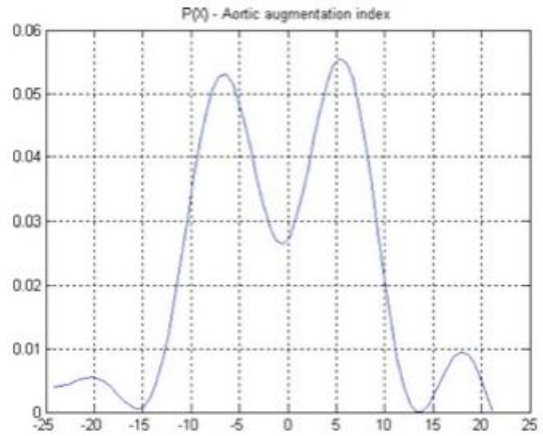


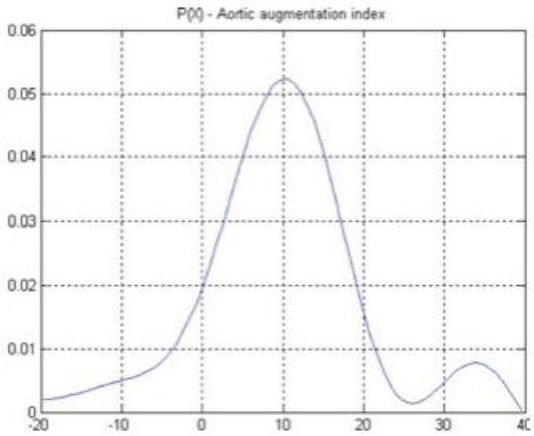
Rozkłady urodzeń z patologią charakteryzują się przesunięciami ekstremów gęstości, w tym polimodalnymi cechami parametru PWVao .



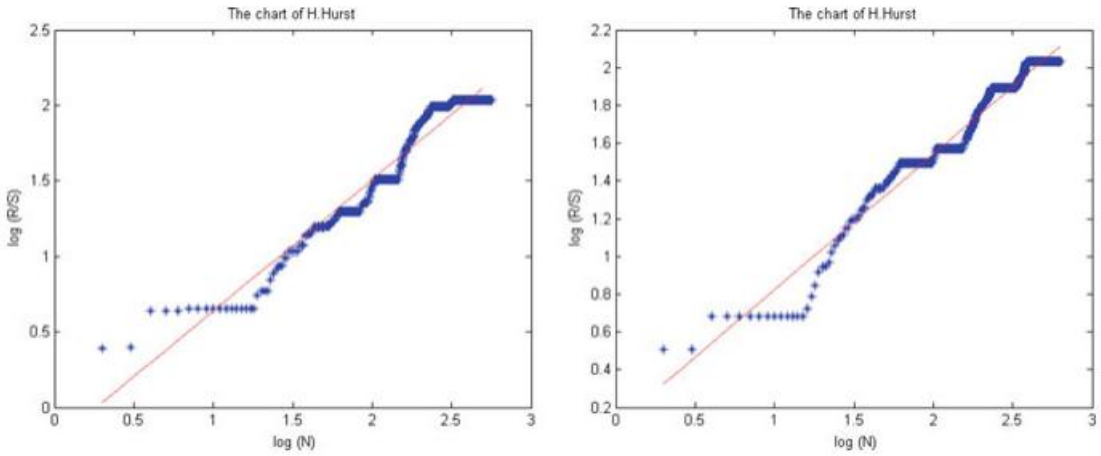


Rysunek 3 ilustruje istotną zmienność innych ważnych wskaźników.





Zidentyfikowane gęstości rozkładu wskaźnika Aixao (wskaźnik powiększenia aorty) dla dwóch pacjentek, które urodziły z powodzeniem (próby odpowiednio 72 i 56 zliczeń) wykazują pewne wyraźne cechy polimodalne i różne centra klastrowania najbardziej prawdopodobnych wartości. Wskazuje to na wpływ kilku istotnych czynników ciążowych na ten wskaźnik. Zatem, dla prawidłowej interpretacji wskaźnika Aixao, kluczowe jest uwzględnienie polimodalności i dodatkowej analizy. Gęstości rozkładu wskaźnika Aixao w grupie kobiet, które urodziły z powodzeniem, są generalnie bimodalne. Markery fali odbitej – RWTT i szacowanej prędkości fali tętna w aorcie – PWVao zidentyfikowane metodą stochastyczną zostały pomyślnie potwierdzone analizą R/S (metoda Hursta). Rysunek 4 przedstawia wykresy zależności lg(R/S) od lg(N) u pacjentek, które pomyślnie porodziły (17 kobiet).



N jest tutaj liczbą zliczeń w połączonej próbie. Zależności te pozwalają nam obliczyć wykładnik Hursta H: dla RWTT wynosi on 0,868, dla PWVao H = 0,717. Należy zauważyć, że dla grupy pacjentek z patologią okołoporodową uzyskano odpowiednio wartości H = 0,828 i H = 0,986. Biorąc pod uwagę wysoką czułość indeksu Hursta, różnice w wybranych grupach można uznać za istotne. Ponadto, zgodnie z teorią fraktali, wskaźnik ten oznacza przy H > 0,5 trwałość analizowanych wskaźników w okresie obserwacji. Wdrożenie tej nowej, sprawdzonej metody diagnostycznej planowane jest w Obwodowym Szpitalu Klinicznym im. N.A. Siemaszki (Katedra Położnictwa i Ginekologii Nadwołżańskiego Uniwersytetu Medycznego).

4 Wyniki i wnioski

Aby wcześnie przewidzieć ryzyko powikłań ciążowych, obiecujące wydaje się badanie stanu elastyczności ścian naczyń w oparciu o analizę stochastyczną, co pozwoli na optymalizację postępowania w ciąży i poprawę wyników okołoporodowych u kobiet w ciąży z nadciśnieniem tętniczym. Podstawą podejścia do probabilistycznej (stochastycznej) analizy wyników klinicznych u kobiet w ciąży będą metody prawidłowej identyfikacji praw rozkładu wartości losowych w biomedycynie. W celu zbadania danych z 24-godzinnego monitorowania ciśnienia krwi (BP) w trakcie nadzoru nad kobietami w ciąży, zidentyfikowano prawa rozkładu (gęstości rozkładu prawdopodobieństwa), które między innymi posiadają cechy polimodalne odzwierciedlające aktualny stan pacjentek. Na podstawie analizy stochastycznej zidentyfikowano dwa wskaźniki markerowe (czas przejścia fali odbitej RWTT oraz szacowaną prędkość fali tętna w aorcie PWVao), które charakteryzują sztywność układu naczyniowego i mają istotne znaczenie dla przewidywania przebiegu ciąży i pomyślnego porodu.

Hierarchiczne, rozmyte wnioskowanie o adekwatności wysoce informacyjnych objawów diagnostycznych raka piersi

Streszczenie Niniejsza sekcja przedstawia podejście do określania wpływu wysoce informacyjnych cech diagnostycznych na wynik klasyfikacji w cyberfizycznych systemach diagnostyki pacjentów w medycynie. Wprowadzono koncepcję adekwatności jako kryterium generalizujące, łączące takie sprzeczne cechy, jak poziom wpływu na poprawność i błędność diagnozy. Przeprowadzono eksperyment obliczeniowy w celu ustalenia poziomu adekwatności cech w ramach określonej metody klasyfikacji. Dostępny zbiór danych zawierał silnie skorelowane cechy i umożliwiał analizę czynnikową dopiero po usunięciu pewnej liczby cech.

Problem ten rozwiązano w oparciu o model czynnikowy o zredukowanej liczbie cech, do którego zastosowano algorytm wnioskowania rozmytego. Umożliwiło to aproksymację ładunków czynnikowych dla pozostałych cech, które nie zostały uwzględnione w modelu czynnikowym.

Słowa kluczowe Systemy cyberfizyczne · Termometria mikrofalowa · Rak piersi · Analiza czynnikowa · Wnioskowanie rozmyte

1 Wstęp

Obecnie systemy cyberfizyczne znajdują szerokie zastosowanie w opiece zdrowotnej. Podobne systemy są również wykorzystywane w diagnostyce raka piersi na podstawie danych termometrycznych. Metoda mikrofalowej termometrii radiowej (RTM) polega na określaniu temperatury powierzchni (skóry) i głębokiej ciała poprzez pomiar natężenia promieniowania elektromagnetycznego. Pomiary przeprowadzane są za pomocą aplikatorów antenowych, instalowanych na powierzchni ciała pacjentki. Uzyskane dane temperaturowe są wykorzystywane przez system sztucznej inteligencji, którego celem jest identyfikacja anomalii termicznych wewnątrz piersi.

Przedstawiono problemy diagnostyczne i ich rozwiązanie metodami radiotermometrii. Do tej pory w ramach prowadzonych badań specjaliści zbudowali:

1. Model koncepcyjny zachowania pól temperaturowych, opisujący aspekty charakterystyczne dla różnych klas diagnostycznych;

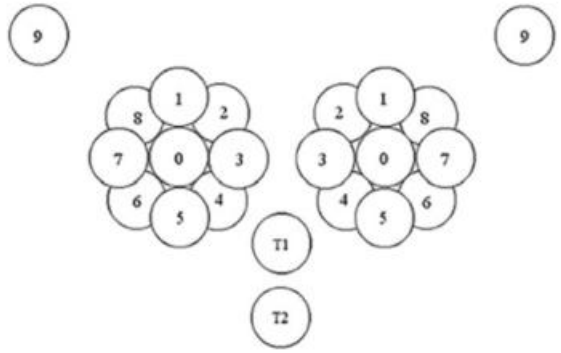
2. Model matematyczny formalizujący hipotezy heurystyczne dotyczące zachowania temperatur;

3. Przestrzeń cech, która jest wykorzystywana przy klasyfikacji pacjentek z wykorzystaniem metod uczenia maszynowego.

Wyniki etapów 1–3 można znaleźć na przykład w pracach A.G. Loseva, V.V. Levshinsky'ego, E.A. Mazepy. Należy również zwrócić uwagę na szereg badań, które dowodzą, że zastosowanie radiotermometrii mikrofalowej w połączeniu z metodami sztucznej inteligencji wykazuje wysoką skuteczność we wczesnej diagnostyce chorób. Powyższe podejścia umożliwiają zastosowanie metod uczenia maszynowego do budowy modeli klasyfikatorów, które rozwiązują problemy diagnozowania grupy ryzyka choroby u pacjentów. Rodzi to pytanie o wpływ każdej cechy na wynik klasyfikacji. Wiedza ta pozwala, po pierwsze, zidentyfikować przyczyny uzyskanego wyniku, co może stanowić podstawę do stworzenia systemu rekomendacji wspomagającego podejmowanie decyzji przez specjalistę. Po drugie, skorygować model klasyfikatora w celu uzyskania bardziej adekwatnych wyników. Problem ten można rozwiązać poprzez łączne zastosowanie analizy czynnikowej i algorytmów wnioskowania rozmytego.

2 Materiały i metody badań

Kompleks pomiarowy RTS-01-RES służy do diagnostyki gruczołów piersiowych. Podczas pomiarów rejestrowane są 44 temperatury w zakresie RTM (głębokiej) i IR (skórnej). Schemat pomiaru temperatury gruczołów piersiowych pokazano na rysunku.



Przykładowe dane temperaturowe można przedstawić w następujący sposób:

$$S = \begin{pmatrix} t_0^1 & \cdots & t_{43}^1 \\ \vdots & \ddots & \vdots \\ t_0^k & \cdots & t_{43}^k \\ t_0^{k+1} & \cdots & t_{43}^{k+1} \\ \vdots & \ddots & \vdots \\ t_0^n & \cdots & t_{43}^n \end{pmatrix},$$

gdzie t^j_i , i = 0, ..., 9 — i-ta temperatura punktu prawej piersi j-tej pacjentki (zakres RTM);

t^j_i , i = 10, ..., 19 — [i–10]-ta temperatura punktu prawej piersi j-tej pacjentki (zakres IR);

t^j_i , i = 20, ..., 29 — [i–20]-ta temperatura punktu lewej piersi j-tej pacjentki (zakres RTM);

t^j_i , i = 30, ..., 39 — [i–30]-ta temperatura punktu lewej piersi j-tej pacjentki (zakres IR);

t^j_{40} = T1, t^j_{41} = T2 — temperatury punktów odniesienia j-tej pacjentki (zakres RTM z);

t^j_{42} = T1, t^j_{43} = T2 — temperatury punktów odniesienia j-tej pacjentki (zakres IR).

W wyniku analizy przeprowadzonej przez specjalistów, na podstawie danych temperaturowych określono jakościowe objawy raka piersi. Wśród nich można zauważyć:

• zwiększoną symetrię termiczną między poszczególnymi punktami w dotkniętym chorobą gruczole piersiowym;

• zwiększoną zmienność temperatury między poszczególnymi punktami w porach piersi;

• stosunek temperatury skóry do temperatury wgłębnej;

• różnicę temperatury brodawki sutkowej;

• podwyższona temperatura sutka w dotkniętym chorobą gruczole piersiowym w porównaniu ze średnią temperaturą piersi, z uwzględnieniem zmian temperatury związanych z wiekiem.

Dla każdego jakościowego objawu raka możliwe jest skonstruowanie zestawu funkcji modelujących na podstawie temperatur, których wartość charakteryzuje ten objaw. Jako przykład podano kilka funkcji symulujących:

Asymetrię pól temperaturowych gruczołów piersiowych:

$$\sum_{i=0}^{9} |t_i - t_{i+20}|;$$

1. Stosunek temperatury skóry do temperatury głębokości:

$$\max_{i=\overline{0,9}} (t_i - t_{i+10}).$$

W rezultacie obliczono wartości cech p_{jr} , gdzie r = 1, . . . , s, dla każdej pacjentki. Znaki te oznaczono jako MG001, ..., MG062, czyli s = 62. W niniejszej pracy wykorzystano bazę danych termometrycznych zawierającą informacje o pacjentkach ośrodków onkologicznych. Do przeprowadzenia badania bazę danych przekształcono w taki sposób, aby każdy element odpowiadał oddzielnemu gruczołowi piersiowemu pacjentki. Stopień zaawansowania choroby piersi ocenia się w 6-stopniowej skali:

• th0 — brak anomalii temperatury; gruczoł piersiowy zdrowy;

• th1 — wykryto anomalie temperatury, ale nie są one typowe dla raka piersi (niższe temperatury);

• th2 — wykryto anomalie temperatury, ale nie są one typowe dla raka piersi (wyższe temperatury);

• th3 – wysoki poziom temperatury skóry i głębokiej;

• th4 – stwierdzono liczne anomalie temperatury;

• th5 – termogram jest typowy dla pacjentek z ostrym zaostrzeniem lub rakiem piersi.

W tej skali diagnoza jest tłumaczona na typ numeryczny. Mianowicie, w przypadkach th0–th2 zakładamy, że gruczoł piersiowy jest zdrowy, w przypadkach th3–th5 – gruczoł piersiowy należy do grupy ryzyka. Cała próbka danych H zawiera 15 116 elementów. Klasa diagnostyczna „Zdrowa” zawiera 13 570 elementów, klasa diagnostyczna „Grupa ryzyka” – 1546 elementów. Stan diagnostyczny pacjentki opisują 62 funkcje, które są matematyczną formalizacją jakościowych objawów raka. Aby przeprowadzić analizę czynnikową, konieczne było sklasyfikowanie wszystkich rekordów z bazy danych za pomocą jednego z algorytmów uczenia maszynowego. Jako taki algorytm wybrano regresję logistyczną. Do klasyfikacji każdego rekordu postanowiono zastosować metodę walidacji krzyżowej. Metoda ta polega na sekwencyjnym uczeniu i testowaniu algorytmu klasyfikacji na różnych podpróbach, wymieszanych ze sobą w próbie uczącej. W tym celu podzielono całą bazę danych na 10 równych i nienakładających się na siebie części (maksymalna różnica w ich wielkości wynosi dwa rekordy). Każda część zawierała równą liczbę rekordów każdej klasy. Następnie 9 z 10 części umieszczono w próbie, na której trenowano regresję logistyczną. Pozostałą część, niewłączoną do próby uczącej, sklasyfikowano. Wyniki klasyfikacji zapisano w ostatecznej bazie danych H. W ten sposób, oprócz klasy rzeczywistej, przechowuje ona również klasę przypisaną przez regresję logistyczną. Po sklasyfikowaniu wszystkich rekordów z rozważanej części, algorytm został ponownie wytrenowany na nowym zbiorze uczącym. Nowa próba została sporządzona w taki sposób, że jedna z części poprzedniej próby została zamieniona z próbą sklasyfikowaną. Proces klasyfikacji powtórzono 10 razy, aż wszystkie rekordy z bazy danych zostały sklasyfikowane. W ten sposób regresja logistyczna poprawnie sklasyfikowała 90% zdrowych gruczołów piersiowych i 88% gruczołów z grupy ryzyka. Na podstawie uzyskanego wyniku klasyfikacji, bazę H podzielono na dwie próby:

• H^0, składającą się z 2766 elementów, reprezentującą dane dotyczące piersi, dla których algorytm dokonał błędnej klasyfikacji.

• H^1, składającą się z 12 350 elementów, reprezentującą dane dotyczące piersi, dla których algorytm klasyfikacji ustalił prawidłową diagnozę.

Dla każdej próby przeprowadzono analizę czynnikową oddzielnie, stosując estymację metodą największej wiarygodności. Do wyboru czynników wystarczy, aby macierz korelacji była dodatnio określona, a wyznacznik tej macierzy był różny od zera. Wymagania te nie są jednak spełnione dla tego zestawu cech, co potwierdza analiza tego zestawu za pomocą programu statystycznego. Wynika to z faktu, że niektóre cechy są ze sobą silnie skorelowane. W związku z tym postanowiono usunąć takie cechy z późniejszym przywróceniem ich ładunków czynnikowych za pomocą wnioskowania rozmytego. Celem analizy czynnikowej jest uzyskanie ładunków czynnikowych cech. Ładunki czynnikowe to korelacja każdego znaku z każdym zidentyfikowanym czynnikiem. Im bliższy związek tej cechy z rozpatrywanym czynnikiem, tym wyższa wartość ładunków czynnikowych. Zatem po skonstruowaniu modeli czynnikowych dla każdej cechy definiuje się dwie charakterystyki:

• poziom wpływu na błędność diagnozy postawionej przez klasyfikator (odpowiadający wartości ładunku czynnikowego w próbie H^0);

• poziom wpływu na poprawność diagnozy postawionej przez klasyfikator (odpowiadający wartości ładunku czynnikowego w próbie H^1).

Proponuje się połączenie tych dwóch, ogólnie rzecz biorąc, sprzecznych charakterystyk adekwatności, również za pomocą wnioskowania rozmytego.

Przedstawmy definicję adekwatności w tym kontekście. Adekwatność jest rozumiana jako oczekiwanie badacza, że cecha będzie miała pozytywny wpływ na poprawność diagnozy i odwrotnie, negatywny wpływ na błędną klasyfikację. Dla poprawnego opisu wielowymiarowych zależności „wejściowo-wyjściowych" zadania określania poziomu adekwatności cechy w ramach określonej metody klasyfikacji proponuje się zastosowanie hierarchicznego wnioskowania rozmytego. W niniejszej pracy wykorzystano następującą metodę: przeprowadzenie wnioskowania rozmytego dla zmiennych pośrednich (przywrócenie ładunków czynnikowych) z późniejszym przeniesieniem jednoznacznych wartości tych zmiennych do układów rozmytych kolejnego poziomu hierarchii (określenie poziomu adekwatności cechy) .

3 Eksperyment obliczeniowy i jego omówienie

Dla wybranych 22 cech (w próbie H^0) i 25 cech (w próbie H^1) przeprowadzono analizę czynnikową z wykorzystaniem pakietu oprogramowania Statistic (aby przeprowadzić analizę czynnikową z wykorzystaniem estymacji największej wiarygodności, pozostawiono tylko te cechy, których wartość współczynnika korelacji jest mniejsza niż 0,7). W rezultacie uzyskano macierz składowych dla wybranych cech. Wysokie ładunki czynnikowe uzyskane w próbie H^1 mogą oznaczać, że zmienne te przyczyniły się do tego, że instancje próby testowej znalazły się w klasie z poprawnymi wynikami klasyfikatora. W przypadku próby H0, przeciwnie, cechy te mogą oznaczać, że klasyfikator wyciągnął niepoprawne wnioski na temat przykładów testowych w próbie. Możliwe jest obliczenie poziomu wpływu dla cech brakujących w modelu czynnikowym, wykorzystując informacje o poziomie korelacji, za pomocą wnioskowania rozmytego. W szczególności, dla zmiennych, które nie zostały wykorzystane w procesie konstruowania modeli

czynnikowych, poziomy wpływu na poprawność/błędność zostały przywrócone za pomocą wnioskowania rozmytego Mamdaniego. Wprowadzono następujące zmienne lingwistyczne:

(1) $\widetilde{C_i^1}$ — „poziom wpływu", który jest ładunkiem czynnikowym cechy i, gdzie i ∈ I — liczba cech wykorzystanych w konstrukcji modelu czynnikowego;

(2) $\widetilde{C_i^2}$ — „poziom wpływu", który jest ładunkiem czynnikowym cechy i, gde i ∈ J — liczba cech, które nie zostały wykorzystane w konstrukcji modelu czynnikowego;

(3) $\widetilde{R_{ij}}$ — „korelacja" — wartość współczynnika korelacji między p_{ir} a p_{jr} , gde r = 1, . . . , s.

Zmienne językowe charakteryzują się zestawem znaczeń (terminów): „średnia siła", „silna siła", „znaczna siła". Formalny opis zmiennej językowej „korelacja" można przedstawić następująco.

$$\widetilde{R^1} = (\beta = \text{“}\,correlation\,\text{”}, T = \{t_1 = \text{“}\,medium\,strength\,\text{”}, t_2 = \text{“}\,strong\,\text{”}, t_3 = \text{“}\,significant\,strength\,\text{”}\}, X = [0.7, 1], G = \emptyset, M = \emptyset),$$

gdzie t_1 — termin opisany przez zmienną rozmytą.

$\hat{t}_1 = (\alpha = \text{“}medium\,strength\text{”}, X = [0.7, 1], \mu_1(x) = \frac{-25x}{3} + \frac{41}{6}, 0.7 \le x \le 0.82);$
t_2—*a term described by a fuzzy variable* $\hat{t}_2 = (\alpha = \text{“strong”}, X = [0.7, 1], \mu_2(x) = \begin{cases} \frac{100x}{13} - \frac{72}{13}, & 0.72 \le x \le 0.85, \\ \frac{-25x}{3} + \frac{97}{12}, & 0.85 < x \le 0.97. \end{cases});$

t_3—*a term described by a fuzzy variable* $\hat{t}_3 = (\alpha = \text{“}significant\,strength\text{”}, X = [0.7, 1],$

$$\mu_3(x) = \frac{100x}{13} - \frac{87}{13}, \quad 0.87 \le x \le 1).$$

Formalny opis zmiennej językowej „poziom wpływu" można przedstawić następująco:

$$\widetilde{C^1} = (\beta = \text{“}\,level\,of\,influence\,\text{”}, T = \{t_1 = \text{“}\,medium\,strength\,\text{”}, t_2 = \text{“}\,strong\,\text{”}, t_3 = \text{“}\,significant\,strength\,\text{”}\}, X = [0.5, 1], G = \emptyset, M = \emptyset),$$

gdzie t_1 — termin opisany przez zmienną rozmytą.

$\hat{t}_1 = (\alpha = \text{“}medium\,strength\text{”}, X = [0.5, 1], \mu_1(x) = -5|x| + 3.5, 0.5 \le x \le 0.7);$
t_2—*a term described by a fuzzy variable* $\hat{t}_2 = (\alpha = \text{“}strong\text{”}, X = [0.5, 1], \mu_2(x) = \begin{cases} \frac{100|x|}{21} - \frac{18}{7}, & 0.54 \le x \le 0.75, \\ -5|x| + 4.75, & 0.75 < x \le 0.95 \end{cases});$
t_3—*a term described by a fuzzy variable* $\hat{t}_3 = (\alpha = \text{“}significant\,strength\text{”}, X = [0.5, 1],$

$$\mu_3(x) = \frac{100|x|}{21} - \frac{79}{21}, 0.79 \le x \le 1).$$

Fragment wykorzystanej bazy reguł, składający się z 9 reguł:

Reguła 1: Jeżeli ($\widetilde{C_i^1}$ „znaczna siła") oraz ($\widetilde{R_{ij}}$ „znaczna siła"), to ($\widetilde{C_i^2}$ i „znaczna siła").

Reguła 2: Jeżeli ($\widetilde{C_i^1}$ i „silne") oraz ($\widetilde{R_{ij}}$ j „znaczna siła"), to ($\widetilde{C_i^2}$ „silne").

Reguła 3: Jeżeli ($\widetilde{C}_i^1$ „średnia siła”) oraz ($\widetilde{R}_{ii}$ „znaczna siła”), to ($\widetilde{C}_i^2$ „średnia siła”).

Do implementacji algorytmu wnioskowania rozmytego Mamdaniego wykorzystano program Matlab. Po defuzzyfikacji zmiennych wyjściowych uzyskano wartości poziomu wpływu cech, które nie brały udziału w budowie modelu czynnikowego. Dzięki analizie czynnikowej możliwe było zatem ustalenie poziomów wpływu cech na poprawność/błędność wyniku klasyfikacji. Dalsze wykorzystanie algorytmu wnioskowania rozmytego Mamdaniego pozwoliło na zniwelowanie ograniczeń estymacji metodą największej wiarygodności, związanych z zależnością liniową i wysoką korelacją cech. Uzyskane poziomy wpływu służą jako zmienne wejściowe do rozmytego wniosku o adekwatności znaków : wartość ładunków czynnikowych w próbie z poprawnymi diagnozami to zmienna lingwistyczna „poprawność”, wartość ładunków czynnikowych w próbie z błędnymi diagnozami to „błędna”. Te zmienne lingwistyczne charakteryzują się wieloma znaczeniami (terminami): „praktycznie nieobecne”, „małe”, „średnie”, „wysokie”. Formalny opis zmiennej lingwistycznej „poprawność” można przedstawić następująco:

$$\widetilde{Cor^1} = \left(\beta = \text{“}\, correctness\, \text{”}, T = \left\{t_1 = \text{“}\, practically\ absent\, \text{”}, \right.\right.$$
$$\left.\left. t_2 = \text{“}\, small\, \text{”}, t_3 = \text{“}\, medium\, \text{”}, t_4 = \text{“}\, high\, \text{”}\right\}, X = [0, 1], G = \emptyset, M = \emptyset\right),$$

where t_1—*a term described by a fuzzy variable* $\hat{t}_1 = (\alpha =$ “*practically absent*”, $X = [0, 1]$, $\mu_1(x) = 2x, 0 \le x \le 0.5)$;

t_2—*a term described by a fuzzy variable* $\hat{t}_2 = (\alpha =$ “*small*”, $X = [0, 1]$,

$$\mu_2(x) = 1 - \frac{x - 0.5}{0.2}, 0.5 \le x \le 0.7);$$

t_3—*a term described by a fuzzy variable* $\hat{t}_3 = (\alpha =$ “*medium*”, $X = [0, 1]$,

$$\mu_3(x) = \begin{cases} 1 - \frac{0.75 - x}{0.2}, & 0.55 \le x \le 0.75, \\ 1 - \frac{x - 0.75}{0.2}, & 0.75 \le x \le 0.95 \end{cases};$$

t_4—*a term described by a fuzzy variable* $\hat{t}_4 = (\alpha =$ “high”, $X = [0, 1]$, $\mu_4(x) = 1 - \frac{1 - x}{0.2}, 0.8 \le x \le 1)$.

Podobnie jak w poprzednim przykładzie, zbudowano bazę reguł rozmytych. W wyniku wnioskowania rozmytego, w szczególności etapów rozmycia zmiennych wejściowych, agregacji podwarunków, aktywacji podwniosków, akumulacji wniosków oraz defuzzyfikacji zmiennych wyjściowych , uzyskano jednoznaczne wartości adekwatności cech.

4. Wnioski

Osiągnięty poziom adekwatności cech do wyniku klasyfikacji pozwoli na:

• po pierwsze, interpretację wyników klasyfikatora;

• po drugie, udoskonalenie klasyfikatora poprzez zmianę jego modelu matematycznego w kierunku, który zapewni większą rolę tym cechom, których poziom adekwatności zostanie oznaczony jako wysoki.

Opracowanie spersonalizowanej diety z wykorzystaniem optymalizacji strukturalnej

Opracowanie spersonalizowanej diety człowieka z uwzględnieniem różnorodnych czynników wiąże się z analizą systemową i formalizacją zgromadzonych danych i wiedzy, a także cyfrowymi technologiami inteligentnymi do ich przetwarzania i podejmowania optymalnych decyzji. Przedstawiono metodologię optymalizacji i tworzenia spersonalizowanych diet opartą na modelowaniu strukturalno-parametrycznym. Proponowane podejście pozwala na rozwiązanie następujących zadań: (1) analizę dziennego jadłospisu lub poszczególnych posiłków (śniadania, obiadu, podwieczorku, kolacji, posiłków dodatkowych (przekąsek)) ze znanym zestawem produktów gotowych pod względem wartości energetycznej i składu chemicznego w celu wykrycia zaburzeń żywieniowych; (2) obliczenie dla określonej listy produktów gotowych ich optymalnej ilości, jak najbardziej zbliżonej pod każdym względem do zalecanej diety referencyjnej w zależności od aktywności psychicznej i fizycznej; (3) optymalizację diety w zależności od zadania poprzez wybór grupy produktów gotowych z pełnej lub wyselekcjonowanej listy danych archiwalnych, z uwzględnieniem wszystkich niezbędnych parametrów jakościowych każdego produktu; (4) dostosowanie diety z uwzględnieniem odchyleń żywieniowych w zakresie niektórych parametrów składu chemicznego i wartości energetycznej poprzez dodatkowe wprowadzenie produktów o podwyższonej wartości biologicznej specjalnego przeznaczenia, suplementów multiwitaminowych i multiwitaminowo-mineralnych, a także naturalnych substancji bioaktywnych.

Słowa kluczowe: Dieta spersonalizowana · Żywienie cyfrowe · Opis parametryczny · Optymalizacja strukturalna

1 Wstęp

Dieta jest kluczowym czynnikiem wpływającym na zdrowie człowieka, a wybór odpowiedniego zestawu produktów spożywczych może znacząco poprawić stan zdrowia . Przestrzeganie zdrowego stylu życia stało się niezwykle ważnym aspektem w życiu ludzi. Wymaga to utrzymania zdrowego sposobu odżywiania, uwzględniającego charakter i ilość spożywanych produktów spożywczych, co pozwala na regulację spożycia kalorii i składników odżywczych. Badania pokazują, że nieprawidłowe odżywianie zwiększa ryzyko chorób niezakaźnych, takich jak otyłość, cukrzyca itp. Podejścia do automatyzacji procesu planowania posiłków wiążą się przede wszystkim z metodami matematycznymi: optymalizacją liniową, metaheurystyką, wnioskowaniem opartym na przypadkach, logiką rozmytą. Wiele badań naukowych dotyczących planowania żywienia wiąże się z wykorzystaniem optymalizacji liniowej. Pierwsze podejście do rozwiązania tego problemu zostało opracowane w 1945 roku przez amerykańskiego ekonomistę Stiglera. Zaproponowano model matematyczny i zastosowano metodę prób i błędów. W ramach tego podejścia wymieniono niestrukturyzowane produkty spożywcze, które spełniają standardowe wymagania żywieniowe, przy jednoczesnej optymalizacji cen produktów spożywczych. W ciągu ostatnich dwóch dekad „model Stiglera" został przyjęty i rozszerzony przy użyciu różnych wariantów paradygmatu programowania liniowego, programowania całkowitoliczbowego, mieszanego programowania całkowitoliczbowo-liniowego i programowania docelowego . Rozwiązując problemy planowania posiłków, przeprowadzono następujące badania Autorzy stosują system wymiany do planowania menu. Oferowana jest opcja substytucji żywności z tej samej grupy żywności (na przykład zastąpienie 1/2 szklanki makaronu 1/3 szklanki białego ryżu, ponieważ oba produkty należą do kategorii skrobiowej). Jednak praca nie bierze pod uwagę kompatybilności produktów z różnych kategorii, jak również wielkości porcji. W [8] autorzy

zwracają uwagę na ludzkie preferencje żywieniowe, rozważając potrzeby żywieniowe, ale ignorują kompatybilność między produktami spożywczymi. To podejście nie przypisuje żywności do planu posiłków (śniadanie, obiad, kolacja), ale przedstawia zestaw zalecanych produktów jako niestrukturalną listę. Problem w stosowaniu optymalizacji liniowej wynika przede wszystkim z faktu, że ze względu na dużą liczbę ograniczeń, rozwiązywalność problemu maleje. Dlatego nie jest możliwe uwzględnienie wielu czynników, takich jak wymagania żywieniowe, preferencje żywieniowe, różnorodność i kompatybilność. Aby stworzyć spersonalizowane plany żywieniowe, autorzy zastosowali podejście metaheurystyczne z algorytmami genetycznymi. Szczegółowy przegląd tych badań przedstawiono w artykule Bianchi oraz El-Ghazi zauważają, że podejście metaheurystyczne nie gwarantuje globalnie optymalnego rozwiązania, ale raczej zapewnia „dobre" rozwiązanie problemu. Wykorzystano techniki rozumowania oparte na studium przypadku. Naukowcy i badacze starają się zalecać plany żywieniowe, identyfikując najlepszy plan żywieniowy spośród zestawu istniejących, a następnie dostosowując go do potrzeb pacjenta docelowego poprzez rewizję istniejącego menu.Jako dane wejściowe należy określić produkty spożywane na śniadanie i obiad. W rezultacie pacjent otrzymuje pozostałe spożycie kalorii i niezbędne porcje jedzenia na kolację. Wykorzystując rozmyte ontologie żywności i profili osobistych przedstawiono rozwiązanie oferujące kilka opcji dań z każdej kategorii produktów, które spełniają wymagania dotyczące porcji obiadowych. Nie jest jednak konieczne gotowanie pełnego obiadu, ponieważ rozwiązanie nie łączy produktów w celu utworzenia kompletnego obiadu, lecz oferuje opcje produktów. Istnieje wiele aplikacji mobilnych i internetowych związanych z żywieniem i zdrowiem. Zazwyczaj są to narzędzia do śledzenia kalorii, takie jak MyFitnessPal , MyPlate , MyNetDiary i inne. Pomagają one pacjentom monitorować dzienne spożycie kalorii i składników odżywczych. Aplikacje te przyjmują spożyte przez pacjenta pożywienie jako dane wejściowe i obliczają ilość kalorii i składników odżywczych zawartych w spożywanych produktach spożywczych. Narzędzia do planowania posiłków, takie jak EatThisMuch , Fitness Meal Planner , MakeMyPlate , Yum-Me , pomagają tworzyć codzienne plany posiłków w oparciu o wymaganą wartość kaloryczną podaną przez pacjenta. Jednym z takich narzędzi jest MakeMyPlate , aplikacja mobilna, która zapewnia pacjentom wstępnie optymalnie zbilansowane, zalecane pod względem odżywczym codzienne diety, które spełniają określone przez użytkownika poziomy kalorii. Każdy plan posiłków pozwala znaleźć alternatywy, jednocześnie zachowując równowagę na talerzu. Pacjent wybiera alternatywę z dostępnej bazy danych, nie sprawdzając, czy zastępczy produkt/posiłek jest równoważny kalorycznie z oryginalnym (co może prowadzić do nadmiaru lub zmniejszenia zalecanej zawartości kalorycznej i ilości składników odżywczych w pożywieniu). Ponadto aplikacja mobilna MakeMyPlate również nie bierze pod uwagę preferencji żywieniowych pacjenta. Inne rozwiązanie, EatThisMuch, przyjmuje jako dane wejściowe podstawowe informacje o zdrowiu pacjenta (płeć, wiek, wzrost, waga i poziom aktywności) oprócz procentowej zawartości tkanki tłuszczowej. Oprócz tego wprowadzane są informacje tekstowe: (1) „co chcesz zrobić ze swoją wagą" — utrzymać, zredukować, przytyć i zwiększyć masę mięśniową); (2) preferowany rodzaj diety — śródziemnomorska, wegetariańska itp.; (3) czy masz jakieś preferencje żywieniowe i jak często chcesz „widzieć swoje preferencje" w tygodniowej diecie. W rezultacie aplikacja podaje użytkownikowi dzienne racje żywnościowe jako wniosek. Chociaż jest to potężna aplikacja mobilna, ma kilka ograniczeń. Po pierwsze, pozwala użytkownikowi tworzyć plany posiłków tylko na bieżący dzień. Planowanie posiłków na dwa lub więcej dni wymaga wykupienia subskrypcji premium. Po drugie, nie uwzględnia preferencji użytkownika co do konkretnego dania w formie oceny gradientowej, na przykład od „niepożądanego" do „bardzo preferowanego", co byłoby bardziej przydatne przy komponowaniu dostosowanych planów posiłków. Inna internetowa aplikacja fitnessowa Fitness Meal Planner powtarza większość funkcji i ograniczeń aplikacji

EatThisMuch. Jednak we wszystkich badaniach i aplikacjach mobilnych z reguły stosuje się kryterium wartości energetycznej. Trudność w podejmowaniu optymalnych decyzji polega na spersonalizowanym podejściu, doborze produktów i dań w diecie, uwzględnieniu genetyki i wymagań biomedycznych, uwzględnieniu cech genetycznych i parametrów danej osoby (wiek, płeć, stan fizjologiczny, pracochłonność, cechy regionalne, etniczne i ekologiczne, a także stan strawności poszczególnych produktów spożywczych), zależności strukturalnych i ograniczeń na poziomie składników, pierwiastków i monostrukturalnym. W tym kontekście interesujące jest opracowanie metodologii optymalizacji spersonalizowanych diet, która pozwoli na rozwiązanie następujących zadań: (1) analizę dziennego jadłospisu lub poszczególnych posiłków (śniadania, obiadu, podwieczorku, kolacji, posiłków dodatkowych (przekąsek)) ze znanym zestawem ilościowym produktów gotowych pod względem wartości energetycznej i składu chemicznego w celu wykrycia zaburzeń żywieniowych; (2) obliczenie dla określonej listy produktów gotowych ich optymalnej ilości, jak najbardziej zbliżonej pod każdym względem do zalecanej diety referencyjnej w zależności od aktywności umysłowej i fizycznej; (3) optymalizację diety w zależności od zadania poprzez wybór grupy produktów gotowych z pełnej lub wyselekcjonowanej listy danych archiwalnych, z uwzględnieniem w równym stopniu wszystkich niezbędnych parametrów jakościowych każdego produktu; (4) dostosowanie diety z uwzględnieniem odchyleń żywieniowych w zakresie niektórych parametrów składu chemicznego i wartości energetycznej poprzez dodatkowe wprowadzenie produktów o podwyższonej wartości biologicznej o specjalnym przeznaczeniu, suplementów multiwitaminowych i multiwitaminowo-mineralnych, a także naturalnych substancji bioaktywnych.

2. Strukturalno-parametryczny model spersonalizowanego żywienia człowieka

Implementacja oprogramowania do optymalizacji spersonalizowanej diety opiera się na bazie wiedzy w postaci opisu strukturalno-parametrycznego wielu czynników genotypowych i zmiennych stanu fizjologicznego organizmu żywego, relacji między nimi, składu i właściwości surowców spożywczych i produktów spożywczych, odzwierciedlającego wyniki badań naukowych oraz doświadczenie dietetyków, lekarzy, technologów i innych ekspertów, z uwzględnieniem wymagań biomedycznych i potrzeb fizjologicznych organizmu żywego. Cechy genetyczne i fizjologiczne człowieka, jego metabolizm węglowodanów, lipidów i białek można praktycznie określić w procesie badań genetycznych i medycznych, w szczególności na podstawie biochemicznego badania krwi. Model strukturalno-parametryczny (SPM) spersonalizowanego, odpowiedniego żywienia człowieka powstaje w oparciu o wiedzę zawartą w opisie parametrycznym człowieka i jego preferencji żywieniowych. SPM jest macierzą komórkową i odzwierciedla całą różnorodność istniejących znanych i nieznanych zależności między ludzkim genotypem, zmiennymi stanu fizjologicznego człowieka oraz parametrami diety i reżimu żywieniowego. Aby skomponować odpowiednio zbilansowaną dietę z należytym uwzględnieniem „paszportu zdrowia" danej osoby, ryzyka chorób i stanu przewodu pokarmowego, stosuje się algorytm dialogowy. Na podstawie opisów stanu medycznego i biologicznego danej osoby, a także biomarkerów, tworzony jest parametryczny model jej/jego odpowiedniego odżywienia jako określone parametry, normy i stosunki wymaganych składników odżywczych i składników na dzień. Konstruując model parametryczny osoby, należy uwzględnić następujące grupy właściwości:

1. Dane osobowe (wiek, płeć, aktywność fizyczna, region zamieszkania, rodzaj aktywności, stan psycho-emocjonalny itp.);

2. Właściwości krytyczne (wskaźnik masy ciała, ciśnienie krwi, poziom glukozy, indeks aterogenny itp.);

3. Właściwości drugorzędne (metabolizm węglowodanów, metabolizm białek, metabolizm lipidów itp.).

Optymalizacja strukturalna diety odbywa się z wykorzystaniem kryterium kwadratowego minimalnego odchylenia od struktury referencyjnej wskaźników wartości odżywczej i energetycznej dla określonej, określonej grupy osób. Obecnie wskaźniki referencyjne istnieją w formie uśrednionej dla różnych, określonych grup osób. Brak jest spersonalizowanych zaleceń. Działanie algorytmu polega na sekwencyjnym określeniu nierównowagi w k-tym pierwiastku z ograniczeniami na całkowitą objętość dziennej racji pokarmowej, na dopuszczalnych granicach zmiany ułamka masowego (objętości) j-tego produktu w diecie i jego minimalizacji (w przypadku niedoboru) poprzez zwiększenie go do górnej granicy udziału masowego jednego produktu y_l o maksymalnej właściwej zawartości b^{max}_{kl} niedoborowego pierwiastka i odpowiadającego mu zmniejszenia do minimalnego ułamka masowego innego produktu yr o minimalnej właściwej zawartości b^{min}_{kr} k-tego pierwiastka. W przypadku nadmiaru następuje przesunięcie strukturalne w przeciwnym kierunku. W wyniku redystrybucji ułamków masowych wybranej pary produktów do y^{max}_{l} i y^{min}_{r} otrzymujemy nową, ulepszoną wartość kryterium optymalizacji. Wartość redystrybucji ułamków masowych składników dobiera się na podstawie ograniczeń określonych granic y^{max}_{l} i y^{min}_{r}, a także ograniczeń dotyczących całkowitej objętości diety, jako

$\delta = \min\{\Delta y_k, \Delta y_{kmax}, \Delta y_{kmin}\}$

gdzie $\Delta yk_{max} = y^{max}_{l} - y_l$ — dopuszczalny wzrost ułamka masowego l-tego produktu;

$\Delta y_{kmin} = y_p - y^{min}p$ — dopuszczalny spadek ułamka masowego p-tego produktu.

Nowe wartości ułamków masowych wyznacza się jako $y'_l = y_l + \delta$; $y'_p = y_p - \delta$. Procedura jest powtarzana, aby znaleźć kolejną parę produktów w celu redystrybucji ich ułamków masowych, aż do znalezienia lokalnego minimum dla k-tego pierwiastka. Procedura jest kontynuowana aż do wyczerpania wszystkich możliwości redystrybucji ułamków masowych składników, co skutkuje alternatywną opcją diety z minimalnym odchyleniem k-tego pierwiastka od podanej wartości standardowej w strukturze odpowiedniego asortymentu żywieniowego. Tworząc dietę dla studentów z problemami przewodu pokarmowego, należy uwzględnić zasady profilaktyki oparte na badaniach naukowych i określone przez wymagania biomedyczne. Dysponując tą wiedzą, przeprowadzono klasteryzację w bazie danych „Produkty spożywcze” i uzyskano listę produktów i dań według kategorii (sałatki, dania pierwsze, drugie, dodatki itp.), którą wykorzystano w przygotowaniu diety. W wyniku optymalizacji strukturalnej uzyskano opcje dietetyczne w profilaktyce chorób przewodu pokarmowego.

3 Wnioski

Komponowanie spersonalizowanej diety dla człowieka jest bezpośrednio związane z technologiami cyfrowymi. Tylko one pozwalają na uwzględnienie całego niezbędnego zakresu czynników: indywidualnych cech fizjologicznych (grupy krwi, poziomu glukozy we krwi, wskaźnika masy ciała, mikrobiomu, kwasowości i płynów żołądkowo-jelitowych itp.), danych osobowych (wieku, płci, wzrostu, wagi itp.), poziomu i charakteru przemian składników odżywczych pożywienia pod wpływem działania pepsyny i trypsyny w przewodzie pokarmowym oraz ich strawności.

Opracowanie algorytmu przygotowania półproduktów do pakowania

Przedstawiono algorytm przygotowania półproduktów do pakowania i oceniono jego efektywność. Półproduktami są schłodzone nuggetsy, które przemieszczają się wzdłuż szybkiego odcinka linii przenośnika w otwartych kartonach. Przygotowanie polega na wyrównaniu opakowań polietylenowych z nuggetsami tak, aby nie wystawały poza ścianki kartonu. Jest to konieczne, aby zapewnić prawidłowe zamknięcie kartonu z nuggetsami na końcowym etapie pakowania. Poziomowanie półproduktów odbywa się za pomocą robota-manipulatora DR-1, produkcji Intelligent Robotics LLC, zawieszonego nad taśmociągiem na specjalnej ramie. System automatyki obejmuje również detektor kartonów z schłodzonymi półproduktami oraz czujnik prędkości taśmy przenośnika. Przedstawiono oszacowanie błędów podczas przesyłania danych między czujnikami a elementami wykonawczymi. Zastosowanie proponowanego algorytmu pozwala na automatyczne wyrównywanie kartonów z nuggetsami, zwiększając tym samym wydajność i zmniejszając liczbę odrzutów.

Słowa kluczowe Produkcja żywności · Ramię robota · Pakowanie półproduktów · Zautomatyzowany system sterowania

1 Wstęp

W firmie przetwórstwa spożywczego na linii pakowania produktów gotowych istnieje zadanie wyrównywania półproduktów znajdujących się w otwartym opakowaniu. Konieczność tego procesu wynika z faktu, że jeśli półprodukty wystają poza opakowanie, maszyna pakująca nie będzie działać prawidłowo. Obecnie wyważanie jest wykonywane przez ludzi. Proponuje się zautomatyzowanie procesu wyrównywania. W publikacjach opisano proces rozwoju robota-manipulatora DR-1, który posiada niezbędne cechy do realizacji wyrównywania. Wykorzystuje on kinematykę prostą i odwrotną do pozycjonowania chwytaka. Do sortowania odpadów wykorzystywane są roboty Delta o tej samej konstrukcji [5]. Celem jest opracowanie algorytmu, który na podstawie sygnału z czujnika ruchu opakowania na taśmie przenośnika wydaje polecenia sterujące do manipulatora robota. Zdefiniowane zadania to:

1. Zbudowanie instalacji do wyrównywania półproduktów w warunkach laboratoryjnych.

2. Stworzenie matematycznego opisu procesu wyrównywania.

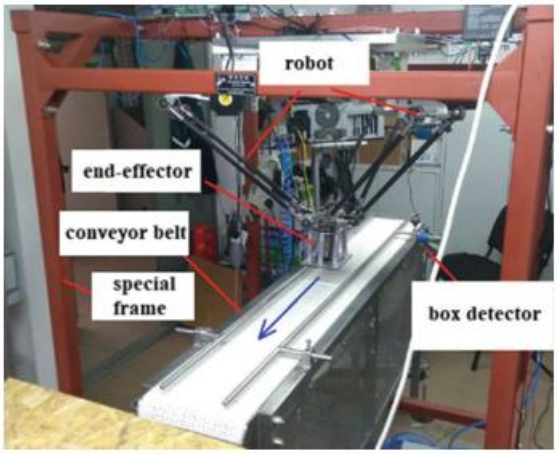
3. Opracowanie algorytmu procesu wyrównywania.

4. Obliczenie błędów wykonania algorytmu.

5. Przeprowadzenie testów powstałego systemu automatyki.

2 Opis sprzętu

Linia produkcyjna do pakowania bryłek składa się z przenośnika taśmowego i umieszczonego nad nim ramienia robota DR-1. Robot zamontowany jest na ramie. Do manipulatora robota, który jest platformą wibracyjną, przymocowany jest korpus roboczy. Linia przenośnikowa charakteryzuje się maksymalną prędkością ruchu 0,55 m/s i regulowaną szerokością, co umożliwia przesuwanie pudełek z bryłkami jeden po drugim .



Pudełko z samorodkami porusza się w kierunku niebieskiej strzałki i ma charakterystyki pokazane w tabeli .

Cecha: Wartość

Długość, m: 0,18

Szerokość, m: 0,13

Wysokość, m: 0,04

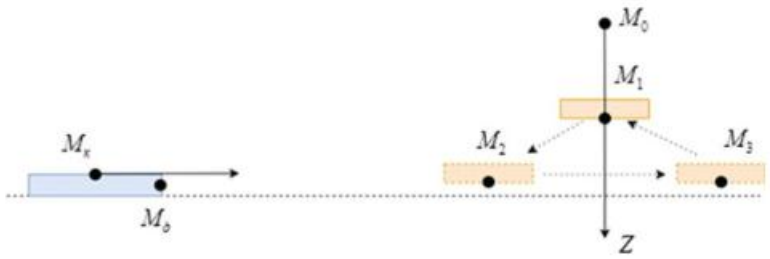
Liczba bryłek, sztuk: 14

Waga, kg: 0,3

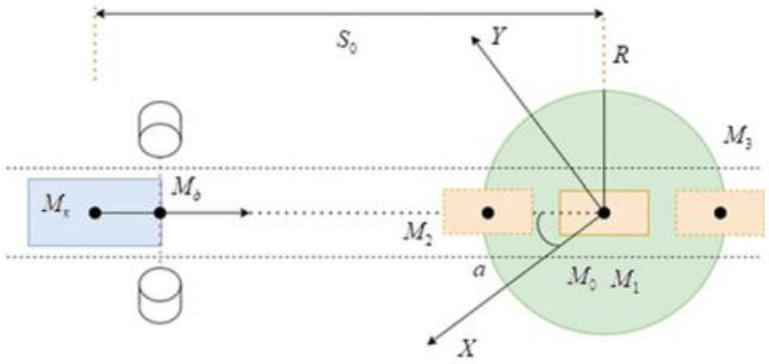
Wskaźnik laserowy i czujnik fotoelektryczny zostały użyte jako detektor, zamocowane na początku linii przenośnika, w odległości około 0,7 m od środka robota. Jest to najlepszy sposób, ponieważ używanie kamer RGB-D [8] i sieci neuronowych do określania położenia pudełek zajmuje zbyt dużo czasu i jest w tej sytuacji zbędne. Zastosowano bardziej klasyczny sposób. Czujnik fotoelektryczny wyprowadza dyskretny sygnał 0 V, jeśli pudełko nie jest obecne, i 3,3 V, jeśli pudełko przerywa wiązkę. Czujnik został podłączony do dyskretnego wejścia mikrokontrolera pokładowego robota DR-1.

3. Matematyczny opis procesu wyrównywania

Dla uproszczenia, instalacja przedstawiona na rysunku jest przedstawiona schematycznie z boku



i z góry



Przedstawiono układ współrzędnych, którego środek znajduje się w punkcie M_0, oś Z jest skierowana w dół, a oś X tworzy kąt a z linią środkową przenośnika. Oś Y uzupełnia prawą trójkę. Punkt M_0 jest środkiem geometrycznym robota. Kąt a pojawia się, ponieważ robot jest przymocowany do ramy w taki sposób, że nie może jej dotknąć podczas ruchu. Dlatego oś środkowa przenośnika nie jest współosiowa z żadną z osi układu współrzędnych robota. Kąt a jest stały. Podczas eksperymentów nad debugowaniem systemu planowania ruchu manipulatora robota DR-1 stwierdzono, że najwyższą prędkość można uzyskać, wyznaczając ścieżkę ruchu z minimalną liczbą punktów. Wynika to z faktu, że manipulator robota musi zatrzymać się w danym punkcie. Dlatego podczas operacji spłaszczania bryłek stosuje się trajektorię w kształcie trójkąta .Na rysunku 2 niebieski prostokąt przedstawia pudełko z bryłkami, a punkt $M_к$ stanowi środek jego górnej krawędzi. Pudełko pojawia się w linii widzenia manipulatora robota w momencie przejścia przez detektor (punkt $M_д$). M_0 jest początkiem układu współrzędnych manipulatora robota. Punkt M_1 jest współrzędną środka dolnej płyty korpusu roboczego w pozycji oczekiwania. Punkt M_2 jest współrzędną, w której dolna płyta manipulatora robota po raz pierwszy styka się z pudełkiem. W tym momencie następuje uderzenie w bryłki w pudełku przez sam manipulator robota i wibrator pneumatyczny umieszczony na organie roboczym. Uderzenie to trwa do punktu M_3, gdzie korpus roboczy towarzyszy pudełku. W punkcie M_3 narzędzie przestaje uderzać w pudełko i rozpoczyna ruch do pozycji oczekiwania – punktu M_1. Wysokości punktów M_1, M_2, M_3 są dobierane eksperymentalnie. Od nich zależy jakość nuggetsów oraz bezpieczeństwo opakowań i półproduktów. Podczas testów zaobserwowano, że najlepszym sposobem na wypoziomowanie nuggetsów jest ustawienie wysokości punktów M_2 i M_3 na wysokości górnej krawędzi pudełka. Taka wysokość minimalizuje wpływ implementacji na karton. Wysokość punktu M_1 jest o kilka centymetrów niższa (oś Z skierowana w dół) niż M_2 i M_3, co wystarcza, aby zapobiec zahaczeniu pudełka i znajdujących się w nim półproduktów podczas ruchu narzędzia. W momencie przejścia pudełka przez detektor (punkt $M_д$, moment czasowy t0), znane stają się współrzędne środka pudełka $M_к$ (1) oraz rzut odległości $M_к M_0$ na płaszczyznę XY, oznaczone jako S_0. Okazuje się, że S_0 to odległość, o jaką pudełko przesunie się od momentu wykrycia do momentu, gdy znajdzie się dokładnie pod środkiem robota.

$$\begin{cases} M_{к_x} = M_{д_x} + \frac{l * \cos(a)}{2} \\ M_{к_y} = M_{д_y} + \frac{l * \sin(a)}{2} \end{cases} \qquad (1)$$

gdzie l jest długością pudełka, indeksy x i y oznaczają rzuty na odpowiednie osie. Skrzynia porusza się do rzutu punktu M_0 ze stałą prędkością v, równą prędkości przenośnika. Możemy zatem zapisać równanie ruchu środka skrzyni w rzutach na osie X i Y, zakładając, że t to czas bieżący:

$$\begin{cases} s_x(t) = s_0 * \cos(a) + \mathrm{v} * \cos(a) * (t - t_0) \\ s_y(t) = s_0 * \sin(a) + \mathrm{v} * \sin(a) * (t - t_0) \end{cases} \quad (2)$$

Należy wyjaśnić, że prędkość v będzie miała wartość ujemną w układzie współrzędnych robota. Zdefiniujmy na płaszczyźnie XY okrąg o promieniu R i środku w punkcie M_0, który będzie efektywnym obszarem roboczym manipulatora robota. Oznacza to, że robot nie osiągnie równowagi, dopóki pudło nie znajdzie się wewnątrz okręgu:

$$\sqrt{s_x(t)^2 + s_y(t)^2} \leq R \quad (3)$$

Promień R został wybrany eksperymentalnie. Przeprowadzono eksperyment, w którym R zmniejszono z 0,2 do 0,03 m. Kryterium było jakościowe wyrównanie bryłek (pakiet bryłek nie wystaje poza górną krawędź pudełka) i fakt, że robot ma czas na wyrównanie wszystkich pudełek na taśmie przenośnika. Wysoka wartość R jest nadmierna i nie wyrównuje wszystkich pudełek. Mała wartość R, która degeneruje się do pchania pudełka od góry bez towarzyszącego temu, jest mniej skuteczna z punktu widzenia poziomowania. Wprowadzono wartość delty do śledzenia wielu pudełek. Jest to czas potrzebny, aby jedno pudełko całkowicie przesunęło się wzdłuż taśmy przenośnika mijając detektor pudełek. Delta jest obliczana ze wzoru (4).

delta = l/v (4)

gdzie l to długość pudełka, v — prędkość przenośnika. Wartość delta pozwala wykluczyć fałszywe wyzwolenia detektora, gdy jedno pudełko przesuwa się, a jednocześnie zasygnalizować początek nowego pudełka, jeśli przesuwają się jedno po drugim bez rozdzielenia. Przypisz każdemu pudełku liczbę porządkową i (i = 1 . . .), a dla każdego i-tego pudełka zapamiętamy t0i — czas przejścia każdego pudełka przez detektor, na podstawie którego możemy obliczyć położenie pudełka w chwili t.

4 Algorytm wyrównywania półproduktów

Na podstawie matematycznego opisu procesu wyrównywania półproduktów opracowano algorytm, którego opis podano w języku naturalnym. Algorytm podzielony jest na 2 bloki, działające równolegle. Pierwszy blok odpowiada za przetwarzanie sygnałów z detektora pudełek i rejestrowanie czasu t_{0i} w kolejce pierścieniowej: W przypadku odebrania sygnału z detektora pudełek, wykonaj następujące czynności. Jeśli kolejka jest pusta lub różnica między bieżącym czasem t a ostatnim elementem kolejki jest większa lub równa delcie, dodaj bieżący czas t do końca kolejki. Pusta kolejka oznacza, że żaden moduł nie przeszedł jeszcze przez detektor. Różnica czasu musi zostać określona, aby rozdzielić 2 moduły, które następują po sobie. Drugi blok odpowiada za obliczenie pozycji modułu, którego czas t_0 jest rejestrowany na początku kolejki, oraz za wysłanie poleceń sterujących.

1. Ustaw sygnał „Operacja dozwolona” na FAŁSZ.

2. Jeśli sygnał „Operacja dozwolona” jest ustawiony na FAŁSZ, przejdź do kroku 3, w przeciwnym razie przejdź do kroku 5.

3. Jeśli rozmiar kolejki jest większy od 0, oblicz bieżącą pozycję pola zgodnie ze wzorem (2), przejdź do kroku 4, w przeciwnym razie przejdź do kroku 2.

4. Jeśli bieżąca pozycja pola znajduje się w obszarze roboczym robota (wzór (3)), ustaw sygnał „Operacja dozwolona” na PRAWDA i usuń pierwszy element z kolejki, przejdź do kroku 2.

5. Wykonaj sekwencyjny ruch z punktu M_1 do M_2, następnie z M_2 do M_3. Następnie z M_3 do M_1. Ustaw sygnał „Operacja dozwolona” na FAŁSZ. Przejdź do kroku 2.

Przedstawiony algorytm został zaimplementowany w standardowym języku programowania C++ 11 i zintegrowany z istniejącym rozwiązaniem programowym jako węzeł w platformie programowej Robot Operating System (ROS) . Ten węzeł komunikuje się z węzłami solwera kinematyki prostej i odwrotnej za pośrednictwem tematów i informuje, że działa w czasie rzeczywistym. W celu zwiększenia wydajności zaleca się użycie poprawki jądra Linux do pracy w czasie rzeczywistym, jak w tym przypadku . Połączenie TCP/IP przez Ethernet jest wykorzystywane w robotyce i jest dość wydajne . Mikrokontroler pokładowy wykorzystuje system operacyjny FreeRTOS, który działa w czasie rzeczywistym.

5. Oszacowanie błędów

Szczególnie ważne jest oszacowanie błędów podczas pracy nad wyrównywaniem pudełek z bryłkami, ponieważ rozważamy odcinek linii produkcyjnej o dużej prędkości, gdzie błędy mogą prowadzić do uszkodzenia półproduktów i opakowań. Dopuszczalny błąd pozycjonowania elementu roboczego wynosi e = 0,02 m. Jest to różnica między długością i szerokością pudełka a podobnymi parametrami dolnej części elementu roboczego. Przy prędkości przenośnika v = 0,55 m/s błąd w czasie nie może być większy niż terr = 36 ms. Obliczenia wykonano według wzoru (5).

$t_{err} = e/v$ (5)

Tabela 2 przedstawia główne opóźnienia, które mogą wpływać na błąd pozycjonowania.

Źródło opóźnienia: Wartość, ms

Odczyt detektora przez mikrokontroler pokładowy: 0,333

Transfer danych między mikrokontrolerem pokładowym a komputerem sterującym robota: Mniej niż 10

Transfer danych w ramach platformy oprogramowania ROS między węzłami systemu: Mniej niż 3

Częstotliwość sprawdzania danych przez opracowany algorytm: 10

Zatem w najgorszym przypadku opóźnienie nie będzie większe niż 23 ms, co spełnia warunki.

6 Testy

Do testów wykorzystano trzy pudełka z bryłkami, z których każda wystaje poza opakowanie. Głównym warunkiem wyrównywania jakości było takie wypoziomowanie bryłek, aby nie wystawały poza opakowanie. Oceniono również integralność opakowania i półproduktów, przy czym należy wziąć pod uwagę, że wyrównywanie musi być wykonane dokładnie raz w warunkach produkcyjnych, a w warunkach eksperymentalnych kilkakrotnie. Pudełka z bryłkami umieszczono na taśmie przed detektorem w różnych kombinacjach:

1. Jedno pudełko.

2. Dwa pudełka ustawione w poprzek szczeliny.

3. Dwa pudełka ustawione obok siebie.

4. Trzy pudełka ustawione obok siebie.

Przypadek z trzema pudełkami ustawionymi jeden za drugim uznano za najbardziej „dokładny" i najtrudniejszy do wykonania. Przeprowadzono dwadzieścia eksperymentów. Pudełka podawano kolejno na przenośnik, tak aby nie było między nimi żadnych przerw.

Następnie pudełka poddano działaniu czynników zewnętrznych i sprawdzono wynik.

Wszystkie pudełka z nuggetsami zostały prawidłowo wyrównane. Zaobserwowano uszkodzenia kartonu i niektórych nuggetsów. Nie zaobserwowano uszkodzeń opakowania polietylenowego.

7 Wnioski

W toku badań osiągnięto wszystkie założone cele. Cel badań został osiągnięty. Opracowany algorytm wykazał skuteczność w wyrównywaniu skrzynek z bryłkami, może być stosowany w przemyśle po dłuższych testach w kompleksie z manipulatorem robota i korpusem roboczym. Jednocześnie zidentyfikowano wady konstrukcji manipulatora robota:

1. Montaż sztywniejszego połączenia z taśmociągiem.

2. Zmniejszenie błędu pozycjonowania poprzez przeniesienie algorytmu z komputera sterującego do mikrokontrolera pokładowego, który jest systemem czasu rzeczywistego.

3. Dodanie możliwości szybkiej kalibracji systemu we własnym zakresie.

Takie udoskonalenia pozwolą na bardziej efektywne i tańsze wykorzystanie algorytmu wyrównywania i całego systemu automatyki.

Komputerowy system szkoleniowy do planowania wieloasortymentowej produkcji ciągłej dyskretnej

Przedstawiono specjalistyczny system komputerowy do szkolenia personelu produkcyjnego wieloasortymentowych przedsiębiorstw przemysłowych w zakresie umiejętności rozwiązywania problemu optymalnego planowania produkcji. Zadanie to charakteryzuje się koniecznością uwzględnienia szerokiego asortymentu produktów oraz różnorodności typów i konfiguracji

urządzeń produkcyjnych, które można rekonfigurować do różnych typów zleceń. Aby skutecznie rozwiązać problem szkoleniowy, opracowano strukturę funkcjonalną symulatora, który jest elastycznym systemem komputerowym skonfigurowanym pod konkretną produkcję. Struktura obejmuje ergonomiczne interfejsy uczestnika szkolenia i instruktora, wsparcie informacyjne: bibliotekę modeli matematycznych funkcji docelowych i bazę danych reguł ich tworzenia, bazę danych zamówień, urządzeń i klientów, a także implementację programową różnych nowoczesnych metod optymalizacji. Na podstawie wsparcia informacyjnego i matematycznego opracowano scenariusz i protokół szkoleniowy, w których modelowane są różne sytuacje planowania produkcji. Skuteczność symulatora komputerowego potwierdzają wyniki testów w rosyjskich i zagranicznych przedsiębiorstwach przemysłowych, a także jego wdrożenie do procesu edukacyjnego. Słowa kluczowe Symulatory komputerowe · Szkolenie personelu · Planowanie kalendarzowe · Harmonogramy produkcji · Produkcja dyskretna i ciągła

1 Wstęp

Efektywność funkcjonowania nowoczesnych przedsiębiorstw wieloasortymentowych zależy bezpośrednio od skuteczności procesu planowania produkcji i sporządzania harmonogramów. Zadanie optymalizacji planowania jest istotne dla różnorodnych innowacyjnych przedsiębiorstw cywilnych i przemysłowych, produkujących produkty wieloasortymentowe na sprzęcie high-tech. Szczególnie pilne jest zadanie efektywnego szkolenia personelu produkcyjnego w procesie planowania .Szkolenie w zakresie znajdowania optymalnego harmonogramu produkcji jest zadaniem dość trudnym ze względu na konieczność uwzględnienia wielu cech procesu produkcyjnego podczas rozwiązywania: rodzaju i konfiguracji urządzeń produkcyjnych, parametrów produktów wieloasortymentowych, a także wymagań dotyczących jakości i terminowości realizacji zamówień. Ze względu na duży rozmiar problemu, aby skutecznie znaleźć optymalny harmonogram, konieczne jest opracowanie implementacji programowych nowoczesnych metod optymalizacji, które umożliwiają wyszukiwanie rozwiązań w dużej przestrzeni rozwiązań akceptowalnych w akceptowalnym czasie maszynowym [6]. W związku z tym dynamicznie rozwijająca się digitalizacja produkcji wymaga opracowania specjalistycznych symulatorów, które umożliwiają personelowi trenowanie scenariuszy i algorytmów tworzenia optymalnych harmonogramów produkcji w oparciu o nowoczesne metody optymalizacji, uwzględniające szereg ograniczeń technologicznych. W związku z tym wskazane jest opracowanie komputerowego systemu szkoleniowego, który będzie operacyjnie skonfigurowany pod kątem różnych charakterystyk procesu produkcyjnego każdego przedsiębiorstwa, będzie posiadał inteligentne interfejsy zapewniające zmienność wykorzystania funkcjonalności pakietu oprogramowania w różnych sytuacjach planistycznych i zapewni efektywne rozwiązanie problemów planistycznych o różnych wymiarach poprzez zastosowanie nowoczesnych metod optymalizacji. Wprowadzenie takiego komputerowego systemu szkoleniowego do systemu zarządzania przedsiębiorstwem jest konieczne ze względu na wysokie wymagania dotyczące kwalifikacji personelu produkcyjnego oraz znaczenie wyników planowania dla całego procesu technologicznego.

2 Informacyjny i matematyczny opis procesu planowania produkcji

2.1 Sformalizowany opis obiektu badań

Pierwszym etapem wdrażania procesu edukacyjnego jest opracowanie sformalizowanego opisu planowania produkcji przemysłowej w sposób ciągły dyskretny jako obiektu badań. Podano wektor danych wejściowych $X = (O, E, Pd)$, składający się z następujących parametrów:

$$\tilde{O} = \left\{ \underline{O_i} : \underline{O} = \left(N^O, K^O, P^O, Q^m, \tilde{R}, \tilde{G}, Cs \right), i = \underline{1, N} \right\}$$

-wektor charakteryzujący zamówienia: N^O - numer zamówienia; K^O - kod zamówienia; P^O - rodzaj produktów zawartych w zamówieniu; Q^m - ilość materiału; $\tilde{R}$, - wektor charakteryzujący formulację materiału; $\tilde{G} = \{W^m, Th^m\}$ — wektor charakteryzujący geometryczne cechy materiału, takie jak W^m — szerokość materiału, Th^m — grubość materiału; Cs — klient; i — numer kolejnego zamówienia; N — liczba zamówień.

$$\tilde{E} = \left\{ \underline{E_j} : \underline{E} = \left(N^E, K^E, V^P, \underline{To^E}, \underline{C_E} \right), j = \underline{1, Me} \right\}$$

-wektor charakteryzujący linie produkcyjne (agregaty, gniazda robocze): N^E - nazwa urządzenia; K^{Rc} - kod urządzenia; V^P - wydajność;

$\underline{To^E} = \left\{ \underline{To_r^E}, r = \underline{1, R} \right\}$ -operacje technologiczne, które można wykonać na tym sprzęcie; C_E — wektor opisujący ustawienia sprzętowe; j — numer seryjny linii produkcyjnej (jednostki, stanowiska roboczego); Me — liczba linii produkcyjnych (jednostek, stanowisk roboczych). Aby sformalizować opisane powyżej parametry w aplikacji do rzeczywistych produkcji przemysłowych i w celu dostosowania operacyjnego do nowego typu przedsiębiorstwa, reprezentacja wiedzy jest wykorzystywana w postaci map inteligencji, które odzwierciedlają wiedzę o sprzęcie produkcyjnym i produktach objętych planowaniem, ich cechach i hierarchicznych relacjach między nimi. Karty inteligencji są również wykorzystywane w rozwoju rozszerzonych baz danych jako element wsparcia informacyjnego systemu komputerowego.

$\widetilde{Pd} = (\tau^o, Y^{cr})$ - wektor charakteryzujący parametry planowania: $\tau^o = [\tau^b, \tau^e]$ - okres planowania;

τ^b - data rozpoczęcia planowania; τ^e - data zakończenia planowania; Y^{cr} - kryteria optymalizacji. Wektor zmiennych parametrów to

$$Q = \left\{ Q : Q = (j, k, \tau_{oi}, \tau_i), i = \underline{1, N}, j = \underline{1, Me}, k = \underline{1, L_j}, L \in \tilde{N} \right\}$$

wektor opisujący rozkład i zamówień na j liniach produkcyjnych (agregaty, stanowiska robocze), k = 1, L_j - numer sekwencji wykonania i-tego zamówienia na j-tej linii produkcyjnej (jednostce, stanowisku roboczym) w bieżącym harmonogramie Q, L_j — liczba zamówień wykonanych na j-tej linii produkcyjnej (jednostce, stanowisku roboczym) w harmonogramie Q. Wynikiem planowania (wektor Y) będzie taki wektor rozkładu zamówień na liniach produkcyjnych (agregaty, stanowiska robocze)

$$Qopt = \left\{ Q_i : Q = (j, k, \tau_{oi}, \tau_i)\, j = \underline{1, Me}, k = \underline{1, L_j}, L_j \subset N, i = \underline{1, N} \right\}$$

która wartość funkcji celu dla każdej linii produkcyjnej (jednostki, stanowiska roboczego) i całego cyklu produkcyjnego będzie optymalna.

2.2 Wyznaczenie zadania optymalizacji planowania produkcji

Należy znaleźć taką wartość wektora zmiennych parametrów Q^{opt} =

$$\left\{ Q_i : Q = (j, k, \tau_{oi}, \tau_i)\, j = \underline{1, Me}, k = 1, L_j, L_j \subset N, i = \underline{1, N} \right\}$$

, czyli rozmieszczenie i kolejność realizacji N zamówień na Me liniach produkcyjnych w okresie planowania $[\tau^b, \tau^e]$, co zapewni optymalną wartość funkcji celu. W przypadku rozwiązania

problemu dla jednej linii produkcyjnej (jednostki, stanowiska roboczego) łączny czas realizacji zamówienia i konfiguracji urządzeń jest minimalizowany:

$$F_j(Q^{opt}) = \min_{Q \in D \subset R^{|Q|}} \left(\sum_{k=1}^{L_j - 1} \omega^{Ch}_{O_{j,k},\, O_{j,k+1}} * \tau^{ch}(O_{j,k},\, O_{j,k+1}) + \sum_{k=1}^{L_j} \omega^{O}_{k} * \tau^{O}_{j,k} \right) \quad (1)$$

W celu zapewnienia minimalnego cyklu produkcyjnego zleceń produkcyjnych na kilku liniach produkcyjnych (agregatach, gniazdach roboczych) minimalizuje się czas realizacji zleceń na linii produkcyjnej (agregacie, gnieździe roboczym), dla której jest on maksymalny:

$$F_C(Q^{opt}) = \min_{Q \in D \subset R^{|Q|}} (F_j(Q)) \quad (2)$$

gdzie $\tau^{ch}(O_{j,k}, O_{j,k+1})$ -czas przezbrajania urządzeń od wykonania zlecenia k do następnego zlecenia

o numerze k+1, τ^O_k -czas wykonania k-tego zlecenia, ω_k -współczynnik wagowy k-tego zlecenia, $R^{|Q|}$ —przestrzeń poszukiwań, D —zbiór dopuszczalnych wartości wektora Q. Rozwiązaniem zadania optymalizacyjnego jest taki wektor rozkładu zleceń wzdłuż linii produkcyjnych (agregatów, gniazd roboczych) w przedziale planowania Q^{opt} =

$$\left\{ Q_i : Q = (j,\, k,\, \tau_{oi},\, \tau_i)\; j = \underline{1,\, Me},\; k = \underline{1,\, L_j},\; L_j \subset N,\; i = \underline{1,\, N} \right\}$$

, że wartość funkcji celu dla każdej linii produkcyjnej (jednostki, gniazda roboczego) $F_j(Q^{opt})$ i całkowity czas cyklu produkcyjnego $F_C(Q^{opt})$ będą optymalne, co zapewni równomierne obciążenie urządzeń produkcyjnych. Informacje eksperckie uzyskane od specjalistów przedsiębiorstwa są formalizowane w procesie wypełniania specjalnych tablic pojęciowych i stanowią podstawę do tworzenia bazy reguł rekonfiguracji urządzeń z produkcji jednego typu na inny. Reguły ustawień urządzeń są przedstawiane w postaci modelu produkcyjnego i mogą być wykorzystane do prawidłowego tworzenia typu funkcji celu. Przy obliczaniu wartości funkcji celu całkowity czas rekonfiguracji opiera się na stwierdzeniach implikatywnych w postaci „IF [parametr i–1-tego rzędu P^O = . . .] AND [parametr i-tego rzędu P^O = . . .] AND [parametr linii j-tej P^E = . . .], THEN czas rekonfiguracji τ^{Ch}_m = . . .”.

2.3 Metody rozwiązywania problemu optymalnego planowania produkcji

Aby skutecznie poszukiwać optymalnego planu produkcji w systemie komputerowym, stosuje się różne metody optymalizacji dla zadań o różnej skali, ponieważ wraz ze wzrostem liczby zleceń, jednostek wyposażenia i liczby możliwych rekonfiguracji w produkcji wyrobów wieloasortymentowych, liczba wariantów planu (zakres dopuszczalnych wartości) rośnie wykładniczo. Aby podjąć decyzję o wyborze optymalnej metody planowania, oblicza się łączną liczbę wariantów planu dla danej liczby zleceń, jednostek wyposażenia i ich konfiguracji. Następnie, na podstawie liczby znalezionych wariantów, podejmuje się decyzję o stosowalności metody pełnego poszukiwania, która gwarantuje znalezienie optymalnego rozwiązania, a jeśli dopuszczalny wymiar problemu zostanie przekroczony – $C_{var} > 10^6$ – problem planowania rozwiązuje się za pomocą bardziej złożonej metody optymalizacji – algorytmu genetycznego.

Genetyczny algorytm optymalizacji ewolucyjnej w zastosowaniu do rozpatrywanego problemu ma szereg zalet, które zdecydowały o jego zastosowaniu i badaniu. Dzięki połączeniu zalet metod gradientowych i losowych (z wykorzystaniem operatorów selekcji, krzyżowania i mutacji), poszukiwanie wartości optymalnej jest kierunkowe, ale odbywa się w szerokim zakresie możliwych wartości . Rozpoczęcie poszukiwań z różnych punktów startowych (utworzenie początkowej populacji rozwiązań o różnej wielkości), a także zastosowanie operatora mutacji i połączenie suboptymalnych rozwiązań z różnych populacji pozwalają uniknąć popadania w ekstrema lokalne i zapewnić znalezienie rozwiązania bliskiego ekstremum globalnego . Parametry wejściowe problemu optymalnego planowania są przekształcane w elementy algorytmu genetycznego. Porównując dwa harmonogramy, porównuje się wartości funkcji dopasowania (w zależności od kryterium optymalizacji) i oblicza składowe funkcji celu – czas produkcji zamówienia i czas rekonfiguracji sprzętu. Po wystąpieniu kryterium zatrzymania algorytm kończy swoją pracę, a użytkownikowi dostarczany jest otrzymany plan produkcji/produkcji z najlepszą wartością funkcji dopasowania w ostatniej populacji.

3. Struktura komputerowego systemu szkoleniowego w zakresie planowania produkcji

3.1 Struktura funkcjonalna systemu komputerowego

Na podstawie przedstawionego powyżej sformalizowanego opisu matematycznego procesu planowania opracowano strukturę funkcjonalną komputerowego systemu szkoleniowego symulatora. Struktura obejmuje cztery główne interfejsy: inżyniera wiedzy, administratora bazy danych, instruktora i uczestnika szkolenia. Interfejs administratora bazy danych umożliwia aktualizację wsparcia informacyjnego pakietu oprogramowania. Wsparcie informacyjne obejmuje bazy danych sprzętu i produktów; bazę reguł rekonfiguracji sprzętu; biblioteki modeli matematycznych funkcji celu, ograniczeń kryterialnych i metod optymalizacji, a także bazę teoretycznych materiałów edukacyjnych. Wsparcie informacyjne jest konfigurowane dla różnych obiektów badań i planowania za pomocą specjalnego interfejsu inżyniera wiedzy, który umożliwia również integrację z systemami automatycznego sterowania dla innowacyjnych produkcji ciągło - dyskretnych. Interfejs instruktora umożliwia formułowanie zadań, monitorowanie postępów i wyników ich realizacji, opracowywanie scenariuszy edukacyjnych, przeprowadzanie testów i tworzenie protokołu rezultatów uczenia się. Interfejs uczestnika szkolenia zawiera moduły, które umożliwiają zapoznanie się z wiedzą ekspercką na temat obiektu planowania, zapoznanie się z procedurą importu danych, stworzenie sytuacji planistycznej (wybór zamówień i sprzętu do planowania, a także ustalenie terminów budowy planu), zdawanie testów i zapoznanie się z materiałami referencyjnymi. Otwarta architektura systemu komputerowego pozwala na dołączanie nowych modułów programowych, co pozwala na rozszerzenie funkcjonalności symulatora. Rozwój systemu komputerowego został zrealizowany z wykorzystaniem obiektowego języka programowania C# w środowisku Microsoft Visual Studio na platformie APS.NET 2.0. Bazę danych opracowano z wykorzystaniem systemu sterowania relacyjną bazą danych SQLite.

3.2 Proces nauki planowania produkcji

Proces szkolenia z zakresu planowania produkcji realizowany jest zgodnie z opracowanymi scenariuszami i metodami. Interfejs instruktora umożliwia realizację różnych opcji szkolenia: wydawanie rekomendacji, udzielanie odpowiedzi na pytania studenta, przeprowadzanie testów (zgodnie z wybranym modelem sterowania wiedzą), tryb egzaminacyjny. System komputerowy zapewnia studentowi następujące formy opanowania materiału: zapoznanie się z materiałem teoretycznym (przeglądanie systemu informacji i odniesień), zaliczenie testów, konfiguracja systemu dla konkretnego sprzętu, zapoznanie się z procesem tworzenia harmonogramów

kalendarzowych w trybie szkolenia samodzielnego (swobodnego). Procedury te stanowią podstawę kształtowania kompetencji zawodowych i umiejętności efektywnego harmonogramowania produkcji. Zadania praktyczne do rozwiązania w systemie komputerowym sporządzono na podstawie danych dostarczonych przez firmy produkujące materiały polimerowe w Rosji i Niemczech, a także przez rosyjski przemysł metalurgiczny (Czelabińskie Zakłady Konstrukcji Metalowych) [1, 20]. Na podstawie wyników badania sporządzono protokół szkolenia, który zawiera dane uczestnika szkolenia, datę i godzinę szkolenia, scenariusz oraz nazwę symulowanych sytuacji. Protokół jest do dyspozycji instruktora i nauczyciela i stanowi dokument oceny kwalifikacji studenta. Rozważmy sytuację planowania opartą na danych producenta wielowarstwowych folii polimerowych „Maria Soell HTF" (Niemcy): konieczne jest znalezienie optymalnej kolejności produkcji 58 różnych zamówień asortymentowych, planowanie odbywa się przez 1,5 miesiąca, liczba różnych kombinacji parametrów produktu przekracza 200. Po rozwiązaniu problemu za pomocą interfejsów systemu komputerowego, stażysta musi porównać uzyskany plan z planem utworzonym w produkcji. Na rys. 2 przedstawiono wizualizację skonstruowanego optymalnego planu w wyniku opanowania kompetencji planowania produkcji. Na diagramie widoczna jest kolejność produkcji zamówień, data rozpoczęcia i zakończenia produkcji każdego zamówienia, czas jego wytworzenia oraz czas rekonfiguracji z poprzedniego zamówienia na bieżące. Skuteczność symulatora komputerowego potwierdzają wyniki szkolenia studentów i personelu produkcyjnego w rosyjskich i zagranicznych przedsiębiorstwach przemysłowych, a także wprowadzenie do procesu edukacyjnego.

4. Wnioski

Opracowano symulator komputerowy i system szkoleniowy, aby przeszkolić kadrę produkcyjną wieloasortymentowych przedsiębiorstw przemysłowych w zakresie umiejętności rozwiązywania problemu optymalnego planowania produkcji. Wprowadzenie opracowanego systemu komputerowego do rozwoju wyników nauczania opartego na kompetencjach pozwala znacząco podnieść poziom zawodowy kadry produkcyjnej i studentów w zakresie badania zadań optymalnego planowania produkcji. Opracowany system komputerowy został pomyślnie przetestowany w rosyjskich i zagranicznych przedsiębiorstwach przemysłowych, wdrożony do procesu edukacyjnego i może być wykorzystany do doskonalenia umiejętności kadry kierowniczej i produkcyjnej w innowacyjnych branżach.

Inteligentny System Szkoleniowy do Kontroli Procesów Wytłaczania Polimerów w Sytuacjach Awaryjnych

Opracowano inteligentny system szkoleniowy do Kontroli Procesów Wytłaczania w Produkcjach Wieloasortymentowych Folii Polimerowych w Sytuacjach Awaryjnych. Wytłaczanie jest jednym z głównych etapów produkcji, który decyduje o wydajności systemu oraz o parametrach użytkowych produktu (liczbie różnego rodzaju wad powierzchniowych, kolorze). Parametry te zależą od stanu termicznego, jednorodności i koloru wytłoczyny. Dlatego sytuacje awaryjne związane z odchyleniami w wydajności wytłaczarki i wskaźnikach jakości wytłoczyny, wynikającymi z ograniczeń regulacyjnych, prowadzą do spadku wydajności i wzrostu liczby wad folii. System szkoleniowy zawiera bazę wiedzy o sytuacjach awaryjnych, ich przyczynach i zaleceniach dotyczących ich eliminacji, bibliotekę modeli matematycznych do obliczania

wskaźników wydajności i jakości w zależności od działań kontrolnych, bazy danych typów folii, właściwości polimerów, parametrów geometrycznych wytłaczarek i parametrów technologicznych wytłaczania, moduł do tworzenia scenariusza szkolenia, moduł do obsługi sytuacji awaryjnych i tworzenia wskazówek sterujących, moduł do tworzenia protokołu szkolenia, interfejsy uczestnika szkolenia (operatora wytłaczarki) i instruktora. Biblioteka modeli matematycznych pozwala na dostosowanie systemu do szkolenia z zakresu sterowania wytłaczarkami różnych typów. Scenariusz szkolenia ustalony przez instruktora obejmuje rodzaj folii, wydajność systemu, markę wytłaczarki, charakterystykę symulowanej sytuacji awaryjnej. Protokół szkolenia rejestruje działania uczestnika szkolenia oraz osiągnięte wartości parametrów wyjściowych. Testy przeprowadzone w procesach wytłaczania w procesie kalandrowania folii farmaceutycznych i spożywczych w fabrykach w Rosji i Niemczech potwierdziły sprawność działania systemu komputerowego.

Słowa kluczowe System inteligentny · Baza wiedzy ekspertów · Model ramowy produkcji · Bazy danych surowców · Sprzęt · Produkty · Modele matematyczne · Szkolenia z zakresu sterowania · Wytłaczanie · Sytuacje awaryjne · Wady produktów · Produkcja folii polimerowych

1 Wstęp

Współczesna produkcja folii polimerowych do pakowania produktów farmaceutycznych i spożywczych odbywa się w ciągłych, wieloetapowych, wielkotonażowych (wydajność do 3000 kg/h) systemach technologicznych . Złożoność tych systemów wynika z obecności surowców wtórnych, które są niezbędne do zawrócenia odpadów (obciętych i rozdrobnionych krawędzi folii) do produkcji jako surowców wtórnych . Produkcja folii polimerowych charakteryzuje się różnorodnością urządzeń, szerokim asortymentem produktów (ponad 60 rodzajów folii) i w rezultacie, częstymi przezbrojeniami w celu wytworzenia nowego asortymentu produktów (nawet kilka razy na zmianę). Właściwości użytkowe folii zależą w złożony sposób od wskaźników jakościowych materiału polimerowego na etapach produkcji (jakość półproduktów często nie jest monitorowana), charakterystyki surowców i aparatów oraz parametrów technologicznych etapów. Liczba takich wiązań w kalandrowanej produkcji folii polimerowych przekracza 800. Wygląd i kolor charakteryzują główne walory folii jako materiałów opakowaniowych dla konsumentów. Różnego rodzaju wady powierzchni folii kształtują jej wygląd. Defekty spowodowane przegrzaniem i zniszczeniem polimeru (czarne punkty, destrukcyjne paski), niepełnym uplastycznieniem polimeru w masie (niestopione cząstki polimeru), niedostatecznym wymieszaniem stopu polimeru (wady typu „rybie oko") występują najczęściej i negatywnie wpływają na przydatność folii jako opakowania. Wady koloru folii obejmują niedopasowanie współrzędnych kolorów do wartości odniesienia i nierównomierne zabarwienie (niedostateczne wymieszanie stopu jest również przyczyną nierównomiernego zabarwienia). Proces wytłaczania odgrywa kluczową rolę w zapewnieniu wymaganych przez konsumenta cech folii (ograniczona liczba wad powierzchni, zgodność koloru z normą, równomierne zabarwienie). Jest ona przeznaczona do wytwarzania jednorodnej lepkiej masy (wytłoczyny) poprzez stopienie fazy stałej poruszającej się w aparacie ślimakowym (wytłaczarce), ogrzanie, wymieszanie i przepchnięcie stopu przez głowicę formującą wytłaczarki. Następnie (na kalandrze) wytłoczyna jest formowana w folię. Odchylenia reżimu wytłaczania od przepisów przyczyniają się do naruszenia warunków hydrodynamicznych, termicznych i mechanizmu przemiany fazowej, powstawania stref stagnacji w wytłaczarce. W rezultacie jakość wytłoczyny znacznie się pogarsza, co uniemożliwia jej dalsze formowanie. Na przykład, jedna z głównych wad wytłoczyny — destrukcja termiczna (destrukcja łańcuchów makrocząsteczek polimeru [5]) — jest konsekwencją nadmiernego (z powodu stref stagnacji) czasu przebywania polimeru w wytłaczarce w wysokiej temperaturze. Degradacja

termiczna zmienia kolor ekstrudatu (od żółknięcia do ciemnienia), zwiększa jego pełzanie i przyleganie do przenośnika między ekstruderem a kalandrem. Odpady zwrotne podawane do ekstrudera przyczyniają się do zwiększenia odchylenia koloru ekstrudatu od standardu w produkcji kolorowych folii polimerowych. Wada ekstrudatu nie jest eliminowana na kolejnych etapach produkcji (formowanie ekstrudatu, utrwalanie struktury folii) i prowadzi do pojawienia się odpowiednich defektów na powierzchni folii. Do 20% defektów folii polimerowych jest spowodowanych naruszeniem reżimu wytłaczania. Brak synchronizacji ekstrudera i kalandra może prowadzić nie tylko do spadku wydajności produktów wysokiej jakości, ale także do zatrzymania systemu produkcyjnego . Przy nadmiernej przepustowości ekstrudera ekstrudat gromadzi się w szczelinie zasilającej kalandra, przegrzewa się z powodu długiego (~5 min) przebywania w materiale i zapada się pod wpływem wysokiej temperatury. W przypadku „głodu" kalandra, w szczelinie podającej kalandra nie formuje się materiał wytłaczany, co prowadzi do sytuacji awaryjnej (kolizji rolek podających kalandra). Aby zapewnić brak wad i wypadków, operator musi zmniejszyć (zwiększyć) przepustowość wytłaczarki, opierając się na wizualnej ocenie poziomu materiału wytłaczanego. W produkcji folii kolorowych (z wykorzystaniem metody barwienia stopu polimeru ciekłymi barwnikami dozowanymi do wytłaczarki) taka regulacja przepustowości jest zaburzeniem prowadzącym do odchylenia koloru wytłaczanego materiału od standardowego. Wymaga to odpowiedniej zmiany w działaniach kontrolnych, tj. natężenia przepływu barwników. Jeśli właściwości użytkowe folii odbiegają od wartości granicznych określonych przepisami, folia nie może być używana jako opakowanie (ponieważ nie zapewnia bezpieczeństwa i atrakcyjności produktów dla konsumentów) i jest utylizowana jako odpad niepodlegający zwrotowi. Wzrost ilości wadliwych folii prowadzi do spadku konkurencyjności produkcji i wzrostu jej zagrożenia dla środowiska. Dlatego też, terminowe wykrywanie i korygowanie sytuacji awaryjnych związanych z odchyleniami w przepustowości wytłaczarki, wskaźnikami jakości wytłaczanego materiału poza progowe wartości regulacyjne, pomaga zapobiegać pojawianiu się istotnych wad folii (zmniejsza ilość odpadów nie podlegających zwrotowi) i zwiększa wydajność produktów wysokiej jakości. Jednakże złożoność sprzętowa i technologiczna projektu procesu wytłaczania , brak systemu monitorowania przepustowości wytłaczarki i jakości wytłaczanego materiału (z reguły jakość ocenia się wizualnie na podstawie wyglądu wytłaczanego materiału), wiele przyczyn sytuacji awaryjnych znacznie komplikuje sterowanie wytłaczaniem w sytuacjach awaryjnych. Prowadzi to do podejmowania przez operatorów błędnych decyzji sterujących. W związku z tym konieczne jest zapewnienie personelowi operacyjnemu możliwości wcześniejszego, za pomocą komputerowego systemu szkoleniowego, wypracowania umiejętności sterowania wytłaczaniem w sytuacjach awaryjnych dla różnych rodzajów folii polimerowych oraz opcji projektowania sprzętowego i technologicznego procesu. Analiza systemów komputerowych do szkolenia personelu produkcyjnego w zaawansowanych technologicznie zakładach przemysłu chemicznego, rafineryjnego, hutniczego i energetycznego wykazała, że najskuteczniejszym sposobem rozwiązania problemów szkoleniowych w zakresie sterowania złożonymi procesami technologicznymi w różnych trybach ich funkcjonowania (dostosowanie do nowego zadania, trybu rutynowego, sytuacji awaryjnych) jest opracowanie i wdrożenie inteligentnych systemów szkoleniowych . Systemy te oparte są na wielowymiarowych modelach opisu obiektów sterowania, dostosowanych do ich zmiennych charakterystyk. Modele szkoleniowe (symulacyjne) obejmują:

• modele reprezentacji nieformalnej wiedzy eksperckiej do badania sposobów sterowania procesami technologicznymi, sposobów eliminacji sytuacji awaryjnych i kształtowania intelektualnych wskazówek dotyczących sterowania;

• modele matematyczne umożliwiające obliczanie wskaźników efektywności procesów (przepustowości, zużycia energii, wskaźników jakości półproduktów i produktów) w zależności od działań kontrolnych i zapewniające możliwości aktywnego szkolenia;

• modele informacyjne surowców, urządzeń, trybów technologicznych, produktów, zapewniające dostosowanie systemów szkoleniowych do różnych modyfikacji sprzętowych i technologicznych w projektowaniu procesów.

Zatem, aby utrzymać na odpowiednim poziomie i doskonalić kompetencje zawodowe operatorów procesów wytłaczania w produkcji folii polimerowych, istotne jest opracowanie elastycznego, inteligentnego systemu szkoleniowego. System taki pozwala (w oparciu o model reprezentacji wiedzy eksperckiej o sytuacjach awaryjnych procesu wytłaczania oraz model matematyczny obliczania przepustowości wytłaczarki i wskaźników jakości wytłoczyny) kształtować umiejętności sterowania wytłaczaniem w sytuacjach awaryjnych dla różnych scenariuszy szkoleniowych (rodzaje folii, marki wytłaczarek).

2. Sformułowanie problemu szkolenia z zakresu sterowania wytłaczaniem

Analiza procesu wytłaczania polimerów jako obiektu badań i sterowania pozwoliła na sformułowanie jego opisu informacyjnego w postaci zbioru wektorów parametrów wejściowych X, akcji sterujących U, zakłóceń F oraz parametrów wyjściowych Y (parametry stanu S i wskaźniki wydajności Q):

$$
\begin{aligned}
&Y = F(X,\ U,\ F),\\
&Y = \{S,\ Q\},\\
&S = \{v_{\mathrm{x}},\ v_{\mathrm{z}},\ P,\ T,\ \eta,\ \tau_{\mathrm{av}}\},\\
&Q = \{G,\ Q_{\mathrm{ext}}\},\\
&Q_{\mathrm{ext}} = \{I_{\mathrm{d}},\ \varphi_{\mathrm{s}},\ \gamma_{\mathrm{av}},\ L_{\mathrm{ext}},\ a_{\mathrm{ext}},\ b_{\mathrm{ext}},\ \Delta E_{\mathrm{ext}}\},\\
&X = \left\{T_{\mathrm{film}},\ H_{\mathrm{polym}},\ M_{\mathrm{extrud}},\ \Gamma_{\mathrm{extrud}}\right\},\\
&M_{\mathrm{extrud}} = \left\{T_{\mathrm{extrud}},\ D,\ L/D\right\},\\
&\Gamma_{\mathrm{extrud}} = \{n_{\mathrm{scr}},\ \Gamma_{\mathrm{scr}},\ \Gamma_{\mathrm{die}},\ S_0,\ A_{\mathrm{scr}}\},\\
&U = \{N,\ T_{bk},\ k = 1 \ldots n_{\mathrm{T}},\ G_{ci},\ i = 1 \ldots n_{\mathrm{c}}\},\\
&F = \{G_{\mathrm{wast}}\}.
\end{aligned}
$$

Do produkcji folii polimerowych charakteryzujących się typami folii T stosuje się wytłaczarki różnych typów wytłaczarki T (wytłaczarki jednoślimakowe, posuwisto-zwrotne, dwuślimakowe). Wytłaczarki różnych typów różnią się liczbą ślimaków nscr oraz charakterem ich ruchu (obrotowy lub jednoczesny obrót i ruch posuwisto-zwrotny osiowy). Wytłaczarki tego samego typu różnią się markami Mextrud, charakteryzującymi się średnicą D oraz względną długością L/D ślimaka/ślimaków. Parametry geometryczne wytłaczarki tej marki obejmują parametry geometryczne ślimaka/ślimaków Γ_{scr} i głowicy formującej (matrycy) Γ, a także parametry charakteryzujące cechy typu wytłaczarki (amplitudę ruchu posuwisto-zwrotnego ślimaka wytłaczarki posuwisto-zwrotnej S_0, odległość osiową ślimaków wytłaczarki dwuślimakowej Ascr). Typ folii determinuje typ polimeru filmotwórczego T polym, przetwarzanego w wytłaczarce. Parametry właściwości termofizycznych fazy stałej i stopu Hpolim zależą od rodzaju polimeru. Intensywność i czas trwania oddziaływań termicznych i mieszania na polimer w wytłaczarce zmieniają się wraz ze zmianą parametrów sterowania. Należą do nich prędkość ślimaka/ślimaków N, temperatury stref termicznych cylindra T_{bk} oraz natężenie przepływu ciekłych barwników G_{ci}. Natężenie przepływu odpadów zwrotnych G_{wast} jest zaburzeniem, które prowadzi do odchyleń

trybu wytłaczania od regulacji. Stan procesu charakteryzuje się prędkościami przepływu poprzecznego i dolnego kanału v_x, vz, ciśnieniem P, temperaturą T, lepkością η stopu polimeru oraz średnim czasem przebywania w wytłaczarce $\tau_{śr}$. Wydajność wytłaczarki G i wskaźniki jakości wytłoczonego materiału Q_{ext} są wskaźnikami wydajności wytłaczania Q. Jakość wytłoczonego materiału ocenia się na podstawie następujących wskaźników:

• wskaźnik zniszczenia termicznego Id (odwzorowujący zależność stopnia zniszczenia polimeru od czasu w reżimie temperaturowo-czasowym w wytłaczarce od zależności uzyskanej eksperymentalnie w warunkach izotermicznych odpowiadających początkowi zmiany barwy w wyniku zniszczenia);

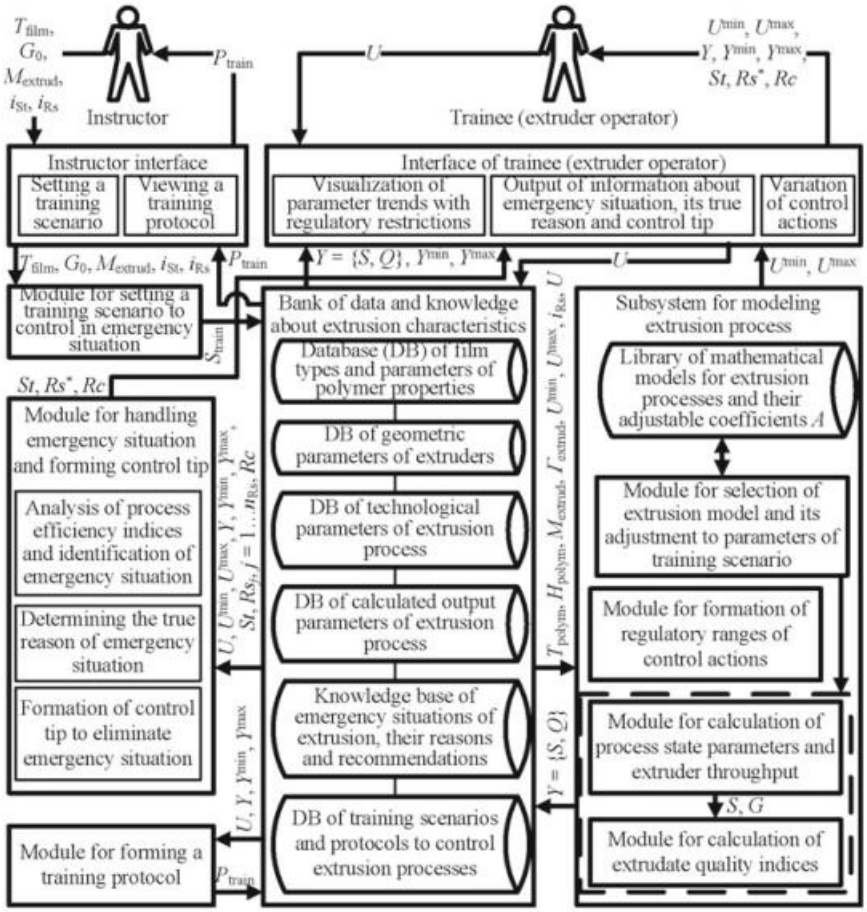
• charakterystyka jednorodności materiału – udział fazy stałej ϕ_s, średni stopień zmieszania γ_{av} (odkształcenie ścinające kumulujące się podczas czasu przebywania w wytłaczarce);

• współrzędne barwy (w przestrzeni CIELab) L_{ext}, a_{ext}, b_{ext}, odchylenie barwy od wzorca ΔE_{ext} .

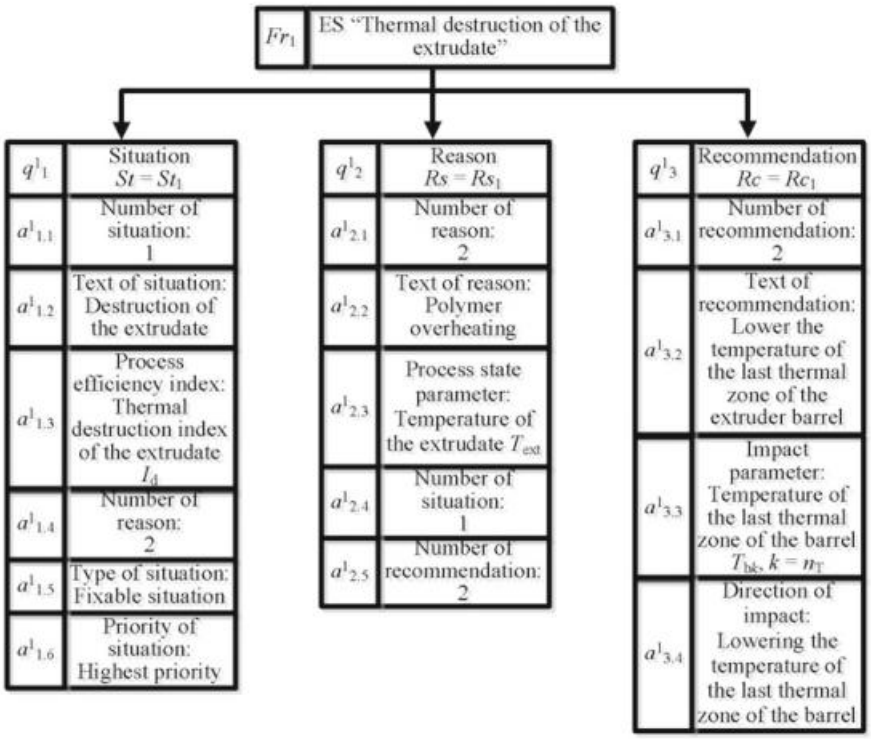
Sytuacje awaryjne wytłaczania są związane przede wszystkim z defektami wytłoczonego materiału (zniszczenie termiczne, niejednorodność materiału, odchylenia barwy). Aby wykluczyć wady folii wynikające z przygotowania ekstrudatu o złej jakości, jego wskaźniki jakości muszą spełniać wymaganie $Q_{ext} \subset [Q_{ext}^{min}; Q_{ext}^{max}]$. Wymaganie to wynika z warunku przydatności ekstrudatu do formowania na kalandrze. Gdzie Q_{ext}^{min}, Q_{ext}^{max} są wektorami wartości progowych wskaźników jakości (na przykład maksymalny wskaźnik zniszczenia I_d^{max}, minimalny stopień wymieszania γ_{av}^{min}). Zależą one od rodzaju folii T_{film}. Na przykład w przypadku folii farmaceutycznych punkty czarne są najmniej akceptowalne, więc ograniczenie wskaźnika zniszczenia jest bardziej rygorystyczne w porównaniu z foliami do produkcji kart plastikowych. Problem szkolenia w zakresie sterowania procesem wytłaczania w sytuacjach awaryjnych jest następujący: dla scenariusza szkoleniowego $S_{train} = \{T_{film}, G_0, M_{extrud}, i_{St}, i_{Rs}\}$ ustawionego przez instruktora, konieczne jest zmienianie działań kontrolnych w zakresach regulacyjnych $U \subset [U^{min}; U^{max}]$ zgodnie z końcówką sterującą Rc, utworzoną przez inteligentny system szkoleniowy (na podstawie analizy przyczyn sytuacji awaryjnej R_{sj}, $j = 1...n_{Rs}$ i określenia prawdziwej przyczyny Rs^*), w celu wyeliminowania sytuacji awaryjnej St, poprzez wprowadzenie odpowiedniego wskaźnika wydajności do zakresu dopuszczalnych wartości $Q_l \in [Q_l^{min}; Q_l^{max}]$. Gdzie G_0 jest przepustowością systemu; i_{St} jest numerem symulowanej sytuacji awaryjnej; Rs i_{Rs} jest numerem symulowanej przyczyny awarii; U^{min}, U^{max} są wektorami wartości progowych akcji sterujących, zależnych od rodzaju T_{folii} i marki wytłaczarki M_{extrud}; Rc jest końcówką zawierającą kierunek zmiany akcji sterujących na wytłaczarce; Q_l jest wskaźnikiem wydajności procesu, którego odchylenie jest niedopuszczalne według przepisów ($\Delta l = Q_l^{min} - Q_l > 0 \vee \Delta l = Q_l - Q_l^{max} > 0$) jest oznaką sytuacji awaryjnej St.

3. Struktura inteligentnego systemu szkoleniowego

Strukturę inteligentnego systemu szkoleniowego opracowanego w celu rozwiązania problemu szkoleniowego przedstawiono na rysunku.



Inteligentny system szkoleniowy zawiera moduł do tworzenia scenariusza szkoleniowego, bazę danych i wiedzy o charakterystyce procesu wytłaczania, podsystem modelowania procesu, moduł obsługi sytuacji awaryjnych i tworzenia wskazówek dotyczących sterowania procesem, moduł tworzenia protokołu szkoleniowego oraz interfejsy między uczestnikiem szkolenia (operatorem wytłaczarki) a instruktorem. System tworzy wskazówki dla uczestnika szkolenia dotyczące kierunku zmiany działań kontrolnych w celu wyeliminowania sytuacji awaryjnych określonych przez instruktora. Sytuacje awaryjne są związane z odchyleniami wskaźników wydajności (przepustowości, wskaźników jakości wytłaczanego materiału) poza wartości progowe. Wiedzę ekspercką na temat sytuacji awaryjnych w procesie wytłaczania uzyskuje się poprzez wywiady z personelem kontroli produkcji, studiowanie przepisów technologicznych i instrukcji dotyczących sterowania produkcją folii polimerowych w sytuacjach awaryjnych. Wiedza ta ma złożoną strukturę. Dlatego też, w celu ich sformalizowania, opracowano bazę wiedzy o sytuacjach awaryjnych, ich przyczynach i zaleceniach opartą na modelu reprezentacji wiedzy opartym na ramce produkcyjnej. Wiedza podmiotowa o sytuacji awaryjnej (ES) jest przedstawiona w postaci ramki prototypowej, którą opisano w notacji Backusa–Naura: Fr:: = < ES, Q_{Fr}, A_Q > , $Q_{Fr} = \{q_1, q_2, q_3\}$, $A_Q = \{a_{1.1}, \ldots, a_{3.4}\}$, gdzie Q_{Fr} jest zbiorem atrybutów ramki, które pozwalają na ustrukturyzowanie wiedzy o sytuacjach awaryjnych (q_1), przyczynach ich wystąpienia (q_2) i zaleceniach dotyczących eliminacji (q_3), A_Q jest zbiorem cech atrybutów, które składają się na ich opis językowy i informacyjny. W celu przedstawienia wiedzy o konkretnych sytuacjach awaryjnych procesu wytłaczania, wykonano przejście od ramki prototypowej do systemu ram przykładowych



Aby uzyskać przykładową ramkę konkretnej sytuacji awaryjnej, przypisano charakterystyki każdego atrybutu ramki prototypowej wartości $a_{1.1}$, ..., $a_{3.4}$ opisujące tę sytuację. Baza wiedzy zawiera opis sytuacji awaryjnych, które są związane z defektami wytłaczania (zniszczenie termiczne, niestopione wtrącenia itp.) i odchyleniami w przepustowości wytłaczarki, prowadzącymi do wahań wsadu wytłaczanego podczas podawania do kalandra. Sytuacje te są naprawialne, ponieważ są eliminowane poprzez celową zmianę działań kontrolnych bez zatrzymywania systemu produkcyjnego. Priorytety sytuacji awaryjnych są ustalane w zależności od stopnia negatywnego wpływu sytuacji na działanie systemu produkcyjnego i cech użytkowych folii polimerowych. Priorytety te określają kolejność eliminacji sytuacji awaryjnych w przypadku wystąpienia kilku sytuacji. Każda sytuacja awaryjna w określonych warunkach produkcyjnych występuje z jednej z kilku możliwych przyczyn. Dlatego jednym z zadań inteligentnego systemu szkoleniowego jest określenie prawdziwej przyczyny sytuacji awaryjnej. Każdemu powódowi przypisano konkretne zalecenie opisujące działania operatora wytłaczarki, które należy podjąć, aby wyeliminować przyczynę, a tym samym spowodowaną nią sytuację awaryjną. Wiedza sterująca niezbędna do identyfikacji sytuacji awaryjnych, określenia ich prawdziwych przyczyn i sformułowania wskazówek kontrolnych jest reprezentowana w postaci reguł produkcyjnych: Pr:: = IF (warunek), THEN (konsekwencja). Na przykład, aby zidentyfikować sytuację awaryjną „Termiczne zniszczenie wytłoczyny" (St_1) i sformułowanie wskazówki kontrolnej, zaimplementowano następujące reguły: prostą regułę IF ($I_d > I_d^{max}$), THEN($St = St_1$); reguły skupiające IF ($St = St_1$) ∧ ($T_{ext} > T_{ext}^{max}$), THEN ($Rs^* = Rs_1$), IF ($St = St_1$) ∧ ($Rs^* = Rs_1$), THEN ($Rc = Rc_1$). Symulacja występowania sytuacji awaryjnych jest przeprowadzana w systemie poprzez modelowanie przyczyn sytuacji określonych w scenariuszach szkoleniowych Odkształcenie. W tym celu wartości odpowiednich parametrów modelu matematycznego procesu wytłaczania (parametry wejściowe X, działania sterujące U lub regulowane współczynniki równań modelu A) są zmieniane w stosunku do ich wartości nominalnych Xnom, Anom lub zakresów regulacyjnych [Umin; Umax], zapisanych w bazie danych i bibliotece modeli. Różnorodność typów wytłaczarek i polimerów, złożoność procesów hydrodynamicznych oraz procesów wymiany ciepła i masy

zachodzących w kanale ślimakowym wytłaczarki, a także nienewtonowskie zachowanie się stopów wymagają zintegrowanego podejścia do modelowania procesów wytłaczania. Jednocześnie istniejące modele matematyczne opisują procesy w wytłaczarkach niektórych typów lub w oddzielnych strefach funkcjonalnych wytłaczarki. Z reguły nie pozwalają one na uwzględnienie wszystkich cech wytłaczania (przemian fazowych, anomalii lepkości, wycieku stopu przez szczeliny, nacięć w ślimakach, rozpraszania energii, wymiany ciepła polimeru z cylindrem, ślimakiem) i na dokonanie kompleksowej ilościowej oceny charakterystyki procesu, z uwzględnieniem jakości wytłoczyny. W związku z tym opracowano bibliotekę modeli matematycznych, aby dostosować system do szkolenia w zakresie sterowania procesami wytłaczania wieloasortymentowego w różnych typach urządzeń. Zawiera ona modele przetwarzania polimerów w wytłaczarkach jednoślimakowych, posuwisto-zwrotnych i dwuślimakowych, posiadających ślimaki o typowej konfiguracji i wyposażonych w matryce. Modele są deterministyczne i oparte na prawach zachowania masy, pędu, energii (do konstruowania równań równowagi fazowej i międzyfazowej), prawach reologii (do konstruowania równań zależności lepkości od szybkości odkształcenia ścinającego i temperatury). Przy konstruowaniu modeli przyjmuje się następujące założenia ogólne (dla wszystkich typów wytłaczarek): odwrócony ruch cylindra i ślimaka; niewielka krzywizna i całkowite wypełnienie kanału ślimaka; ustalony charakter procesu w czasie i wzdłuż długości kanału; stałość właściwości polimeru; brak przepływu promieniowego, sił bezwładności i masy, konwekcyjnego przenoszenia ciepła przez szerokość kanału i przenoszenia ciepła przez przewodnictwo cieplne wzdłuż długości kanału; adhezja stopu do ścianek kanału. Model procesu wytłaczania jest wybierany z biblioteki modeli w zależności od typu wytłaczarki T_{extrud}. Dopasowanie modelu do typu polimeru T_{polym} i marki wytłaczarki M_{extrud} polega na utworzeniu nominalnych wartości parametrów właściwości faz wytłaczanego polimeru H_{polym}, parametrów geometrycznych wytłaczarki Γ_{extrud}, współczynników modelu A. Regulowanymi współczynnikami modelu są współczynniki równania reologicznego stanu stopionego (na przykład indeks potęgowy n), współczynniki przenoszenia ciepła między fazami polimeru, cylindrem i ślimakiem wytłaczarki, współczynniki modelu zniszczenia termicznego (czas τ_d, temperatura T_d i energia aktywacji E_d zniszczenia), współczynniki modeli regresyjnych do obliczania współrzędnych kolorów (c_p, p = 0...n_L, d_q, q = 0...na). Ze względu na nieliniowość równań modelu do ich rozwiązania stosuje się metody numeryczne (metody teorii płaskich przepływów asymetrycznych, metody różnic skończonych). Skorygowany model wytłaczania pozwala obliczyć rozkłady prędkości przepływu vx, vz, ciśnienia P, temperatury T, lepkości η stopu w kanale ślimaka, średniego czasu przebywania τ_{av} oraz przepustowości $G = \rho Q_E$. Gdzie ρ to gęstość stopu, a Q_E to objętościowe natężenie przepływu w kanale dolnym przez układ „kanał ślimaka-matryca", obliczane poprzez iteracyjne rozwiązanie równania utworzonego ze stanu sprzężenia ślimaka i matrycy (równe różnice ciśnień w kanale ślimaka i matrycy). Zatem dla wytłaczarki posuwisto-zwrotnej równanie to ma następującą postać:

$$(\pi N)^n \left(1 + \frac{1}{n}\right)^n \int_0^Z (D\cos\varphi + S_0 \sin\Phi_{osc} \sin\varphi)^n \frac{B_z \mu}{H^{n+1}} \frac{F_d - Q_E/Q_d}{(1 - Q_E/Q_d) F_P} dz - \mu_{die} \left(\frac{Q_E}{k_{die}}\right)^n = 0, \quad (1)$$

gdzie Z, H to długość i głębokość kanału ślimaka; ϕ to kąt nachylenia zwojów ślimaka; Φ_{osc} to faza ruchu posuwisto-zwrotnego ślimaka; B_z to bezwymiarowy gradient ciśnienia w przepływie stopu w kanale dolnym; μ, μ_{die} to współczynniki konsystencji stopu w kanale ślimaka i matrycy; F_d, F_P,

k_{die} to współczynniki kształtu geometrycznego kanału ślimaka i matrycy; Q_d to objętościowe wymuszone natężenie przepływu w kanale ślimaka. Na podstawie parametrów stanu S obliczane są wskaźniki jakości wytłoczyny:

$$I_d = \frac{\tau_{av}}{\tau_d}\exp\left[\frac{E_d}{8.31(T_{ext}+273)(T_d+273)}(T_{ext}-T_d)\right]100, \qquad (2)$$
$$T_{ext} = T|_{z=Z},$$

$$\gamma_{av} = \tau_{av}(ZH)^{-1}\int_0^Z\int_0^H\sqrt{(\partial v_x/\partial y)^2+(\partial v_z/\partial y)^2}\,dydz, \qquad (3)$$

$$L_{ext} = f_L(k_i,\ i=1\ldots n_c,\ c_p,\ p=0\ldots n_L), \qquad (4)$$
$$a_{ext} = f_a(k_i,\ i=1\ldots n_c,\ d_q,\ q=0\ldots n_a),$$

$$\Delta E_{ext} = \sqrt{(L_{ext}^* - L_{ext})^2 + (a_{ext}^* - a_{ext})^2 + (b_{ext}^* - b_{ext})^2}, \qquad (5)$$
$$k_i = G_{ci}/G,$$

gdzie L_{ext}^*, a_{ext}^*, b_{ext}^* to współrzędne kolorów wzorca, zależne od rodzaju folii T. Opracowane modele matematyczne pozwalają zatem obliczyć parametry stanu S i wskaźniki wydajności Q procesu wytłaczania w zależności od działań kontrolnych U dla różnych typów wytłaczarek T i marek M_{extrud}, typów przetwarzanych polimerów T_{polym}. Aby sprawdzić adekwatność modeli matematycznych, porównano zmierzone i obliczone wartości przepustowości procesu, temperatury i współrzędnych kolorów wytłoczyny. Dane uzyskano dla wytłaczarek posuwisto-zwrotnych i dwuślimakowych różnych marek podczas przetwarzania sztywnego polichlorku winylu (PCW) i realizacji różnych trybów wytłaczania (przez zróżnicowanie działań kontrolnych). Wyniki statystycznej obróbki uzyskanych danych potwierdziły adekwatność modeli zgodnie z kryterium Fishera, współczynnikiem determinacji i odchyleniem standardowym, które nie przekracza 3–5% dla różnych parametrów wyjściowych procesu. Dopasowanie inteligentnego systemu szkoleniowego do parametrów scenariusza szkoleniowego odbywa się z wykorzystaniem modelu informacyjnego procesu wytłaczania. Model ten jest zaimplementowany jako zbiór relacyjnych baz danych. Baza danych typów folii i parametrów właściwości polimerów umożliwia tworzenie nominalnych wartości parametrów właściwości termofizycznych faz polimerowych Hpolym (w zależności od rodzaju polimeru T_{polym}) oraz wartości progowych wskaźników jakości wytłoczyny Q_{ext}^{min}, Q_{ext}^{max} (w zależności od rodzaju folii T_{film}). Baza danych parametrów geometrycznych wytłaczarek umożliwia tworzenie nominalnych wartości parametrów geometrycznych Γ_{extrud} (w zależności od marki wytłaczarki M_{extrud}). Baza danych parametrów technologicznych procesu umożliwia tworzenie ograniczeń regulacyjnych na czynności regulacyjne U^{min}, U^{max}, parametry stanu procesu S^{min}, S^{max}, wydajność wytłaczarki G^{min}, G^{max} (w zależności od rodzaju folii T_{film}, marki wytłaczarki M_{extrud}, wydajności systemu G_0). Tablice wartości akcji sterujących są przechowywane w bazie danych parametrów technologicznych, które są importowane z bazy danych zmierzonych parametrów przemysłowej produkcji folii polimerowych, która jest zawarta w pakiecie oprogramowania do przetwarzania dużych zbiorów danych i kontroli jakości folii. Tablice te służą do symulacji procesu wytłaczania w czasie. Na podstawie wartości akcji sterujących U, które zmieniają się w czasie, obliczane i wyświetlane są parametry wyjściowe Y (w postaci trendów). Wyniki modelowania są przechowywane w bazie danych parametrów wyjściowych. Scenariusze szkoleniowe i wyniki są przechowywane w bazie danych scenariuszy i protokołów. Protokół szkoleniowy P_{train}, wygenerowany przez system, zawiera wartości akcji sterujących U, ustawione przez osobę szkoloną w celu wyeliminowania sytuacji awaryjnej, a także obliczone (dla nich) i progowe wartości parametrów wyjściowych Y, $Y^{min} = \{S^{min}, G^{min}, Q_{ext}^{min}\}$, $Y^{max} =$

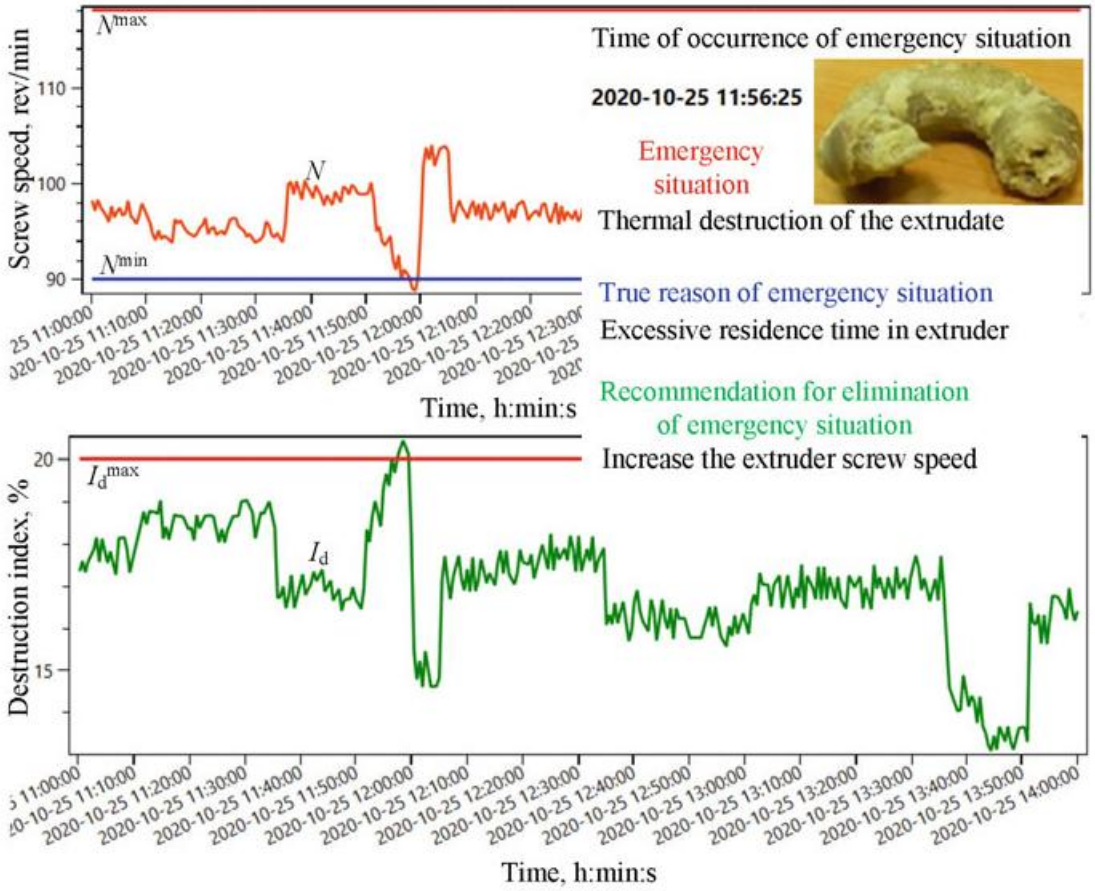
{S^{max}, G^{max}, $Q_{ext}{}^{max}$}. Cel szkolenia jest osiągnięty, jeśli osoba szkolona ustawi tryb wytłaczania $U \subset [U^{min}; U^{max}]$: $Y \subset [Y^{min}; Y^{max}]$.

4 Algorytm modelowania sytuacji awaryjnych i formowania końcówek do sterowania wytłaczaniem

Algorytm formowania końcówek do eliminowania sytuacji awaryjnych przedstawiono na rys. 3. Dla danego scenariusza szkoleniowego, wartości progowe odkształceń akcji sterujących Umin, Umax i parametrów wyjściowych Y min, Ymax procesu wytłaczania, tworzona jest tablica wartości akcji sterujących U (w celu symulacji procesu w czasie), wybierany jest model wytłaczania, a jego parametry X, U, A są dostosowywane w celu symulacji wystąpienia danej sytuacji awaryjnej iSt. Wartości wskaźników efektywności procesu Q obliczone przez model matematyczny są porównywane z wartościami progowymi Q^{min}, Q^{max}. W przypadku odchylenia wskaźnika efektywności Q_l poza wartości progowe identyfikowana jest sytuacja awaryjna St, której opisem informacyjnym jest wskaźnik Q_l. Dla zidentyfikowanej sytuacji, z bazy wiedzy tworzona jest lista możliwych przyczyn jej wystąpienia R_{sj},j = 1...n_{Rs}. Każda przyczyna jest opisana jednym z parametrów procesu (U, S). Aby określić prawdziwą przyczynę Rs^* w przypadku wystąpienia tej sytuacji, wartości tych parametrów są porównywane z wartościami progowymi, a następnie identyfikowany jest parametr, którego wartość narusza ograniczenia. Aby kontrolować poprawność działania systemu, sprawdzana jest równość liczby prawdziwych przyczyn i liczby i_{Rs} określonych w scenariuszu szkoleniowym. Dla znalezionej przyczyny z bazy wiedzy pobierane jest zalecenie Rc. Zawiera ono polecenie zmiany odpowiednich działań kontrolnych U w celu wyeliminowania tej przyczyny, a tym samym sytuacji awaryjnej. Uczestnik szkolenia otrzymuje komunikat o sytuacji awaryjnej St, jej prawdziwej przyczynie Rs^* oraz wskazówce kontrolnej Rc. Zmieniając odpowiednio działania kontrolne w procesie, uczestnik szkolenia eliminuje sytuację awaryjną, przenosząc proces do trybu rutynowego. Proponowany algorytm umożliwia rozwiązanie problemu uczenia sterowania procesem wytłaczania w różnych sytuacjach awaryjnych dla różnych wariantów sprzętowych i technologicznych procesu oraz rodzajów polimerów, ponieważ cechy te są ustalane w scenariuszu szkoleniowym, dla którego algorytm jest dostosowywany.

5 Testowanie inteligentnego systemu szkoleniowego

System został przetestowany na przykładach szkoleniowych w zakresie sterowania procesami wytłaczania w sytuacjach awaryjnych związanych z różnego rodzaju defektami wytłaczanego materiału (destrukcja termiczna, wtrącenia niestopionego polimeru, niedostateczne mieszanie, odchylenia koloru) i powstającymi z różnych przyczyn (zbyt długi/niedostateczny czas przebywania, przegrzanie/niedostateczne ogrzewanie, niewystarczająca intensywność mieszania itp.). Do testów wykorzystano dane z kalandrowanej produkcji folii do pakowania produktów farmaceutycznych i żywności różnego rodzaju (na bazie sztywnego PVC) w fabrykach w Rosji i Niemczech. Jako sprzętową konstrukcję procesów wytłaczania wykorzystano wytłaczarki posuwisto-zwrotne i przeciwbieżne wytłaczarki dwuślimakowe. Przykładowy interfejs uczestnika szkolenia przedstawiono na rysunku.



Przedstawiono na nim trendy prędkości ślimaka N wytłaczarki posuwisto-zwrotnej PR-200 oraz wskaźnik destrukcji termicznej wytłaczanego materiału Id wraz z wartościami progowymi N^{min}, N^{max}, I_d^{max}. Wyświetla również informacje o sytuacji awaryjnej związanej z destrukcją wytłoczyny ($I_d > I_d^{max}$), jej prawdziwej przyczynie (nadmierny czas przebywania polimeru w wytłaczarce: $\tau_{av} > \tau_{av}^{max}$, gdzie τ_{av}^{max} to górna wartość progowa czasu przebywania), modelowanej poprzez zmniejszenie prędkości ślimaka N poniżej dolnej wartości progowej N^{min}, oraz końcówkę kontrolną (zwiększenie prędkości ślimaka N). Wdrażając końcówkę kontrolną, uczestnik szkolenia wprowadza wskaźnik destrukcji do zakresu regulacyjnego (który jest wyświetlany na wykresie trendu wskaźnika destrukcji), eliminując sytuację awaryjną i zapobiegając odpowiadającym jej wadom powierzchni folii polimerowej. Wyniki testów potwierdziły przydatność systemu do kształtowania umiejętności sterowania procesami wytłaczania w sytuacjach awaryjnych.

6. Wnioski

Opracowano konfigurowalny, inteligentny system szkoleniowy, który ma na celu szkolenie operatorów wytłaczarek do wieloasortymentowej produkcji folii polimerowych. System pozwala, w oparciu o wielowymiarowe modele opisujące procesy wytłaczania (model ramowy produkcyjno-przemysłowy do reprezentacji wiedzy eksperckiej, funkcjonalne modele matematyczne, relacyjny model informacyjny), rozwiązać problem kształtowania umiejętności sterowania procesami wytłaczania w sytuacjach awaryjnych dla różnych scenariuszy szkoleniowych (różne rodzaje produkowanych folii, projekt sprzętowy i technologiczny procesu, sytuacje awaryjne i przyczyny ich występowania). Wyniki testów, przeprowadzone na danych z przemysłowej produkcji folii polimerowych do pakowania produktów farmaceutycznych i

spożywczych, potwierdziły przydatność systemu jako narzędzia komputerowego do podnoszenia kompetencji zawodowych operatorów wytłaczarek, przyczyniając się do skrócenia czasu podejmowania decyzji sterujących i wyeliminowania sytuacji awaryjnych w procesie wytłaczania. Prowadzi to do poprawy właściwości folii dla konsumentów, oszczędności zasobów, zwiększenia wydajności produktów wysokiej jakości i zwiększenia bezpieczeństwa produkcji pod względem środowiskowym (dzięki zmniejszeniu ilości odpadów nie podlegających zwrotowi w wyniku terminowego eliminowania sytuacji awaryjnych w procesie wytłaczania i zapobiegania odpowiadającym im wadom folii polimerowej).

Metoda Konstrukcji Symulatorów Rzeczywistości Wirtualnej do Toczenia i Frezowania dla Systemu Kształcenia Inżynierów Budowy Systemów Cyberfizycznych

Opisano opracowanie i wdrożenie trójwymiarowego komputerowego symulatora toczenia i frezowania w rzeczywistości wirtualnej, architekturę kompleksu sprzętowo-programowego i jego funkcjonalność w oparciu o proponowaną metodologię. Rozważono zasady interakcji użytkownika z modelami trójwymiarowymi oraz projektowanie UI/UX. Symulator został zaprojektowany do rozwiązywania zadań wizualizacji modeli trójwymiarowych zgodnie z określonymi scenariuszami podczas szkoleń i oceny kwalifikacji. Praktyczne zastosowanie w realizacji efektywnych szkoleń w przedsiębiorstwach przemysłowych, optymalizacji kosztów rzeczowych i finansowych, minimalizacji liczby błędów i wypadków, braku możliwości uszkodzenia sprzętu, konieczności posiadania specjalnie wyposażonego pomieszczenia, naliczania amortyzacji, kosztów konserwacji i materiałów eksploatacyjnych: frezów, wierteł i przedmiotów obrabianych. Zmienność scenariusza pozwala na wykonywanie różnorodnych operacji w aplikacji. Nowością proponowanego projektu jest: model szkolenia zawodowego i oceny kwalifikacji, umożliwiający ocenę procesów pracy i umiejętności zawodowych studentów; specjalistyczny pakiet oprogramowania z unikalnymi modułami programowymi; metoda optymalizacji modułów programowych i trójwymiarowych modeli kompleksu programowego szkoleń w segmencie przemysłowym w jednej wirtualnej przestrzeni.

Słowa kluczowe: Technologie immersyjne · Rzeczywistość wirtualna · Wizualizacja trójwymiarowa · Symulatory komputerowe · Głębokie uczenie · Grafika komputerowa

1 Wstęp

Usprawnienie procesu szkoleniowego w branży zaawansowanych technologii wymaga szczególnej uwagi. W związku z tym badanie metod i środków zastosowania technologii immersyjnych w celu poprawy jakości i dostępności szkoleń jest niezwykle istotne we współczesnym społeczeństwie . Należy zauważyć, że symulatory komputerowe są z powodzeniem wykorzystywane do szkolenia operatorów w różnych dziedzinach . W tym przypadku wykorzystywany jest efekt rzeczywistości wirtualnej . Systemy rzeczywistości wirtualnej są wykorzystywane w edukacji inżynierskiej , medycznej , dla osób niepełnosprawnych oraz w nauce jazdy . Symulatory rzeczywistości wirtualnej są coraz szerzej wykorzystywane w szkoleniach zawodowych i ocenie kwalifikacji. Istnieją różne opinie na temat tego, czy wprowadzenie technologii rzeczywistości wirtualnej do szkoleń jest zasadne, czy też służy uzyskaniu pięknego efektu wizualnego i nie wpływa bezpośrednio na działania edukacyjne. Jakość szkoleń wymaga wysiłku, a jednym ze sposobów na zwiększenie ich efektywności jest wykorzystanie technologii immersyjnych, a mianowicie prowadzenie części szkoleń praktycznych

na symulatorach rzeczywistości wirtualnej. Obecnie technologia wirtualnej rzeczywistości cieszy się popytem nie tylko w branży rozrywkowej: kinematografii, grach i atrakcjach. W edukacji technologie immersyjne są z powodzeniem stosowane od wielu lat w celu poprawy jakości, efektywności ekonomicznej i dostępności . Możliwości wirtualnej rzeczywistości są dość szerokie, umożliwiając przeniesienie użytkownika w różne miejsca i interaktywne scenariusze. Zalet jest wiele, więc możliwość ponownego uruchomienia procesu technologicznego bez utraty zasobów i ryzyka rzeczywistych awarii w produkcji sprawia, że wykorzystanie wirtualnej rzeczywistości w szkoleniach jest właściwe i konieczne. Widoczność, bezpieczeństwo, immersja i zaangażowanie sprawiają, że systemy wirtualnej rzeczywistości są interesujące i atrakcyjne zarówno dla stażystów, jak i nauczycieli . Zatem szkolenie tokarzy i frezerów na uniwersalnych obrabiarkach dla przedsiębiorstw przemysłowych jest procesem pracochłonnym i kosztownym. Dlatego niniejszy projekt opracowania i wdrożenia symulatora toczenia i frezowania w środowisku wirtualnej rzeczywistości jest istotny. Niniejszy rozdział opisuje wyniki opracowania i wdrożenia trójwymiarowego komputerowego symulatora toczenia i frezowania w wirtualnej rzeczywistości w oparciu o proponowaną metodologię. Sformułowanie problemu. Celem pracy jest zwiększenie efektywności procesu edukacyjnego opartego na wykorzystaniu symulatora rzeczywistości wirtualnej, minimalizacja kosztów materiałów eksploatacyjnych w procesie szkolenia i oceny kwalifikacji, a także dostępność sprzętu. Aby osiągnąć ten cel, w procesie rozwoju symulatora konieczne jest rozwiązanie szeregu zadań:

• przeprowadzenie analizy istniejących metod i środków rozwoju symulatorów rzeczywistości wirtualnej;

• analiza podobnych, opracowanych systemów i ich wpływu na efektywność;

• opracowanie koncepcji działania, zakresu zadań, scenariusza użytkownika, UX/UI symulatora rzeczywistości wirtualnej;

• sformułowanie wymagań dla opracowywanego systemu;

• opracowanie struktury systemu i algorytmów jego komponentów;

• zaprojektować i utworzyć bazę danych do przechowywania informacji o użytkowniku i systemie symulatora rzeczywistości wirtualnej;

• przeprowadzić testy i przeprowadzić kompleksowe debugowanie systemu ;

• opracować dokumentację techniczną symulatora rzeczywistości wirtualnej;

• ocenić efektywność opracowanego symulatora rzeczywistości wirtualnej.

Dane wstępne. Sformułujmy wymagania funkcjonalne dla systemu rzeczywistości wirtualnej do szkolenia na uniwersalnych frezarkach i tokarkach. Po przeanalizowaniu podobnych systemów, opracowaniu koncepcji działania, zakresu zadań, scenariusza użytkownika, UX/UI symulatora rzeczywistości wirtualnej, sformułujmy wymagania funkcjonalne dla systemu do szkolenia na uniwersalnych frezarkach i tokarkach. Aplikacja rzeczywistości wirtualnej ma zostać wdrożona jako symulator szkolenia i oceny kwalifikacji dla kierunku „Inżynieria Mechaniczna". Celem symulatora „Laboratorium Uniwersalnej Obróbki Toczeniem i Frezowaniem w Środowisku Rzeczywistości Wirtualnej (VR)" jest:

• zwiększenie efektywności procesu dydaktycznego;

• minimalizacja kosztów materiałów eksploatacyjnych w procesie szkolenia i oceny kwalifikacji;

• zapewnienie dostępności sprzętu.

Sformułujmy wymagania dla systemu:

• w zakresie struktury i funkcjonowania: stabilne i niezawodne funkcjonowanie modułów systemu, zapewniające pełną wydajność procesów, nieprzerwane zasilanie środków technicznych, korzystanie z licencjonowanego oprogramowania, standardy czasowe konserwacji oprogramowania i sprzętu, ochronę informacji przed szkodliwym oprogramowaniem;

• w zakresie liczebności i kwalifikacji personelu systemu oraz jego sposobu działania: do pracy z oprogramowaniem wymagany jest operator komputera z wiedzą z zakresu symulatorów rzeczywistości wirtualnej oraz użytkownik końcowy. Po uzyskaniu wymaganych umiejętności użytkownik końcowy może pełnić funkcję operatora komputera. Zadaniem operatora komputera jest utrzymanie sprawności sprzętu, instalacja i utrzymanie sprawności oprogramowania i innych powiązanych programów, instalacja programów oraz praca z danymi, w tym tworzenie kopii zapasowych;

• w zakresie niezawodności: symulator rzeczywistości wirtualnej powinien automatycznie sprawdzać poprawność wprowadzanych danych, a także zgodność z dopuszczalnymi zakresami danych, w tym obsługę sytuacji nadzwyczajnych;

• ergonomii i estetyki technicznej: symulator rzeczywistości wirtualnej musi posiadać estetyczny interfejs z dobrą wizualizacją obiektów (UI), a także przyjazny i intuicyjny interfejs użytkownika (UX);

• trybu dialogowego: w przypadku błędów w złożonych modułach oprogramowania powinny być wyświetlane komunikaty o błędach z zaleceniami dotyczącymi ich usunięcia;

• ochrony informacji przed nieautoryzowanym dostępem: dostęp do aplikacji rzeczywistości wirtualnej jest zapewniany za pomocą loginu i hasła i jest ograniczony trybami administratora i użytkownika. Dostęp do aplikacji odbywa się za pomocą klucza licencyjnego. Hasło ma wymagania dotyczące długości i nazwy (od 3 znaków);

• do czystości patentowej: warunki umów licencyjnych muszą być przestrzegane dla wszystkich obiektów technicznych i programowych symulatora rzeczywistości wirtualnej oraz zapewniona czystość patentowa

• do standaryzacji i unifikacji: przy opracowywaniu symulatora rzeczywistości wirtualnej należy stosować standardową metodologię modelowania informacji (GOST 34.601–90 – Technologia informacyjna. Zbiór norm dla systemów zautomatyzowanych), (GOST 19.102.77 – Zunifikowany system dokumentacji programowej. Etapy rozwoju).

2. Metodyka rozwoju i implementacji symulatorów dla wybranego obszaru tematycznego toczenia i frezowania

Celem jest opracowanie symulatora, który poprawi efektywność ekonomiczną procesów edukacyjnych oraz rozwiąże inne problemy aplikacyjne z dziedziny inżynierii mechanicznej. Analiza i opracowanie wymagań. Przeprowadzenie analizy obszaru tematycznego w celu uzasadnienia wykonalności rozwoju: zdefiniowanie celów, zagadnień, przedmiotu i przedmiotu badań, a także istotności i celów niezbędnych do osiągnięcia sformułowanego celu. Następnie,

na podstawie analizy, tworzone są wymagania dotyczące symulatora. Projektowanie. Na podstawie wyników analizy i zadanych wymagań dotyczących symulatora, projektowane jest oprogramowanie, w tym architektura dodatku, algorytmy realizacji żądanych funkcji symulatora. Opracowywana jest koncepcja artystyczna i szkice na podstawie wybranych odniesień do wizualizacji środowiska i modeli projektowanego symulatora. Rozwój. Na podstawie opracowanego modelu symulatora tworzone są modele 3D środowiska i obiektów interakcji wykorzystywanych w symulatorze. Podczas modelowania, kod programu jest opracowywany z uwzględnieniem proponowanej architektury i algorytmów. W procesie realizacji bloków programistycznych modele są wprowadzane do środowiska programistycznego i łączone z programami w celu etapowego tworzenia produktu końcowego. Testowanie. Po etapie opracowywania symulatora przeprowadzane są różnego rodzaju testy w celu sprawdzenia spełnienia określonych wymagań dla produktu i zidentyfikowania błędów pojawiających się dla użytkownika symulatora. Na podstawie wyników testów symulator jest w razie potrzeby udoskonalany, aby przygotować projekt do uruchomienia dla rzeczywistych użytkowników. Wdrożenie i utrzymanie. Produkt końcowy (symulator) jest instalowany na komputerach i przekazywany klientowi do użytkowania. Po instalacji organizowane jest szkolenie personelu klienta w celu zapoznania użytkowników z zarządzaniem i różnymi rodzajami interakcji z systemem. W przypadku wystąpienia trudności, które nie zostały zidentyfikowane w fazie testowania, programiści naprawiają zidentyfikowane problemy. Aktualizacje symulatora są opracowywane w procesie utrzymania, jeśli proces ten jest określony w wymaganiach projektu. Poniżej opisano rezultaty projektu opracowania i wdrożenia komputerowego symulatora 3D do toczenia i frezowania w rzeczywistości wirtualnej w oparciu o zaproponowaną metodykę, od etapu projektowania do etapu wdrożenia i utrzymania.

3. Opis procesu projektowania symulatora toczenia i frezowania w wirtualnej rzeczywistości oraz architektury kompleksu sprzętowo-programowego

Proces projektowania symulatora obejmuje następujące kroki:

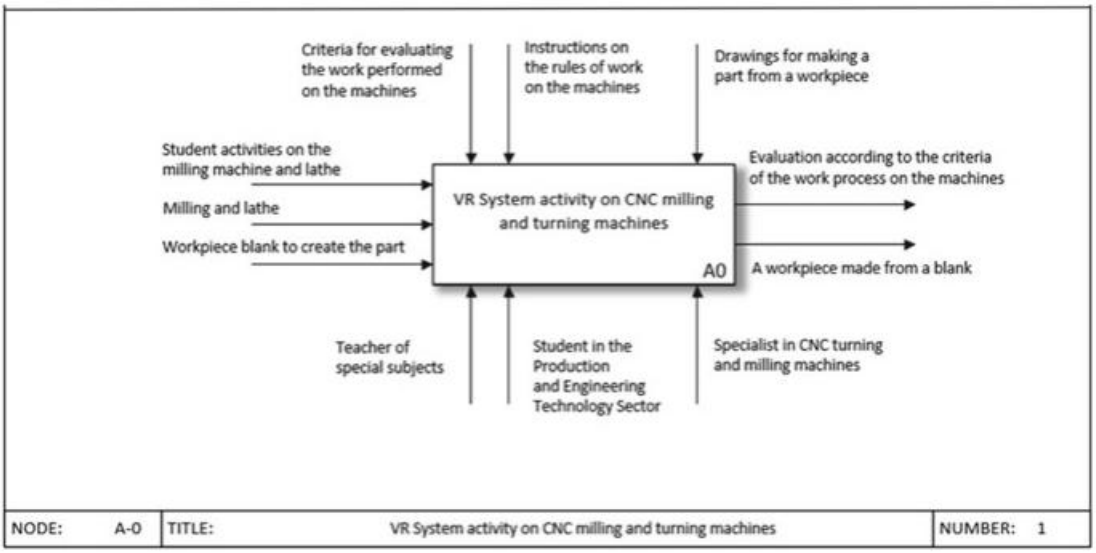
1. analiza i zanurzenie w obszarze tematycznym;

2. opracowanie zakresu zadań wraz z opisem procesów merytorycznych, scenariuszy i modeli;

3. opracowanie szkiców, schematów ekranów z przejściami, projektowanie UI/UX;

4. opracowanie architektury systemu.

Opracowany system wirtualnej rzeczywistości dla frezarek i tokarek ma następującą strukturę

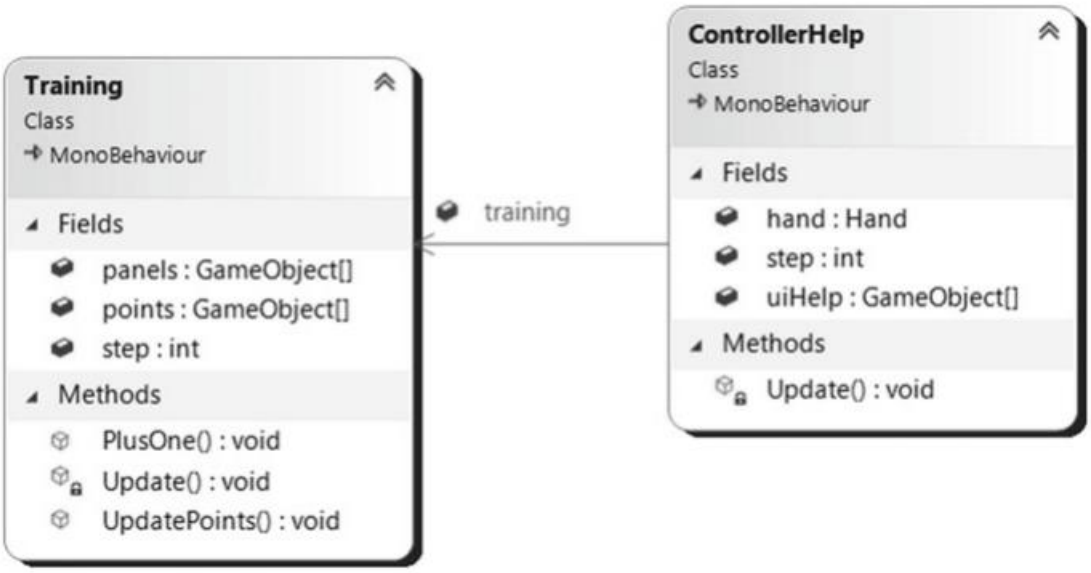


Zasobami wykonującymi pracę są nauczyciele kierunków specjalistycznych, studenci kierunku „Technologia produkcji i inżynieria" oraz specjaliści ds. produkcji pracujący na maszynach uniwersalnych. W trakcie pracy pracują z obiektami, które są wykorzystywane i przetwarzane w celu uzyskania określonego rezultatu. Należą do nich uniwersalne frezarki i tokarki, półfabrykaty do produkcji części na maszynach. Czynności wykonywane przez użytkownika w trakcie pracy na obrabiarkach uniwersalnych. Części są kontrolowane i monitorowane na maszynach przez specjalny system, który monitoruje proces obróbki na podstawie określonych kryteriów, w oparciu o instrukcje obsługi obrabiarki i rysunki

produkowanych części. Działania te skutkują produkcją części na maszynach ogólnego przeznaczenia z punktacją jakości procesu opartą na określonych kryteriach. Procesy są wskazane na diagramie kontekstowym Aktywności Systemu BP dla pracy na frezarkach i tokarkach uniwersalnych.

Architektura aplikacji. Architektura kodu frezarki:

1. Klasy odpowiedzialne za naukę obsługi frezarki przedstawiono na rysunku 2



a. Training — klasa odpowiedzialna za szkolenie.

b. ControllerHelp — klasa odpowiedzialna za pomoc dotyczącą frezarki.

2. Klasy odpowiedzialne za interakcję z interfejsem użytkownika frezarki i jej obsługę. Poniżej pominięto ilustrację graficzną opracowanych klas, podano jedynie ich opisy.

a. MillingManager — klasa odpowiedzialna za stan pracy frezarki.

b. LockDoorButton — klasa odpowiedzialna za zamknięcie drzwi we frezarce.

c. ControllerTakerMilling — klasa odpowiedzialna za znalezienie przedmiotu obrabianego we frezarce.

d. SwitchPos — klasa odpowiedzialna za ruch frezu.

e. Generation — klasa odpowiedzialna za generowanie przedmiotu obrabianego.

f. BindMilling — klasa odpowiedzialna za interfejs frezarki.

g. HandController — klasa odpowiedzialna za śledzenie pras na sterowniku.

h. SliderUp — klasa odpowiedzialna za działanie interfejsu suwaka na frezarce.

i. SheetSwitch — klasa odpowiedzialna za zmianę ekranu w interfejsie wyboru przedmiotu obrabianego.

j. ControllerToolMilling — klasa odpowiedzialna za działanie narzędzia frezarki.

3. Klasy odpowiedzialne za działanie narzędzia, obliczanie odległości ugięcia frezu

a. Jupka — klasa odpowiedzialna za pracę narzędzia, pomiar odchyłki.

b. Pozostałe klasy są potrzebne do korygowania odchyłek z różnych stron.

4. Klasy odpowiedzialne za obróbkę części frezem

a. ControllerCube — klasa odpowiedzialna za wymiar przedmiotu obrabianego.

b. MeshStudy — klasa odpowiedzialna za obróbkę przedmiotu obrabianego.

Architektura kodu tokarki:

1. Klasy odpowiedzialne za działanie tokarki

a. ControllerLathe — klasa odpowiedzialna za stan pracy tokarki i przechowywanie danych o jej ustawieniach.

b. ControllerSupport — klasa odpowiedzialna za działanie suportu tokarki.

c. ControllerSupportMove — klasa odpowiedzialna za ruch suportu.

d. Taker — klasa odpowiedzialna za lokalizację części w obrabiarce.

e. ToolLathe — klasa odpowiedzialna za narzędzie tokarki.

f. ControllerDisplaySpeed — klasa odpowiedzialna za wyświetlanie ustawień maszyny.

g. TakeCutter — klasa odpowiedzialna za znalezienie narzędzia w tokarce.

h. ButtonCoordController — klasa odpowiedzialna za zmianę współrzędnych suportu na wyświetlaczu.

i. CylPartController — klasa odpowiedzialna za ustawienie przedmiotu obrabianego w tokarce.

2. Klasy odpowiedzialne za obróbkę części na tokarce

a. InsideTool — klasa odpowiedzialna za narzędzie obróbcze wewnątrz przedmiotu obrabianego z konikiem.

b. InsideProcessing — klasa odpowiedzialna za kontakt między narzędziem, w przypadku obróbki wewnętrznej, a przedmiotem obrabianym.

c. ControllerInsideProcessing — klasa odpowiedzialna za obróbkę wewnątrz części.

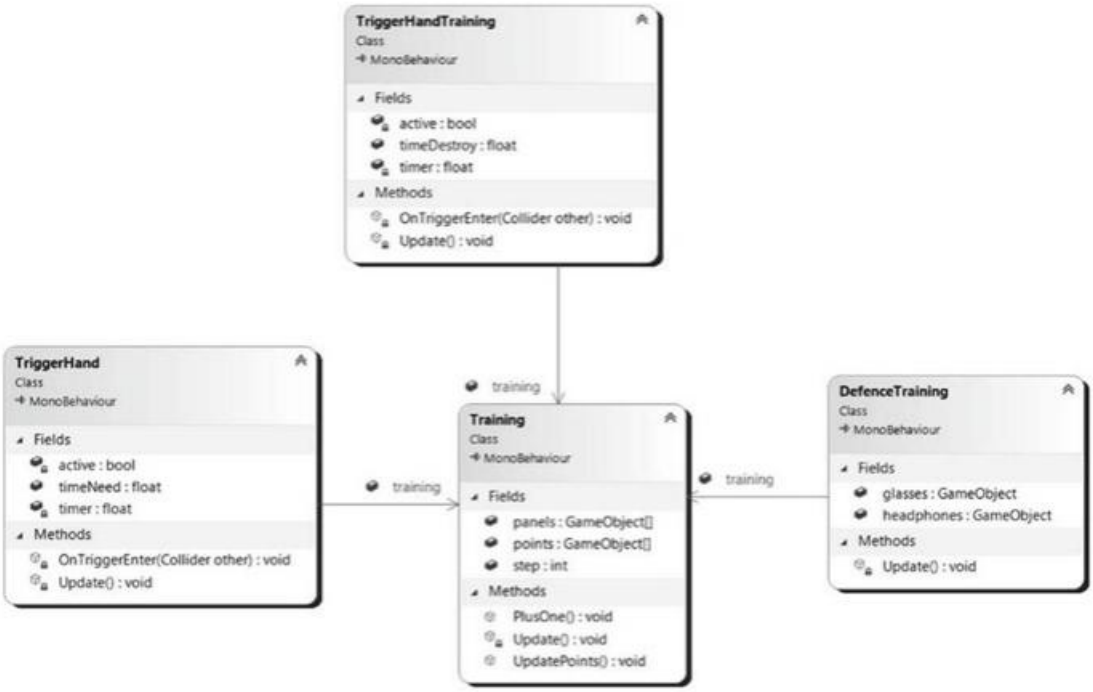
d. CylindController — klasa odpowiedzialna za obróbkę części.

e. CylCamsController — klasa odpowiedzialna za wyszukiwanie kamer do zapisu części jako obrazy.

f. BackForwardCamScreenshot — klasa odpowiedzialna za działanie przedniej i tylnej kamery, które przechowują część.

Architektura kodu ogólnej funkcjonalności:

1. Klasy odpowiedzialne za trening przedstawiono na rysunku 3.



a. Training — klasa odpowiedzialna za zmianę panelu treningowego.

b. TriggerHand — klasa odpowiedzialna za śledzenie wykonywania wskazówek poprzez śledzenie dłoni.

c. TriggerHandTraining — klasa odpowiedzialna za śledzenie wskazówek poprzez śledzenie palca wskazującego.

d. DefenceTraining — klasa śledząca zakładanie osłony.

2. Klasy odpowiedzialne za interfejs suwaka

a. HandController — klasa odpowiedzialna za naciskanie kontrolerów.

b.SliderControllerTrigger — klasa odpowiedzialna za zmiany suwaka.

3. Klasy odpowiedzialne za ruch gracza i laser

a. Move — klasa, która płynnie porusza graczem.

b. MoveController — klasa wybierająca rodzaj ruchu.

c. Laser — klasa kontrolująca laser.

d. SceneManage — klasa odpowiedzialna za zmianę sceny.

4. Klasy odpowiedzialne za system błędów

a. ErrocChecker — klasa monitorująca błędy.

b. ErrorManager — klasa ostrzegająca o błędzie.

5. Klasy odpowiedzialne za mikrometr

a. MekrometrTrigger — klasa odpowiedzialna za pomiar.

b. Micrometr — klasa obsługująca pomiary.

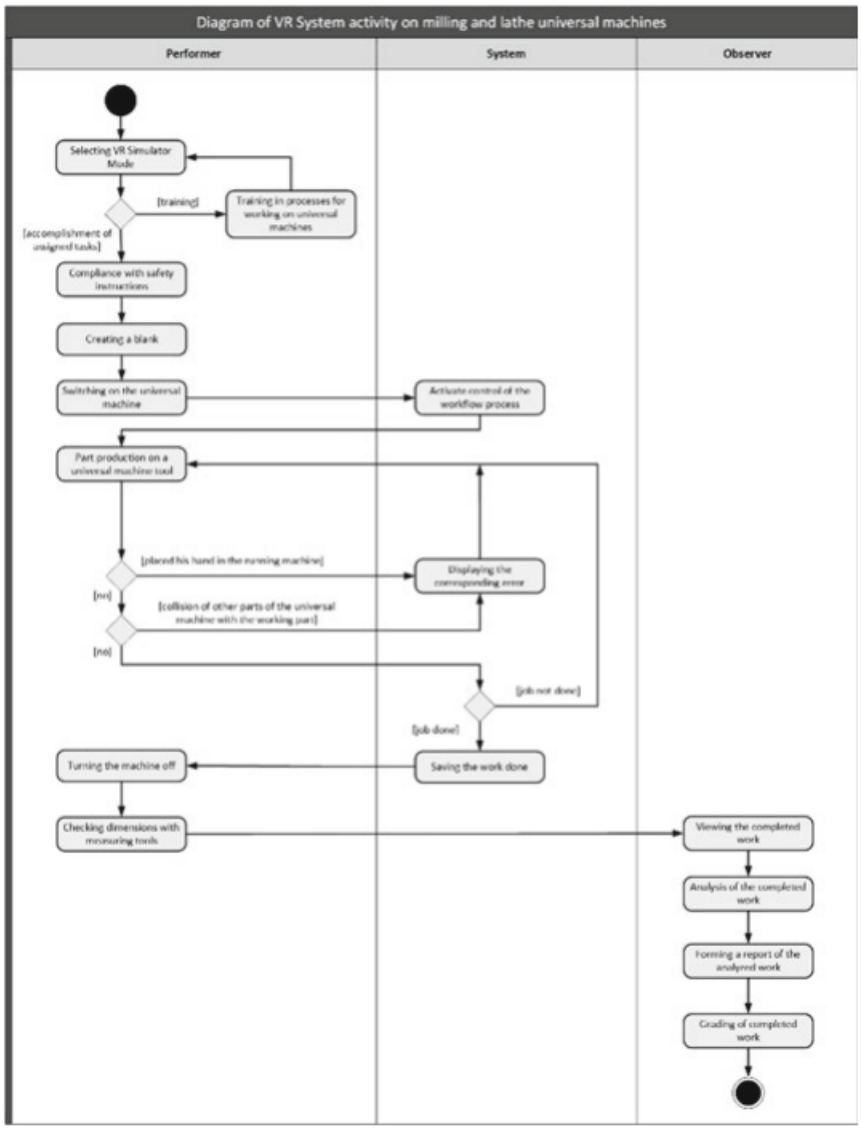
6. Klasy odpowiedzialne za zapisanie części

a. SaveScrin — klasa odpowiedzialna za zapisanie obrazu części.

b. CylindrController — klasa odpowiedzialna za część tokarki.

4. Rozwój funkcjonalnego wsparcia kompleksu programowo-sprzętowego stanowiska

Główną funkcjonalnością systemu jest realizacja dwóch trybów: szkoleń na obrabiarkach uniwersalnych oraz realizacji zadań produkcyjnych. System z kolei kontroluje realizację zadań przez Wykonawcę od momentu włączenia maszyny uniwersalnej do momentu wyprodukowania detalu. W przypadku nieprawidłowych działań Wykonawcy, System wyświetla odpowiednie powiadomienia o błędach. Następnie Obserwator może przeglądać wykonaną pracę i przeprowadzać pomiary w celu oceny jej jakości. Następnie tworzony jest raport i wystawiana jest ocena za wykonaną pracę. Demonstracja sekwencji stanów i działań systemu wraz z opisem głównych funkcji systemu rzeczywistości wirtualnej w pracy na uniwersalnych obrabiarkach frezarskich i tokarskich jest przedstawiona na schemacie działania na rysunku 4.



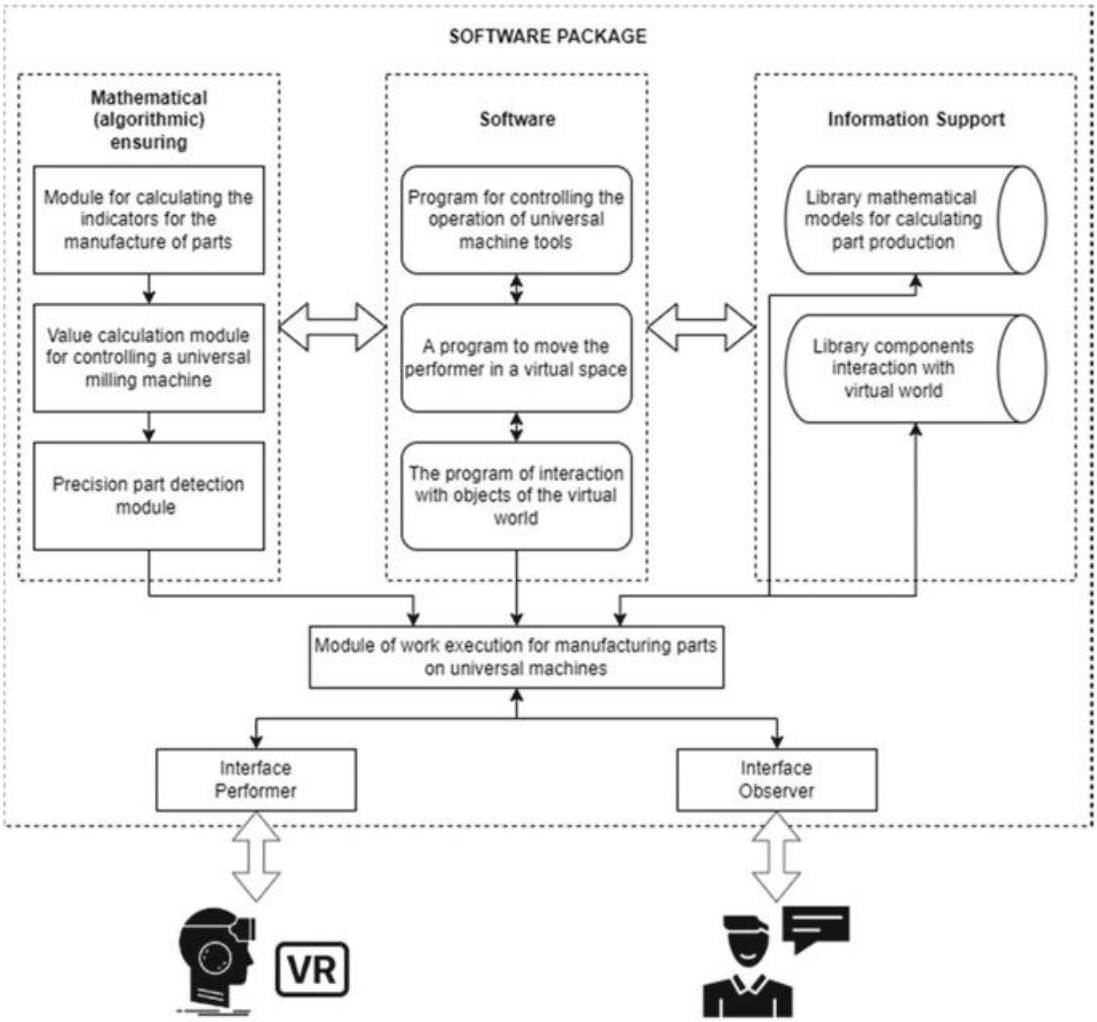
Do rozwiązania wymaganych zadań wykorzystano następujące systemy wsparcia:

1. Wsparcie matematyczne (algorytmiczne) – zestaw metod matematycznych, modeli i algorytmów przetwarzania informacji, wykorzystywanych do kształtowania oddziaływań zarządczych na maszynach uniwersalnych, a także do rozwiązywania problemów obliczania wskaźników jakości procesu wytwarzania części.

2. Oprogramowanie – zestaw programów niezbędnych do wdrożenia funkcjonalności środowiska wirtualnego.

3. Oprogramowanie informacyjne – system klasyfikacji i kodowania informacji, baz danych i tablic danych, które są niezbędne do realizacji funkcji realizowanych w projekcie.

Proponowana struktura funkcjonalna systemu rzeczywistości wirtualnej dla uniwersalnych obrabiarek frezarskich i tokarskich jest przedstawiona na rysunku 5.



Symulator toczenia i frezowania w rzeczywistości wirtualnej prezentuje aplikację z dwoma trybami: egzaminacyjnym i szkoleniowym z wyskakującymi okienkami. Użytkownik określa dane trójwymiarowego detalu, który zostanie poddany dalszej obróbce na maszynach i zmierzony za pomocą specjalnego narzędzia pomiarowego. Modele 3D sceny symulatora są opracowywane zgodnie z ogólnie przyjętymi zasadami, które przedstawiono poniżej:

1. Opracowanie koncepcji modelu, szkicu i inżynierii projektowej w programie Adobe Photoshop.

2. Opracowanie modelu w specjalistycznym oprogramowaniu Autodesk 3ds Max, Blender, Maya, ZBrush.

3. Retopologia, optymalizacja modelu, redukcja wielokątów.

4. Stworzenie przeciągnięcia wszystkich części siatki w celu poprawnego nałożenia tekstury.

5. Scalanie map w programie Substance 3D Painter.

6. Teksturowanie modelu w programie Substance 3D Painter.

7. Animacja i animacja kości dla niektórych modeli, w tym riggowanie i nakładanie skórek dla postaci.

Optymalizacja aplikacji VR. Aby zapewnić prawidłowe działanie aplikacji w środowisku wirtualnej rzeczywistości, prace przeprowadzono zgodnie z następującymi zasadami:

1. zoptymalizowano geometrię poziomu – całkowita objętość modeli na scenie nie przekracza 100 000 wielokątów;

2. zoptymalizowano oświetlenie statyczne – w celu zmniejszenia obciążenia obliczono mapy świetlne obiektów statycznych;

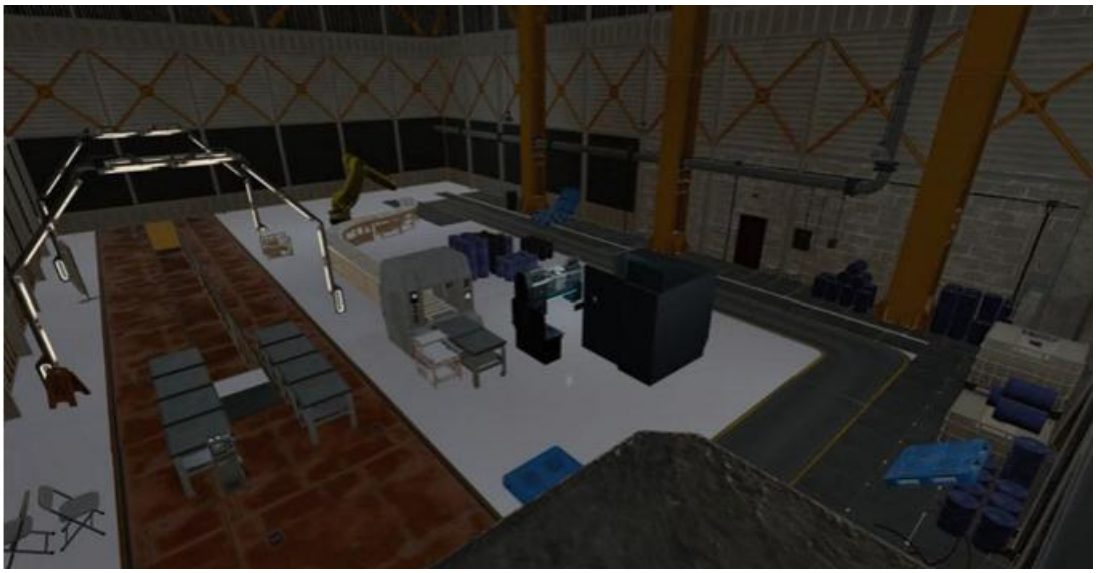
3. OcclusionCulling – w programie tworzone są tylko modele 3D znajdujące się w polu widzenia użytkownika;

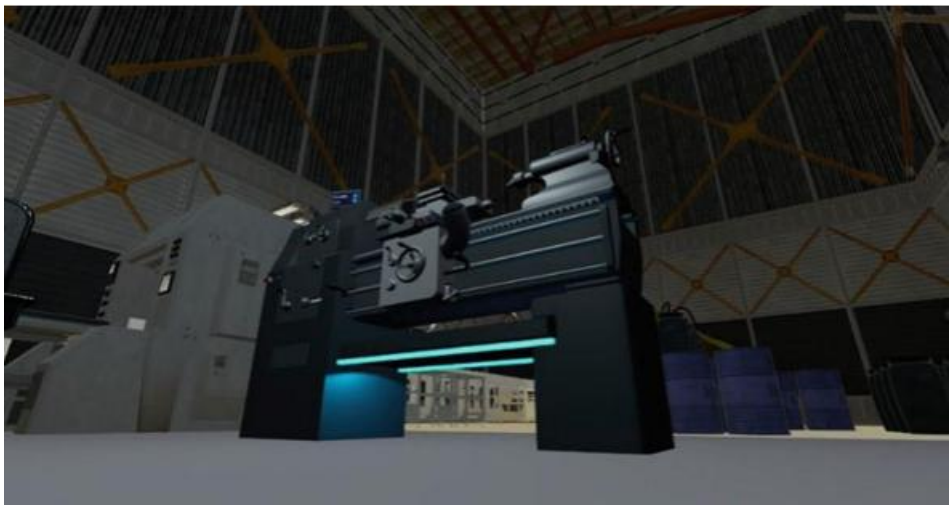
4. ReflectionProbes – wykorzystywane do realistycznego odbicia obiektów na scenie;

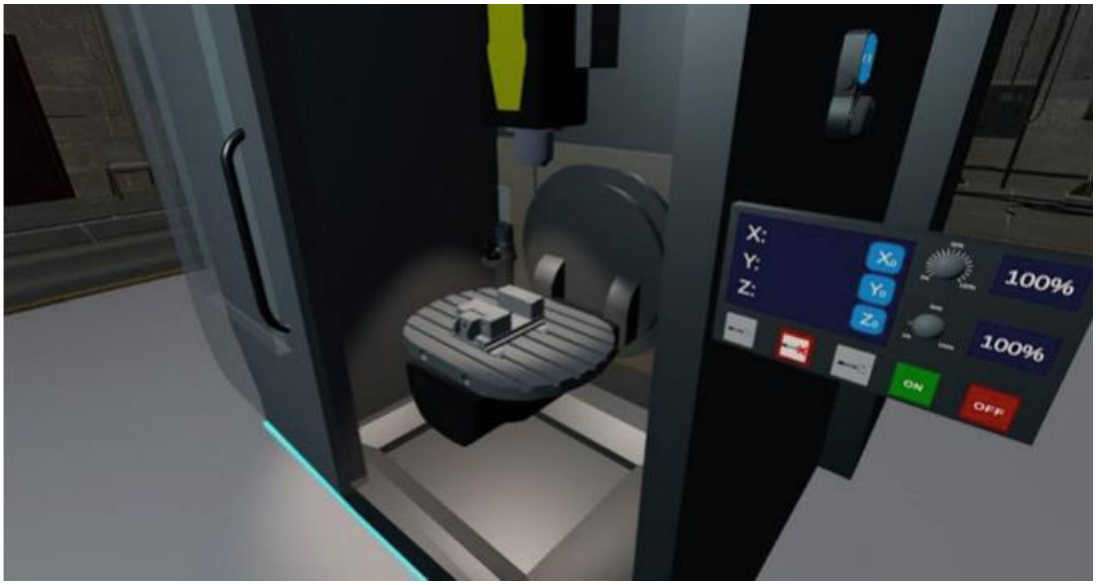
5. Wykorzystanie GPU-Instance do masowego przetwarzania obiektów 3D na scenie;

6. Wykorzystanie Batching Calls – wykorzystanie karty graficznej w celu zmniejszenia obciążenia procesora.

Narzędzia i sprzęt programistyczny. Do stworzenia symulatora wykorzystano wieloplatformowe środowisko programistyczne Unity z połączeniem edytora kodu Visual Studio w językach C# i synchronizacją ze sprzętem wirtualnej rzeczywistości opartym na narzędziu Steam VR. Treść graficzną i część wizualną opracowano przy użyciu programów Autodesk 3ds Max, Blender, Maya, ZBrush, Substance 3D Painter i Adobe Photoshop. Zarządzanie treścią oparto na sprzęcie wirtualnej rzeczywistości HTC Vive, a interakcja z obiektami odbywa się za pomocą kontrolerów. Podczas fazy testowej zidentyfikowano pewne ukryte wady oprogramowania i wprowadzono odpowiednie zmiany w celu ich wyeliminowania. Opis interfejsu UI/UX. Interfejs użytkownika aplikacji został zaimplementowany i dostosowany do technologii wirtualnej rzeczywistości. W menu głównym użytkownik znajduje się w wygodnej odległości od interfejsu i korzysta ze wskaźnika laserowego. Interfejs wykorzystuje głównie czytelne ikony, minimalizując część tekstową. Główny blok aplikacji wykorzystuje palec wskazujący do interakcji z przyciskami, co zapewnia większą precyzję. Animacje poprawiają wizualizację interfejsu. Kolor interfejsu jest niebieski. Interakcja z maszynami została dostosowana i uproszczona do technologii rzeczywistości wirtualnej, aby umożliwić wygodną obsługę tokarki i frezarki, tzn. konfiguracja maszyny odbywa się za pomocą klikalnego interfejsu, a nie za pomocą przekrętów/dźwigni.







5. Wnioski

Opracowany zautomatyzowany system szkoleniowy do uniwersalnego toczenia i frezowania w środowisku wirtualnej rzeczywistości z efektem immersji ma istotne praktyczne zastosowanie w

realizacji efektywnych szkoleń w przedsiębiorstwach przemysłowych, optymalizacji kosztów fizycznych i finansowych oraz minimalizacji liczby błędów i wypadków. Produkt programowy stanowi funkcjonalną symulację działania uniwersalnej tokarki i frezarki z pełnym pakietem funkcji identycznych z oryginalnym sprzętem. Uniwersalna tokarka z cyfrowym wyświetlaczem i zestawem narzędzi – frezów i wierteł, których sterowanie programowe zapewnia ruch uchwytu narzędzia poprzez obrót ramion dzięki obrotowi sterownika. Na tej podstawie symulowane są ruchy wykonywane na rzeczywistej tokarce. Uniwersalna frezarka z systemem kontroli położenia i cyfrowym urządzeniem wskazującym z zestawem narzędzi skrawających, posiadająca mechanizmy i metody obróbki zbliżone do sterowania rzeczywistym sprzętem. Zautomatyzowany system szkoleniowy nie ma ograniczeń wiekowych dotyczących tolerancji, w przeciwieństwie do fizycznych uniwersalnych frezarek i tokarek, które należą do kategorii podwyższonego ryzyka. Istotnym czynnikiem efektywności ekonomicznej jest brak możliwości uszkodzenia sprzętu, konieczność posiadania specjalnie wyposażonego pomieszczenia, naliczanie amortyzacji, koszty konserwacji oraz materiały eksploatacyjne: frezy, przecinaki, wiertła i detale. Należy zauważyć, że podczas praktyki „PM.03 Rozwój procesów technologicznych montażu podzespołów i wyrobów w produkcji podzespołów mechanicznych" jeden ze studentów zużywał detal w postaci „walców stalowych" o średnicy 60 m i długości 2 m. Po wdrożeniu symulatora VR udało się zmniejszyć materiał eksploatacyjny z 2 do 0,5 m. Tym samym udało się czterokrotnie zmniejszyć ilość półfabrykatów. Zmienność scenariusza pozwala na wykonywanie różnorodnych operacji w aplikacji. Zautomatyzowany system szkoleniowy do uniwersalnego toczenia i frezowania w środowisku rzeczywistości wirtualnej został wdrożony i jest wykorzystywany w Państwowej Instytucji Budżetowej Edukacji Zawodowej Miasta Moskwy „Moskiewski Państwowy Kompleks Edukacyjny". Nowością proponowanego projektu jest:

• model szkolenia zawodowego i oceny kwalifikacji, umożliwiający ocenę procesów pracy i umiejętności zawodowych studentów;

• specjalistyczny pakiet oprogramowania z unikalnymi modułami programowymi;

• metoda optymalizacji modułów programowych i modeli trójwymiarowych kompleksu programowego szkoleń w segmencie przemysłowym w jednej przestrzeni wirtualnej.

Rozwój oprogramowania do organizacji szkoleń pracowników elektrowni cieplnych (TPP) w oparciu o cyfrowe bliźniaki urządzeń

Omówiono zadanie opracowania oprogramowania wykorzystującego cyfrowe duplikaty urządzeń do organizacji szkoleń personelu elektrowni cieplnych (TPP). Rozwiązanie to może być również wykorzystane w procesie kształcenia na wyższych uczelniach o specjalnościach inżynierskich. Aby rozwiązać problem wdrożenia prototypowego oprogramowania, opracowano matematyczne i algorytmiczne wsparcie budowy modeli matematycznych różnych urządzeń do wykorzystania w cyfrowych bliźniakach urządzeń i schematów cieplnych elektrowni cieplnych. W toku rozwoju oprogramowania stworzono następujące modele matematyczne: model cylindra turbiny parowej, model podgrzewacza niskociśnieniowego typu powierzchniowego, model skraplacza. Zsyntetyzowano algorytmiczne wsparcie budowy modeli do wykorzystania w systemach szkolenia personelu elektrowni i organizacji procesów kształcenia na wyższych uczelniach o

specjalnościach inżynierskich. Symulacja zjawisk negatywnych, zgodnie z wcześniej zaprojektowanym scenariuszem, pozwala na realizację procesu kształcenia z wysoką efektywnością. Efektywność warunkowana jest nie tylko możliwością generowania różnych trybów pracy urządzeń, ale także uwzględnieniem wpływu różnych czynników negatywnych na ich funkcjonalność. Narzędzia do tworzenia scenariuszy sytuacji nietypowych: ustalanie wartości określonych parametrów, jednokrokowa, skokowa zmiana parametrów, tworzenie liniowej funkcji zmiany parametrów fizycznych obiektu w czasie. Wybrano język programowania do realizacji oprogramowania symulatora.

Słowa kluczowe Cyfrowy bliźniak · Elektrownia cieplna · Model symulacyjny · Schemat cieplny · Szkolenie personelu · Sytuacje awaryjne · Adaptacja

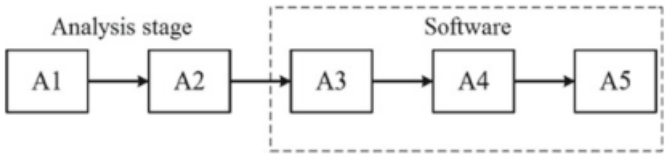
1 Wstęp

Jednym z problemów współczesnej energetyki jest spadek efektywności energetycznej obiektów wytwórczych w trakcie ich eksploatacji . Dlatego rozwiązanie problemu szkolenia personelu do pracy w niestandardowych, nietypowych trybach pracy urządzeń może znacząco poprawić efektywność energetyczną elektrociepłowni. Ze względu na fakt, że w trakcie eksploatacji obiekty energetyczne są narażone na wiele negatywnych wpływów, ich parametry ulegają znacznym zmianom w czasie . To zjawisko prowadzi do zmiany zachowania obiektu w czasie, co może być mylące dla personelu operacyjnego. Niepodejmowanie technicznie poprawnych decyzji może prowadzić do wypadków i sytuacji poprzedzających wypadki [8–12].

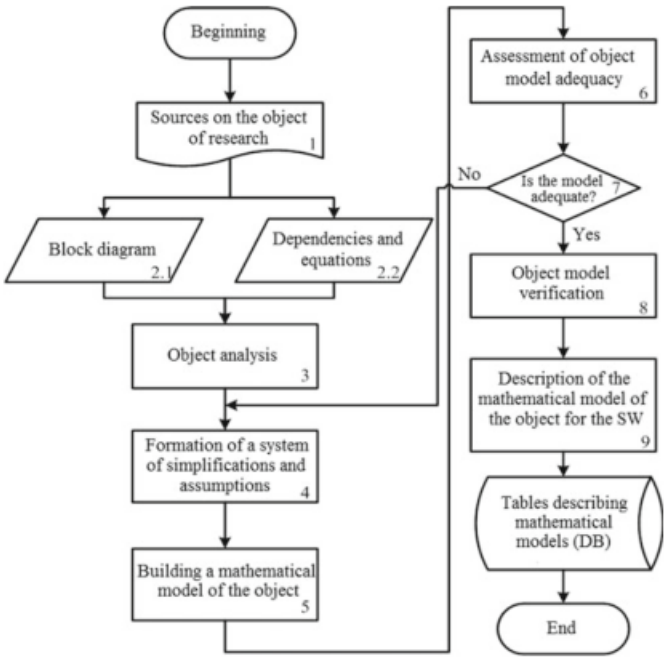
Szkolenie personelu na rzeczywistych urządzeniach jest zbyt kosztowne, dlatego sensowne jest wprowadzenie modeli cyfrowych, które mogą symulować pracę urządzeń w różnych trybach pracy z wymaganą dokładnością. W związku z tym coraz większego znaczenia nabiera zadanie rozwoju i zastosowania cyfrowych systemów symulacyjnych w celu tworzenia symulatorów opartych na cyfrowych bliźniakach, służących do szkolenia studentów w placówkach edukacyjnych oraz specjalistów w zakładach energetycznych.

2. Opracowanie algorytmów

Oprogramowanie do budowy modeli urządzeń cyfrowych bliźniaków opiera się na odpowiednich metodach algorytmicznych i matematycznych. Przed zbudowaniem modeli matematycznych jednostek urządzeń TPP i ich wdrożeniem w oprogramowaniu, konieczne jest opracowanie algorytmów, które pozwolą na wdrożenie koncepcji budowy cyfrowych bliźniaków w urządzeniach TPP, gdy będą one wykorzystywane jako część opracowywanego oprogramowania. Algorytmy zapewniające rozwiązanie zadania składają się z następującego zestawu algorytmów: A1 — algorytm budowy modelu matematycznego obiektu (jednostki urządzeń); A2 — algorytm budowy zależności funkcjonalnych dla cyfrowego bliźniaka obiektu (jednostki urządzeń); A3 — algorytm budowy schematu procesu technologicznego; A4 — opracowanie scenariuszy szkoleniowych; A5 — realizacja scenariuszy (sytuacje awaryjne, kryzysowe, przedwypadkowe itp.). Uogólniony schemat podejścia koncepcyjnego do budowy i zastosowania modeli cyfrowego bliźniaka jako elementu oprogramowania przedstawiono na rysunku



Część kluczowych algorytmów (A1 i A2) jest implementowana w fazie wstępnej (fazie analizy), a ich wynik stanowi podstawę do budowy cyfrowego bliźniaka w oprogramowaniu. Ponadto algorytmy te będą udoskonalane w miarę postępu prac. W obecnej fazie mają one charakter koncepcyjny i są reprezentowane przez uogólnione diagramy strukturalne. Na etapie wstępnym konieczne jest zbudowanie modelu matematycznego konkretnego urządzenia. W tym celu syntetyzuje się algorytm,



który implementuje klasyczne podejście do budowy modelu matematycznego obiektu badawczego w postaci równań bilansu cieplnego i materiałowego, równań różniczkowych i innych. Blok 1 dokonuje wyboru źródeł informacji o obiekcie, dla którego ma zostać zbudowany model matematyczny, zapewniając realizację koncepcji budowy cyfrowych bliźniaków urządzeń. Źródłami tymi mogą być zarówno konkretne schematy technologiczne i cieplne zakładów, jak i inne źródła, które dostarczają różnorodnych opisów procesu technologicznego. Na podstawie analizy źródeł wybierany jest schemat strukturalny (blok 2.1), którego cyfrowa kopia zostanie zaimplementowana w oprogramowaniu, a także wszystkie zależności matematyczne, charakterystyki energetyczne i inne zależności funkcyjne (blok 2.2), które zapewniają budowę cyfrowego bliźniaka obiektu w odniesieniu do natury rozwiązywanego problemu. Blok 3 ma na celu analizę obiektu pod kątem możliwości matematycznego, funkcjonalnego lub eksperckiego opisu procesów i obiektów schematu cieplnego, tak aby w bloku 4 powstał system uproszczeń i założeń, na podstawie którego zbudowana zostanie konstrukcja modelu obiektu, który następnie zostanie zaimplementowany w produkcie programowym i będzie stanowił matematyczną podstawę cyfrowego bliźniaka. W bloku 5, w oparciu o przyjęty system uproszczeń i założeń, dostępny schemat oraz zależności matematyczne i funkcyjne, zbudowany zostanie model matematyczny obiektu. W oparciu o klasyczne metody statystyczne (z wykorzystaniem kryterium Fishera, podejście zostanie opisane szerzej na etapach oceny adekwatności modelu), oceniana jest adekwatność zbudowanego modelu matematycznego. Jeśli model nie jest adekwatny do opisywanego obiektu (sprawdź w bloku 7), sprawdzany jest system uproszczeń i założeń oraz zależności zawarte w modelu (przejście do bloku 4). W przeciwnym razie model jest weryfikowany

i opisywany w wymaganym formacie, który można łatwo zaimplementować w języku programowania. Metody zastosowane w blokach syntetyzowanego algorytmu: blok 1 — dobór źródeł dotyczących tematu badań; bloki 2–3 — analiza tekstowa wybranych źródeł; blok 4 — analiza tekstowa, ankieta ekspercka; blok 5 — metody budowy modeli matematycznych, równania bilansu cieplnego i materiałowego, równania różniczkowe, charakterystyki energetyczne; blok 6 — metody statystyczne oceny adekwatności (ocena oparta na kryterium Fishera); blok 8 — ocena błędu średniokwadratowego, dokładności obliczeń modelu; blok 9 — budowa opisu na podstawie wygenerowanego szablonu. W niniejszym rozdziale przyjęto następujący format (szablon) modelu matematycznego obiektu, który będzie wynikiem opisanego algorytmu A1:

(1) parametry wejściowe: X1 - ...; X2 - ...; ...; Xn - ...

(2) parametry wyjściowe: Y1 - ...; ...; Ym - ...

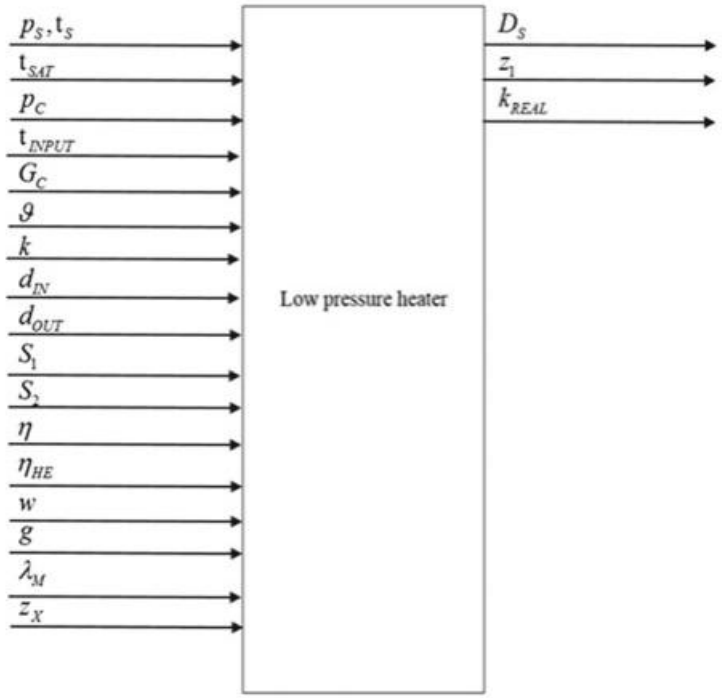
(3) do obliczenia parametru Y1 stosujemy następującą zależność matematyczną: Y1 = f(X1, ...,Xn).

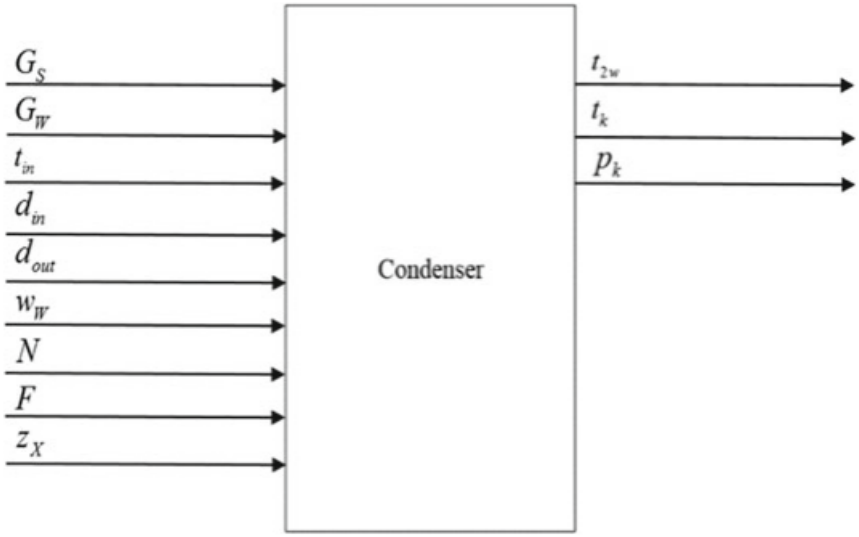
Bazując na opisanym algorytmie, w przyszłości zostaną zbudowane modele jednostek sprzętowych prototypu oprogramowania będącego w trakcie opracowywania.

3. Opracowanie modeli matematycznych dla symulatora personelu

Ponieważ system elektroenergetyczny zawiera zbiór oddzielnych elementów połączonych w określony sposób, model matematyczny systemu musi odtwarzać wszystkie relacje i powiązania w obrębie obiektu, dotyczące relacji wszystkich elementów lub wybranych grup elementów, rozpatrywanych w tym przypadku jako podsystemy . Przy modelowaniu systemów elektroenergetycznych rozróżnia się przypadki, w których podobieństwo ustala się dla wszystkich elementów wpływających na badane funkcje, występujących zarówno w czasie, jak i w przestrzeni (pełne podobieństwo), oraz przypadki, w których podobieństwo ustala się tylko dla części procesów lub funkcji badanego systemu (częściowe podobieństwo), na przykład gdy bada się zmianę parametrów procesu tylko w czasie, bez uwzględniania odpowiadających jej zmian w przestrzeni . Pełne podobieństwo, a co za tym idzie, pełna symulacja systemów elektroenergetycznych, jest realizowana głównie w badaniu systemów lub pojedynczych elementów, których działanie jest istotnie związane z propagacją energii elektromagnetycznej w przestrzeni (projektowanie i badanie takich elementów systemu, jak maszyny elektryczne, transformatory, falowody, długie linie elektroenergetyczne itp.). Modelowanie niekompletne jest zazwyczaj stosowane w badaniu modów systemu elektroenergetycznego . Modelowanie matematyczne systemów elektroenergetycznych jest w praktyce realizowane poprzez tworzenie układu równań przystosowanych do rozwiązywania komputerowego, reprezentowanych w postaci algorytmów i programów, za pomocą których na komputer przekazywane są charakterystyki numeryczne procesów (w postaci wykresu lub tabeli) zachodzących w badanym systemie elektroenergetycznym . Modelowanie matematyczne systemów elektroenergetycznych jest szeroko stosowane w symulatorach i symulacjach, co jest szczególnie ważne w badaniu procesów i szkoleniu personelu. Główną zaletą stosowania symulacji w symulatorach jest możliwość uzyskania niezbędnych danych o obiekcie badań, procesie lub zjawisku bez konieczności bezpośredniej pracy z nim. Pozwala to nie tylko zaoszczędzić czas, pieniądze i koszty pracy, ale także może pomóc w ochronie podmiotu lub osób nad nim pracujących przed niepotrzebnym ryzykiem. Podczas opracowywania symulatora można uwzględnić strukturę umożliwiającą dalszą rozbudowę funkcjonalności (skalowalność symulatora wirtualnego), stworzyć system symulacji systemu wirtualnego (system obiektów współzależnych),

funkcjonalność umożliwiającą szybką podmianę elementów „badanego” obiektu, projektowanie (pobieranie plików z nośnika zewnętrznego, bez zmiany programu uruchamiającego symulator) . Symulator umożliwia zatem symulację różnych scenariuszy, takich jak na przykład scenariusz normalnej pracy urządzeń lub wystąpienia sytuacji awaryjnych. Ponadto, wykorzystując dane o możliwych sytuacjach nietypowych, podczas uruchamiania symulacji w trybie automatycznym, program symulatora może stworzyć unikalny scenariusz do zaliczenia szkolenia. Ponieważ symulator pełni funkcję edukacyjną, pojawia się pytanie, jak oceniać działania personelu w sytuacjach awaryjnych. Należy dobrać kryteria oceny działań personelu w sytuacjach awaryjnych. W naszym symulatorze proponujemy następujące kryteria: czas działań operatora; liczba błędów popełnionych przez operatora; określenie przyczyny sytuacji awaryjnej. Po zakończeniu szkolenia operator zostanie poinformowany o ocenie swoich umiejętności i/lub, w przypadku negatywnych doświadczeń, o konsekwencjach wynikających z jego nieprawidłowych działań. W toku rozwoju oprogramowania stworzono modele matematyczne cylindra turbiny parowej, skraplacza oraz podgrzewacza niskociśnieniowego typu powierzchniowego (LPH). Uogólnione struktury skraplacza i LPH przedstawiono odpowiednio na rysunkach 3 i 4.





Parametry wejściowe skraplacza to p_S, t_S — parametry pary grzewczej, t_{sat} — temperatura nasycenia w podgrzewaczu, p_c — ciśnienie ogrzanego kondensatu, t_{input} — temperatura kondensatu na wlocie do podgrzewacza, G_c — natężenie przepływu kondensatu przez podgrzewacz, θ — niedogrzanie wody w podgrzewaczu do temperatury nasycenia pary grzewczej, k — współczynnik przenikania ciepła dla podgrzewaczy z rurami mosiężnymi w kształcie litery U, d_{in} — średnica wewnętrzna rur w kształcie litery U, d_{out} — średnica zewnętrzna rur w kształcie litery U, S_1 — rozmiar przesuniętego układu rur, S_2 — rozmiar przesuniętego układu rur, η — współczynnik wypełnienia rur, η_{he} — sprawność podgrzewacza, w — prędkość ruchu wody w rurach, g — przyspieszenie swobodnego spadania, λ_m — współczynnik przenikania ciepła metalu, z_x — liczba przepływów wody.

Parametry wyjściowe skraplacza to: D_S – natężenie przepływu pary do podgrzewacza, z_1 – liczba rur wzdłuż cieku wodnego, k_{real} – rzeczywisty współczynnik przenikania ciepła. Parametry wejściowe LPH to: G_S – natężenie przepływu pary w skraplaczu, G_w – natężenie przepływu wody chłodzącej, t_{in} – temperatura wody na wlocie do skraplacza, d_{in} – średnica wewnętrzna rur, d_{out} – średnica zewnętrzna rur, w_w – prędkość przepływu wody w rurach, N – liczba ruchów, F – powierzchnia wymiany ciepła, z_x – liczba rur.

Parametry wyjściowe dla LPH to: t_{2w} — temperatura wody na wylocie ze skraplacza, k — wartość temperatury nasycenia w skraplaczu, p_k — ciśnienie nasycenia w skraplaczu.

4 Opracowywanie symulacji sytuacji awaryjnych

Właściwości obiektów w modelu matematycznym i relacje między nimi muszą być wykorzystane jako wskazówka przy opracowywaniu symulacji awaryjnej. Na przykład, aby stworzyć sytuacje awaryjne, możemy zmienić parametry obiektu jednorazowo, ustalić określone wartości lub utworzyć liniową funkcję zmiany wartości fizycznej w czasie, która będzie symulować zużycie sprzętu. Zatem ustalenie zmiennej może posłużyć do symulacji awarii zaworu regulacyjnego lub odcinającego, położenia RO pozostającego na stałym poziomie (symulującego zablokowany zawór), pęknięcia rury wodnej, poziomu wody w elemencie powinien wzrosnąć i pozostaje on stały. Nagła zmiana parametrów może również opisywać pęknięcie rury lub naruszenie szczelności obiektów. Jednak zastosowanie zmiany wielkości fizycznej w czasie, która musi

pozostać stała, może opisać korozję metalu, odkładanie się soli, ścienianie metalu itp. Wszystkie te narzędzia pozwalają opisać defekty prowadzące do awarii sprzętu i sytuacji awaryjnych.

5 Implementacja oprogramowania

Do implementacji oprogramowania symulatora wybrano język programowania wysokiego poziomu Python; pozwala on zarówno na tworzenie interfejsu, jak i modeli matematycznych. Ze względu na swoją wszechstronność i łatwość nauki doskonale nadaje się do tworzenia symulatora szkoleniowego . Uczestnicy szkolenia mogą spróbować stworzyć własną ewentualność, jeśli chcą, a dzięki funkcji biblioteki Python „maintenance" mogą załadować tę sytuację do modułu oprogramowania bez zmiany głównego kodu programu. Fragment programu przedstawiono na rysunku 5.

```
delta = 5  # начальная мера ошибки

tk = 30 + 273.15  # в первом приближении принимаем температуру насыщения в конденсаторе равной 30°C
while delta > 0.001:
    r = HSST(tk) - HSWT(tk)  # теплота парообразования
    t2a = t1a + ((Gn * r) / (Ga * 4.19))  # температура воды на выходе из конденсатора
    delta_t = (t2a - t1a) / (log((tk - t1a) / (tk - t2a)))  # среднелогарифмическая разность температур
    tcp_a = (t1a + t2a) / 2  # средняя температура в трубках
    tn = (tk + tcp_a) / 2  # температура пленки конденсата на трубках

    # по температуре пленки конденсата определяем:

    # коэффициент теплопроводности
    lambda_n = 0.045 * log(1 + tn - 273.15 - 0.845) + 0.47
    # плотность пленки конденсата
    po_n = DSWT(tn)
    # коэффициент динамической вязкости
    mu_n = (1.78 * 10 ** (-6)) / (1 + 0.0337 * (tn - 273.15) + 0.00021 * (tn - 273.15) ** 2) * DSWT(tn)
    # теплота парообразования
    delta_h_n = HSST(tn) - HSWT(tn)
    # коэффициент теплоотдачи при конденсации практически неподвижного пара на горизонтальных гладких трубках
    alfa_n = 0.728 * (
            delta_h_n * 1000 * lambda_n ** 3 * po_n ** 2 * 9.81 / (mu_n * (tk - tcp_a) * dw * 10 ** (-3))) ** 0.25

    # по средней температуре воды в трубках определяем:

    # коэффициент кинематической вязкости
    nu_a = (1.78 * 10 ** (-6)) / (1 + 0.0337 * (tcp_a - 273.15) + 0.00021 * (tcp_a - 273.15) ** 2)
    # коэффициент теплопроводности
    lambda_a = 0.045 * log(1 + tcp_a - 273.15 - 0.845) + 0.47
    # число Прандля
    Pr = 11.228 * 2.71828 ** (-0.032 * (tcp_a - 273.15)) + 1.212  # Спорный момент
    # плотность воды
    po_a = DSWT(tcp_a)

    Re = W * (dsw * 10 ** (-3) / (nu_a))  # число Рейнольдса
    alfa_tp = 0.023 * Re ** 0.8 * Pr ** 0.4 * (lambda_a / (dsw * 10 ** (-3)))  # коэффициент теплоотдачи
    Rct = 1.15 * ((dsw * 10 ** (-3)) / 29.4) * log10(dw / dsw)
    k = ((1 / alfa_tp) * (dw / dsw) + (1 / alfa_n) + Rct) ** (-1) # коэффициент теплопередачи с гладкими трубками
```

6. Wnioski

Opracowano oprogramowanie do szkolenia inżynierów. Opracowano modele matematyczne podgrzewacza niskiego ciśnienia, skraplacza i cylindra turbiny. Przeprowadzono analizę narzędzi do tworzenia scenariuszy sytuacji nietypowych i awaryjnych, ukrytych i oczywistych usterek urządzeń. Wybrano język programowania do stworzenia oprogramowania. W przyszłości opracowane oprogramowanie stanie się podstawą zestawu symulatorów szkoleniowych dla personelu TPP.

Heterogeniczny System Informacyjny do Integracji Baz Danych Departamentów dotyczących Stanu i Rozwoju Kapitału Ludzkiego

Przedstawiono wyniki prac nad opracowaniem architektury heterogenicznego środowiska informacyjnego do integracji zasobów informacyjnych przedsiębiorstw, rozwiązujących problemy inwestowania w kapitał ludzki, zapewniając dostępność aktualnych danych o stanie i historii rozwoju kapitału ludzkiego zarówno w skali przedsiębiorstw, jak i na poziomie federalnym.

Słowa kluczowe: Środowisko informacyjne · Kapitał ludzki · Integracja zasobów informacyjnych · Integracja baz danych

1 Wstęp

Rozwój gospodarki, społeczeństwa i rodziny, łączący zasoby pracy, wiedzę, narzędzia pracy intelektualnej i menedżerskiej, środowisko i aktywność zawodową, wymaga rozwoju i wdrożenia technologii informatycznych do zarządzania rozwojem kapitału ludzkiego . Postęp w rozwoju technologii informatycznych, cyfryzacja wielu sfer życia determinuje postęp w rozwoju i wdrożeniu technologii informatycznych do zarządzania rozwojem kapitału ludzkiego. Jednocześnie niezwykle ważne jest posiadanie aktualnych danych na temat stanu i historii rozwoju kapitału ludzkiego nie tylko w skali przedsiębiorstw (organizacji, instytucji), ale także na poziomie federalnym, zapewniając wdrożenie jednolitej przestrzeni informacyjnej dla kształtowania i rozwoju gospodarki innowacyjnej i gospodarki opartej na wiedzy. To z kolei wymaga syntezy heterogenicznego środowiska informacyjnego w celu integracji zasobów informacyjnych przedsiębiorstw rozwiązujących problemy inwestowania w kapitał ludzki. Znane są liczne sformułowania rozwiązania problemu stworzenia bazy informacyjnej do monitorowania i zarządzania rozwojem kapitału ludzkiego. Wiele zadań analizy kapitału ludzkiego można by rozwiązać w oparciu o kompleksowy rejestr ludności, który zawierałby informacje o każdej osobie, obejmujące wszystkie aspekty jej życia o znaczeniu społecznym. Idee stworzenia takiego rejestru pojawiły się już w latach 50. XX wieku, a pierwsze decyzje o jego wdrożeniu na szczeblu federalnym sięgają lat 70. XX wieku. Następnie decyzje takie były wielokrotnie podejmowane na najwyższym szczeblu w różnych formach, ale nie zostały w pełni wdrożone. Głównym powodem jest to, że różne departamenty dysponujące spersonalizowanymi informacjami o ludności nie zgodziły się na ich oddanie „w niepowołane ręce". Dlatego na drodze do utworzenia rejestru ludności zawsze pojawiają się nie do pokonania przeszkody technologiczne, prawne i finansowe. Jednocześnie, z dużym powodzeniem, aktywnie rozwijają się różne bazy danych resortów, gromadząc łącznie niemal wszystkie informacje niezbędne do analizy kapitału ludzkiego kraju w różnych sekcjach. Na przykład, Ministerstwo Spraw Wewnętrznych posiada bazy danych właścicieli pojazdów, Fundusz Emerytalny posiada bazę danych wszystkich pracowników wraz z ich zarobkami, urząd skarbowy posiada bazy danych dotyczące dochodów, własności nieruchomości i firm, Ministerstwo Zdrowia posiada bazy danych medycznych itd. Jednocześnie, w większości przypadków, zapewniono mechanizmy umożliwiające użytkownikom zewnętrznym (oczywiście, jeśli posiadają odpowiednie uprawnienia) uzyskiwanie informacji z baz danych. Ponadto, w wielu przypadkach ustanowiono międzyresortową interakcję elektroniczną, umożliwiającą przekazywanie niezbędnych informacji. Zatem brak kompleksowego rejestru niezbędnego do zarządzania kapitałem ludzkim na szczeblu federalnym nie oznacza, że resorty nie są gotowe do dzielenia się posiadanymi informacjami, ale właśnie, że nie są gotowe do ich całkowitego oddania i chcą pozostać ich monopolistycznymi właścicielami. Z tego jasno wynika, że w celu rozwiązania problemu monitorowania i rozwoju kapitału ludzkiego konieczne jest stworzenie systemu pozyskiwania niezbędnych informacji z dużej liczby różnych (heterogenicznych) baz danych. To z kolei wymaga rozwiązania szeregu problemów technologicznych i prawnych.

2 Cechy organizacji centrów danych

Aby monitorować kapitał ludzki, zgodnie z istniejącymi praktykami oceny wskaźników rozwoju społecznego, konieczne jest uzyskanie informacji w co najmniej trzech głównych obszarach: wskaźnik oczekiwanej długości życia uwzględnia komponenty zdrowia (działania mające na celu zapewnienie zdrowia fizycznego i psychicznego społeczeństwa, od których zależy sprawność człowieka) i demografii (długowieczność, wskaźniki średniej długości życia przy urodzeniu itp.) Wskaźnik wykształcenia opiera się na komponencie edukacyjnym (działania na rzecz wytwarzania zasobów intelektualnych i kwalifikowanych poprzez przyswajanie i wykorzystywanie zdobytej wiedzy, umiejętności i doświadczenia, dostęp do edukacji, średni oczekiwany czas trwania edukacji dzieci w wieku szkolnym, średni czas trwania edukacji osób dorosłych itp.) oraz komponencie kulturalnym (interakcja ze światem wartości materialnych i duchowych, zdolność rozumienia i oceniania różnych form kultury, a także zdolność do tworzenia własnych wartości) ; Wskaźnik dochodu narodowego brutto: godziwy standard życia, mierzony wartością dochodu narodowego brutto na mieszkańca według parytetu siły nabywczej. Systemy monitorowania powinny również uwzględniać następujące komponenty: komponent obywatelski – oznacza wypełnianie obowiązków obywatela swojego kraju i działanie jako uczestnik stosunków administracyjnych, prawnych i politycznych. Komponent duchowo-moralny opiera się na historycznie ugruntowanych, społecznie akceptowanych normach moralnych i wartościach, które zapewniają osobistą autonomię i solidarność społeczną. Komponent pracy definiuje się jako potrzebę, zdolność i gotowość do wykonywania czynności zawodowych, przestrzegania norm stosunków pracy oraz entuzjazmu wobec obowiązków pracowniczych. Optymalne wykorzystanie komponentu pracy we współczesnej gospodarce światowej jest uwarunkowane aspektami motywacyjnymi i wartościowymi, które prowadzą do mobilizacji istniejącego potencjału siły roboczej. Oczywiste jest, że trafność oceny wszystkich tych wskaźników zależy bezpośrednio od ilości informacji, na których opierają się obliczenia. W istniejących systemach informatycznych struktur publicznych i prywatnych zgromadzono już sporo danych, co może znacząco poprawić jakość obliczeń. Szybkie gromadzenie informacji stwarza coraz więcej możliwości badań statystycznych i analitycznych, ale często niezbędne informacje są rozproszone w różnych, heterogenicznych źródłach, a dostęp do ich pełnej objętości jest utrudniony. Idea rozwiązywania specjalistycznych problemów analitycznych poprzez ekstrakcję informacji w heterogenicznym, rozproszonym środowisku, gdzie zintegrowanych jest kilka niezależnych baz danych, jest aktualna od ponad trzech dekad. Wraz z pojawieniem się sieci WWW wzrosła różnorodność systemów obliczeniowych i oprogramowania do budowy lokalnych baz danych i systemów informatycznych. To z kolei stawia przed użytkownikiem zadanie połączenia w jeden system zróżnicowanych pod względem implementacji technicznej i różnych typów zasobów komputerowych (zasobów obliczeniowych, zasobów do przechowywania i przesyłania informacji) oraz przekazania całości zasobów użytkownikowi. Pierwsze próby tworzenia heterogenicznych, rozproszonych baz danych opierały się na ujednoliceniu stosunkowo podobnych systemów zarządzania bazami danych (przynajmniej pod względem organizacji danych). Na przykład, już w 1985 roku sformułowano 12 reguł Codda dla relacyjnych systemów zarządzania bazami danych, które umożliwiałyby dość łatwe budowanie rozproszonych baz danych. Jednak w praktyce twórcy systemów zarządzania relacyjnymi bazami danych nie do końca przestrzegają tak ścisłych reguł. Zadanie łączenia lokalnych baz danych w rozproszoną na poziomie transakcji straciło na znaczeniu wraz z pojawieniem się systemów zarządzania bazami danych, takich jak ORACLE, DB2 i podobne, z jednej strony, a systemów ERP z drugiej. Dalszy rozwój metod integracji systemów informatycznych w rozproszonym środowisku heterogenicznym podąża ścieżką podejścia

usługowego. W tym przypadku każdy z systemów lokalnych staje się usługą obsługującą żądania określonego typu. Systemy rozproszone są budowane z wykorzystaniem bram w postaci usług WEB, technologii architektury zorientowanej na usługi (SOA), technologii GRID itp. We wszystkich tych przypadkach medium wymiany jest środowisko internetowe, a protokoły interakcji są powiązane z transakcjami sieciowymi. Wirtualizacja i otwarte standardy pozwalają usługom wykraczać poza granice przedsiębiorstwa (organizacji). Dostawcy usług komunikacyjnych i usługodawcy odgrywają bardzo istotną rolę w zacieraniu granic między przedsiębiorstwami i kształtowaniu globalnych ekosystemów. Podczas budowy heterogenicznych systemów rozproszonych występują poważne problemy z identyfikacją podobnych obiektów podczas integracji niezależnych systemów w jednym środowisku – problemy pewności semantycznej i konflikty semantyczne. Trendem ostatnich lat jest wykorzystanie systemów przesyłania wiadomości i/lub elektronicznych systemów zarządzania dokumentacją do łączenia niezależnych baz danych w heterogeniczne środowisko. Jednocześnie takie środowisko wymiany zapewnia również funkcję monitorowania realizacji oczekujących transakcji. W tym podejściu szczególną rolę odgrywają formularze dokumentów jako podstawa interakcji z bazami danych i wiedzą. Formularze umożliwiają generowanie schematów baz danych, systematyzują lokalizację i przetwarzanie danych w systemach zarządzania bazami danych, co znacznie upraszcza tworzenie baz danych i badania analityczne. Formularz to struktura wielu ramek, elementarnych części wiedzy. Bazy danych i wiedza są budowane w oparciu o strukturę obiektów i pojęć prezentowanych w schematach baz danych; szablony, układy, formularze wyświetlania i procedury przetwarzania wiedzy prezentowanej w formularzach, a także wiedzę deklaratywną, słowniki i klasyfikatory przechowywane w systemach zarządzania bazami danych. Ważnym aspektem w budowie rozproszonych systemów informatycznych jest problem zapewnienia bezpieczeństwa i integralności dokumentów krążących w takich systemach. Rozważając te kwestie, należy mieć na uwadze nowoczesne technologie podpisu cyfrowego i technologie blockchain. Zintegrowany system heterogeniczny może łączyć różne systemy jednego przedsiębiorstwa przemysłowego, agencji rządowej lub branży, a problemy identyfikacji obiektów, kontroli przepływu informacji i dostępu do danych są dość łatwo rozwiązywane dzięki obecności jednego mechanizmu kontroli administracyjnej. W przeważającej większości przypadków metody stosowane do integracji baz danych nie są instrumentalne w tym sensie, że nie pozwalają na opisanie i wdrożenie procesu włączania nowych danych z abstrakcyjnej bazy danych do systemu. Systemy te są budowane w oparciu o specyficzne struktury danych, specyficzne wyspecjalizowane modele obiektów i procesów i nie wymagają niezbędnego poziomu zróżnicowania praw dostępu. W przypadku pracy z danymi osobowymi ten ostatni aspekt staje się nie tylko ważny, ale wręcz prawnie konieczny. Ważnym elementem systemu informatycznego, służącym do integracji resortowych baz danych dotyczących stanu i rozwoju kapitału ludzkiego, są centra przetwarzania danych . Aby takie centra mogły sprostać potrzebom menedżerów i rozwiązywanym przez nich zadaniom, rozwój baz danych wymaga systematycznego podejścia. Spełnienie tego wymogu osiąga się poprzez dobór metodologii budowy baz danych (pod względem składu, objętości i aktualizacji danych), która odpowiednio uwzględniałaby potrzeby ich użytkowników. Jednocześnie jednolitość informacyjna gromadzonych i przechowywanych danych zapewnia jednoznaczność prezentacji i zrozumienia wszystkich wykorzystywanych danych [12, 13]. Aby rozwijać i udoskonalać wsparcie informacyjne dla rozwiązywania problemów rozwoju kapitału ludzkiego, konieczne jest:

• stworzenie perspektywicznej usługi zasobów informacyjnych do zarządzania kapitałem ludzkim, określenie jej struktury, funkcji i zadań do rozwiązania; stworzenie jednolitej przestrzeni informacyjnej (SIS), określenie jej składu, a także źródeł i mechanizmów wypełniania centrum

wymiany danych, struktury i hierarchii przechowywanych informacji w celu racjonalnej organizacji informacyjnego i analitycznego wsparcia działań urzędników;

• opracowanie koncepcji systemu automatyzacji zarządzania, informatyzacji i digitalizacji zadań do rozwiązania oraz etapów jego wdrażania w celu zwiększenia efektywności rozwiązywania problemów rozwoju kapitału ludzkiego;

• opracowanie zagadnień teorii i praktyki stosowania i rozwoju cyfrowych technologii informacyjnych, doprecyzowanie ich miejsca i roli w tworzonym SIP, systemach zarządzania rozwojem kapitału ludzkiego;

• doprecyzowanie aparatu pojęciowego wdrażanej digitalizacji, zadań i problemów wdrażania SIP, a także funkcji zapewniających tworzenie perspektywicznych regionalnych i resortowych systemów informatycznych w ramach SIP dla rozwoju kapitału ludzkiego;

• opracowanie nowoczesnych podsystemów kontroli jakości działań organizacji i urzędników ds. rozwoju kapitału ludzkiego;

• wsparcie procesu decyzyjnego w zakresie zarządzania rozwojem kapitału ludzkiego;

• automatyzacja zarządzania codziennymi czynnościami specjalistów ds. księgowości, raportowania, grupowania i podsumowywania informacji, planowania przygotowywania dokumentacji sprawozdawczo-informacyjnej, analitycznej i administracyjnej.

Tak czy inaczej, rozwiązanie tych zadań jest niemożliwe bez wykorzystania nowoczesnych technologii cyfrowych, które obejmują osobiste wyposażenie techniczne specjalistów i ich miejsc pracy, a także przedstawicieli różnych grup społeczno-zawodowych. To z kolei wiąże się z rozwojem metod i środków przechowywania informacji, technologii internetowych, oprogramowania, czujników i technologii M2M – nowoczesnej interakcji czujników i elementów wykonawczych (interakcja „od maszyny do maszyny") . Technologie M2M mają na celu interakcję lub łączenie technologii telekomunikacyjnych i informatycznych. Ich zastosowanie zapewnia automatyzację różnych procesów, a także tworzenie bardziej zaawansowanych pakietów usług we wszystkich obszarach związanych z rozwojem kapitału ludzkiego. Jednocześnie obserwuje się stały trend rozwoju elektronicznych inteligentnych urządzeń, które mogą ze sobą współdziałać. Sugeruje to, że działania w zakresie rozwoju kapitału ludzkiego, mające charakter interdyscyplinarny, będą w najbliższej przyszłości wiązać się z wykorzystaniem technologii M2M, korporacyjnych technologii chmurowych itp. Powyższe pozwala nam rozważyć możliwość rozwoju i tworzenia usług zarządzania rozwojem kapitału ludzkiego, rozumianego jako zespół procesów zarządczych, organizacyjnych, metodologicznych i klinicznych, zapewniających skoordynowane działania specjalistów z różnych organizacji oraz zautomatyzowane elementy czujników i aktuatorów (zarówno stacjonarnych, jak i mobilnych) .

Powyższe pozwala nam stwierdzić, że istnieje już społeczny nakaz stworzenia heterogenicznego systemu informacyjnego integrującego resortowe bazy danych dotyczące stanu i rozwoju kapitału ludzkiego. Zakłada się, że w nadchodzących latach kwestia obowiązkowego monitorowania jakości życia znacznej części populacji wraz z obowiązkowym, szybkim dostarczaniem niezbędnych danych do wdrażania zarządzania rozwojem kapitału ludzkiego na poziomie organizacji, regionów, ministerstw (departamentów), a także na poziomie federalnym . Analiza literatury, uogólnienie światowych i krajowych doświadczeń w zakresie rozwoju technologii informatycznych, a także własne dane, pozwalają zidentyfikować następujące trendy: w

niedalekiej przyszłości spodziewane jest masowe wyposażenie organizacji i pracowników socjalnych w urządzenia mobilne, pełniące funkcję osobistych asystentów cyfrowych, zawierających „przewodowe" procesy, regulaminy pracy oraz obowiązki zawodowe lub zawodowe poszczególnych specjalistów w zakresie rozwoju kapitału ludzkiego. Zintegrowana osobista karta elektroniczna – zestaw elektronicznych akt osobowych dotyczących jednej osoby, gromadzonych, przesyłanych i wykorzystywanych przez kilka organizacji. Akta zawarte na zintegrowanych osobistych kartach elektronicznych mogą być przechowywane zarówno centralnie, jak i rozproszone. W obu przypadkach dostęp do indywidualnych elektronicznych akt osobowych odbywa się za pośrednictwem numeru identyfikacyjnego. Obecnie opracowywane są kwestie zarządzania zintegrowanymi osobistymi kartami elektronicznymi, przechowywania w nich informacji (w istocie w centrach wymiany danych), praw dostępu oraz standardów wymiany informacji i kompatybilności danych (interoperacyjności). Ogólne wymagania dotyczące zintegrowanych osobistych kart elektronicznych powinny zostać sformułowane w odrębnej normie krajowej. Jednocześnie, pomimo pojawienia się aplikacji mobilnych i gadżetów, a także odpowiedniego doświadczenia we wdrażaniu technologii sensorowych w wielu sektorach w celu zapewnienia dobrobytu społeczno-ekonomicznego ludności, w istniejących systemach informatycznych występuje niewiele wdrożeń tych ostatnich, a złożone rozwiązania systemowe praktycznie nie istnieją. Wielu autorów, firm i organizacji rozwojowych proponuje poważną reorganizację systemu monitorowania kapitału ludzkiego. Propozycje obejmują rozwój (tworzenie i utrzymanie), eksperymentowanie, testowanie praktyczne i wdrażanie w codziennej działalności następujących narzędzi,:

• technologie cyfrowe (sieciocentryzm, centra przetwarzania danych, mobilność lekarzy i pacjentów, systemy cyberfizyczne jako ideologia interakcji czujników i aktuatorów) oraz stały rozwój usług cyfrowych;

• urządzenia do rejestracji, gromadzenia i przetwarzania danych osobowych (wdrożenie w praktyce nowoczesnych czujników, aktorów, kontrolerów itp.);

• repozytorium cyfrowych modeli usług ekosystemu do zarządzania rozwojem kapitału ludzkiego, a także stanowisko głównego projektanta; Zapewnienie mobilności ludności poprzez wyposażenie jej w urządzenia mobilne i agentów programowych połączonych z „chmurą" ekosystemu w celu zarządzania rozwojem kapitału ludzkiego;

• wdrożenie idei sieciocentryzmu i zapewnienie szybkiej migracji centrum podejmowania decyzji operacyjnych na węzły ekosystemu usług zarządzania rozwojem kapitału ludzkiego;

• zatwierdzenie i aktywacja usługi w celu tworzenia grup użytkowników korporacyjnych kart SIM według kryteriów instytucjonalnych, zawodowych, funkcjonalnych, technologicznych, terytorialnych i/lub innych;

• środki komunikacji i komunikacji (mobilność uczestników procesu rozwoju kapitału ludzkiego).

Dalszy rozwój interakcji i wprowadzenie technologii M2M, połączenie agentów programowych i aktuatorów z „chmurą" ekosystemu w celu zarządzania rozwojem kapitału ludzkiego, stworzenie wieloagentowej platformy komunikacyjnej dla rozwoju kapitału ludzkiego wymagają rozważenia problemu automatycznego gromadzenia i przechowywania informacji faktograficznych. Wdrożenie nowoczesnych technologii cyfrowych pozwoli zatem nie tylko na zorganizowanie pracy nad rozwojem kapitału ludzkiego na poziomie światowych standardów, ale także na lepsze wykorzystanie zgromadzonych informacji w celu poprawy jakości życia ludności na szczeblu federalnym (w skali krajowej).

3 Architektura heterogenicznego systemu informacyjnego do integracji baz danych wydziałowych

W wyniku złożonych badań interdyscyplinarnych w kilku dziedzinach wiedzy (analiza systemowa, informatyka, prawoznawstwo, ekonomia, socjologia, psychologia, zarządzanie) uzasadniono architekturę takiego systemu, obejmującego pięć podsystemów, a także zasady interakcji tych podsystemów .

1. Podsystem wprowadzania i rozpoznawania informacji nietekstowych jest przeznaczony do wprowadzania i analizy informacji nietekstowych z wykorzystaniem naukowego i metodologicznego aparatu widzenia technicznego, który implementuje złożony łańcuch technologiczny do pozyskiwania klatek ze strumienia wideo, ekstrakcji i analizy istotnych informacji. Algorytmy zaimplementowane w konstrukcji tego podsystemu rozwiązują problemy rejestracji strumienia wideo zawierającego informacje nietekstowe; adaptacyjnego filtrowania szumu występującego podczas rejestracji strumienia wideo; rozpoznawania strumienia wideo (wykrywanie obiektów, śledzenie obiektów, klasyfikacja obiektów); przetwarzanie strumienia wideo (przygotowanie obrazów do wykrywania poruszających się na nich obiektów).

2. Podsystem wprowadzania i przetwarzania informacji tekstowych, który zapewnia wprowadzanie dokumentów tekstowych w formie papierowej oraz import dokumentów tekstowych w formie elektronicznej.

3. Podsystem automatyzacji pracy z dokumentami i monitorowania ich wykonania, który zapewnia przeglądanie i drukowanie dokumentów; rozpoznawanie obrazów graficznych dokumentów, w tym w trybie skanowania strumieniowego; rozpoznawanie kodu kreskowego wprowadzanych dokumentów i tworzenie pliku opisu; generowanie raportów; analizę morfologiczną reprezentacji tekstowej dokumentu; indeksowanie zawartości plików; obsługę hierarchicznych słowników i katalogów. Podsystem ten powinien wspierać pracę w strukturze rozproszonej geograficznie, zapewniając przesyłanie informacji między odległymi stanowiskami pracy w formie zakodowanej .

4. Podsystem przechowywania operacyjnego i archiwalnego, który rozwiązuje problemy długoterminowego przechowywania skanów dokumentów oryginalnych. Ten podsystem implementuje bezpieczną integrację baz danych za pomocą rozproszonych rejestrów, zgodnie z projektem Peppol, a także obsługuje spójny format wymiany danych, standard szyfrowania oparty na kluczach publicznych i wykorzystuje autonomiczną infrastrukturę urzędów certyfikacji. Bezpieczna integracja baz danych jest zapewniona za pomocą rosyjskiego systemu międzywydziałowego elektronicznego zarządzania dokumentacją oraz systemu międzywydziałowej interakcji elektronicznej .

5. Podsystem danych i komunikacji, który obsługuje elektroniczną wymianę danych za pomocą aplikacji obsługiwanych przez elektroniczną wymianę danych, interfejsu elektronicznej wymiany danych oraz podsystemu. Aplikacja obsługiwana przez EDI musi przetwarzać transakcje otrzymane za pośrednictwem interfejsu EDI. Podsystem musi konwertować komunikaty EDI zgodne z międzynarodowymi standardami (EDIFACT lub X.12) na interfejs EDI i odwrotnie, a także przesyłać komunikaty za pośrednictwem sieci wymiany. Aby uzyskać dostęp do danych w czasie rzeczywistym poprzez pośredni schemat zintegrowanych baz danych, w podsystemie danych i komunikacji zaimplementowano pojedynczy uniwersalny interfejs zapytań, który umożliwia pobieranie informacji bezpośrednio ze źródłowych baz danych zgodnie z modułowym podejściem do tworzenia oprogramowania. W podsystemie danych i komunikatów zaimplementowano następujące poziomy integracji bazy danych:

• warstwa danych, dla której dozwolone są reprezentacje niestrukturalne, półstrukturalne i twarde przy użyciu ;

• poziom standardowych języków zapytań i protokołów (np. SQL, JDBC, OLE DB itp.) ;

• poziom rozszerzony słowników metadanych;

• poziom technologii magazynów danych;

• poziom oprogramowania i interfejsów użytkownika;

• poziom korporacyjnych systemów informatycznych;

• poziom usług sieciowych, który opiera się na zapewnieniu standardowego interfejsu dostępu do danych ze wszystkich baz danych połączonych organizacji za pomocą usług sieciowych opartych na standardowym protokole wymiany ustrukturyzowanych komunikatów w rozproszonym środowisku obliczeniowym (SOAP) oraz przeglądarce internetowej zapewniającej dostęp do aplikacji korporacyjnych różnych organizacji, do których dostęp jest realizowany i kontrolowany przez odpowiednie usługi sieciowe.

W podsystemie danych i komunikatów konieczne jest wdrożenie protokołu dostępu do zasobów rozproszonych Z39.50, który obsługuje pojedynczą hierarchiczną reprezentację rozproszonych zasobów informacyjnych w środowisku architektury klient-serwer i zapewnia użytkownikowi pojedynczy interfejs dostępu do nich, zapewniając dostęp do baz danych w języku SQL. Proponowane opisy i charakterystyki podsystemów funkcjonalnych ogólnej architektury heterogenicznego środowiska informacyjnego, służące integracji zasobów informacyjnych przedsiębiorstw rozwiązujących problemy inwestowania w kapitał ludzki, są nowe i umożliwiają integrację dużych zbiorów danych, zarówno testowych, jak i nietekstowych. Rozwiązanie problemu integracji baz danych w celu budowy heterogenicznych systemów informatycznych jest pilnym zadaniem, będącym w centrum zainteresowania współczesnej nauki światowej, co potwierdzają badania nad metodami integracji baz danych w budowaniu systemów z komponentów różnych architektur, w tym integracji częściowo jednorodnych baz danych oraz integracji baz danych w chmurze.

4. Wnioski

Jednym z zastosowań wyników badania jest ich wykorzystanie w ramach wsparcia informacyjnego ośrodków zdrowia – jednostek działających w trzech głównych obszarach:

• opieka medyczna w zakresie korygowania czynników ryzyka przewlekłych chorób niezakaźnych u pacjentów;

• prace organizacyjne, metodologiczne i analityczne mające na celu koordynację działań podobnych ośrodków na poziomie regionalnym;

• działania terenowe w celu przeprowadzenia kompleksowego badania ankietowego pracowników przedsiębiorstw i różnych organizacji.

Proponowane rozwiązania zapewniają interfejs systemów informatycznych ośrodków zdrowia z systemami informacyjnymi władz federalnych i gminnych – umożliwia to uwzględnienie, przy opracowywaniu i wdrażaniu zestawu działań mających na celu ochronę zdrowia ludności, zestawu cech statusu społecznego, wykształcenia zawodowego, cech warunków życia, a tym samym poprawę jakości i trafności decyzji podejmowanych w interesie rozwoju kapitału ludzkiego.

Zastosowanie modeli matematycznych o charakterze kwantowym opartych na funkcjach stanu do analizy systemów społeczno-ekonomicznych. Część 1. Wprowadzenie

Rozważane są podstawowe koncepcje prawdopodobieństwa klasycznego, prawdopodobieństwa ujemnego oraz zbiorów rozmytych, wraz z odpowiednimi analogiami do podejścia mechaniki kwantowej, w celu tworzenia modeli matematycznych do opisu zjawisk społecznych. Krótko opisano również cechy zastosowania teorii intencjonalności, która wykorzystuje pojęcie instytucjonalności do opisu zjawisk społecznych. Jednocześnie instytucjonalność rozumiana jest jako właściwość świadomości ukierunkowana na obiekty i stany rzeczy w świecie rzeczywistym. Wyróżniono pojęcie instytucjonalności zbiorowej, które definiuje fakty społeczne wynikające ze zbiorowego wpływu członków społeczeństwa. Podano pojęcie statusu, które oznacza zbiór charakterystycznych wartości podmiotu lub obiektu, determinujących jego pozycję w systemie. Opisano cechy i zakres modeli rozmytych, które służą do matematycznego opisu terminów językowych i intencjonalności w szacowaniu prawdopodobieństw zdarzeń w teorii zbiorów rozmytych. Ponadto opisano pojęcie prawdopodobieństwa ujemnego i możliwość jego wykorzystania do opisu zdarzeń społecznych. Poniżej przedstawiono przykład ilustrujący pojęcie funkcji o wartościach zespolonych i jej zastosowanie do pomiaru charakterystyk procesów społeczno-ekonomicznych. Podsumowując, wprowadzono koncepcję funkcji statusu w celu porównania jej z funkcjami przynależności i funkcjami falowymi.

Słowa kluczowe: Model matematyczny o charakterze kwantowym · Funkcja statusu · System społeczno-ekonomiczny · Teoria intencjonalności · Instytucjonalność

1 Wstęp

Zadanie tworzenia modeli matematycznych opisujących procesy społeczno-gospodarcze staje się coraz bardziej istotne. Można wyróżnić główne aspekty, które uzasadniają potrzebę tworzenia modeli matematycznych. Po pierwsze, jest to wzrost szybkości procesów społeczno-gospodarczych . Jednocześnie wzrasta ryzyko decyzji zarządczych. Po drugie, jest to znaczący rozwój technologii informatycznych. Z jednej strony zapewniają one przyspieszenie procesów społeczno-gospodarczych, z drugiej strony wymagają wzrostu zasobów mocy obliczeniowej do rozwiązywania wielu problemów. Po trzecie, jest to pojawienie się synergetyki jako nauki o ogólnych wzorcach rozwoju systemów ożywionych i nieożywionych, co umożliwia tworzenie nowych modeli systemów społeczno-gospodarczych. Początkowo próby opisu właściwości rzeczywistych zdarzeń w działalności człowieka doprowadziły do powstania teorii prawdopodobieństwa w XVII wieku. Analiza takich działów jak teoria prawdopodobieństwa, statystyka matematyczna, teoria mnogości doprowadziła do powstania teorii zbiorów rozmytych. Jednocześnie zbiory rozmyte pozwalają na wprowadzenie skali pomiarowej zjawisk społeczno-ekonomicznych i uregulowanie reguł podejmowania decyzji. Należy zauważyć, że możliwości wykorzystania zbiorów rozmytych są ograniczone. Po pierwsze, duża ilość informacji zawarta jest w intuicyjnych preferencjach, które są trudne do sformalizowania, osoby tworzącej listy szacowanych parametrów i konstruującej funkcje przynależności. Po drugie, wprowadzone korytarze oceny rozmytej są statyczne. Aby zrozumieć mechanizmy regulujące zachowanie złożonego systemu społeczno-ekonomicznego, konieczne jest zmierzenie jego zmiennych stanu i matematyczne modelowanie dynamiki każdego z elementów systemu . Tworzenie modeli może koncentrować się na różnym aparacie matematycznym. Jednym z kluczowych problemów jest złożoność uwzględniania wzajemnego wpływu poszczególnych procesów. W związku z tym trudno

jest określić, jak osiągnąć stan całego mierzonego systemu. Rozwiązania problemów opisu zjawisk w systemach społeczno-ekonomicznych nie należy szukać w aparacie matematycznym mechaniki kwantowej. Sformułowanie problemu. Omówimy koncepcje prawdopodobieństwa klasycznego, prawdopodobieństwa ujemnego, zbiorów rozmytych oraz kwantowo-mechaniczne podejście do tworzenia modeli matematycznych zjawisk społecznych.

2 Modele probabilistyczne

Ta klasa modeli jest szeroko stosowana, gdy badane czynniki są niepewne. Takie sytuacje są typowe dla różnych dziedzin ludzkiej aktywności. Przykładami są: rzut określoną liczbą na kostce , popyt na określony produkt, wynik testu, wyniki głosowania w wyborach, sytuacja polityczna w danym kraju itp. Aby zrozumieć metody probabilistyczne, warto pamiętać, że niepewność może mieć zupełnie inny charakter. W przypadku prawdopodobieństwa klasycznego fundamentalne znaczenie ma pojęcie zdarzenia losowego. Jest to zdarzenie, które może wystąpić lub nie ze względu na konkretny test. W tym przypadku test może być zarówno działaniem celowym, jak i zjawiskiem występującym niezależnie od obserwatora. W wyniku testu może wystąpić jeden ze skończonego zbioru równie możliwych wyników. Na przykład, rzucamy kostką. Prawdopodobieństwo klasyczne definiuje się jako stosunek liczby wyników sprzyjających wystąpieniu zdarzeń m do całkowitej liczby wyników n. Gdy liczba wyników jest nieskończona, wprowadza się pojęcie prawdopodobieństwa geometrycznego. Jeśli liczby m i n nie są znane z góry, konieczne jest wynalezienie nowych metod. Wprowadza się prawdopodobieństwo statystyczne dla zdarzenia, które może wystąpić lub nie ze względu na eksperyment. Dlatego możliwe jest przeprowadzenie określonej liczby eksperymentów n i określenie, ile razy wystąpiło zdarzenie A. Wówczas stosunek m/n będzie nazywany względną częstością występowania zdarzenia A w n próbach i wynosi również $\in [0; 1]$. W wielu rzeczywistych sytuacjach prawdopodobieństwo subiektywne wiąże się z określaniem prawdopodobieństwa zdarzeń, gdy jedna z powyższych metod jest niemożliwa do zastosowania. Wtedy na pierwszy plan wysuwa się wspomniane powyżej rozumienie prawdopodobieństwa jako miary wiarygodności danego zdarzenia. W takim przypadku należy przeprowadzić badanie eksperckie i na podstawie jego wyników uzyskać subiektywne prawdopodobieństwo zdarzenia. Najwcześniejsza teoria intencjonalności zajmuje się zasadami rozróżniania obiektów istniejących w rozumieniu od obiektów istniejących w rzeczywistości. Obecnie koncepcja intencjonalności jest aktywnie dyskutowana w naukach humanistycznych. Na przykład intencjonalność jest rozumiana jako zdolność umysłów do bycia w pobliżu, reprezentowania lub reprezentowania rzeczy, właściwości i stanów rzeczy. W przeciwieństwie do fizyki czy chemii, obszar przedmiotowy nauk społecznych i ekonomicznych zależy od czynników instytucjonalnych. Dzieje się tak, ponieważ niektóre właściwości otaczającego nas świata zależą od obserwatora, a inne nie. Na przykład gwiazdy i planety, prawa ich ruchu, ich czas życia nie zależą od obserwatora. Własność przedsiębiorstw przemysłowych zależy od systemu społecznych stosunków produkcji. Pieniądz papierowy, będąc papierem, na którym jest drukowany, staje się środkiem płatniczym dopiero na mocy norm prawnych regulujących procesy społeczno-gospodarcze. W przypadku braku odpowiednich umów pieniądz nie jest wart więcej niż papier, na którym jest drukowany. Można zatem założyć, że problemy poruszane w naukach humanistycznych zależą od konwencji ustanowionych przez instytucje społeczne.

3 Instytucjonalność

Searle wprowadza pojęcie instytucjonalności i rozróżnia pojęcia obiektywności i subiektywności jako właściwości stwierdzeń. Stwierdzenia obiektywne należy rozpatrywać niezależnie od odczuć i opinii osób formułujących lub interpretujących takie sądy. Sądy subiektywne natomiast zależą

od odczuć i opinii uczestników dyskusji. Pojęcie instytucjonalności obejmuje szereg zagadnień związanych z formalnymi strukturami społecznymi, takimi jak struktura państwa, oraz nieformalnymi, na przykład zwyczajami, wzorcami zachowań, rodziną. Konstrukcję rzeczywistości społecznej można przedstawić w oparciu o trzy proste zasady: intencjonalność zbiorową, funkcje przypisane, funkcje statusu. Intencjonalność zbiorowa obejmuje intencje zbiorowe, przekonania i pragnienia zbiorowe. Osoba może posiadać intencjonalność, czyli intencje, przekonania i pragnienia, takie jak intencje zbiorowe. Intencjonalność należy rozumieć jako właściwość świadomości ukierunkowaną na obiekty i stany rzeczy w świecie rzeczywistym. Zatem przekonania, pragnienia, nadzieje, lęki i emocje można zazwyczaj technicznie scharakteryzować jako intencjonalne. Pojęcie intencjonalności zbiorowej obejmuje inne formy intencjonalności, takie jak zbiorowe przekonania i pragnienia. Intencjonalność zbiorowa kształtuje fakty społeczne, które wyłaniają się ze zbiorowego działania członków społeczeństwa. Przypisane funkcje zależą od obserwatora. Ludzie często przypisują funkcje obiektom, które same w sobie nie mają żadnej funkcji, ale nabywają ją dopiero poprzez takie przypisanie. Na przykład narzędzia pełnią funkcje w oparciu o ich rzeczywiste właściwości fizyczne. Możliwe jest przypisanie funkcji obiektom i zjawiskom, które różnią się od ich właściwości fizycznych (struktury fizycznej) jako rzeczywistych obiektów. Następnie przypisanie im funkcji nadaje tym obiektom właściwości, których w rzeczywistości nie posiadają. Przypisane funkcje, które nadają obiektowi lub zjawisku właściwości nieokreślone przez jego strukturę fizyczną, nadają obiektom (podmiotom lub zjawiskom) status. Status oznacza zestaw charakterystycznych wartości inherentnych podmiotowi (lub przedmiotowi), które określają jego pozycję w systemie. Searle [18] nazywa funkcje statusu pełnione przez obiekt lub podmiot ze względu na istniejące porozumienia instytucjonalne w systemie. Przypisywanie funkcji statusu ma zazwyczaj następującą postać: „X liczy się jako Y w kontekście C” („X liczy się jako Y w C”). We wszystkich tych przypadkach X oznacza jakąś właściwość obiektu, jednostki lub sytuacji, a Y nadaje jednostce, obiektowi lub stanowi rzeczy szczególny status. W tym przypadku C reprezentuje reguły konstrukcyjne. Na przykład pieniądze, stopnie naukowe, stosunki małżeńskie, gra w szachy nie mogą istnieć bez reguł i konwencji społecznych, które je tworzą. Funkcje statusu w postaci „X liczy się jako Y w C” stanowią podstawę konstrukcji relacji społecznych.

4 Zbiory rozmyte

Matematyczny opis stwierdzeń językowych i intencjonalności w szacowaniu prawdopodobieństwa zdarzenia został opracowany przez Zadeha w teorii zbiorów rozmytych. Zadeh wprowadził założenie, że funkcja charakterystyczna zbioru może przyjmować dowolną wartość w przedziale . Funkcja ta nazywana jest funkcją przynależności dla zbioru rozmytego i stanowi podstawową koncepcję logiki rozmytej. Możliwość tworzenia modeli matematycznych przykładowego ludzkiego rozumowania, które charakteryzuje się niezwykłą zdolnością do podejmowania właściwych decyzji w środowisku niepełnych i rozmytych informacji, spowodowała ogromny wzrost zainteresowania zastosowaniem zbioru rozmytego w niemal wszystkich dziedzinach nauki i techniki. Prace L. Zadeha położyły podwaliny pod modelowanie ludzkiej aktywności intelektualnej. Typowym przykładem systemu, który dobrze nadaje się do implementacji z wykorzystaniem logiki rozmytej, jest ABS – układ zapobiegający blokowaniu kół. Istnieje wiele implementacji ABS, jednak w ogólnym przypadku sterowanie odbywa się na podstawie dwóch zmiennych wejściowych: poślizgu kół (stosunku prędkości pojazdu do chwilowej prędkości liniowej punktu na zewnętrznym promieniu koła względem jego środka) oraz przyspieszenia promieniowego koła. Obie zmienne są reprezentowane jako zmienne logiczne z zestawem 5–8 członów, z których każdy, na przykład „brak”, „słaby”, „średni”, „silny”, „bardzo silny” itd., na podstawie których kalkulator, stosując ustalone reguły (ich liczba jest równa

iloczynowi liczby członów zmiennych wejściowych), otrzymuje wartość ciśnienia w cylindrze hamulcowym, dążąc do utrzymania optymalnego poślizgu. Podobny problem rozwiązują jednak również klasyczne kalkulatory wykorzystujące trójwymiarowe tablice opisujące płaszczyznę wartości wyjściowej w zależności od dwóch wartości wejściowych. W ten sposób Lotfi A. Zadeh wprowadził funkcję przynależności, którą można uznać za słabo sformalizowany analog prawdopodobieństwa wystąpienia zdarzeń. W tym przypadku prawdopodobieństwa są rozszerzone na pewne zbiory, zwane rozmytymi. Zbiory te są zdefiniowane przez funkcje i mają dziedzinę geometryczną.

5 Prawdopodobieństwo ujemne

Skomplikujmy rozważany przypadek wystąpienia lub niewystąpienia zdarzenia. Jeśli zastosujemy klasyczną definicję prawdopodobieństwa lub częstości wystąpienia zdarzenia, mierząc stan na podstawie pewnej próbki z ogólnego zbioru zdarzeń, wówczas prawdopodobieństwo wystąpienia zdarzenia $p \in [0;1]$. Zatem prawdopodobieństwo niewystąpienia zdarzenia wynosi $1 – p$. Czy może istnieć coś więcej niż tylko niewystąpienie zdarzenia? Sukces, porażka, antysukces: (1 0 − 1)? Czy może istnieć coś więcej niż tylko niewystąpienie zdarzenia? Na przykład osoba A nie ma własnego zdania na pewien temat. Osoba B próbuje przekonać A do zaakceptowania jego punktu widzenia, ale uzyskuje odwrotny efekt komunikacji. Na przykład, ujemne prawdopodobieństwo? Jeśli użyjemy klasycznej definicji prawdopodobieństwa lub częstości wystąpienia zdarzenia, mierząc stan na podstawie pewnej próbki z ogólnego zbioru zdarzeń, to prawdopodobieństwo wystąpienia zdarzenia wynosi $p \in [0;1]$. Zatem prawdopodobieństwo niewystąpienia zdarzenia wynosi $1 − p$. Dirac powiązał ujemne prawdopodobieństwo z ilością pieniędzy. Na przykład, czekając na pensję, pożyczył pieniądze. Nie otrzymał zapłaty i został zadłużony. Wtedy okazuje się, że uniwersalny zbiór możliwych zdarzeń musi zostać rozszerzony. Takie zdarzenia, które tworzą prawdopodobieństwo niezajścia zdarzenia (nie jest negacją zajścia zdarzenia). Lub charakteryzują czynniki z nim związane z drugiej strony. W eksperymentach Bielińskiego wykazano potrzebę wprowadzenia ujemnego prawdopodobieństwa. Użyto interferometru z parametrycznymi źródłami promieniowania dla dwóch obserwatorów. Skorelowane fotony są wysyłane do dwóch obserwatorów w dwóch modach, z których jeden jest opóźniony. Następnie mody są mieszane w proporcji 50% w dzielnikach wiązki i są wykrywane przez cztery odbiorniki, po dwa dla każdego z obserwatorów. Proces losowy jest opisany czterema zmiennymi dychotomicznymi przyjmującymi pojedyncze wartości ± 1. Sugeruje się, że istnieje dodatnio określona znormalizowana funkcja rozkładu prawdopodobieństwa. Dla nich tworzonych jest $2^4 = 16$ prawdopodobieństw łącznych dla dyskretnej funkcji rozkładu prawdopodobieństwa. Tworzenie funkcji prawdopodobieństwa składa się z wyrazów, z których połowa jest ujemna. W rezultacie udowodniono istnienie dodatnio określonej znormalizowanej funkcji rozkładu prawdopodobieństwa. Jednak w praktyce wynik jest inny. Ze względu na interferencyjny charakter eksperymentu, prawdopodobieństwa na detektorach zależą od opóźnień fazowych formy harmonicznej. Łączne prawdopodobieństwa wyznaczone na podstawie eksperymentu dają wynik odpowiadający opisowi kwantowemu. Jedynym wyjaśnieniem istnienia łącznej funkcji prawdopodobieństwa uzyskanej w eksperymencie jest nieprzestrzeganie oczywistej zasady dodatniości prawdopodobieństwa. Belinsky A.V. uważa, że ujemne prawdopodobieństwo oznacza, że prawdopodobieństwo zdarzenia oczekiwanego maleje, podczas gdy prawdopodobieństwo zdarzenia przeciwnego rośnie. Prawdopodobieństwa ujemne i prawdopodobieństwa o wartościach zespolonych były wielokrotnie omawiane w literaturze.

6 Prawdopodobieństwo zespolone

Koncepcja prawdopodobieństwa ujemnego została uzyskana w mechanice kwantowej. Jednak nawet w mechanice kwantowej pojęcie prawdopodobieństwa ujemnego nie jest oczywiste, a prawdopodobieństwo ujemne jest wprowadzane jako konieczność do interpretacji niektórych eksperymentów. W tym przypadku prawdopodobieństwo jest w rzeczywistości określone przez funkcję falową o wartościach zespolonych ψ, będącą funkcją zespoloną. Eksperymentalnie można zmierzyć jedynie gęstość prawdopodobieństwa $|\psi|^2$. W ortodoksyjnym rozumieniu mechaniki kwantowej ujemne prawdopodobieństwa nie istnieją ze względu na niemożność przeprowadzenia eksperymentu w celu ich bezpośredniego zmierzenia. Zatem ujemne prawdopodobieństwo jest sprzeczne z ortodoksyjną teorią prawdopodobieństwa i ortodoksyjną mechaniką kwantową, ale nie jest sprzeczne z kilkoma eksperymentami i zdrowym rozsądkiem. Załóżymy, że wynik zastosowania tego operatora do przypisanej funkcji losowej pozwoli nam obliczyć intencjonalną charakterystykę funkcji przypisanej podmiotowi. Wyobraźmy sobie znaczenie użycia wielkości zespolonych do przypisania ich do losowych funkcji oszacowań. Zbiór możliwych stanów układu jest scharakteryzowany przez wprowadzenie zmiennych losowych w celu opisania zmiennych tworzących ten zbiór. Zmienna losowa jest losowa w tym sensie, że odpowiada rzeczywistemu stanowi układu. Na przykład, opiszmy strzelanie do celu. Zdarzeniem losowym jest trafienie w środek celu. Załóżmy, że każda zmienna losowa przypisana stanowi obiektu (lub podmiotu) składa się z faktu, że jest on Z i jego percepcji Z':

Z = Z + Z'. (1)

Następnie, aby wprowadzić system pomiarów stanów, obiekt (podmiot) musi wprowadzić uporządkowaną parę rzeczywistych zmiennych losowych S = {S1, S2} należących do różnych zestawów oszacowań obserwowanych „zdarzeń" odpowiadających pomiarowi obiektu (podmiotu). Tutaj S1 jest zmienną losową oceny obiektu (podmiotu), którą można zmierzyć w praktyce, S2 odzwierciedla percepcję zdarzenia S1 odpowiadającą urojonej części zdarzenia. Na przykład wartość binarna przyjmuje losowe wartości 0 i 1. Przypiszmy im terminy lingwistyczne interpretacji semantycznej „Prawda" i „Fałsz". Rozważmy zdarzenie trafienia w cel S1 = 1, chybienie – S1 = 0. Dla wartości S2 = 1 zdarzenie przyjmuje się jako wartość prawdziwą S1, S2 = 0, zdarzenie przyjmuje się jako wartość fałszywą S1. W parach S = {S1, S2}: (1,1) – trafienie przyjmuje się jako trafienie;

(1,0) – trafienie uznaje się za chybienie;

(0,1) – chybienie przyjmuje się jako chybienie;

(0,0) – chybienie traktuje się jako trafienie.

Możliwe jest zatem proste i wizualne wprowadzenie złożonych zmiennych losowych przypisanych do stanu obiektu (podmiotu). Zauważmy, że funkcje przypisane do stanu systemu zawsze zawierają składnik percepcji zdarzenia obserwowanego. Pokażmy celowość wprowadzania wielkości zespolonych do pomiaru stanów obiektu (podmiotu). W mechanice kwantowej problem wprowadzenia złożonego prawdopodobieństwa wynika z niemożności określenia, w którą stronę przebyła cząstka, oraz z występowania interferencji. „Przenosząc" problem wyznaczania drogi do fizyki klasycznej, można spróbować określić, przez który z równoległych przewodników przepłynął prąd elektryczny. W potocznej interpretacji można zadać pytanie: „skąd pochodzi informacja wpływająca na decyzję": z Internetu, reklamy banerowej, od znajomych, czy skądinąd? Często nie sposób wybrać jednej z nich. Zatem pytanie „w którą stronę?" może być również istotne dla wspomagania decyzji. Załóżmy, że stanowi obiektu (podmiotu) przypisana jest losowa funkcja f, która jest opisana rozkładem prawdopodobieństwa ρ(f). Często wystarczy znać wartość charakterystyczną f i przedział, w którym jej losowe wartości najprawdopodobniej się rozchodzą.

Aby uzyskać oszacowanie wartości środkowej i szerokości przedziału, zazwyczaj używa się średniej f i odchylenia standardowego od średniej σ:

$$\langle f \rangle = \int f\rho(f)\mathrm{df} \Big/ \int \rho(f)\mathrm{df}, \qquad (2)$$

$$\sigma = \sqrt{\langle f^2 \rangle - \langle f \rangle^2}. \qquad (3)$$

Obiekt (podmiot) może znajdować się w stanach A i B, które są określone przez funkcje atrybutów ψ_1 i ψ_2. Jeżeli stany te oddziałują na siebie, rozkład prawdopodobieństwa obiektu (podmiotu) opisuje się następująco:

$$P = \psi_1 + \psi_2. \qquad (4)$$

Taka reprezentacja matematyczna jest w pełni spójna z wieloma powiązanymi procesami w wewnętrznej strukturze obiektu (przedmiotu). Do opisu zachowań grup społecznych w naukach humanistycznych używa się terminu „synergia grupowa". Taki opis matematyczny odpowiada intuicyjnemu rozumieniu procesów zachodzących w strukturach społeczno-ekonomicznych. Na przykład aspektów zrównoważonej karty wyników przedsiębiorstwa (wewnętrzne procesy biznesowe, relacje z klientami, wyniki finansowe oraz uczenie się i rozwój pracowników), które są wewnętrznie powiązane. Tradycyjny pomiar stanów pozwala uzyskać następujący rozkład prawdopodobieństwa oszacowań stanów. Taka reprezentacja matematyczna jest w pełni spójna z wieloma powiązanymi procesami w wewnętrznej strukturze obiektu (przedmiotu). Do opisu zachowań grup społecznych w naukach humanistycznych używa się terminu „synergia grupowa". Taki opis matematyczny odpowiada intuicyjnemu rozumieniu procesów zachodzących w strukturach społeczno-ekonomicznych. Na przykład aspektów zrównoważonej karty wyników przedsiębiorstwa (wewnętrznych procesów biznesowych, relacji z klientami, wyników finansowych oraz uczenia się i rozwoju pracowników), które są wewnętrznie powiązane. Tradycyjny pomiar stanów pozwala uzyskać następujący rozkład prawdopodobieństwa oszacowań stanów. Stosowane metody estymacji wprowadzają błędy do mierzonych stanów, co nieuchronnie wpływa na zasady zarządzania opartego na faktach. Wprowadzenie złożonych funkcji oceny pozwoli przezwyciężyć tę rozbieżność.

7 Funkcje statusu

Wprowadzono i opisano funkcje statusu w celu odwzorowania instytucjonalnych cech procesów społeczno-ekonomicznych na problemy związane z konstruowaniem modeli matematycznych procesów społeczno-ekonomicznych. Funkcję statusu (SF) nazwiemy operacją, która ustala regułę określającą zgodność pewnej uporządkowanej pary argumentów z pewną wartością im odpowiadającą. Taką funkcję można przedstawić jako

$$(r, k) = A(r)e^{i2\pi kr} \qquad (5)$$

lub

$$\begin{aligned}(r, k) &= A(r)\cos(2\pi kr) + iA(r)\sin(2\pi kr) \\ &= A(r)\left(\cos^2(\pi kr) - \sin^2(\pi kr)\right) \\ &\quad + 2iA(r)\cos(\pi kr)\sin(\pi kr). \qquad (6)\end{aligned}$$

Funkcje statusu stanowią rozszerzenie teorii zbiorów rozmytych, tworząc kwantowo-mechaniczny odpowiednik funkcji przynależności.

7.1 Porównanie funkcji statusu z funkcjami przynależności i funkcjami falowymi

Porównajmy funkcje przynależności, funkcje statusu i kwantowo-mechaniczne funkcje falowe .

(a) Funkcje. Funkcje statusu (SF) ψ(r). Podstawowe SF stanowią ortonormalną bazę badanego systemu społeczno-gospodarczego. Rodzaj SF jest wybierany na podstawie utworzenia ortonormalnej bazy systemu oceny, pod warunkiem, że odpowiadająca mu zmienna lingwistyczna mierzonej cechy jest adekwatna do procesu oceny stanu obiektu sterowania.

(b) Znaczenie/znaczenie fizyczne. SF ma znaczenie fizyczne funkcji falowej. Kwadrat modułu SF jest definiowany jako prawdopodobieństwo, że wartość szacowanej cechy odpowiada danym współrzędnym.

(c) Normalizacja. SF ψ musi spełniać tzw. warunek normalizacji, na przykład w reprezentacji współrzędnych mającej postać:

$$\int_V \psi * \psi \mathrm{dV} = 1. \qquad (7)$$

Warunek ten wyraża, że prawdopodobieństwo oszacowania układu z daną SF we wprowadzonym systemie oceniania jest równe jedności. W ogólnym przypadku całkowanie powinno być wykonane po wszystkich zmiennych, od których SF w każdej reprezentacji zależy.

(d) Zasada superpozycji/unii. Jeżeli układ może znajdować się w stanach opisanych przez SF ψ_1 i ψ_2, to może również znajdować się w stanie opisanym przez WF $\psi_\Sigma = \sum_{i=1}^{n} c_i \psi_i$ dla dowolnego zespolonego c_1 i c_2. Superpozycja dowolnej liczby możliwych stanów jest opisana przez SF $\mathrm{SF}\,\psi_\Sigma = \sum_{i=1}^{n} c_i \psi_i$. W takim stanie kwadrat modułu c_i określa prawdopodobieństwo, że podczas pomiaru układ znajdzie się w stanie opisanym przez SF ψ_i.

(e) Dla części rzeczywistej i urojonej stosuje się dwie zmienne. Zmienną dla części rzeczywistej SF jest niezależna bezwymiarowa zmienna lub zmienne wprowadzonego układu współrzędnych. Zmienną dla części urojonej jest wynik intencjonalności. Jest to ukryta ocena kierunku zmiany stanu i energii wydatkowanej na osiągnięcie rezultatu. W przypadku osobowości jest to bezpośrednia ocena intencjonalności; w przypadku systemów społeczno-ekonomicznych jest to zgodność stanu obserwowanego ze stanem oczekiwanym.

(f) Proces pomiaru stanu. Uzyskiwanie informacji o stanie systemu poprzez przeprowadzanie czynności oceny. Wyniki pomiaru są interpretowane jako wartości oszacowanych wskaźników. Pomiar odbywa się poprzez interakcję. Podczas interakcji mierzony obiekt jest poddawany wpływowi, dlatego wyniki pomiaru są zniekształcane, a stan mierzonego systemu ulega zmianie.

7.2 Porównanie teorii zbiorów rozmytych z metodą funkcji stanu i funkcjami falowymi mechaniki kwantowej

Porównajmy teorię zbiorów rozmytych, funkcje stanu i funkcje falowe mechaniki kwantowej.

(a) Funkcje. Funkcje przynależności (FC) μ(r) są wybierane na podstawie prostoty ich reprezentacji i obliczenia, pod warunkiem, że odpowiadająca im zmienna lingwistyczna jest adekwatna do

procesu. Analogia FP—SF: używana jako wyrażenie matematyczne zamiast zmiennej (na przykład lingwistycznej), która jej nie posiada.

(b) Znaczenie/znaczenie fizyczne. FP to „nieprawdopodobny subiektywny pomiar niedokładności". Ponadto FP różni się od gęstości prawdopodobieństwa i od funkcji rozkładu prawdopodobieństwa. Czasami interpretowana jako możliwość lub użyteczność zdarzenia.

(c) Normalizacja. Zbiór rozmyty A nazywa się normalnym, jeśli istnieje takie r, że maks $\mu_A(r) = 1$. W przeciwnym razie nazywa się go podnormalnym. Należy zauważyć, że $\mu(r) \in [0;1]$.

(d) Zasada superpozycji/unii. Znaczenie jest najbliższe operacji łączenia zbiorów rozmytych, na przykład A i B, a zbiór rozmyty A∨B nazywa się FP, którego jest równe: $\mu A \vee B(r) = \mu A(r) \vee \mu B(r)$, gdzie ∨ znak maksymalnej operacji.

(e) Zmienna jest wprowadzana tylko dla części rzeczywistej — jest to zmienna bazowa; może być wybrana na podstawie warunków problemu. Nie ma analogii dla zmiennej części zespolonej SF lub WF. (f) Proces pomiaru warunków. Procedura rozmycia (wprowadzenie rozmycia) polega na ustaleniu zgodności między wartością liczbową zmiennej wejściowej rozmytego systemu wnioskowania a wartością FP.

7.3 Porównanie funkcji falowych mechaniki kwantowej z metodą funkcji stanu

Porównajmy funkcje stanu i funkcje falowe mechaniki kwantowej.

(a) Funkcje. Funkcje falowe (WF) $\psi(r, t)$ lub $\psi(r)$ – w stanie stacjonarnym odpowiadają poziomom energii. WF jest tutaj funkcją zespoloną opisującą stan układu mechaniki kwantowej. Pozwala uzyskać najpełniejszą informację o układzie, fundamentalnie osiągalną w mikroświecie. Analogia WF – SF: są harmoniczne, spełniają równanie Laplace'a. Jeśli system społeczno-ekonomiczny ma dobrze zdefiniowane stany i można wybrać stany zdefiniowane jako bazowe, to istnieje wszelkie uzasadnienie dla formalizmu matematycznego teorii kwantów. Zatem SF odpowiadają WF stanów czystych. Czyste stany układu są stanem w pełni opisanym. Stan to pełny opis układu zamkniętego w wybranej bazie.

(b) Znaczenie/znaczenie fizyczne. Funkcja WF nie ma znaczenia fizycznego. Znaczenie fizyczne kwadratu modułu WF interpretuje się jako gęstość prawdopodobieństwa (w przypadku widm dyskretnych po prostu prawdopodobieństwo) wykrycia układu w położeniu opisanym przez dane współrzędne w danym czasie.

(c) Normalizacja. Funkcja WF ψ musi spełniać tzw. warunek normalizacji, na przykład w reprezentacji współrzędnych mającej postać (7). Warunek ten wyraża fakt, że prawdopodobieństwo znalezienia cząstki o danej funkcji falowej w przestrzeni o objętości V jest równe jedności. W ogólnym przypadku całkowanie powinno być przeprowadzone po wszystkich zmiennych, od których funkcja WF w danej reprezentacji zależy.

(d) Zasada superpozycji/unii. Jeżeli układ może znajdować się w stanach opisanych przez funkcję WF ψ_1 i ψ_2, to może również znajdować się w stanie opisanym przez funkcję WF $\psi_\Sigma = c_1\psi_1 + c_2\psi_2$ dla dowolnego zespolonego c_1 i c_2. Superpozycja dowolnej liczby stanów kwantowych jest opisana przez $\mathrm{WF}\,\psi_\Sigma = \sum_{i=1}^{n} c_i \psi_i$. W takim stanie kwadrat modułu ci określa prawdopodobieństwo, że podczas pomiaru układ znajdzie się w stanie opisanym przez WF ψ_i. Zatem dla znormalizowanych WF $\sum_{i=1}^{n} |c_i|^2 = 1$.

(e) Zmienne. Dla części rzeczywistej są to niezależne zmienne WF, współrzędne cząstek lub pędy. Dla części urojonej jest to liczba falowa wprost proporcjonalna do pędu. Wybór układu jednostek umożliwia uzyskanie dokładnej równości liczby falowej i pędu. Dla światła w próżni i innym polu bezmasowym jest ona proporcjonalna do energii.

(f) Proces pomiaru warunków. Uzyskanie informacji o stanie układu poprzez przeprowadzenie eksperymentu fizycznego. Wyniki pomiaru interpretuje się jako wartości mierzonej wielkości fizycznej. Jeśli wynik pomiaru pozostaje nieznany, układ kwantowy przechodzi w stan opisany macierzą gęstości. Pomiar odbywa się poprzez interakcję. Podczas interakcji następuje oddziaływanie na mierzony obiekt, co powoduje zniekształcenie wyników pomiaru i zmianę stanu układu kwantowego.

8. Wnioski

Uzasadniono zatem wykorzystanie kwantowych modeli matematycznych opartych na funkcjach statusu do analizy systemów społeczno-ekonomicznych. Jednocześnie przeprowadzono analizę epistemologiczną pojęć prawdopodobieństwa klasycznego i ujemnego, funkcji przynależności dla zbiorów rozmytych, a także teorii mechaniki kwantowej. Podano główne zależności matematyczne dla funkcji statusu w celu porównania z funkcjami przynależności i funkcjami falowymi. Ponadto zaproponowano wykorzystanie funkcji statusu do przedstawienia instytucjonalnych cech procesów społeczno-ekonomicznych w konstrukcji modeli matematycznych procesów społeczno-ekonomicznych. Opisano również cechy teorii zbiorów rozmytych w celu porównania z metodą funkcji statusu i funkcjami falowymi mechaniki kwantowej. Ponadto, podano charakterystyki funkcji falowych mechaniki kwantowej w celu porównania z metodą funkcji statusu.

Zastosowanie kwantowych modeli matematycznych opartych na funkcjach stanu do analizy systemów społeczno-ekonomicznych. Część 2: Modelowanie oparte na funkcjach stanu

Proponuje się wykorzystanie modeli kwantowych z wykorzystaniem funkcji stanu do modelowania matematycznego i późniejszej analizy złożonych systemów społeczno-ekonomicznych. Opisano ograniczenia metod klasycznej teorii prawdopodobieństwa i statystyki matematycznej, a także teorii zbiorów rozmytych, algorytmów Mamdaniego, Suzukiego i innych do rozwiązywania podobnych problemów. Przedstawiono opis głównych założeń stosowanych w matematycznym modelowaniu obiektów społeczno-ekonomicznych opartych na funkcjach stanu. Rozważono przykłady opisujące cechy ścieżek przejść przez stany pośrednie. Przedstawiono operator przejścia systemu społeczno-ekonomicznego do różnych stanów, podobny do hamiltonianu. Wprowadzono spektrum możliwych trajektorii wirtualnych opisujących przejścia do różnych stanów. Zaproponowano model matematyczny oparty na funkcjach stanu opisujący przejście systemu do stanu mierzalnego. W proponowanym hamiltonianie pierwszy człon reprezentuje podsystem wskaźników, drugi jest analogią energii wskaźników w środowisku informacyjnym. Jednocześnie wyróżnia się człony będące analogami energii układu dwóch kontrolowanych wskaźników: interakcji, kinetycznej i potencjalnej. Przedstawiono opis wyników modelowania matematycznego oraz analizę interakcji dwóch

hipotetycznych wskaźników systemu społeczno-gospodarczego. Wskaźniki pochodzą ze statystyk wskaźników innowacyjności jednego z regionów Federacji Rosyjskiej.

Słowa kluczowe: Funkcja statusu · Model matematyczny · System społeczno-gospodarczy · Model kwantowy

1 Wstęp

Istnieją różne podejścia do opisu metodami matematycznymi procesów i zjawisk zachodzących w systemach społeczno-ekonomicznych . Historycznie rzecz biorąc, metody rachunku prawdopodobieństwa były pierwszymi, które wykorzystano do zrozumienia istoty zachodzących zjawisk i prób predykcyjnych ocen procesów w celu wsparcia procesu decyzyjnego. Rozwój metod rachunku prawdopodobieństwa i statystyki matematycznej doprowadził do powstania teorii tolerancji błędów, stworzenia algorytmów wspomagania decyzji itp. Należy wyróżnić teorię zbiorów rozmytych, algorytmy Mamaniego, Sugeno itp. Z drugiej strony, rozwój fizyki i pojawienie się fizyki kwantowej doprowadziły do powstania i rozwoju matematycznych podejść do opisu zjawisk, których wcześniej nie można było wyjaśnić metodami wcześniejszych modeli matematycznych. Opisując zjawiska społeczno-ekonomiczne, takie jak podejmowanie decyzji przez jednostkę czy zarządzanie konkurencyjnością regionów i krajów, należy podejmować decyzje w warunkach „głębokiej niepewności", działać w warunkach głębokiej niepewności, zarówno w odniesieniu do możliwych działań innych decydentów, jak i otaczającego złożonego środowiska informacyjnego. Środowisko informacyjne obejmuje badane obiekty, a także warunki informacyjne ich istnienia i kanały wymiany informacji. To środowisko informacyjne może również obejmować stany pewności decyzji. To ostatnie stwierdzenie jest bardzo ważne przy podejmowaniu decyzji zarządczych i nie jest uwzględniane w większości stosowanych modeli matematycznych. Pewność podejmowania decyzji należy przypisać ukrytym czynnikom, które wpływają na wyniki eksperymentu dotyczącego wystąpienia zdarzenia, ale nie są bezpośrednio mierzone. Na przykład, jak faza oscylacji podczas pomiaru interferencji światła. W związku z tym możliwe staje się wykorzystanie modeli matematycznych mechaniki kwantowej w badaniu zjawisk i procesów zachodzących w środowisku społeczno-ekonomicznym i informacyjnym. W badaniu procesów decyzyjnych w społeczno-ekonomicznym środowisku informacyjnym można oddzielnie wprowadzić i stosować kryteria dotyczące pierwotnej informacji oficjalnej, kanałów informacyjnych i informacji wtórnej powstającej w wyniku przetwarzania informacji pierwotnej. Możliwe jest wykorzystanie pojęć „stopnia informacji": informacji pierwszego rzędu, która łączy informację ze zdarzeniem, oraz informacji „wyższego rzędu", która odnosi się do samej informacji . Pojęcie „wiedzy powszechnej" lub „wiedzy instytucjonalnej", używane przy wprowadzaniu funkcji statusu, jest przykładem takiej informacji wyższego rzędu. Model przedstawiony w tym rozdziale wykorzystuje modele kwantowe do oceny wskaźników otoczenia społeczno-ekonomicznego w oparciu o funkcje statusu. Funkcje statusu wykorzystano do oceny stanów wskaźników. Wskaźniki są ze sobą powiązane poprzez otoczenie informacyjne. Do opisu procesu adaptacji wykorzystano aparat matematyczny kwantowej teorii pola. Proces adaptacji oszacowań wskaźników do otoczenia jest reprezentowany przez dynamikę operatorów kwantowych. Z ogólnej oceny stanu otoczenia wyodrębnia się pewne wartości, które są prawdopodobieństwami stanu danego zbioru wskaźników stanu obiektów w systemie. W tym przypadku ośrodek jest reprezentowany jako pole kwantowe. Przedstawiono scenariusze ewolucji systemu z uwzględnieniem stopni swobody otoczenia. Zaletą zastosowanego podejścia jest możliwość opisu złożonej dynamiki systemu wyłącznie w kategoriach stanów czystych. Do reprezentacji stanów czystych wykorzystano funkcje stanu wprowadzone w pracach autorów rozdziału. Funkcje stanu reprezentują maksimum informacji o stanie opisywanego systemu. Opisują one wskaźnik

w chwili początkowej. Następnie stan ten ulega zmianie pod wpływem interakcji z innymi wskaźnikami i otoczeniem. Superpozycja różnych stanów koduje ukrytą, głęboką niepewność systemu, której nie da się modelować w ramach klasycznego rachunku prawdopodobieństwa i statystyki matematycznej.

2 Obiekty społeczno-ekonomiczne i ich modelowanie w oparciu o stan

Zdefiniujemy system społeczno-ekonomiczny S jako zbiór obiektów społecznych i relacji między nimi. W tym przypadku przyjęto następujące założenia.

1. Tutaj s_1, ..., s_N to obiekty społeczno-ekonomiczne, które są podsystemami; każdy podsystem jest opisany przez swoje wewnętrzne charakterystyki. Wyniki pomiaru tych charakterystyk można opisać na podstawie wektorów bazowych przestrzeni. W tym celu wskazano regułę, która pozwala wyznaczyć dla dowolnych dwóch elementów przestrzeni ich iloczyn skalarny.

2. Stan obiektu społeczno-ekonomicznego można opisać za pomocą zbioru cech w momencie pomiaru, które są uporządkowane i tworzą wektor $|x\rangle$. Podstawowe „czyste stany" układu są określone jedynie przez kierunek wektora $|x\rangle$, a nie przez jego długość. Stany te tworzą ortonormalną bazę układu lub stan układu można opisać na ich podstawie. W tym przypadku nie ma stanów, których opis na podstawie tych wektorów jest niemożliwy. Dlatego ewolucję układu można opisać operatorem unitarnym, który opisuje jedynie zmianę kierunku wektora stanu układu. Ewolucję układu w czasie również opisuje operator unitarny:

$$|x(t)\rangle = \hat{U}(t, t_0)|x(t_0)\rangle. \qquad (1)$$

3. Jeżeli mierzony stan układu $|y_i\rangle$ można opisać na podstawie wektorów charakterystycznych stanu $|x\rangle$, to stan ten należy przedstawić jako superpozycję:

$$|x\rangle = \sum_i c_i|y_i\rangle. \qquad (2)$$

Jest to kluczowa różnica w porównaniu z tradycyjnymi pomiarami, które poddawane są obróbce statystycznej opartej na aparacie matematycznym klasycznej teorii prawdopodobieństwa. W tym przypadku przejście z jednego stanu do drugiego $|x\rangle \rightarrow |y_i\rangle$ jest zbudowane w oparciu o odwzorowania, które ustanawiają powiązania między wewnętrznymi charakterystykami opisanych podsystemów $s_i \rightarrow s_j$ (i, j = 1, ..., n). Przejście ze stanu $|x\rangle$ do stanu $|y_i\rangle$ odpowiada amplitudzie prawdopodobieństwa:

$$c_i = \langle y_i \mid x\rangle, \qquad (3)$$

która jest liczbą zespoloną, a kwadrat jej modułu jest prawdopodobieństwem przejścia $|x\rangle \rightarrow |y\rangle$:

$$P(x \rightarrow y) = \left|c_{xy}\right|^2 \equiv |\langle y \mid x\rangle|^2. \qquad (4)$$

W ten sposób tworzone są operatory projekcji opisujące odwzorowanie obserwowanego stanu układu na stany czyste układu.

4. Jeśli mierzony układ znajdował się w stanie $|x\rangle$, a po pomiarze przechodzi do zbioru stanów $|yi\rangle$, reprezentowanego przez kombinację $|x\rangle = \sum_i c_i|y_i\rangle$, to wybór $|y_i\rangle$ może być dokonany zupełnie przypadkowo. W tym przypadku prawdopodobieństwo warunkowe znalezienia układu w stanie $|y_i\rangle$:

$$w_i = |c_i|^2 = \langle x \mid y_i\rangle\langle y_i \mid x\rangle = \left\langle x\left|\widehat{P_{y_i}}\right|x\right\rangle, \quad (5)$$

gdzie $\hat{P}_{y_i}$ jest projektorem, czyli operatorem odwzorowującym mierzony stan $|y_i\rangle$ na stany czyste $|x\rangle$. Należy zauważyć, że zgodnie z matematycznym modelem stanu, baza jest wybierana losowo, wynik pomiaru może być losowy. Obserwowane wzorce są wynikiem manifestacji ukrytych parametrów układu. Jest to przejaw losowości – jednego z fundamentalnych praw natury. Zatem wartości obserwowane w praktyce, będące wynikiem pomiaru danych statystycznych dotyczących parametrów systemu społeczno-gospodarczego, powinny być przedstawione w postaci projektora-operatora odwzorowującego mierzony stan $|y_i\rangle$ na stany czyste $|x\rangle$.

5. Jeżeli przejście ze stanu $|x\rangle$ do stanu $|yi\rangle$ jest możliwe dwiema różnymi ścieżkami: ścieżką 1 i ścieżką 2, to

$$\langle y_i \mid x\rangle = \langle y_i \mid x\rangle_{\text{path 1}} + \langle y_i \mid x\rangle_{\text{path 2}} \quad (6)$$

Jeżeli ścieżki 1 i 2 odnoszą się odpowiednio do stanów „pośrednich" $|1\rangle$ i $|2\rangle$, wówczas wzór (1) można zapisać jako

$$\langle y_i \mid x\rangle = \langle y_i \mid 1\rangle\langle 1 \mid x\rangle + \langle y_i \mid 2\rangle\langle 2 \mid x\rangle, \quad (7)$$

Oznacza to, że prawdopodobieństwo P_{yi} przejścia $|x\rangle \rightarrow |y_i\rangle$ generalnie nie jest równe sumie prawdopodobieństw $P^{path\ 1}{}_{yi} + P^{path\ 2}{}_{yi}$ przejść $|x\rangle \rightarrow |y_i\rangle$, przez 1 i $|x\rangle \rightarrow |y_i\rangle$ przez 2, jak opisano na przykład w eksperymentach Bielińskiego . Prawdopodobieństwo zawiera człon interferencji, który odpowiada za możliwość niezrealizowania się pewnego zdarzenia i możliwość interakcji między podsystemami. W naukach społecznych zjawisko to jest dobrze znane i nazywane jest synergią. Stan systemu po pomiarze

$$|y_i\rangle = \frac{\hat{P}_{y_i}}{\sqrt{w_i}} e^{i\varphi}. \quad (8)$$

Zmiana różnicy faz wpływa na wzór interferencyjny. Wynik interakcji podsystemów i sposoby przejścia tych podsystemów z jednego stanu do drugiego stają się dostępne do zrozumienia, obserwacji i sterowania .

3 Reprezentacja instrukcji skoku

Stan układu oznaczymy węzłem. Niech $|x\rangle$, $|y\rangle$, . . . będzie stanem układu w węzłach, opisanym przez zbiór charakterystyk pomiaru. Po zaimplementowaniu w układzie oddziaływania $u(u_1, u_2, \ldots)$ układ przechodzi z jednego stanu do drugiego $|x\rangle \rightarrow |y\rangle$. Założymy, że czas pomiaru jest stały. Nie da się tego zrealizować. Załóżmy jednak, że przedział czasu rzeczywistego jest podzielony na N przedziałów i $N \rightarrow \infty$. Załóżmy, że operator energii jest podobny do hamiltonianu $\hat{H}$. Następnie możemy zapisać amplitudy przejścia jako sumę wszystkich możliwych ścieżek wirtualnych, co oznaczamy symbolem $\hat{A}$. Otrzymujemy:

$$
\begin{aligned}
&\left\langle x \left| -\frac{i\hat{H}t}{h} \right| y \right\rangle \\
&= \lim_{N\to\infty} \sum_{k_1,\, k_2,\, \ldots k_{N+1}} \langle x | a_{k_{N+1}} \rangle \\
&\left\langle a_{k_{N+1}} \left| -\frac{i\hat{H}t}{h} \right| a_{k_N} \right\rangle \langle a_{k_N} | \ldots | a_{k_2} \rangle \qquad (9) \\
&\left\langle a_{k_2} \left| -\frac{i\hat{H}t}{h} \right| a_{k_1} \right\rangle \langle a_{k_1} \mid y \rangle \\
&= \sum_{\text{path}} P^{x \leftarrow y},
\end{aligned}
$$

gdzie h jest stałą wynikającą z normalizacji iloczynu skalarnego wektorów w danej przestrzeni Hilberta, a_k i $|a_k\rangle$ są wartościami własnymi i wektorami własnymi zmiennej $\hat{A}$ opisującej przejście:

$$
f^{\text{path}} = \int_0^t \beta(t')a(t')dt', \qquad (10)
$$

gdzie β(t) jest znaną funkcją, której wybór jest dowolny, i jest funkcją bazową rozwinięcia kanonicznego. W naszym przypadku i w dalszej części założymy, że są to funkcje ortogonalnej bazy układu, reprezentowane przez funkcje stanu. Niech zostanie przeprowadzone grupowanie możliwych ścieżek i utworzona funkcja ścieżek, dla której wartość f ścieżki jest równa pewnemu F:

$$
\phi^{x \leftarrow y}(t|F) = \sum_{\text{path}} \delta\left(F - f^{\text{path}}\right) A^{\text{path}}_{x \leftarrow y}. \qquad (11)
$$

Następnie wygładzamy tę grupę za pomocą pewnej funkcji, np. rozkładu Gaussa, i uzyskujemy amplitudę rozproszenia

$$
\psi^{x \leftarrow y}(t|F) = \sum G\left(F - F'\right) \phi^{x \leftarrow y}(t|F) dF'. \qquad (12)
$$

Jeżeli wykorzystamy bazę {β(t)} zawierającą obserwowany stan $|\psi(t, F)\rangle$, wówczas możemy użyć równania takiego jak:

$$
\begin{aligned}
&i\partial/\partial t |\psi(t|F)\rangle \\
&\qquad = \left[\hat{H}\beta(t)\hat{A}|\psi(t|F)\rangle\right] \\
&\qquad - ih\partial/\partial F \beta(t)\hat{A}|\psi(t|F)\rangle. \qquad (13)
\end{aligned}
$$

W ten sposób można opisać ruch stanu podsystemu wzdłuż możliwych trajektorii. Podobne argumenty R. Feynman przeprowadza w odniesieniu do całkowania bliskich trajektorii .

4 Model opisu przejścia układu do stanu mierzalnego

Model opisu przejścia układu do stanu mierzalnego Początkowym ogólnym pomysłem było formalne wykorzystanie funkcji statusu zamiast funkcji przynależności w algorytmach wspomagania decyzji opartych na algorytmie Mamdaniego. Zaletą funkcji statusu wprowadzonych przez autorów jest to, że można ich używać jak elementów składowych kwantowej teorii pola w układach makroskopowych, dzięki czemu układy dynamiczne można rozpatrywać w oparciu o kwantowe modele matematyczne. Funkcje statusu są analogiczne do wektora stanu układu i mogą być używane do reprezentowania bardzo złożonych procesów

dynamicznych. Stany te mogą mieć dowolną naturę: fizyczną, poznawczą, technologiczną, społeczno-ekonomiczną. Jest to minimalna interpretacja, która pozwala przejść do kwantowej teorii pola. Z tego formalnego punktu widzenia zastosowanie formalizmu matematycznego teorii kwantów jest po prostu kwestią wygody i w rzeczywistości okazało się przydatne w wielu zastosowaniach w różnych kontekstach . Zauważmy, że konstrukcję przestrzeni wektorów stanu systemu można rozpocząć od dowolnego wektora należącego do wprowadzonej przestrzeni Hilberta. Wektory te umożliwiają opis systemu metod wprowadzania operatorów typu operatora Hamiltona i zastosowanie postaci macierzowej do opisu stanów. Zastosowanie operatora podobnego do hamiltonianu. Załóżmy, że dwa mierzalne wskaźniki systemu społeczno-gospodarczego nie oddziałują bezpośrednio. Niech istnieje nieskończona liczba sposobów przejścia z jednego stanu mierzalnego do drugiego. Jednakże, ze względu na interakcję ukrytych parametrów, przejście to odbywa się wzdłuż wybranych bliskich trajektorii. Ponadto, $\sum_{\text{path}} P^{x \leftarrow y}$ może być przedstawiona jako całka po przestrzeni bliskich trajektorii. Zbudujmy model, w którym sterowanie stanem systemu społeczno-gospodarczego odbywa się za pomocą dwóch kontrolowanych grup wskaźników, reprezentowanych przez stany ψ_1 i ψ_2. Zauważmy, że model nie odzwierciedla złożonych procesów wewnętrznych w systemie, które prowadzą do przejść między stanami. Zapewnia jedynie formalną reprezentację operatorową tych przejść. Załóżmy, że region może redystrybuować zasoby zarządcze między tymi grupami wskaźników, które w tym kontekście okazują się konkurować ze sobą poprzez środowisko informacyjno-komunikacyjne. W ten sposób wskaźniki oddziałują na siebie. Hamiltonian przedstawimy w postaci:

$$\hat{H} = \hat{H}_1 + \hat{H}_2. \tag{14}$$

Tutaj $\hat{H}_1$ jest hamiltonianem podsystemu wskaźnika; $\hat{H}_2$ jest hamiltonianem analogu energii wskaźnika w środowisku informacyjnym.

$$\hat{H}_1 = \frac{\partial}{\partial t} - ih\nabla^2; \tag{15}$$

$$\begin{aligned}\hat{H}_2 = &\begin{pmatrix} 0 & 0 \\ 0 & \text{inter}_{22} * \psi_1 \end{pmatrix} \\ &+ \begin{pmatrix} t_{11} * \psi_1 & 0 \\ t_{21} * \psi_2 & 0 \end{pmatrix} \\ &+ \begin{pmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{pmatrix},\end{aligned} \tag{16}$$

gdzie ∂/∂t jest pochodną czasową, i jest jednostką urojoną, h jest stałą normalizacji wymiaru przestrzennego, ∇^2 jest gradientem dywergencji. W naszym przypadku jest to druga pochodna funkcji stanu wskaźnika względem zmiennej bazowej. Równanie (18) zawiera człony będące analogami energii układu dwóch kontrolowanych wskaźników: odpowiednio interakcji, kinetycznej i potencjalnej. Założymy, że współczynniki analogów energii inter_{22}, t_{11}, t_{21}, v_{12}, v_{21} są stałymi. Definiujemy $|\psi\rangle$ na podstawie funkcji stanu w postaci:

$$\begin{aligned}|\psi\rangle &= \begin{pmatrix} \psi_1(r) \\ \psi_2(r) \end{pmatrix} \\ &= \begin{pmatrix} 2.26\exp\left(-2i\pi r - 41.32(0.13 + r)^2\right) \\ 2.33\exp(2i\pi r)\begin{pmatrix} \exp\left(-41.32(-0.13 + r)^2\right) \\ -0.24\exp\left(-41.32(0.13 + r)^2\right) \end{pmatrix} \end{pmatrix},\end{aligned} \tag{17}$$

gdzie r jest wprowadzoną zmienną bazową. Części rzeczywiste funkcji stanu przedstawiono na rys. 2. W operatorze analogu energii potencjalnej zmienimy wartości v_{11} i v_{22} w taki sposób, aby porównać przypadek abstrakcyjny z przypadkiem, który spróbujemy powiązać z danymi statystycznymi. Ponadto zmienimy współczynnik $inter_{22}$ w operatorze interakcji i przeanalizujemy przejaw wyniku interakcji w wynikach eksperymentu numerycznego.

5 Wyniki symulacji i dyskusja

Rozważmy wyniki obliczeń dla $t_{11} = 0, 1$, $t_{21} = 0, 25$, $v_{12} = 0, 04$, $v_{21} = 0, 05$, $v_{11} = 1$ i $v_{22} = 1$. Zakładamy, że przeprowadzana jest analiza interakcji dwóch hipotetycznych wskaźników systemu społeczno-gospodarczego. Wykorzystamy funkcje statusu do stworzenia modelu nieskończonego zbioru trajektorii, wzdłuż których system przechodzi z jednego stanu do drugiego. Korzystając z równania (9), otrzymujemy następujące wyrażenia do obliczenia prawdopodobieństwa przejścia z jednego stanu do drugiego (fragment):

$$\begin{aligned}
&\frac{\partial \psi_1[t]}{\partial t} \\
&= \int_{-1}^{1} \mathrm{dr}\Big(0.0566182e^{-41.3223r^2-2i\pi r-10.7438r-0.698347}\psi_1[t] \\
&+ 1.0258\mathrm{e}^{-82.6446r^2-21.4876r-1.39669}\psi_1[t]^2 \\
&+ 0.5129\mathrm{e}^{-82.6446r^2-4i\pi r-21.4876r-1.39669}\psi_1[t]^2 \\
&+ 0.578304\mathrm{e}^{-41.3223r^2-2i\pi r-10.7438r-0.698347}\psi_2[t]\Big);
\end{aligned} \tag{18}$$

$$\begin{aligned}
&\frac{\partial \psi_2[t]}{\partial t} \\
&= \int_{-1}^{1} \mathrm{dr}\Big(2.26473\mathrm{e}^{-41.3223r^2-2i\pi r-10.7438r-0.698347}\psi_1[t] \\
&+ 0.0144576\mathrm{e}^{-41.3223r^2-2i\pi r-10.7438r-0.698347}\psi_2[t] \\
&+ 0.058435\mathrm{e}^{-41.3223r^2-2i\pi r+10.7438r-0.698347}P2(t) \\
&- 2.64679\mathrm{e}^{-82.6446r^2-1.39669}\psi_1[t]\psi_2[t] \\
&+ 1.90965\mathrm{e}^{-82.6446r^2-21.4876r-1.39669}\psi_1[t]\psi_2[t] \\
&+ \Big(0.540691\mathrm{e}^{-82.6446r^2-1.39669} \\
&+ 0.0668871\mathrm{e}^{-82.6446r^2-21.4876r-1.39669} \\
&+ 1.09269\mathrm{e}^{-82.6446r^2+21.4876r-1.39669}\ldots\Big)\psi_2[t]^2.
\end{aligned} \tag{19}$$

Zintegrujmy po zmiennej bazowej i uzyskajmy wyrażenia na $\hat{H}|\psi\rangle$ po zintegrowaniu po zmiennej bazowej, co usuwa trajektorie interakcji wskaźników w środowisku informacyjno-komunikacyjnym. Istnieje moment, w którym zachodzi „interakcja". Wzrost współczynnika interakcji prowadzi do przesunięcia momentu interakcji w kierunku krótszego czasu. Jednocześnie „siła interakcji", z którą można porównać maksymalną wartość szczytu prawdopodobieństw warunkowych funkcji statusu, nie zależy bezpośrednio od współczynnika interakcji. Pozostałe współczynniki w symulacji interakcji pozostają niezmienione. Początkowe wartości funkcji statusu na początku modelowania są ustalane na podstawie oficjalnych statystyk oceny konkurencyjności regionów Rosji: $〚∂ψ〛_1 [0]= 0,51$; $〚∂ψ〛_2 [0]= 0,58$. Załóżmy, że wyniki danych statystycznych świadczą o pewnej ścieżce przebytej przez podsystem mierzonego wskaźnika. Wówczas dane statystyczne odniesiemy do „potencjalnej" energii układu. Weźmy dane statystyczne i dodajmy je do współczynników v_{12} i v_{21}. Aby wykonać tę procedurę, weźmy dane statystyczne dla dwóch wskaźników należących do grupy wskaźników

innowacyjności . Wskaźniki grupowe są znormalizowane do 1. W rezultacie wykorzystujemy dane dla wskaźnika I1{0,5832, 0,6024, 0,5928, 0,6584, 0,6224, 0,844, 0,8536, 0,7952, 0,8112, 0,7512} oraz dla wskaźnika I2{0,512, 0,44, 0,50, 0,56, 0,56, 0,384, 0,9432, 0,892, 0,488}. Wyniki interpolacji przedstawiono na rys. 4 krzywej ciągłej. Odpowiednie wyrażenia:

$$\begin{aligned} I_1[t] &= \mathrm{v}_{12} \\ &= 111.823\mathrm{i}\mathrm{e}^{-0.02it} - 111.823\mathrm{i}\mathrm{e}^{0.02it} \\ &- 219.357\mathrm{i}\mathrm{e}^{-0.01it} + 219.357\mathrm{i}\mathrm{e}^{+0.01it} \\ &+ 0.524102\mathrm{e}^{-0.1t} + 0.0257454\mathrm{e}^{-it} + 0.0257454\mathrm{e}^{it}; \end{aligned} \tag{20}$$

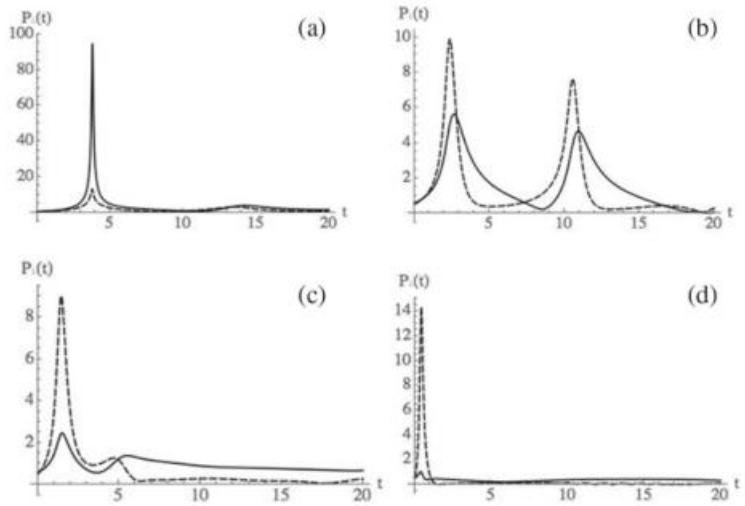
$$\begin{aligned} I_2[t] &= \mathrm{v}_{21} \\ &= \left(137.058\mathrm{i}\mathrm{e}^{-0.02it} - 137.058\mathrm{i}\mathrm{e}^{0.02it} - 270.354\mathrm{i}\right)\mathrm{e}^{-0.01it} \\ &+ 270.354\mathrm{i}\mathrm{e}^{0.01it} + 0.39192\mathrm{e}^{-0.1t} \\ &- 0.0429692\mathrm{e}^{-it} - 0.0429692\mathrm{e}^{it}; \end{aligned} \tag{21}$$

Zakładamy, że mierzone wskaźniki w pewnym momencie czasu oddziałują na siebie poprzez środowisko informacyjno-komunikacyjne. Po całkowaniu po trajektoriach otrzymujemy wyrażenia:

$$\begin{aligned} &\frac{\partial\psi_1(t)}{\partial t} \\ &= -3.69228\mathrm{i}\mathrm{e}^{-0.01it} - 1.88223\mathrm{i}\mathrm{e}^{0.02it} \\ &+ 1.88223\mathrm{i}\mathrm{e}^{-0.02it} + +3.69228\mathrm{i}\mathrm{e}^{0.01it} \\ &+ 0.192211\psi_1(t)^2 + 0.522969\psi_2(t) \\ &+ 0.00882184\mathrm{e}^{-0.1t} + 0.000433354\mathrm{e}^{-it} + 0.000433354\mathrm{e}^{it} \end{aligned} \tag{22}$$

$$\begin{aligned} &\frac{\partial\psi_2(t)}{\partial t} \\ &= (0.564687 - 0.230465i)\psi_1(t)\psi_2(t) \\ &- 182.027\mathrm{i}\mathrm{e}^{(0.-0.01\mathrm{i})t} - 92.2798\mathrm{i}\mathrm{e}^{0.02it} \\ &+ 92.2798\mathrm{i}\mathrm{e}^{-0.02it} + 182.027\mathrm{i}\mathrm{e}^{0.01it} \\ &+ 0.175019\psi_2(t)^2 + 0.0130742\psi_2(t) \\ &+ 0.263877\mathrm{e}^{-0.1t} - 0.0289308\mathrm{e}^{-it} - 0.0289308\mathrm{e}^{it} \end{aligned} \tag{23}$$

Wyniki obliczeń przedstawiono na rysunku dla różnych współczynników inter_{22}.



W tym przypadku wzrost interakcji $inter_{22}$ prowadzi do wcześniejszego czasu interakcji/jonu i nie ma wyraźnej zależności od maksymalnej amplitudy piku interakcji. Dla pewnego zakresu wartości $inter_{22}$ interakcja jest złożona, a prawdopodobieństwo warunkowe wynikowej wartości wskaźnika wskazuje na możliwość wystąpienia „beatów". Nie występuje stabilizacja modalna. Można założyć, że w warunkach rywalizacji dwóch podsystemów wewnętrznych o ograniczone wspólne zasoby mogą pojawić się rozwiązania takie jak rozwiązania równań Lotki-Voltera. Może to prowadzić do powstania obszaru niestabilności systemu, w którym niewielkie działanie sterujące może prowadzić do szybkich zmian w całym systemie. W eksperymencie wykorzystano model dwóch wskaźników komponentu innowacyjnego oceny konkurencyjności regionu. Stworzenie modelu z wieloma wskaźnikami doprowadzi do bardziej złożonych rozwiązań i umożliwi zbadanie zależności reżimów od parametrów. Dalsze badanie systemu obejmuje, po pierwsze, stworzenie modelu działania sterującego w systemie, a po drugie, utworzenie sieci oddziałujących ze sobą wskaźników. Następnie wprowadzimy wskaźniki różnych regionów do danych interpolacyjnych i zbadamy, jak zmienia się dynamika prawdopodobieństw warunkowego ryzyka, reprezentowanych przez wskaźniki dla różnych regionów. Jednocześnie, właśnie stworzenie warunków do ewentualnego istnienia różnych reżimów, zagrażających utratą konkurencyjności regionów Federacji Rosyjskiej, jest przedmiotem szczególnego zainteresowania. Dostępność danych statystycznych umożliwi określenie reżimów odpowiadających stabilizacji systemu oraz wystąpieniu cykli granicznych w przestrzeni fazowej systemu. Są to tryby i wskaźniki, do których dążą struktury zarządzania. W tym przypadku system ulega zaburzeniu po ustabilizowaniu się pomiaru przy określonych wartościach ryzyka. Możliwe jest jednak pojawienie się bardziej złożonych reżimów, których badanie prowadzi się w przestrzeni fazowej systemu.

6. Wnioski

Tu uzasadniono celowość wykorzystania zaproponowanego przez autorów aparatu funkcji statusu do matematycznego modelowania złożonych obiektów społeczno-ekonomicznych. Tradycyjne matematyczne metody opisu, takie jak teoria prawdopodobieństwa, statystyka matematyczna, modele rozmyte, nie są w stanie opisać specyfiki badanych zjawisk. Zjawiska te, z jednej strony, mają charakter losowy, z drugiej strony, konieczne jest sformalizowanie subiektywnej oceny osoby podejmującej decyzje zarządcze. W tym celu proponuje się wykorzystanie funkcji statusu reprezentujących funkcje czasu o wartościach zespolonych. Jednocześnie część rzeczywista odzwierciedla obiektywny aspekt pomiaru wskaźników procesów społeczno-ekonomicznych, a część urojona charakteryzuje aspekt subiektywny. Opisano główne założenia, które są wykorzystywane przy konstrukcji proponowanych kwantowych modeli matematycznych. Wśród nich znajduje się reprezentacja stanu systemu społeczno-ekonomicznego jako zbioru cech w momencie pomiaru, które są uporządkowane i tworzą pewien wektor. Ponadto, stany podstawowe systemu tworzą bazę ortonormalną, na podstawie której opisany jest dowolny stan systemu. W tym przypadku zmiana stanu systemu jest opisana operatorem unitarnym, który charakteryzuje zmianę kierunku wektora stanu systemu. W oparciu o funkcje stanu zaproponowano model matematyczny opisujący różne przejścia systemu społeczno-gospodarczego do stanu mierzalnego. Jednocześnie, na podstawie funkcji stanu, zrealizowano możliwość tworzenia modeli matematycznych podobnych do kwantowych. Badane stany mogą mieć różny charakter: fizyczny, poznawczy, technologiczny, społeczno-ekonomiczny. Ostatnia część rozdziału zawiera opis przykładu. Przedstawiono wyniki symulacji oraz ich analizę. Przeprowadzono analizę interakcji dwóch hipotetycznych wskaźników systemu społeczno-

gospodarczego. Uzyskano wyrażenia umożliwiające określenie prawdopodobieństwa przejścia systemu z jednego stanu do drugiego. Wartości początkowe funkcji stanu w momencie modelowania ustalono na podstawie oficjalnych danych statystycznych w celu oceny konkurencyjności regionów Federacji Rosyjskiej. Przeprowadzono analizę tworzenia warunków zagrożenia utratą konkurencyjności na przykładzie regionu Federacji Rosyjskiej. Dane w tym zakresie zaczerpnięto z oficjalnych statystyk wskaźników innowacyjności.

Opracowanie systemu wspomagania decyzji do oceny i monitorowania ryzyka konkurencyjności regionalnej

Wykorzystanie technik inteligentnych w analizie złożonych systemów społeczno-gospodarczych pozwala na rozszerzenie możliwości diagnostycznych i prognostycznych systemów wspomagania decyzji. Przedstawiono wdrożone moduły systemu wspomagania decyzji do oceny i prognozowania poziomu ryzyka konkurencyjności regionów. Funkcjonalność aplikacji umożliwia wykorzystanie zarówno metod logiki rozmytej do kompleksowej oceny konkurencyjności regionów, jak i narzędzi do modelowania dynamiki ryzyka w oparciu o równania Chapmana-Kołmogorowa. Opracowana metodyka stosowania tych modeli została przetestowana w 78 regionach Federacji Rosyjskiej. W wyniku analizy benchmarkingowej zidentyfikowano regiony wiodące zarówno pod względem aktualnego poziomu ryzyka konkurencyjności, jak i z uwzględnieniem jego dynamiki.

Słowa kluczowe: Ocena konkurencyjności · Regiony · Ryzyko · Modelowanie matematyczne · Równania Kołmogorowa-Chapmana · Wielokryterialne metody podejmowania decyzji · Techniki rozmyte · System wspomagania decyzji

1 Wstęp

Koncepcja Społeczeństwa 5.0 to w dużej mierze strategiczny kierunek długoterminowego rozwoju, utrzymania lub zdobycia pozycji konkurencyjnej gospodarki zarówno na poziomie krajowym, jak i regionalnym. Zalety Społeczeństwa 5.0 wiążą się z możliwością stworzenia zorganizowanej, systemowo zrównoważonej gospodarki, która poprzez rozwój zaawansowanych technologii dostarcza rozwiązań dla szeregu problemów społecznych. Rozwój i wdrażanie inteligentnych systemów informatycznych jest istotne dla diagnozowania poziomu rozwoju regionalnego i procesu strategicznego wspierania konkurencyjności regionów. Rozwiązanie tak złożonych zadań w warunkach niepełnej informacji wymaga wykorzystania technik rozmytych, obliczeń miękkich oraz kontroli dynamiki strukturalnej w przetwarzaniu informacji. Obecnie nie istnieje specjalistyczne oprogramowanie do oceny konkurencyjności regionów. Celem badań jest stworzenie aplikacji, która zapewni wsparcie decyzyjne w ocenie ryzyka konkurencyjności regionów z wykorzystaniem systemowej analizy wielowymiarowej i prognozowania, uwzględniając specyfikę dostępnych danych statystycznych dla regionów. Na różnych poziomach gospodarki koncepcja konkurencyjności stale się rozwija. Teoretyczne podejście do analizy konkurencyjności na poziomie regionalnym opierało się początkowo na koncepcji czterech etapów rozwoju narodów . Problem empirycznej oceny konkurencyjności regionalnej jest obecnie przedmiotem dyskusji w ramach braku przyjętego podejścia. W szczególności istnieją różne

techniki identyfikacji systemu czynników determinujących złożoną koncepcję konkurencyjności na poziomie regionalnym . W warunkach gospodarki opartej na wiedzy następuje rozbudowa tej koncepcji, uwzględniająca nie tylko koszty produkcji, ale także szeroki wachlarz czynników jakościowych, innowacyjnych i instytucjonalnych. Powszechnym sposobem oceny konkurencyjności regionalnej jest oparcie jej na indeksach (wskaźnik konkurencyjności globalnej (Global Competitiveness Index, European Regional Competitiveness Index, Global sustainable Competitiveness index). W praktyce krajowej stosuje się kompleksowy indeks konkurencyjności regionalnej (AV RCI-2020), który obejmuje siedem odrębnych grup wskaźników [9]. Jednakże takie podejście nie pozwala na budowę modeli predykcyjnych i nie uwzględnia złożonych zależności między czynnikami. Wybór czynników w badaniu konkurencyjności regionalnej jest szczególnie ważny, należy zapewnić obszerny zestaw wskaźników. Jednocześnie w zadaniu oceny ilościowej pojawia się pytanie o dostępność ich wartości ilościowych. Ponadto, czynniki te są w złożonej dynamicznej interakcji, konieczne jest uwzględnienie ich wzajemnego wpływu. Niniejsze badanie umożliwia zapewnienie uwzględnienia określonych cech konkurencyjności regionalnej za pomocą odpowiednich modeli matematycznych zaimplementowanych jako oprogramowanie.

2 Modele i metody zaimplementowane w aplikacji

Funkcjonalność opracowanego oprogramowania zapewnia zastosowanie dwóch modeli: modelu opartego na wielokryterialnym podejmowaniu decyzji (MCDM) oraz modelu opartego na równaniach Kołmogorowa-Chapmana. Wielokryterialna metoda podejmowania decyzji Fuzzy ELECTRE II pozwala na wykorzystanie nie tylko ilościowych, ale i jakościowych wartości kryteriów, do implementacji liczb rozmytych, co pozwala na podejmowanie trafnych i spójnych decyzji, a tym samym na zmniejszenie subiektywności oceny . Metoda ta i jej modyfikacje nie opierają się na klasycznej teorii użyteczności; zawierają one względne oszacowania alternatyw. Binarne relacje dominacji są budowane na zbiorze alternatyw. W metodzie ELECTRE II zdefiniowano wskaźniki spójności i niespójności, a dla tych wskaźników istnieje kilka poziomów. Metoda ta klasyfikuje alternatywy na podstawie progowych poziomów spójności i niespójności. Standardowy algorytm Fuzzy ELECTRE II wykorzystuje jedną metrykę do oceny alternatyw, taką jak odległość euklidesowa lub odległość Hamminga. Wybór metryki przy rankingu alternatyw jest kluczowy dla tej metody MCDM. Proponowany algorytm rankingowy obejmuje różne metryki (euklidesową, Manhattanu, Hamminga, Hausdorffa, Zhuravleva, Tonimoto–Jaccarda, Braya–Curtisa). Ostateczna ocena opiera się na wynikach uzyskanych dla różnych metryk. Metody dynamiki systemów są wykorzystywane do oceny rozwoju obiektów społeczno-ekonomicznych. Rozważana metoda, oparta na równaniach Kołmogorowa-Chapmana, zakłada na etapie początkowym ukształtowanie się zbioru czynników ryzyka dla konkurencyjności regionalnego systemu społeczno-ekonomicznego. Pogorszenie się poszczególnych wskaźników odpowiadających zidentyfikowanym czynnikom prowadzi do spadku konkurencyjności regionu jako całości. Taki układ wskaźników jest hierarchicznym modelem logiczno-probabilistycznym. Dla skonstruowanego drzewa czynników ryzyka opisuje on graf przyczynowy. Cechą takiego grafu są funkcje logiczne w węzłach konsekwencji. Zatem każda konsekwencja występuje, gdy wykonywana jest operacja logiczna dla odpowiadających jej przyczyn – indywidualnych czynników ryzyka konkurencyjności. Krawędzie grafu reprezentują możliwe przejścia między stanami systemu, które powstają w wyniku operacji logicznych na przyczynach. Na podstawie tego grafu utworzono wszystkie kombinacje zdarzeń krytycznych, które tworzą fragmenty grafu dla realizacji zdarzenia źródłowego – utraty konkurencyjności regionu. Minimalny przekrój to zbiór wierzchołków liści grafu przyczynowego, połączonych na podstawie relacji binarnych i tworzących podgraf grafu przyczynowego, prowadzący do realizacji zdarzenia źródłowego. Dla każdego minimalnego przekroju identyfikowane są stany regionalnego systemu społeczno-

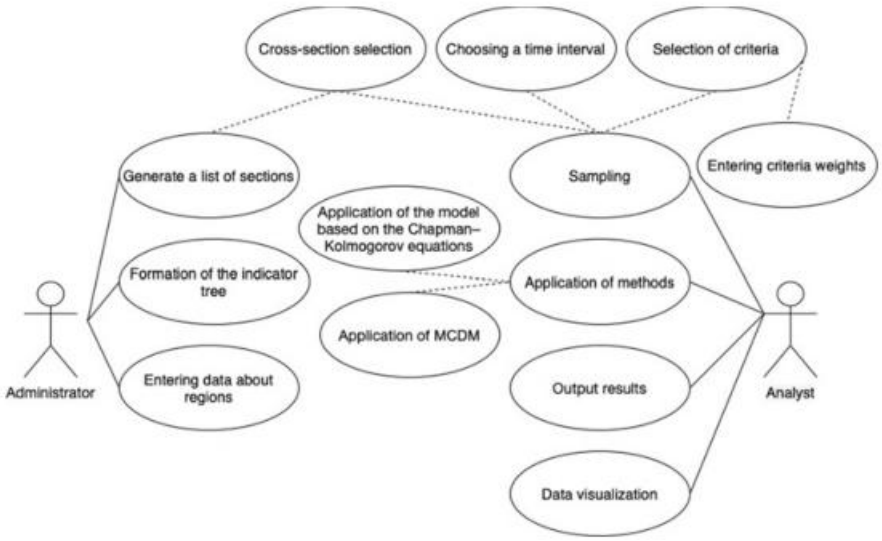
gospodarczego, odpowiadające różnym kombinacjom zdarzeń krytycznych, które determinują ryzyko konkurencyjności regionów. Następnie budowana jest struktura sieciowa możliwych przejść między tymi stanami. Dla powstałej struktury stanów wyznaczany jest układ równań różniczkowych. W wyniku rozwiązania układu równań różniczkowych metodami numerycznymi możliwe jest wyznaczenie prawdopodobieństw realizacji ryzyka w węzłach grafu stanów w danym momencie. Dla minimalnego przekroju s o wymiarze N zbiór stanów S badanego układu zawiera $S = 2^N - 1$ elementów. Układ równań różniczkowych Kołmogorowa–Chapmana można zapisać w postaci macierzowej:

$$\frac{dP_k(t)}{dt} = R \cdot \begin{pmatrix} P_0(t) \\ P_1(t) \\ \dots \\ P_S(t) \end{pmatrix}, \qquad (1)$$

gdzie $P_k(t)$, $k = 1, \dots, S$, jest prawdopodobieństwem, że elementy układu znajdują się w stanie k w czasie t, $R = \{r_{lk}\}$ jest nieskończenie małą macierzą przejścia o wymiarze $S \times S$, zbudowaną na podstawie grafu stanu. Dla stanu k równanie różniczkowe ma postać:

$$\frac{dP_k(t)}{dt} = \sum_{l=0}^{S} r_{lk} P_l(t) - \sum_{l=0}^{S} r_{kl} P_k(t), \qquad (2)$$

gdzie elementy r_{lk} macierzy R odpowiadają prawdopodobieństwom przejścia ze stanu l do stanu k, $r_{kk} = -\sum_{l,l\neq k} r_{kl}$. Macierz przejść R jest tworzona na podstawie parametrów Poissona dla realizacji zdarzeń odpowiadających elementom minimalnego cięcia: d_i to poziom ryzyka, g_i to poziom działania kontrolnego dla poszczególnych wskaźników $i = 1, \dots, N$. Powstały układ równań różniczkowych jest rozwiązywany numerycznie metodą Runge-Kutty z automatycznym dopasowaniem kroku. Opracowana aplikacja zawiera możliwości dla użytkowników przedstawione na schemacie.

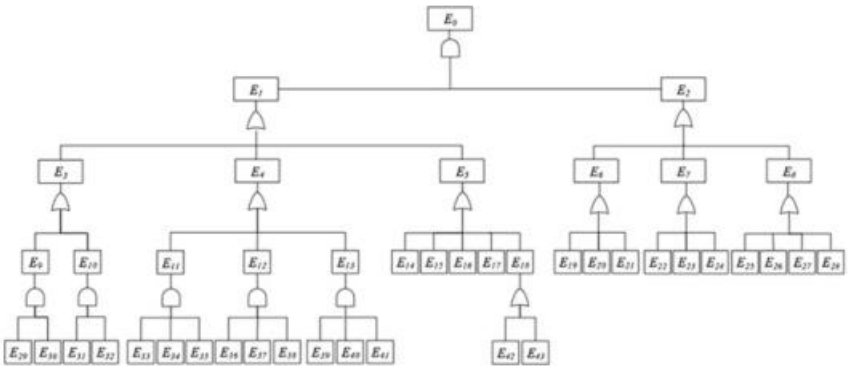


W aplikacji użytkownicy podzieleni są na dwie grupy: administratorów i analityków. Druga kategoria użytkowników posiada uprawnienia do pracy z programem nadane przez administratora aplikacji.W przypadku pierwszego modelu funkcjonalność obejmuje utworzenie zbioru regionów i zbioru kryteriów oceny, uszeregowanie wybranych regionów za pomocą zmodyfikowanej metody

Fuzzy ELECTRE II, przedstawienie wyniku w postaci listy rankingowej regionów i wybranie wiodących regionów. W przypadku drugiej metody opracowana aplikacja pozwala na określenie minimalnych cięć na wykresie związków przyczynowo-skutkowych, wprowadzenie poziomów ryzyka działań kontrolnych dla poszczególnych wskaźników, a także na odczyt z pliku tekstowego. Funkcjonalność aplikacji zapewnia wizualizację wyników. Aplikacja umożliwia wyświetlanie wykresów prawdopodobieństwa zdarzeń krytycznych i ich kombinacji dla zadanego przedziału czasu. Obliczone dane odpowiadające prawdopodobieństwom dla różnych kombinacji zdarzeń krytycznych dla zadanych przedziałów czasu można zapisać w pliku tekstowym. Aplikacja internetowa jest aplikacją typu klient-serwer, w której użytkownik komunikuje się z serwerem za pomocą przeglądarki. Ponieważ logika takiej usługi internetowej jest rozproszona między serwerem a klientem, dane są przechowywane na serwerze, a wymiana informacji odbywa się za pośrednictwem sieci. Aby uruchomić aplikację internetową, użytkownik nie musi instalować żadnego dodatkowego oprogramowania; wystarczy dostęp do Internetu i przeglądarka internetowa na dowolnym urządzeniu. Działanie aplikacji nie jest uzależnione od systemu operacyjnego zainstalowanego na komputerze użytkownika. Usługi internetowe są łatwe w instalacji, utrzymaniu i aktualizacji interfejsu klienta. Po wprowadzeniu zmian aplikacja jest aktualizowana do najnowszej wersji przy następnym załadowaniu strony. Zazwyczaj aplikacje internetowe składają się z co najmniej trzech głównych komponentów: części klienckiej i serwerowej aplikacji internetowej oraz bazy danych. Opracowana aplikacja wykorzystuje język programowania Python rozszerzony o zestaw bibliotek (Itertools, Matplotlib, PyPlot, Math, NumPy, SciPy, Tkinter). Python to skryptowy język programowania wysokiego poziomu, znany ze swojej wszechstronności. Python oferuje zestaw narzędzi z dużą liczbą bibliotek, w tym modele i metody matematyczne oraz różnorodne narzędzia wizualizacyjne. Jako bazę danych wybrano PostgreSQL. Ta baza danych to obiektowo-relacyjny system zarządzania bazą danych, który oferuje wiele zalet w porównaniu z innymi bazami danych typu open source. Podstawową cechą obiektowo-relacyjnej bazy danych jest obsługa obiektów zdefiniowanych przez użytkownika, a także ich zachowań, w tym typów danych, funkcji, operacji, domen i indeksów. PostgreSQL umożliwia tworzenie, przechowywanie i pobieranie złożonych struktur danych. Jednocześnie jest to bezpieczna i rozszerzalna baza danych, a także oferuje rozbudowany ekosystem dostępnych narzędzi, który pozwala na wykorzystanie PostgreSQL w różnych scenariuszach. Klasyczną metodą tworzenia interfejsów internetowych jest użycie HTML (Standardized Document Markup Language) z wykorzystaniem formalnego języka opisu wyglądu dokumentu – CSS. Jednak różne implementacje HTML, CSS i innych specyfikacji w różnych przeglądarkach powodują problemy w tworzeniu aplikacji internetowych, a także komplikują ich późniejszą obsługę. Spośród wielu frameworków, Django został wybrany, aby umożliwić tworzenie skalowalnych aplikacji. Do tworzenia front-endu wybrano framework Bootstrap – darmowy zestaw narzędzi umożliwiający tworzenie aplikacji internetowych. Wybrane narzędzia pozwalają na wdrożenie skalowalnej, nowoczesnej i mobilnej aplikacji z atrakcyjnym, przyjaznym dla użytkownika interfejsem.

3. Podejście metodologiczne do oceny ryzyka konkurencyjności

Do systemowej analizy ryzyka konkurencyjności na poziomie regionalnym zastosowano następującą hierarchiczną strukturę czynników) :



konkurencyjność regionalna (E_0); konkurencyjność transformacyjna (E_1); konkurencyjność transakcyjna (E_2); konkurencyjność techniczna (E_3); konkurencyjność społeczna (E_4); konkurencyjność zasobów naturalnych (E_5); konkurencyjność instytucjonalna (E_6); konkurencyjność informacyjna (E_7); konkurencyjność innowacyjna (E_8) . Na kolejnym poziomie zastosowano rozszerzony zestaw wskaźników: wykorzystanie środków trwałych (E_9); rozwój infrastruktury transportowej (E_{10}); poziom dochodów ludności (E_{11}); demografia (E_{12}); jakość życia (E_{13}); górnictwo (E_{14}); powierzchnia gruntów rolnych (E_{15}); wyposażenie w zasoby leśne (E_{16}); wytwarzanie energii elektrycznej (E_{17}); środowisko (E_{18}); udział organizacji nierentownych (E_{19}); poziom opodatkowania (E_{20}); nieuczciwa konkurencja (E_{21}); wykorzystanie komputerów osobistych (E_{22}); wykorzystanie Internetu w organizacjach (E_{23}); wykorzystanie elektronicznych systemów zarządzania dokumentacją w organizacjach (E_{24}); personel badawczo-rozwojowy (E_{25}); wewnętrzne wydatki na badania i rozwój (E_{26}); koszty działalności innowacyjnej organizacji (E_{27}); wolumen innowacyjnych dóbr (E_{28}); koszt środków trwałych (E_{29}); amortyzacja środków trwałych (E_{30}); gęstość torów kolejowych (E_{31}); gęstość dróg utwardzonych (E_{32}); średni dochód pieniężny na mieszkańca (E_{33}); ludność o dochodach poniżej minimum socjalnego (E_{34}); liczba oficjalnie bezrobotnych (E_{35}); oczekiwana długość życia (E_{36}); współczynnik przyrostu naturalnego ludności (E_{37}); współczynnik wzrostu migracji (E_{38}); poprawa warunków życia ludności (E_{39}); personel medyczny na 10 000 mieszkańców (E_{40}); liczba zgłoszonych przestępstw (E_{41}); emisja zanieczyszczeń do atmosfery (E_{42}); zrzut zanieczyszczonych ścieków (E_{43}). W celu kompleksowej analizy obecnego poziomu i dynamiki konkurencyjności regionalnej zaproponowano podejście metodologiczne, które pozwala, w oparciu o benchmarking, na identyfikację najlepszych praktyk, stworzenie szacunków referencyjnych wskaźników wewnętrznych oraz zbadanie i adaptację najlepszych strategii rozwoju regionalnego. Technika ta obejmuje wdrożenie następujących kroków w oparciu o opracowaną aplikację.

Etap 1. Wybór minimalnego zestawu cięć.

Etap 2. Zastosowanie modelu opartego na zmodyfikowanym programie Fuzzy ELECTRE II do oceny obecnego stanu konkurencyjności regionalnej.

Etap 3. Zastosowanie modelu opartego na równaniach Kołmogorowa-Chapmana do prognozowania ryzyka dla konkurencyjności regionalnej.

Etap 4. Sformułowanie ostatecznej oceny regionów w oparciu o metodę odległości.

Rozważmy minimalny przekrój s, który obejmuje wskaźniki E^s_i , i = 1, . . . , N. Istnieje macierz E^s_{ij} , i = 1, . . . , N , j = 1, . . . , M, wartości N wskaźników uwzględnionych w minimalnym cięciu s, dla M regionów. Do kompleksowej oceny obecnego poziomu konkurencyjności regionalnej, wartość

$\overline{E_j^s}, \ j = 1, \ldots, M$, została wykorzystana za pomocą zmodyfikowanej metody Fuzzy ELECTRE II. Wartość $P_0(t)$ jest używana jako ocena predykcyjna $\overline{\overline{E_i^s}}', \ j = 1, \ldots, M$, ryzyka konkurencyjności regionu. Jest ona uzyskiwana w wyniku rozwiązania utworzonego układu równań Kołmogorowa-Chapmana, dla regionu j, j = 1, . . . , M, oraz określonej przez użytkownika wartości t. Metoda odległości pozwala uwzględnić bliskość obiektów do obiektu odniesienia. Do oszacowania odległości $\overline{E_j^s}'\left(\overline{\overline{E_j^s}}'\right)$ od regionu j do regionu wiodącego pod względem wskaźnika $\overline{E_j^s}\left(\overline{\overline{E_j^s}}\right)$, wykorzystuje się region o maksymalnej (minimalnej) wartości wskaźnika, biorąc pod uwagę jego wielokierunkowy wpływ na ocenę końcową:

$$\left(\overline{E_j^s}'\right) = \max(j = 1, M)\left(\overline{E_j^s}\right) - \left(\overline{E_j^s}\right), \quad j = 1, \ldots, M, \tag{3}$$

$$\overline{\overline{E_j^s}}' = \overline{\overline{E_j^s}} - \min_{j=1,M} \overline{\overline{E_j^s}}, \quad j = 1, \ldots, M \tag{4}$$

Całkowitą ocenę konkurencyjności regionów $\widetilde{E_j^s}$ przeprowadza się w następujący sposób:

$$\left(\widetilde{E_j^s}\right) = a_1\left(\overline{E_j^s}'\right) + a_2\left(\overline{\overline{E_j^s}}'\right), \quad j = 1, \ldots, M, \tag{5}$$

gdzie a_1, a_2 są współczynnikami wagowymi określającymi istotność wskaźników $\overline{E_i^s}'$ and $\overline{\overline{E_i^s}}$.

Etap 5. Analiza ryzyka konkurencyjności regionalnej w oparciu o benchmarking.

4. Zastosowanie aplikacji do analizy konkurencyjności regionalnej regionów

Rozważmy etapowe zastosowanie proponowanego podejścia. W procesie testowania opracowanej aplikacji do oceny konkurencyjności regionalnej uwzględniono minimalne cięcia s_1 i s_2: E_{19}–E_{20}–E_{21}–E_{28}–E_{30}, E_{25}–E_{26}–E_{27}–E_{29}–E_{30}. Minimalne cięcie E_{19}–E_{20}–E_{21}–E_{28}–E_{30} obejmuje materialne i techniczne czynniki ryzyka: udział organizacji nierentownych (E_{19}); poziom opodatkowania (E_{20}); nieuczciwą konkurencję (E_{21}); wolumen dóbr innowacyjnych (E_{28}); koszt środków trwałych (E_{29}); stopień amortyzacji środków trwałych (E30). Minimalne cięcie E_{25}–E_{26}–E_{27}–E_{29}–E_{30} pozwala ocenić takie czynniki ryzyka ekonomicznego, jak liczba personelu zaangażowanego w badania i rozwój (E_{25}); wewnętrzne wydatki na badania i rozwój (E_{26}); koszty działalności innowacyjnej organizacji (E_{27}); wolumen dóbr innowacyjnych (E_{28}); koszt środków trwałych (E_{29}); stopień amortyzacji środków trwałych (E30). Wskaźniki te stanowią otwarte dane statystyk regionalnych za rok 2020 . W badaniu przeanalizowano 78 regionów Federacji Rosyjskiej, biorąc pod uwagę dostępne dane dotyczące wybranych zestawów wskaźników. Moskwa nie znajduje się na liście regionów. Według większości analizowanych wskaźników, ten podmiot Federacji Rosyjskiej jest obserwacją odstającą. W procesie stosowania metody wielokryterialnej przeprowadzono dyskretyzację wartości wskaźników ryzyka konkurencyjności regionów. Aby zademonstrować proponowaną metodologię kompleksowej oceny konkurencyjności regionalnej, rozważono minimalne cięcia s1 i s2 oraz odpowiadające im modele M_1, M_2, M_3, M_4, M_5, M_6 . Dla modeli opartych na równaniach Kołmogorowa-Chapmana obliczono wartości prawdopodobieństwa $P_0(t)$ ryzyka konkurencyjności dla t = 3. Do integralnej oceny konkurencyjności regionów zastosowano współczynniki wagowe a_1 = 0,4, a_2 = 0,6. Ocena

konkurencyjności regionów według grupy wskaźników minimalnego cięcia s_1 z wykorzystaniem modelu M_1 pozwala na stworzenie rankingu regionów pod względem aktualnego poziomu konkurencyjności w 2020 roku. Jako regiony wiodące można wskazać: Sankt Petersburg, obwód omski, obwód samarski, obwód moskiewski, obwód wołgogradzki, Jamalsko-Nieniecki Obwód Autonomiczny, obwód pskowski, obwód wołogodzki, obwód włodzimierski, obwód orenburski. Zastosowanie modelu M_2 pozwoliło na wyodrębnienie regionów o najniższym ryzyku w porównaniu z grupą wskaźników s1. Pierwsze dziesięć regionów obejmuje: Baszkirię, Buriację, obwód tulski, obwód pskowski, obwód kurski, Komi, obwód smoleński, Karelię, obwód permski, Czuwaszję. Zgodnie z ostateczną oceną opartą na modelu M_3, w tej kategorii zidentyfikowano następujące wiodące regiony: obwód pskowski, obwód samarski, autonomiczny obwód jamalsko-nieniecki, obwód wołgogradzki, obwód moskiewski, obwód tomski, Udmurcja, obwód omski, miasto Petersburg, Komi. Wykorzystanie oszacowań $\overline{\overline{E_i^s}'}$ and $\overline{\overline{\overline{E_i^s}}}$, j = 1, . . . , M uzyskanych w wyniku zastosowania modeli M_1 i M_2, które określają odległość do regionu wiodącego, oraz wskaźnika całkowego $\widetilde{E_j^s}$, j = 1, . . . , M, umożliwia identyfikację cech rozwoju regionalnego . Na przykład miasto Sankt Petersburg i obwód omski pozostają daleko w tyle za wiodącymi regionami pod względem wartości predykcyjnych ryzyka konkurencyjności. Doprowadziło to do pogorszenia końcowej oceny tych regionów. Rozważmy minimalny zestaw wycięć s_2. Ocena konkurencyjności regionów za pomocą wskaźników s_2 przy użyciu modelu M_4 pozwala na utworzenie listy rankingowej regionów według poziomu konkurencyjności w 2020 roku. W tym przypadku jako wiodące regiony można wskazać: miasto Sankt Petersburg, obwód moskiewski, obwód krasnojarski, Tatarstan, obwód woroneski, obwód rostowski, obwód krasnodarski, obwód niżnonowogrodzki, obwód nowosybirski, obwód swierdłowski. Używając modelu M_5 można wyróżnić regiony o najniższym ryzyku w stosunku do grupy wskaźników zawartych w minimalnym zestawie wycięć s_2. Wiodącymi regionami są: Petersburg, obwód moskiewski, obwód krasnojarski, Tatarstan, obwód woroneski, obwód rostowski, obwód krasnodarski, obwód niżnonowogrodzki, obwód nowosybirski, obwód swierdłowski. Ostateczna ocena z wykorzystaniem modelu M6 dla minimalnego zbioru cięć s_2 pozwala nam zidentyfikować takie wiodące regiony, jak: obwód tulski, Republika Baszkirii, obwód moskiewski, obwód samarski, obwód woroneski, obwód kemerowski, obwód permski, Buriacja, obwód leningradzki, Udmurcja. Do analizy benchmarkingowej przedstawiamy odległości między regionami a próbkami referencyjnymi zgodnie z modelami M_5 i M_6 . Dla minimalnych wskaźników cięć s_2, opóźnienie w stosunku do regionów wiodących pod względem predykcyjnych poziomów ryzyka jest znacznie mniejsze. Należy zauważyć, że pod względem poziomu konkurencyjności w 2020 r. występuje niewielkie opóźnienie w takich stosunkowo rozwiniętych regionach, jak obwód leningradzki, terytorium permskie i obwód woroneski. W rezultacie możliwe jest porównanie szacunków bieżącego poziomu konkurencyjności regionów na okres 2020 r. i prognozowanych ocen ryzyka z pozycją regionów wiodących. Pomaga to zidentyfikować kluczowe problemy rozwoju regionalnego.

5. Wnioski

Podejście zastosowane w aplikacji obejmuje ocenę według określonych zestawów wskaźników konkurencyjności regionalnej. Wielokryterialna metoda podejmowania decyzji Fuzzy ELECTRE II pozwala oszacować aktualny stan regionalnego systemu społeczno-gospodarczego w danym okresie. Metody pomiaru dynamiki ryzyka oparte na równaniach Kołmogorowa-Chapmana prognozują poziom konkurencyjności regionalnej. Proponowana metodologia oparta na opracowanej aplikacji umożliwia analizę doświadczeń regionów, stworzenie benchmarków do

porównywania ich wyników oraz wprowadzenie najlepszych praktyk w procesie opracowywania strategicznych kierunków rozwoju regionów Federacji Rosyjskiej. Kolejnym kierunkiem doskonalenia systemu wspomagania decyzji w zakresie oceny ryzyka konkurencyjności regionalnej jest wykorzystanie hybrydowych modeli eksploracji danych do predykcyjnej oceny czynników konkurencyjności.

Analiza mentalna przywództwa Stackelberga w nieliniowym oligopolu. Model systemów cyberfizycznych rynku ropy naftowej

Rozważono problem analizy przywództwa Stackelberga w modelu oligopolu z nieliniowymi funkcjami popytu rynkowego i kosztów firm. Na podstawie trendów w dynamice rynku ropy naftowej badane są równowagi Cournota i Stackelberga w przypadku jednego lub kilku liderów. Ponieważ rynek ropy naftowej łączy konsumentów, usługi, sprzęt i kadrę zarządzającą firmami, rynek ten jest powiązany z systemami cyberfizycznymi. W tym przypadku mechanizm decyzyjny kadry zarządzającej opiera się na analizie wzajemnych założeń firm dotyczących możliwych strategii konkurentów, z uwzględnieniem technicznych cech sprzętu. Udowadniamy następujące wyniki. Po pierwsze, pomimo regulacji rządowych, rosyjski rynek ropy charakteryzuje się malejącą, nieliniową krzywą popytu. Po drugie, modelowanie pokazuje, że równowaga Stackelberga w grze o przywództwo jest najbliższa rzeczywistym wskaźnikom rynkowym, co pozwala nam wnioskować o asymetrii świadomości firm. Po trzecie, w równowagach z liderami drugiego poziomu występują dwa atraktory, tj. model nieliniowy zapewnia bifurkacje.

Słowa kluczowe: Oligopol · Nieliniowe funkcje popytu i kosztów · Równowaga Cournota-Nasha · Równowaga Stackelberga · Rynek ropy naftowej

1 Wstęp

Liniowy model wyboru optymalnych działań agentów rynkowych oligopolu, w którym odwrotna funkcja popytu i funkcje kosztów agentów są liniowe, jest szeroko stosowany do analizy równowag w grze oligopolistów, ponieważ model ten pozwala nam na proste rozwiązanie analityczne w postaci równowagi Cournota-Nasha .Nieliniowy model wyboru optymalnych działań, w którym odwrotna funkcja popytu jest liniowa, a funkcje kosztów agentów są nieliniowe, był stosowany rzadziej , ponieważ nie pozwala na znalezienie rozwiązania analitycznego. W tym przypadku szczególne trudności pojawiają się przy obliczaniu równowagi z przywództwem Stackelberga . W konsekwencji złożoności obliczeniowej modelowania, jeszcze rzadziej badacze skupiają się na analizie modelu oligopolu z nieliniową funkcją popytu. Tymczasem rzeczywiste rynki często oferują sytuacje z nieliniowymi funkcjami popytu i kosztów firm. Ostatnie badania wykazały interesujące konsekwencje tej nieliniowości. W przypadku hiperbolicznej funkcji popytu ustalono bifurkację jednego ustalonego punktu Cournota , co jest rezonansowym przypadkiem bifurkacji Neimarka-Sakera , a gdy ustalony punkt Cournota jest niestabilny, mogą istnieć stabilne cykle dowolnego okresu. Analiza hiperbolicznej funkcji popytu dla przypadku duopolu na portrecie fazowym produkcji firm wykazała cykle okresowe i chaotyczne atraktory, między którymi powstają bifurkacje. Dodatkowo, w przypadku wykładniczej funkcji popytu, badania australijskiego rynku energii elektrycznej wykazały wysoki stopień zmienności cen . W przypadku trygonometrycznej funkcji popytu możliwa jest złożona dynamika z dwoma typami bifurkacji: kaskadą podwójnej bifurkacji i chaosu oraz bifurkacją Neimarka-Sakera i, w konsekwencji, chaosem. Racjonalne funkcje popytu ułamkowego mocy mogą prowadzić do destabilizacji równowagi Nasha i

bifurkacji, co powoduje jakościowe zmiany w strukturze atraktorów i ich zakresach. Analiza hiperbolicznej funkcji popytu wykazała, że istnieje wiele złożonych scenariuszy dynamicznych i chaotycznych trajektorii, a także multistabilność, gdy współistnieją różne atraktory . Co ciekawe, w przypadku funkcji popytu firm liniowo niejednorodnych pod względem produkcji [22] wykazano bifurkacje punktów równowagi, zmiany parametrów modelu prowadzą do chaosu, a utrata stabilności może być spowodowana bifurkacjami podwajającymi okres. Nasze badania mają następujące cele. Po pierwsze, formułujemy sytuację na rynku ropy naftowej jako model gry oligopolu. Po drugie, obliczamy równowagę Cournota w tej grze, a także równowagę Stackelberga pierwszego i drugiego poziomu w oparciu o wzory . Po trzecie, analizujemy konsekwencje nieliniowości w kategoriach możliwości występowania wielu atraktorów. Z reguły na rynkach towarowych powstają korporacje zintegrowane pionowo. Firmy te tworzą sieci składające się ze sprzętu górniczego i personelu administracyjnego (tj. kierownictwa). W konsekwencji, te sieci człowiek-maszyna należą do systemów cyberfizycznych . W tym przypadku badamy system cyberfizyczny pod kątem wpływu procesów zarządzania mentalnego na takie parametry techniczne jak objętość transferu paliwa.

2. Sformułowanie problemu

Model wyboru optymalnych strategii rynkowych (wolumenów sprzedaży) firm na rynku oligopolu uwzględnia funkcję celu (zysk firmy).

$$\Pi_i = P(Q)Q_i - C_i(Q_i), \quad Q_i \geq 0, \quad i = 1, \ldots, I, \qquad (1)$$

odwrotna funkcja popytu

$$P(Q) = AQ^{\alpha}, \ A > 0, \ \alpha < 0, \ |\alpha| < 1, \ Q = \sum_{i=1}^{I} Q_i, \qquad (2)$$

i funkcje potęgowe kosztów agenta

$$C_i(Q_i) = C_{Fi} + B_i Q_i^{\beta_i}, \ C_{Fi} > 0, \ B_i > 0, \ 0 < \beta_i < 2, \quad i = 1, \ldots, I, \qquad (3)$$

gdzie Q_i i $\prod_i$ są wolumenem sprzedaży i zyskiem i-tego agenta; I jest liczbą agentów rynkowych; P(Q) jest funkcją odwrotną popytu; A, α są współczynnikami funkcji popytu; Bi, β_i, CF_i są współczynnikami funkcji kosztów. Parametry funkcji kosztów mają następujące znaczenie: CF_i jest stałym kosztem, B_i jest kosztem krańcowym przy βi = 1, βi charakteryzuje efekt skali, który może być dodatni ($0 < \beta_i < 1$) lub ujemny ($1 < \beta_i < 2$). W modelu (1)–(3) równowaga Nasha jest obliczana przez rozwiązanie układu koniecznych warunków optymalności dla danego wektora hipotetycznych wariacji:

$$\frac{\partial \Pi_i\left(Q_i, Q'_{iQ_j}\right)}{\partial Q_i} = 0, \quad i, j = 1, \ldots, I, \qquad (4)$$

gdzie Q'_{iQj} jest hipotetyczną wariacją, tj. oczekiwaną zmianą wolumenu sprzedaży j-tego agenta w odpowiedzi na jednostkowy wzrost sprzedaży i-tego agenta. Udowodniono następujące wzory układu (4) w różnych sytuacjach gry.

1. Przypadek równowagi Cournota, który odpowiada nieświadomości wszystkich agentów (tj. zwolenników) co do działań otoczenia i daje zerowe hipotetyczne wariacje $Q'_{iQj} = 0$:

$$A Q^{\alpha}\left[1+\alpha \frac{Q_i}{Q}\right]-B_i \beta_i Q_i^{\beta_i-1}=0, \quad i=1, \ldots, I \tag{5}$$

2. Przypadek równowagi z przywództwem Stackelberga. W tym przypadku pewien agent (tj. lider) o liczbie k uważa, że jest poinformowany o działaniach otoczenia, tj. istnieje asymetria domniemanej świadomości na rynku. Możliwe jest również, że istnieje kilku liderów, którzy są poinformowani, że pozostali agenci są naśladowcami. W tym przypadku nie zakłada się wzajemnej reakcji liderów, tj. w odniesieniu do siebie liderzy są naśladowcami. W rezultacie układ (4) dla trzech naśladowców (których oznaczamy jako i = 1, ..., I_0) ma postać (5), a dla liderów Stackelberga (których oznaczamy jako k = 1, ..., I_1), równanie (4) ma następującą postać:

$$A Q^{\alpha}\left[1+\alpha \frac{Q_k}{Q}\left(1+\sum_{i=1, i \neq k}^{I} \frac{F_i}{u_i-F_i}\right)\right]-B_k \beta_k Q_k^{\beta_k-1}=0, k=1, \ldots, I_1, \tag{6}$$

gdzie

$$u_i = u_i(Q_i) = B_i \beta_i (\beta_i-1) Q_i^{\beta_i-2}, \quad F_i = F_i(Q, Q_i) = A \alpha Q^{\alpha-1}\left(1+(\alpha-1) \frac{Q_i}{Q}\right),$$

I_1 to liczba liderów pierwszego szczebla.

3. Przypadek równowagi z dwupoziomowym przywództwem Stackelberga. W tym przypadku ze zbioru agentów-liderów wyróżniani są agenci-liderzy drugiego poziomu (oznaczmy liczbę tych agentów przez I_2), informowani, że pozostali liderzy $I_1 = I - I_2 - I_0$ wybierają działania zgodnie z reakcjami (6). Indeksujemy liderów drugiego poziomu symbolem „m", gdzie m = 1, ..., I_2. W rezultacie, system (4) dla agentów Cournota ma postać (5), dla liderów Stackelberga pierwszego poziomu ma postać (6), a dla liderów drugiego poziomu, równanie (4) ma następującą postać:

$$A Q^{\alpha}\left[1+\alpha \frac{Q_m}{Q}\left(1-\sum_{k=1, k \neq m}^{I} \frac{F_k+A \alpha Q^{\alpha-2} Q_k\left[\Sigma_1+Q \frac{F_{m Q_m}^{/} u_m-W_m F_m}{\left(u_m-F_m\right)^2}\right]}{F_k[1+\Sigma_1]-A \alpha Q^{\alpha-1}[1+Q_k \Sigma_2]-W_k}\right)\right] - B_m \beta_m Q_m^{\beta_m-1}=0, \quad m=1, \ldots, I_2, \tag{7}$$

gdzie

$$W_m = W_m(Q_m) = u_{m Q_m}^{\prime} = B_m \beta_m (\beta_m-1)(\beta_m-2) Q_m^{\beta_m-3}, \; F_{m Q_m}^{/} = F_{m Q_m}^{/}(Q, Q_m) = A \alpha(\alpha-1) Q^{\alpha-2}\left(2+(\alpha-2) \frac{Q_m}{Q}\right), \; \Sigma_1 = \Sigma_1\left(F_j, u_j\right) = \sum_{j=1, j \neq k}^{I} \frac{F_j}{u_j-F_j}, \; F_{i Q_k}^{/} = F_{i Q_k}^{/}(Q, Q_i) = A \alpha(\alpha-1) Q^{\alpha-2}\left(1+(\alpha-2) \frac{Q_i}{Q}\right), \; \Sigma_2 = \Sigma_2\left(F_j, u_j\right) = \sum_{j=1, j \neq k}^{I} \frac{F_{j Q_k}^{/} u_j}{\left[u_j-F_j\right]^2}.$$

Rozważamy problem obliczenia współczynników funkcji popytu (2), funkcji kosztów agentów (3) dla rynku ropy naftowej i na podstawie tych danych obliczamy równowagi (5)–(7) w przypadkach 1–3 dla różnych kombinacji liderów i naśladowców.

3 Wyniki i dyskusja

Na rosyjskim rynku ropy naftowej największy udział w rynku posiada spółka naftowa Tatnieft, ponieważ inne firmy dostarczają 2–5% całkowitej sprzedaży ropy na rynek krajowy, podczas gdy Tatnieft sprzedaje 35–50% na rynku krajowym. W 2019 roku największe firmy Rosnieft, Tatnieft i Gazpromnieft posiadały odpowiednio 24,5%, 35,1% i 19,6% udziałów w rynku, co łącznie daje 79,2%. W związku z tym charakterystyka tych firm adekwatnie odzwierciedla trendy i czynniki wpływające na stan rosyjskiego rynku ropy naftowej. Na podstawie danych statystycznych największych firm działających na rosyjskim rynku naftowym (Rosnieft, Tatnieft i Gazpromnieft) analizujemy wolumen sprzedaży ropy naftowej na rynku krajowym, ceny i koszty w latach 2009–2019 w podziale na kwartały (44 okresy). Obliczamy średnią ważoną cenę na podstawie przychodów firm. Stosując metodę najmniejszych kwadratów, definiujemy modele regresji dla funkcji popytu (2) i funkcji kosztów pośredników (3) w następującej postaci:

$$P(Q) = \begin{cases} 9236\,Q^{-0.91}, & t = 2009, \ldots, 2012, \\ 264196770\,Q^{-1.96}, & t = 2013, \ldots, 2019, \end{cases} \tag{8}$$

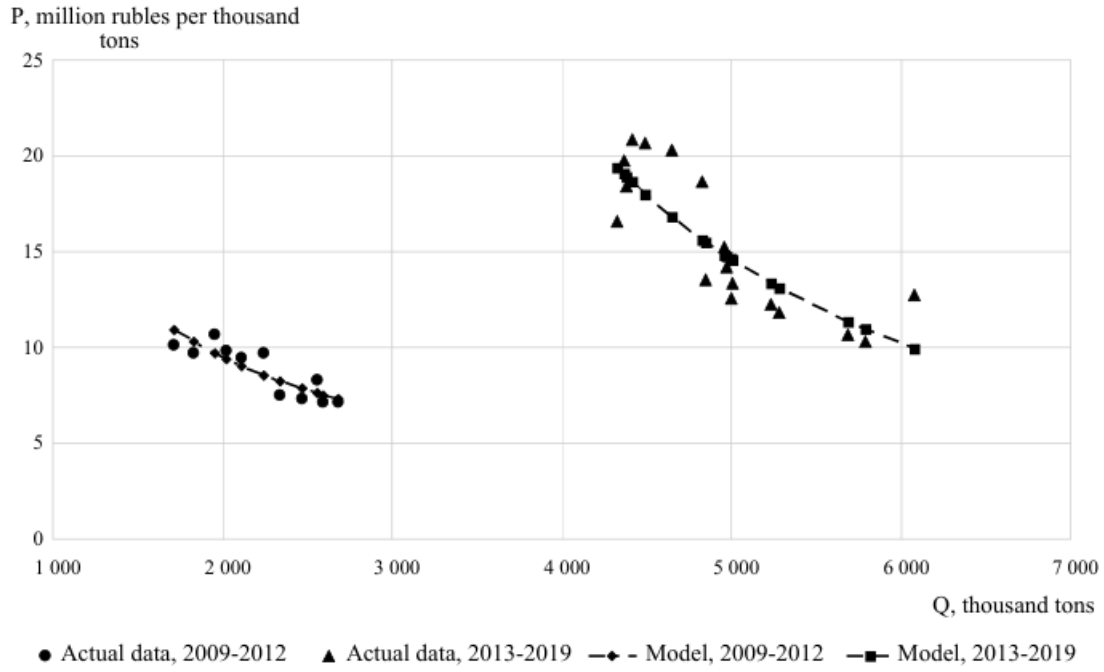
$$C_1(Q_1) = 1.61\,Q_1^{1.13},\ C_2(Q_2) = 2.65\,Q_2^{1.15},\ C_3(Q_3) = 1.04\,Q_1^{1.15}, \tag{9}$$

gdzie t jest rokiem, a wprowadzono następującą indeksację agentów: 1 — Rosnieft, 2 — Tatnieft, 3 — Gazpromnieft. Charakterystyki statystyczne modeli regresji (8), (9) są wysokie (tabele 1 i 2):

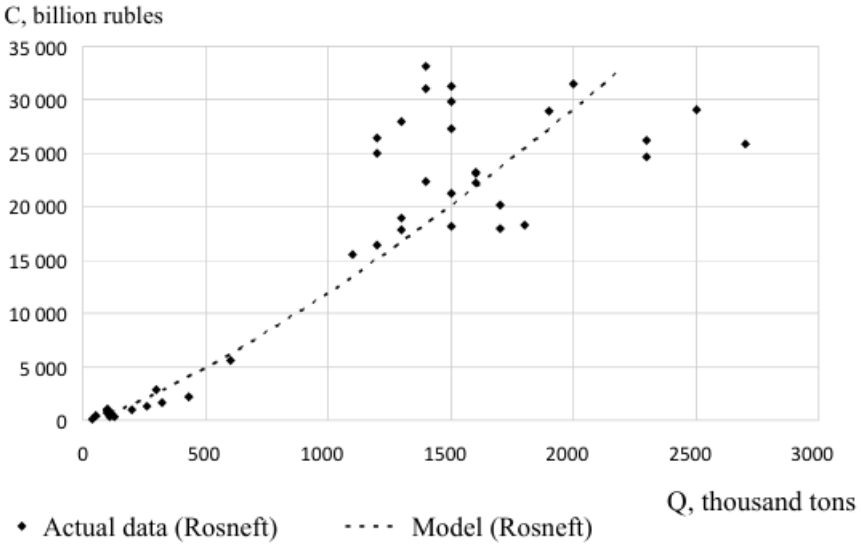
Period	R^2	Fisher criterion	
		Table value ($\alpha = 0.05$)	Actual
2009–2012	0.71	4.45	19.59
2013–2019	0.66	3.74	27.3

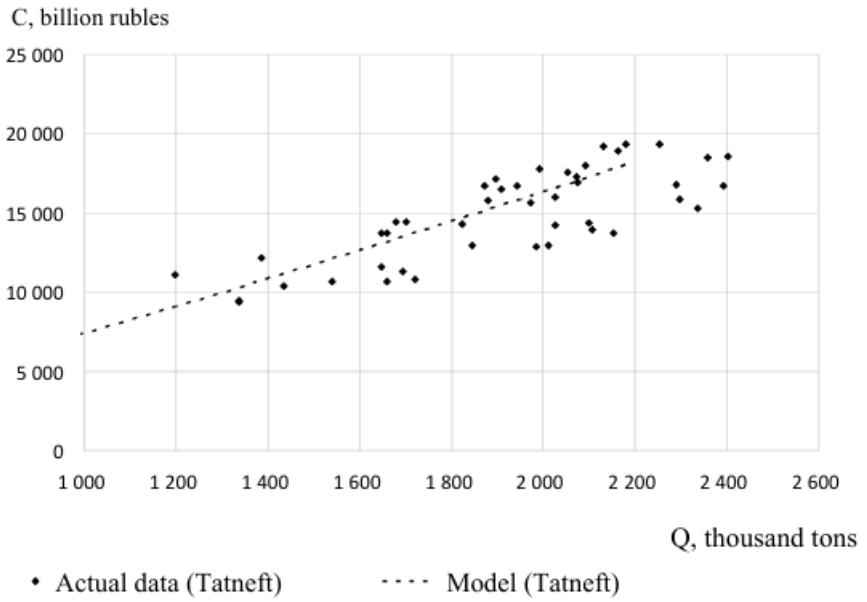
Firm	R^2	Fisher criterion	
		Table value ($\alpha = 0.05$)	Actual
Rosneft	0.95	3.23	774.11
Tatneft	0.65	3.23	76.55
Gazpromneft	0.96	3.23	1 002.26

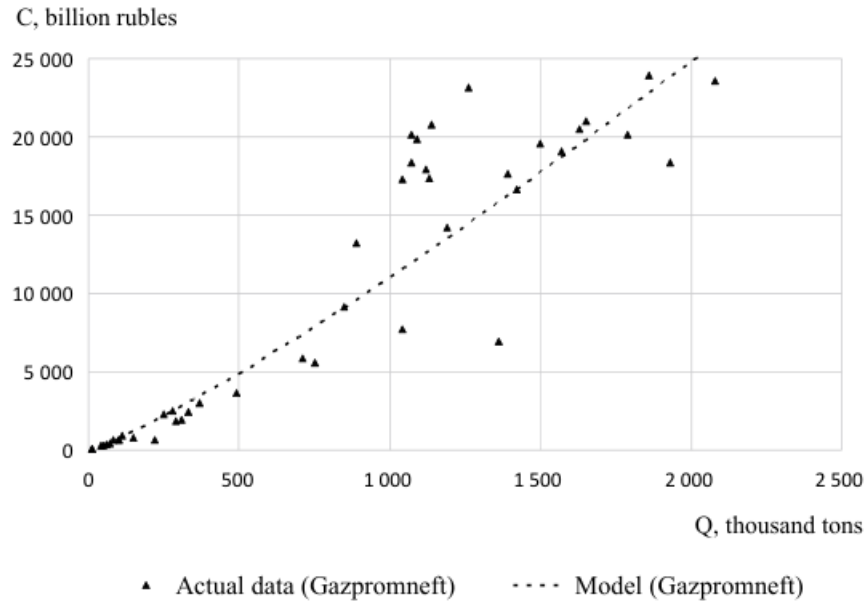
współczynniki determinacji funkcji popytu wynoszą 0,66–0,71, dla funkcji kosztów wynoszą one 0,95, z wyjątkiem funkcji Tatnieft, dla której współczynnik ten wynosi 0,65; obliczona wartość kryterium Fishera jest co najmniej 4-krotnie wyższa od wartości tabelarycznej przy poziomie istotności 0,05. Potwierdza to adekwatność i wiarygodność opracowanych modeli. Analiza szeregów czasowych średnich ważonych cen i wolumenów sprzedaży ropy naftowej na rynku rosyjskim



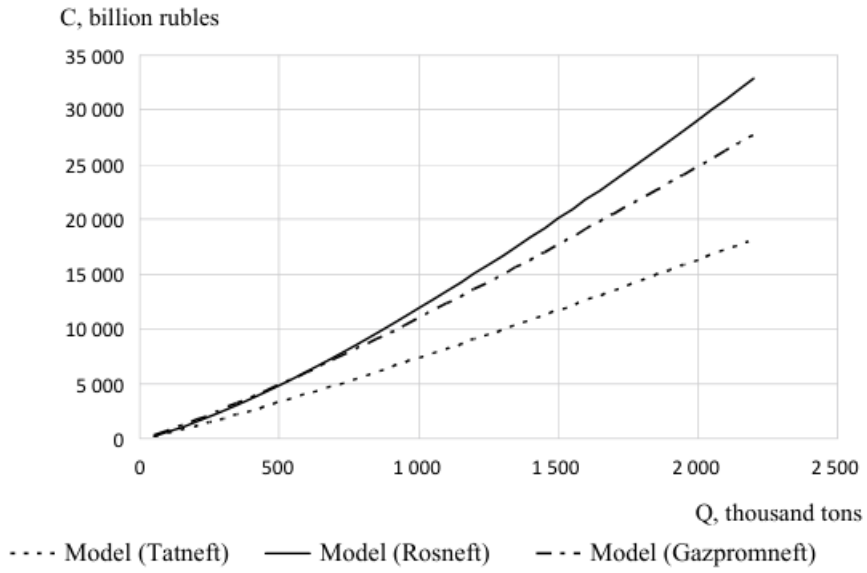
pokazuje, że pomimo państwowej regulacji rynku ropy naftowej, występują tendencje spadkowe cen. Identyfikujemy dwa trendy, z których pierwszy opisuje dynamikę cen ropy naftowej w latach 2009–2012, a drugi – w latach 2013–2019. Co ciekawe, dla tego drugiego przedziału czasowego nachylenie odwrotnej krzywej popytu jest wyższe, tj. popyt stał się mniej elastyczny (współczynnik elastyczności cenowej wyniósł 0,51) niż w latach 2009–2012 (elastyczność cenowa popytu wyniosła 1,01). Ponadto krzywa popytu drugiego przedziału przechodzi nad i na prawo od krzywej pierwszego przedziału, tj. w latach 2013–2019 ceny i wolumeny sprzedaży na rynku były wyższe niż w latach 2009–2012. W konsekwencji rynek się rozszerzył. Przy obliczaniu równowag wykorzystamy model regresji odpowiadający późniejszemu przedziałowi. Aproksymacje funkcji kosztów agentów (9), przedstawione na rysunkach 2, 3 i 4, są funkcjami o rosnącej mocy i malejących korzyściach skali ($\beta > 1$).







Analiza porównawcza trendów kosztów agentów



pokazuje, że w przedziale Q ∈ [500, +∞] krzywa kosztów Tatnieftu znajduje się poniżej krzywej kosztów Gazpromnieftu, która z kolei znajduje się poniżej krzywej Rosnieftu. W konsekwencji Tatnieft ma względną przewagę kosztową, a następnie Gazpromnieft, podczas gdy Rosnieft przegrywa z dwiema pierwszymi firmami pod względem kosztów. Obliczona równowaga Cournota (tabela 3) wykazuje znaczne odchylenie od rzeczywistych wskaźników rynkowych, ponieważ całkowita wielkość sprzedaży w równowadze jest o 31% mniejsza niż rzeczywista wielkość w 2019 r., a cena równowagi przewyższa rzeczywiste ceny firm o 58–121%. W związku z tym firmy nie stosują klasycznych strategii naśladowców, tj. występuje asymetria świadomości. Analiza równowagi Cournota pokazuje, że rozkład wolumenów sprzedaży odpowiada względnej przewadze kosztowej, tj. prognozy modelu są zgodne z logiką. W rzeczywistym rozkładzie rynku firmy te są również rozwarstwione według zysku, zgodnie z ich względną przewagą kosztową, chociaż Rosnieft zajmuje drugie miejsce pod względem udziału w rynku, a Gazpromnieft trzecie. Oznacza to, że model Cournota wykazuje częściową adekwatność rzeczywistości, częściowo w odniesieniu do rozwarstwienia agentów pod względem maksymalnego zysku. W grze walki o przywództwo Stackelberga (tabela 4),

Indicators	Agent	Q_i, thousand tons	Q, thousand tons	P, RUB mln per thousand tons	R, RUB mln	C, RUB mln	П, RUB mln
Optimal	Rosneft	1319	3781	25	33,329	17,009	16,323
	Tatneft	1202			30,366	9096	21,278
	Gazpromneft	1260			31,858	14,441	17,419
Actual	Rosneft	1500	4629	22	33,600	31,213	2387
	Tatneft	2089		16	33,468	16,780	16,688
	Gazpromneft	1040		19	19,885	17,281	2603
Deviation	Rosneft	−12%	−18%	13%	−1%	−46%	584%
	Tatneft	−42%		58%	−9%	−46%	28%
	Gazpromneft	21%		32%	60%	−16%	569%

jeśli wszyscy trzej agenci są liderami pierwszego poziomu, rozkład sprzedaży jest najbliższy rzeczywistemu rozkładowi rynku. W tym przypadku odchylenie obliczonej wartości całkowitego wolumenu sprzedaży od wartości rzeczywistej w 2019 r. wynosi 18%, odchylenie cen równowagi mieści się w przedziale 13–58%, odchylenia przychodów wynoszą odpowiednio 1%, 9% i 60% dla pierwszego, drugiego i trzeciego agenta. W konsekwencji asymetria świadomości agentów wyraża się w tym, że zachowania firm podążają za zasadą postrzegania kontrahentów jako naśladowców. Najciekawsze wyniki uzyskujemy w przypadku liderów drugiego poziomu (tabele 5 i 6).

Indicators	Agent	Q_i, thousand tons	Q, thousand tons	P, RUB mln per thousand tons	R, RUB mln	C, RUB mln	П, RUB mln
Optimal	Rosneft	2315	6637	8	19,398	35,148	−15,750
	Tatneft	2970			24,888	18,705	6183
	Gazpromneft	1352			11,327	15,673	−4346
Actual	Rosneft	1500	4629	22	33,600	31,213	2387
	Tatneft	2089		16	33,468	16,780	16,688
	Gazpromneft	1040		19	19,885	17,281	2603
Deviation	Rosneft	54%	43%	−63%	−42%	13%	−760%
	Tatneft	42%		−48%	−26%	11%	−63%
	Gazpromneft	30%		−56%	−43%	−9%	−267%

Indicators	Agent	Q_i, thousand tons	Q, thousand tons	P, RUB mln per thousand tons	R, RUB mln	C, RUB mln	П, RUB mln
Optimal	Rosneft	2238	5537	12	26,760	33,642	−6883
	Tatneft	2782			33,268	16,844	16,424
	Gazpromneft	517			6181	5079	1102
Actual	Rosneft	1500	4629	22	33,600	31,213	2387
	Tatneft	2089		16	33,468	16,780	16,688
	Gazpromneft	1040		19	19,885	17,281	2603
Deviation	Rosneft	49%	20%	−47%	−20%	8%	−388%
	Tatneft	33%		−25%	−1%	0%	−2%
	Gazpromneft	−50%		−37%	−69%	−71%	−58%

Opisujemy grę, w której liderem drugiego poziomu jest Tatnieft, a liderami pierwszego poziomu są Rosnieft i Gazpromnieft, ponieważ sytuacja ta odpowiada rzeczywistemu rynkowi. W tej grze poszukujemy dwóch atraktorów, z których każdy lider drugiego poziomu (Tatnieft) ma przewagę w sprzedaży i osiąga największy maksymalny zysk w porównaniu z konkurentami. W drugim atraktorze łączna sprzedaż agentów jest niższa, a obliczona cena równowagi wyższa, w wyniku czego zysk lidera drugiego poziomu jest wyższy niż w pierwszym. W konsekwencji w tym przypadku występują bifurkacje trajektorii poszukiwania równowagi. Ponadto, w drugim atraktorze, odchylenie obliczonej wartości całkowitej wielkości sprzedaży od wartości rzeczywistej w 2019 roku wynosi 20%, odchylenie cen równowagi mieści się w przedziale 48–63%, a odchylenia przychodów wynoszą odpowiednio 20%, 1% i 69% dla pierwszego, drugiego i

trzeciego agenta. Zatem drugi atraktor gry z dwoma liderami pierwszego poziomu i liderem drugiego poziomu również dość ściśle odpowiada rzeczywistemu rynkowi, podobnie jak wspomniana wcześniej gra o przywództwo z liderami pierwszego poziomu.

4. Wnioski

Badamy modele gier oligopolu z nieliniowymi funkcjami popytu rynkowego i kosztów firmy. Rozważamy następujące modele zachowań agentów: naśladowcy, czyli agenci odpowiadający hipotezie Cournota, oraz liderzy Stackelberga pierwszego i drugiego poziomu. Analiza statystyczna rosyjskiego rynku ropy naftowej dowodzi malejącej nieliniowej krzywej popytu. W konsekwencji okazało się możliwe modelowanie stanów tego rynku w postaci gry agregacyjnej z nieliniowymi funkcjami użyteczności agentów. Główną różnicą pomiędzy równowagami obliczonymi w niniejszym badaniu w przypadku nieliniowej krzywej popytu a równowagami w przypadku liniowej zależności popytu jest niejednoznaczność decyzji w grze w warunkach liderów drugiego poziomu. Ponieważ w grze występują co najmniej dwa atraktory, nie jest wykluczona możliwość istnienia innych atraktorów, przy przechodzeniu między nimi powstają bifurkacje. Zatem modelowanie oparte na rzeczywistych statystykach rynkowych potwierdziło wnioski teoretyczne, które zostały wyprowadzone wcześniej na podstawie parametrów testowych modelu i opisane w przeglądzie literatury.

Planowanie optymalnego obciążenia transportu kolejowego w przedsiębiorstwie górniczym

Tu rozpatrywany jest problem planowania optymalnego obciążenia pociągów towarowych w ograniczonej sieci kolejowej – sieci przedsiębiorstwa górniczego. Algorytm rozwiązania stanowi kombinację rozwiązań prywatnych problemów harmonogramowania dla pojedynczej lokomotywy, uwzględniając ograniczenia dotyczące możliwości ich wspólnego ruchu. W rozdziale przedstawiono sformułowanie problemu, w tym obecność kilku lokomotyw różnych typów, algorytm rozwiązania i jego zastosowanie w szczególnym przypadku.

Słowa kluczowe: Programowanie liniowe całkowitoliczbowe · Problemy planowania kolejowego · Teoria harmonogramów

1 Wstęp

Funkcjonowanie transportu kolejowego w przedsiębiorstwach górniczych bezpośrednio wpływa na wielkość i ciągłość dostaw surowców mineralnych. Złożoność zadania planowania optymalizacji obciążenia transportu kolejowego zależy od wielkości sieci kolejowej, jej zamknięcia, kierunków przewozu ładunków, dostępności dużej floty pojazdów kolejowych, konieczności transportu różnych rodzajów materiałów itp. Chociaż inżynieria zarządzania jest obecnie dojrzałą dyscypliną, nadal pozostaje obszarem, w którym prowadzi się wiele badań . Optymalizacja obciążenia transportu kolejowego pozwala na obniżenie kosztów przewozu towarów, tworzenie rezerw przepustowości i dostarczanie towarów w określonym czasie, co poprawia efektywność funkcjonowania sieci kolejowej. Stworzono i udoskonalono szereg zautomatyzowanych systemów służących efektywnemu zarządzaniu transportem w przemyśle . Zwiększa się możliwości sieci centrów informatycznych i komputerowych . Wszystko to ma na celu podejmowanie racjonalnych decyzji dotyczących wykorzystania taboru kolejowego. Doskonalenie technologii eksploatacji kolei, a przede wszystkim zarządzania eksploatacją taboru, stanowi ważne zadanie dla poprawy efektywności funkcjonowania kolei . Dlatego zadania

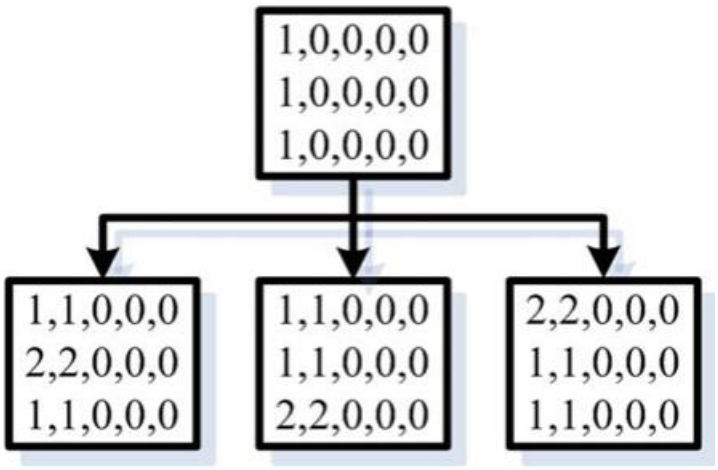
planowania i optymalizacji kolei pozostają aktualne i wymagają opracowania nowych podejść do ich rozwiązywania .Przedstawiono dwa zagadnienia problemowe, pierwsze dotyczy kilku lokomotyw tego samego typu, drugie lokomotyw różnych typów, a także algorytm rozwiązywania takich problemów i jego zastosowanie w konkretnym przypadku.

2. Sformułowanie problemu z jednym typem lokomotyw

Rozważono formalizację i rozwiązanie problemu planowania ruchu dla pojedynczej lokomotywy. Poniżej przedstawiono formalizację uwzględniającą kilka lokomotyw o tych samych parametrach, przewożących ładunki między kilkoma stacjami. Rozważmy następujące oznaczenia. Niech h będzie liczbą lokomotyw, h = 1...,m, gdzie m jest całkowitą liczbą lokomotyw. Wówczas powiemy, że lokomotywa h znajduje się w stanie $S^h(s^h, t^h, k^h)$, jeśli w chwili $t^h \in T$ znajduje się na stacji s^h i do tej chwili dostarczyła $k^h = \{k^h_{ij}, i \neq j, i, j = 1, ..., n\}$ zamówień, gdzie n jest całkowitą liczbą stacji, a k^h i j jest liczbą zamówień przetransportowanych z i-tej stacji do j-tej. Istnieją również zlecenia określające ilość ładunku do przetransportowania z i-tej stacji wyjazdu do j-tej stacji przeznaczenia. Zlecenia pojawiają się w określonych momentach i nie mogą być zrealizowane wcześniej. Konieczne jest opracowanie harmonogramu ruchu lokomotyw, który zapewni realizację wszystkich zleceń. Harmonogram jest określony stanem lokomotyw w każdym momencie. Całkowity czas realizacji wszystkich zleceń przez wszystkie lokomotywy w ostatnim stanie (momencie realizacji wszystkich zleceń) jest wykorzystywany jako kryterium optymalności harmonogramowania:

$$F = \sum_{h=1}^{m} t^h. \qquad (1)$$

Należy zauważyć, że w tej sekcji rozważamy rozwiązanie problemu z kilkoma lokomotywami tego samego typu. W tym przypadku stan przestaje być wektorem i staje się macierzą, gdzie h-ty wiersz jest stanem h-tej lokomotywy. Ze względu na fakt, że rozważane są lokomotywy tego samego typu, będziemy uważać stany zawierające ten sam zestaw wierszy za identyczne, jak na rysunku, i również je ze sobą skleimy.



Tutaj stany mają format (s, t, k_{12}, k_{13}, k_{23}). Przedstawiono algorytm sklejania wierzchołków do rozwiązania problemu formowania pociągów i harmonogramowania ich ruchu dla jednej

lokomotywy. W tej części przedstawiono rozwój tego algorytmu dla przypadku z kilkoma lokomotywami tego samego typu, a także z kilkoma lokomotywami różnych typów. Same algorytmy i przykłady ich zastosowania przedstawiono poniżej.

3 Algorytm dla lokomotyw jednego typu

Drzewo stanów konstruuje się poprzez wyliczenie różnych wariantów ruchu lokomotywy. W każdej chwili, uwzględniając ograniczenia , lokomotywa może znajdować się w jednym z następujących stanów.

1 Oczekiwanie to stan, w którym lokomotywa oczekuje na nowe zamówienie na jednej stacji; dopuszcza się kilka kolejnych okresów oczekiwania.

2 Bezczynność to stan, w którym lokomotywa pracuje na pusto; niedopuszczalne jest więcej niż jedno kolejne bezczynność.

3 Dostawa to stan, w którym lokomotywa dostarcza zamówienie z punktu odjazdu do miejsca przeznaczenia.

4 Pusty to stan, w którym lokomotywa nie działa.

Aby zmniejszyć wymiarowość drzewa stanów, zastosowano następujący algorytm sklejania jego wierzchołków.

Krok 1. Dodajemy nowy stan do drzewa i ustalamy go jako macierz S_j, której każdy wiersz reprezentuje stan pojedynczej lokomotywy.

Krok 2. Spośród skonstruowanych stanów szukamy takiego stanu, w którym zbiór wierszy pokrywa się ze zbiorem wierszy macierzy S_j, porównując je wiersz po wierszu. Oznacza to, że dla każdego k-tego wiersza w macierzy Sj istnieje podobny h-ty wiersz w macierzy S_i, $S^h_i = S^k_j$. Jeśli taki warunek nie występuje, przechodzimy do kroku 1. W przeciwnym razie przechodzimy do kroku 3. Następnie znajdujemy stan nadrzędny S_{i-1}, poprzedzający stan S_i, oraz stan S_{j-1}, poprzedzający S_j. Porównujemy je, licząc wartości stanów c_{i-1} i c_{j-1} używając wzoru:

$$c_z = \sum_{h=1}^{m} \sum_{l,q=1}^{n} k^h_{z_{lq}} \qquad (2)$$

gdzie z jest numerem stanu równym (i−1) lub (j−1), l jest stacją odjazdu, q jest stacją przeznaczenia. Wówczas możliwe są dwie opcje.

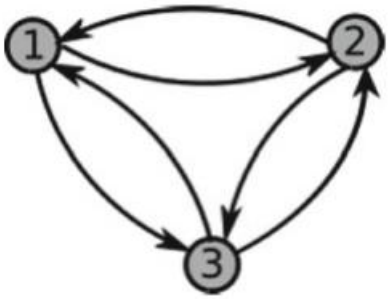
Krok 3a. Jeśli $c_{i-1} \geq c_{j-1}$ przerwij budowanie gałęzi zawierającej stan S_j i usuń ją. Przejdź do kroku 1.

Krok 3b. Jeśli $c_{i-1} < c_{j-1}$, usuń gałąź zawierającą stan S_i, aż do S_i włącznie, i przenieś stany z kolejnych poziomów do stanu S_j. Przejdź do budowy kolejnej gałęzi (krok 1).

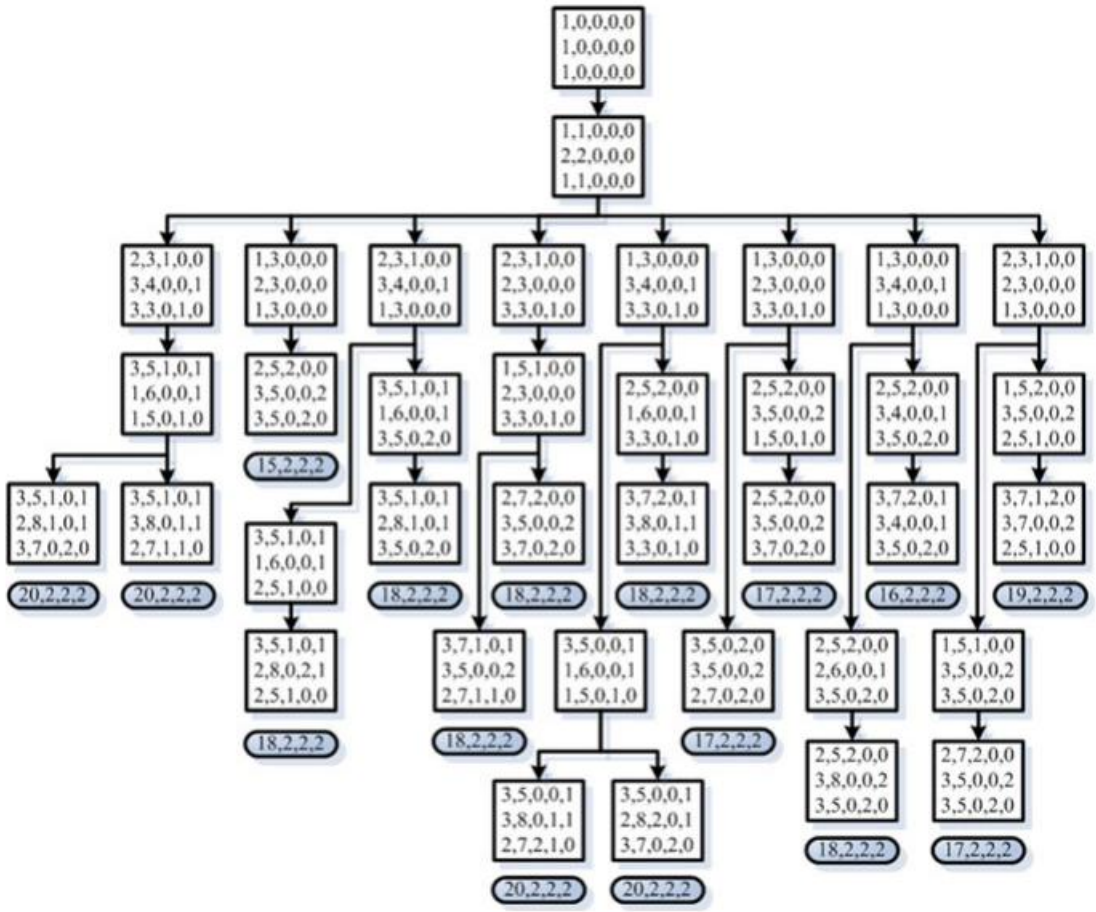
Wzór (2) to suma wszystkich wykonanych zleceń dla danego stanu. Za bardziej opłacalny uznamy harmonogram, w którym w danym momencie wykonano więcej zleceń.

4 Przykład z kilkoma lokomotywami jednego typu

Rozważmy przykład rozwiązania problemu harmonogramowania z trzema lokomotywami dla trzech stacji rozmieszczonych w okręgu. Użyjmy oznaczeń wprowadzonych powyżej. Układ wierzchołków pokazano na rysunku 2.



Ustawmy stacje 1 i 2 jako stacje odjazdowe, 2 i 3 jako stacje docelowe, trasy 12 (ze stacji 1 do stacji 2), 13, 23. Zamówienia docierają do stacji odjazdowych w czasie 1 i 3, wielkość jednego zamówienia wynosi 1 jednostkę, a pojemność każdej lokomotywy wynosi 2 jednostki. Zadanie polega na znalezieniu harmonogramu, w którym wszystkie zamówienia zostaną zrealizowane w najkrótszym czasie. Zbudujmy drzewo stanów dla tego stwierdzenia, wykorzystując algorytm sklejania wierzchołków. Wynik pokazano na rysunku 3.



Na dole, po każdym harmonogramie, obliczana jest wartość kryterium optymalności zgodnie ze wzorem (1) i podawana jest liczba zrealizowanych zamówień dla każdej trasy. Jak widać na rysunku 3, otrzymano 16 możliwych rozkładów jazdy, przy czym najniższa wartość kryterium

optymalności wynosi 15. Wynik ten uzyskano poprzez oczekiwanie na zamówienia do momentu zapełnienia lokomotyw.

5. Sformułowanie problemu z różnymi typami lokomotyw

W porównaniu z poprzednim sformułowaniem, gdy problem obejmuje lokomotywy o różnych pojemnościach, wymiarowość tego problemu wzrasta, a w konsekwencji wzrasta również złożoność. Rozważmy następujące oznaczenia. Niech h_l będzie liczbą lokomotyw typu l, l = 1,..., L, L jest liczbą typów lokomotyw, $\sum_{l=1}^{L} h_l = m$, gdzie m to całkowita liczba lokomotyw. Załóżmy, że lokomotywa h_{lg} znajduje się w stanie $S^{h_{lg}}(s^{h_{lg}}, t^{h_{lg}}, k_{ij}^{h_{lg}}), h_{lg} = 1, ..., h_l$, g jest numerem lokomotywy danego typu, jeśli w chwili $t^{hlg} \in T$ znajduje się ona na stacji s^{hlg} i do tego momentu dostarczyła k^{hlg}_{ij} zamówień ze stacji i do stacji j, i ≠ j, i, j = 1,..., n, gdzie n jest całkowitą liczbą stacji. Istnieją również zamówienia, które określają ilość ładunku do przetransportowania z i-tej stacji wyjazdu do j-tej stacji przeznaczenia. Zamówienia pojawiają się w określonych momentach i nie mogą być zrealizowane wcześniej. Konieczne jest sporządzenie harmonogramu ruchu lokomotyw, który zapewni realizację wszystkich zamówień. Harmonogram jest określony przez stan lokomotyw w każdej chwili. Jako kryterium optymalności przyjęto minimalny czas realizacji wszystkich zleceń przez wszystkie lokomotywy, czyli minimalny czas spośród maksymalnych czasu realizacji wszystkich zleceń w stanie ostatnim (moment realizacji wszystkich zleceń):

$$F = \min(\max_i t_{end}^i) \qquad (3)$$

gdzie t^i_{end} to czas zakończenia i-tego ruchu, i = 1,..., m. W tym sformułowaniu należy zauważyć, że stan systemu transportowego jest tu opisany trójwymiarową tablicą, gdzie każda macierz jest stanem wszystkich lokomotyw odpowiedniego typu. Ponadto algorytm konstrukcji drzewa stanów jest podobny do poprzedniego, różniąc się dużym wymiarem ze względu na ograniczenia . Dla opracowanego algorytmu wyszukiwania optymalnego harmonogramu zaproponowano następującą modyfikację algorytmu sklejania wierzchołków.

6 Algorytm dla różnych typów lokomotyw

Krok 1. Dodajemy nowy stan do drzewa i ustalamy go jako bieżący $S^{h_{lg}}(s^{h_{lg}}, t^{h_{lg}}, k_{ij}^{h_{lg}})^{(cur)}$. Jest to trójwymiarowa tablica, której każdy dwuwymiarowy wycinek jest stanem wszystkich lokomotyw odpowiedniego typu l, tj. linia w niej opisuje stan jednej lokomotywy.

Krok 2. Znajdź wśród skonstruowanych stanów $S^{h_{lg}}(s^{h_{lg}}, t^{h_{lg}}, k_{ij}^{h_{lg}})^{(alt)}$, dla którego spełniony jest warunek: dla dowolnej lokomotywy g, d = 1, ..., hl, l = 1,..., L

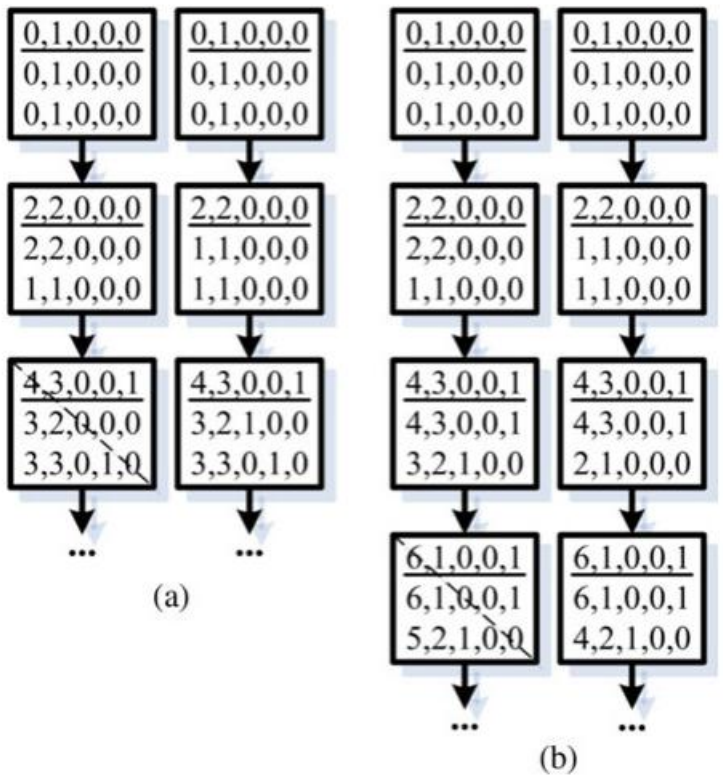
$$S^{h_{lg}} = s^{h_{ld}}, k_{ij}^{h_{lg}(cur)} \geq k_{ij}^{h_{ld}(alt)}$$

Jeśli taki stan nie zostanie znaleziony, przejdź do kroku 1. W przeciwnym razie zapisz go jako alternatywę i przejdź do kroku 3. Krok 3. Znajdź czasy przejazdu lokomotyw w stanie bieżącym i w stanie alternatywnym i wybierz najkrótszy z nich. Jeśli $\max_i t^{i(cur)} > \max_i t^{i(alt)}$, usuwamy bieżącą trasę i przechodzimy do konstruowania następnej gałęzi (krok 1). Jeśli

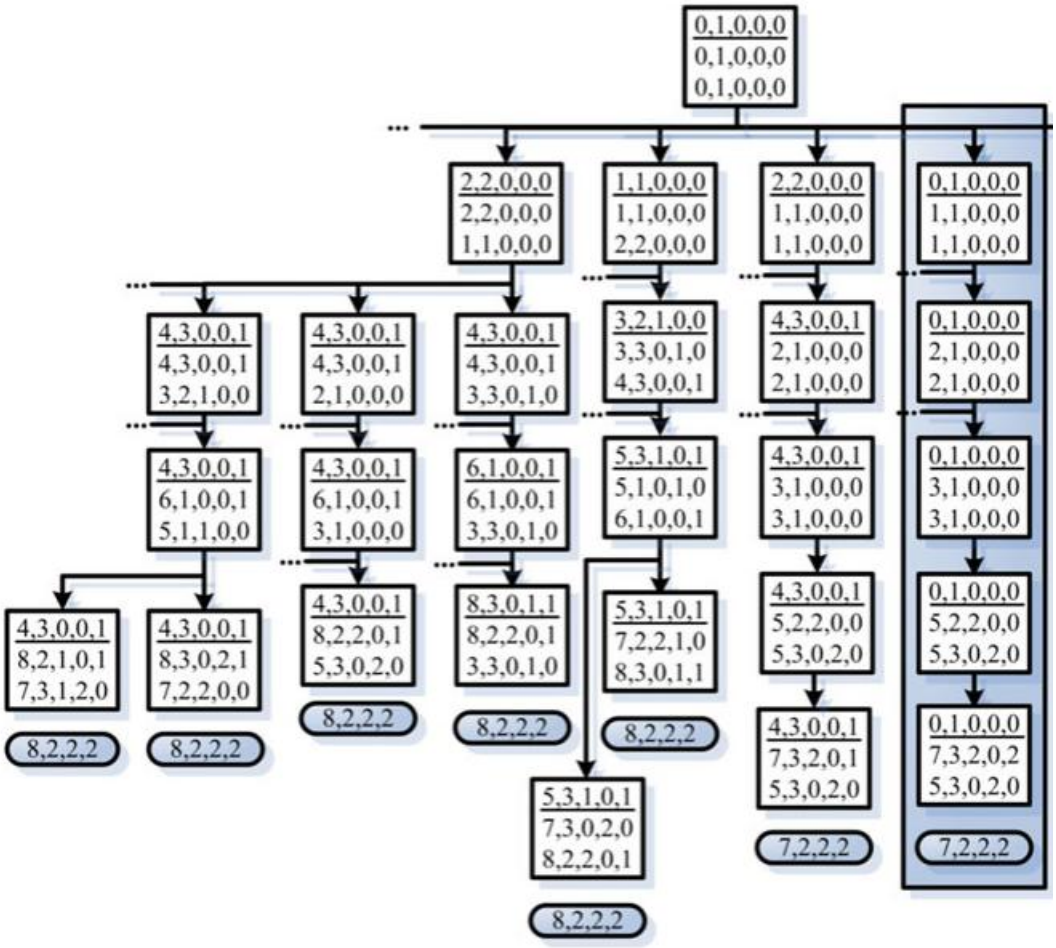
$\max_i t^{i(cur)} = \max_i t^{i(alt)}$, porównujemy całkowity czas przejazdu lokomotyw: $\sum_{i=1}^{m} t^i$ ti. Pozostawiamy trasę z krótszym czasem przejazdu. Jeśli $\max_i t^{i(cur)} < \max_i t^{i(alt)}$, trasa zawierająca stan (alt) jest usuwana, a bieżąca trasa jest budowana (krok 1). Zatem uznamy za bardziej opłacalny harmonogram, w którym nie mniej zleceń zostanie zrealizowanych w krótszym czasie. W przypadkach, gdy taka sama liczba zleceń zostanie zrealizowana w równym czasie, priorytet otrzyma trasa, która została już w pełni zbudowana.

7 Przykład z kilkoma lokomotywami różnych typów

Jako przykład rozważmy sieć kolejową z powyższego przykładu, układ stacji pokazano na rysunku 1. Załóżmy, że mamy dwa typy lokomotyw o różnej pojemności: jedną lokomotywę o pojemności 1 jednostki i dwie lokomotywy o pojemności 2 jednostek. Ustawmy stacje odjazdowe 1 i 2, stacje docelowe 2 i 3, trasy 12, 13,23 jak w powyższym przykładzie. Zamówienia docierają do stacji odjazdowej w czasie 1 i 3, a rozmiar jednego zamówienia wynosi 1 jednostkę. Należy znaleźć harmonogram, w którym wszystkie zamówienia zostaną wykonane w najkrótszym czasie. Drzewo stanów dla tego problemu będzie miało dużą liczbę gałęzi, dlatego poniżej przedstawiono przykłady sklejania węzłów i przykłady gotowych tras. Zatem rysunek 4 przedstawia przykłady sklejania wierzchołków.



Stany mają format (t, s, k_{12}, k_{13}, k23). Pierwszy wiersz przedstawia stan lokomotywy typu 1, drugi i trzeci – stany dwóch lokomotyw typu 2. W przypadku 4a mamy do czynienia z sklejeniem, gdy różna liczba zleceń jest transportowana w równym czasie, a w przypadku 4b – z osiągnięciem tego samego stanu w różnym czasie. Rysunek 5 przedstawia fragment ostatecznego drzewa harmonogramu, uwzględniający najbardziej optymalny harmonogram.



Wartość kryterium optymalności, obliczona ze wzoru (3), oraz liczba zrealizowanych zleceń dla każdej trasy są podane na dole, po każdym harmonogramie. Najmniejsza wartość kryterium optymalności wynosi 7 i została osiągnięta w kilku harmonogramach, a my wybieramy ten, który charakteryzuje się krótszym łącznym czasem pracy lokomotyw. Pierwszy harmonogram ma łączny czas pracy równy 16, a drugi 12. Na rysunku 5 jest on oznaczony prostokątem. W procesie budowy sklejono ze sobą 24 pary tras.

8. Wnioski

Tu przedstawiono zatem dwa algorytmy planowania optymalnego obciążenia transportu kolejowego – dla lokomotyw jednego typu i dla lokomotyw różnych typów. Stan systemu opisany jest w pierwszym przypadku macierzą, której każdy wiersz odpowiada oddzielnej lokomotywie, a w drugim przypadku trójwymiarową tablicą, gdzie macierze opisują stany lokomotyw odpowiednich typów. Harmonogram przedstawia stan każdej lokomotywy w każdej chwili. Podano algorytmy znajdowania optymalnego harmonogramu. Rozwiązanie przedstawiono jako strukturę hierarchiczną odpowiadającą wszystkim możliwym wariantom stanu dla każdej chwili.

Sklejenie identycznych stanów według kryteriów optymalności redukuje wymiarowość problemu. Przedstawiono rozwiązanie szczególnego przypadku opisywanego problemu, a mianowicie harmonogramowanie ruchu trzech lokomotyw między trzema stacjami rozrządowymi, które znajdują się w formie okręgu. W przyszłości planowane jest wykorzystanie tych algorytmów jako podstawy inteligentnego systemu harmonogramowania ruchu kolejowego.

Inteligentny System Diagnostyki Złączy Kolejowych

Poprawa niezawodności transportu kolejowego jest zapewniona dzięki stałemu monitorowaniu stanu torów. W rozdziale zaproponowano głęboką, splotową sieć neuronową do rozpoznawania i klasyfikowania uszkodzeń styków szynowych na podstawie obrazów szyn. Nowoczesne wagony pomiarowe są wyposażone w kamery wideo do odbioru klatek torów. Jednak przetwarzanie danych wideo przez operatorów w czasie rzeczywistym jest niemożliwe ze względu na dużą ilość informacji i dużą prędkość wagonu pomiarowego. Styki szynowe są elementami najbardziej narażonymi na wpływy zewnętrzne, które powodują awarie w ich montażu. Opisano strukturę kompleksu diagnostycznego do monitorowania torów kolejowych. Podczas przejazdu wagonu pomiarowego, na nagraniu wideo szyny, tworzona jest tablica obrazów styków szynowych. Rozważono utworzenie zbioru klas styków szynowych, obejmujących stany uszkodzone oraz obrazy bez szyn. Zbiór ten obejmuje standardowe styki szynowe ze złączami, złącza izolacyjne oraz złącza spawane. Zidentyfikowano dodatkowe klasy, odpowiadające różnym anomaliom konfiguracji styków szynowych na obrazach. Zaproponowano zmodyfikowaną strukturę wstępnie wytrenowanej głębokiej sieci splotowej do klasyfikacji styków szynowych i wykrywania stanów wadliwych. Rzeczywiste obrazy styków szynowych uzupełniono obrazami sztucznymi uzyskanymi za pomocą transformacji afinicznych podczas przygotowywania próbek treningowych. Opisano proces trenowania i testowania klasyfikatora w oparciu o opracowaną sieć splotową. Do trenowania sieci neuronowej wykorzystano metodę transferu uczenia (Transfer Learning). W wyniku przeprowadzonych eksperymentów dokładność klasyfikacji styków szynowych i wykrywania usterek wyniosła co najmniej 96%. Zaproponowano wykorzystanie metody wizualizacji akcentowanej z wykorzystaniem rzeczywistości rozszerzonej w połączeniu ze sztuczną siecią neuronową do szkolenia operatorów wózków pomiarowych w zakresie wykrywania usterek w mocowaniach szynowych i stykach na klatkach wideo toru kolejowego.

Słowa kluczowe: Diagnostyka techniczna · System cyberfizyczny · Sztuczna sieć neuronowa · Uczenie maszynowe · Tor kolejowy · Złącza

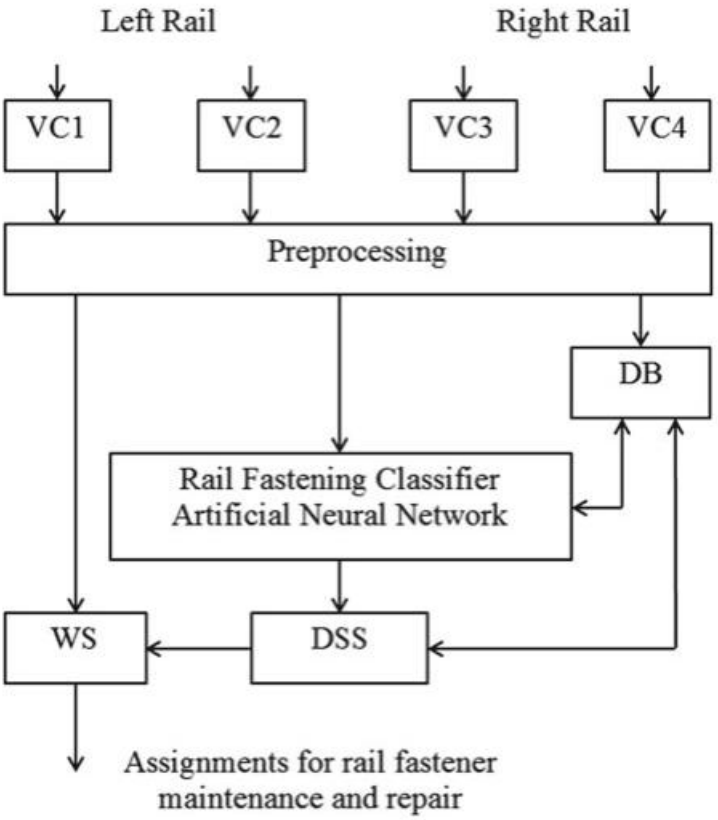
1 Wstęp

Współczesne koleje dużych prędkości to systemy cyberfizyczne, składające się z oddziałujących na siebie podsystemów: torów kolejowych, taboru kolejowego, systemu energetycznego, systemów informacyjno-pomiarowych i systemów sterowania. Kluczowymi kryteriami funkcjonowania kolei są niezawodność i rentowność . Jednocześnie najpoważniejsze problemy pojawiają się w obszarze intensywnej interakcji energetycznej zestawów kołowych pojazdów kolejowych z torem kolejowym. Styk szynowy to połączenie dwóch sąsiednich końców szyn za pomocą mocowania. Styk szynowy musi zapewniać przejście koła z szyny na szynę przy minimalnym tarciu toru. Styk szynowy, najsłabszy element toru, jest stale badany i udoskonalany. Podczas eksploatacji w stykach szynowych mogą wystąpić usterki, prowadzące do niepożądanych, a czasem katastrofalnych skutków. Diagnostyka wibracyjna i statystyczne

przetwarzanie sygnałów oraz obserwacja wizualna to najskuteczniejsze metody określania stanu technicznego elementów toru. Ponadto styk szynowy bierze udział w tworzeniu i działaniu elektrycznego obwodu kolejowego. Podstawowym celem obwodu torowego jest wykrywanie obecności taboru na odcinku toru tworzącym obwód. Łańcuchy szynowe są głównym elementem automatycznej blokady wykorzystywanej do różnych celów w scentralizowanych urządzeniach sterowania ruchem, automatycznych bramkach przejazdowych i zmechanizowanych stacjach rozrządowych . Jakość elektrycznego obwodu szynowego zależy od stanu styku i wartości jego rezystancji elektrycznej. W tym kontekście konieczna jest kontrola integralności konstrukcji styku szynowego. Występuje problem z operacyjnym przetwarzaniem i analizą danych z monitorowania toru. Kontrola wizualna polega na filmowaniu z dużą prędkością setek kilometrów toru podczas jednego przejazdu wagonu pomiarowego. Rezultatem są dziesiątki tysięcy obrazów, których operatorzy nie są w stanie przetwarzać w czasie rzeczywistym. Rozwiązanie tego problemu zaproponowano wykorzystując inteligentne metody przetwarzania dużych tablic danych. W tym kontekście kontrola i diagnostyka stanu styku szynowego z wykorzystaniem sztucznych sieci neuronowych stanowi problem priorytetowy.

2 Klasyfikacja stanów technicznych styków szynowych

Zespół diagnostyczny jako część wagonu pomiarowego został opisany . Kompleks składa się z czterech kamer wideo VC1–VC4 do ciągłego rejestrowania obrazu wizualnego szyn, jednostki wstępnego przetwarzania obrazu oraz systemu wspomagania decyzji DSS o stanie technicznym toru kolejowego .



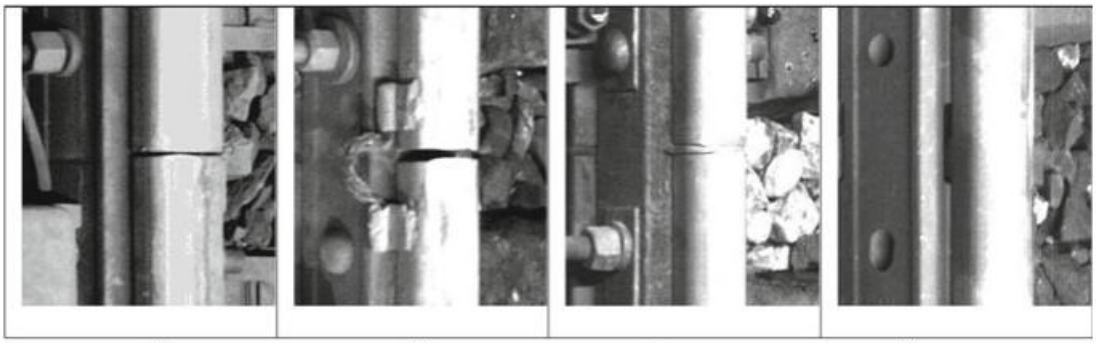
Dane z nagrań wideo i wyniki przetwarzania obrazu są przechowywane w bazie danych DB. Klasyfikacja mocowań szynowych i identyfikacja ich usterek zostały przeprowadzone za pomocą splotowej sieci neuronowej i przeanalizowane przez operatora na stanowisku roboczym WS. Ten kompleks diagnostyczny był już wykorzystywany wyłącznie do kontroli mocowań szynowych. W niniejszym rozdziale proponujemy modernizację kompleksu diagnostycznego poprzez budowę głębokiej sieci neuronowej do wykrywania uszkodzeń styków szynowych. Rozważane są trzy główne typy styków szynowych:

• styk szynowy z nakładkami stykowymi;

• styk izolacyjny;

• styk spawany.

Pierwszy typ odpowiada połączeniu szynowemu z nakładkami stykowymi i śrubami stykowymi. Szczelina stykowa w takim styku szynowym może wynosić zero. Nakładka stykowa jest zazwyczaj mocowana czterema lub sześcioma śrubami. Istnieją dwa warianty styku szynowego z nakładkami stykowymi: bez łącznika i ze łącznikiem.



Rysunki a i b przedstawiają przykłady styków szynowych z nakładkami stykowymi. Obraz styku izolacyjnego przedstawiono na rysunek c, a obraz styku spawanego na rysunek d. W niniejszym rozdziale rozważane są dwa główne typy uszkodzeń, które mogą wystąpić w typowym styku ze łącznikiem:

• styk jest odcięty z jednej strony;

• styk jest całkowicie odcięty.

Przykłady wykrywalnych uszkodzeń przedstawiono odpowiednio na rysunku a i b.

a) b)

Na rysunku a jeden z końców łącznika nie jest zamocowany i wisi w powietrzu. Ślady łącznika są widoczne na szynie. Na rysunku b nie ma żadnych elementów łącznika. W tym przypadku co najmniej jedna szyna musi mieć ślady końca łącznika. Należy zauważyć, że obecność takich śladów na złączu izolacyjnym lub spawanym nie jest wadą, ponieważ w tych złączach nie stosuje się łączników. Aby uzyskać pełny opis możliwych klas obrazów złączy szynowych, proponuje się wprowadzenie dwóch dodatkowych typów obrazów: (a) „fake_rail" — szyna bez złącza i (b) „fake" — obraz bez szyny. Przykłady takich sytuacji pokazano na rysunku.

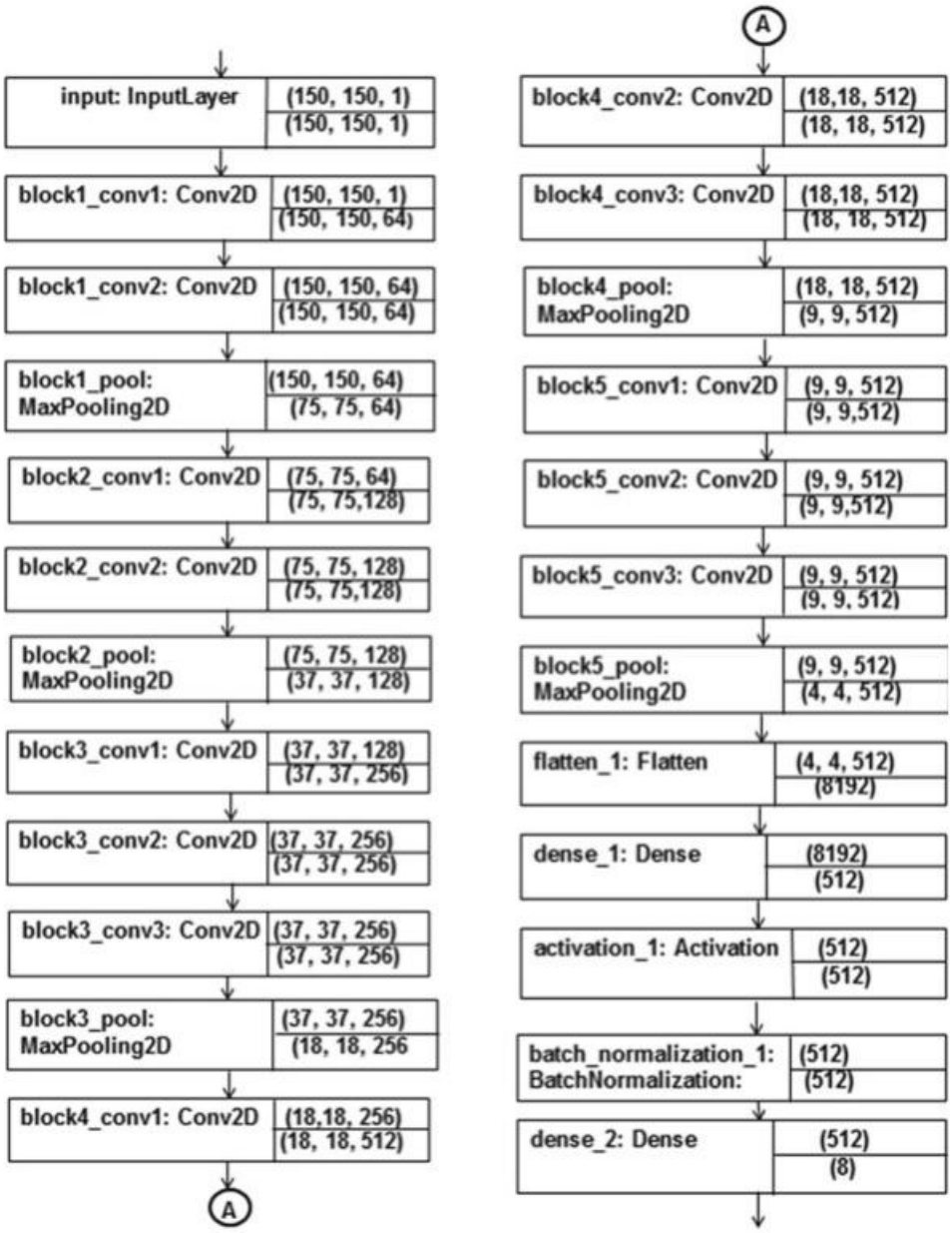
a) b)

W pierwszym przypadku odcinek drogi został zamocowany, gdzie na szynie nie było złącza (a). Takie obrazy można uzyskać, ponieważ wideo jest nagrywane w sposób ciągły, a odległości między szynami mogą się zmieniać. Druga opcja jest związana z prawdopodobieństwem uderzenia w ramkę odcinka toru kolejowego z zakrętem. W takim przypadku może nie być żadnego

połączenia na szynach lub w ramie może pojawić się kilka szyn naraz. Tu proponujemy modernizację kompleksu diagnostycznego poprzez budowę głębokiej sieci neuronowej do wykrywania uszkodzeń styków szynowych. Materiał wideo z odcinków kolejowych jest rejestrowany w specjalnych plikach i umieszczany w archiwum. Każda kamera wideo odpowiada jednemu plikowi, którego nazwa zawiera numer identyfikacyjny wagonu, kamery, odcinka przejazdu oraz datę i godzinę wykonania zdjęcia. Tworzonych jest 10 klatek czarno-białych obrazów połączeń o rozdzielczości 300 × 300 pikseli. Nazwy klatek zawierają te same informacje, co pliki. W przypadku wykrycia usterki pozwala to na dokładne określenie miejsca wystąpienia problemu z połączeniem szynowym. Klatki są dystrybuowane do katalogów z nazwami klas opisującymi stan połączenia i usterkę. Baza danych przechowuje wiele obrazów (do kilku milionów klatek).

3 Struktura splotowej sieci neuronowej VGG16 do klasyfikacji stanów połączeń szynowych

System klasyfikacji połączeń szynowych wykorzystuje obrazy w skali szarości o rozdzielczości 300 × 300 pikseli. Każdy obraz zbioru danych jest trójwymiarową tablicą, przy czym długość tablicy odpowiada wysokości, liczba elementów w drugim wymiarze szerokości, a liczba elementów w trzecim wymiarze liczbie kanałów zapisu wideo. Obraz jest zmniejszany o 2, aby uzyskać klatkę o wymiarach 150 × 150. Każdy element tablicy pikseli w klatce jest podzielny przez 255, co pozwala na zmniejszenie wartości jasności pikseli do zakresu od 0 do 1. Zaleca się użycie języka programowania Python do opracowania systemu klasyfikacji stanów połączeń szynowych. Kod programistyczny został zaprojektowany w środowisku programistycznym PyCharm. Najpopularniejszymi narzędziami do projektowania i trenowania sieci neuronowych są biblioteki Tensorflow i Keras .Tensorflow implementuje Koncepcje uczenia maszynowego i głębokiego pozwalają na wyszukiwanie i klasyfikowanie obrazów, umożliwiają precyzyjne dostrojenie dowolnych parametrów procesu trenowania sieci oraz wykorzystują istniejące charakterystyki (funkcje strat dostępne w bibliotece, optymalizatory i funkcje aktywacji). Keras to biblioteka języka Python wysokiego poziomu działająca na platformie TensorFlow. Keras pomaga rozwiązywać problemy klasyfikacji obrazów, wykorzystując wygodne narzędzia do konfigurowania warstw sieci, parametrów i charakterystyk treningowych, prezentowania struktur i wyników. Sieć oparta na wstępnie wytrenowanej sieci splotowej VGG16 została zaprojektowana do rozwiązywania problemów klasyfikacji połączeń szynowych . Sieć została wytrenowana na wielu obrazach ze strony internetowej ImageNet: zwierzęta i rośliny, domy, samochody i inne obiekty. Architekturę sieci neuronowej używanej do klasyfikacji połączeń szynowych przedstawiono na rys. 5. Model sieci zawiera warstwy: InputLayer, Conv2D (warstwa splotu), MaxPolling2D (warstwa maksymalnego łączenia), BatchNormalization, Flatten, Dense (w pełni połączona) i Activation. Strzałki na diagramie wskazują kolejność Warstwy. Każdy blok na rys. 4 zawiera informacje o nazwie i typie warstwy. Pole w prawym górnym rogu bloku zawiera wymiary tablicy wejściowej, pole w prawym dolnym rogu – wymiary tablicy wyjściowej. Warstwa wejściowa sieci neuronowej VGG16 otrzymuje tablice o wymiarach 150 × 150 × 3 pikseli. Ponieważ używane są klatki w skali szarości o wymiarach 300 × 300, wymagana jest zmiana obrazów i warstwy wejściowej sieci. Zmieniany jest wymiar warstwy wejściowej, a następnie dostosowywane są następujące wagi. Architektura sekcji splotu VGG16 pozostaje niezmieniona. Zmodyfikowane wagi są ładowane do drugiej warstwy sieci, a oryginalne wagi są ładowane do pozostałych warstw. Klasyfikator w oryginalnej sieci VGG16 został całkowicie zastąpiony klasyfikatorem naszego projektu (ostatnie cztery bloki na rysunku).



Wynika to z zastosowanej metody szkolenia, która jest opisana poniżej. Klasyfikator zawiera dwie w pełni połączone warstwy: warstwę aktywacji i warstwę normalizacji. Początkowe wagi w wstępnie wytrenowanej sekcji VGG16 sieci splotowej nie są zapisywane, ponieważ klasyfikator naszego projektu został przygotowany do nowego rozwiązywanego problemu, a jego wagi muszą zostać uzyskane poprzez ponowne trenowanie. Każdy obraz z pewnym prawdopodobieństwem należy do określonej klasy. Warstwa wyjściowa sieci neuronowej zawiera 8 neuronów, z których każdy odpowiada określonej klasie w Tabeli 1.

Class number	Class designation	Type and state of rail joint	Number of images in the class
1	conn	Fish plate rail joint with the connector	7010
2	no_conn	Fish plate rail joint without the connector	5010
3	iso	Insulating joint	5024
4	weld	Welded joint	5008
5	cut	The connector is torn off at one end	1446
6	missing	The connector is completely torn off	5024

Sygnał neuronu wyjściowego jest równy prawdopodobieństwu, że obraz należy do danej klasy po przejściu przez sieć neuronową. W ramach rozwoju głębokiej sieci neuronowej zastosowano jedną z odmian metody gradientu zstępującego, Adam (Adaptive Moment Estimation) . Metoda ta w szczególności oblicza wykładniczą średnią ruchomą gradientu i gradient kwadratowy. Reguła aktualizacji wag dla metody Adam jest określana na podstawie oszacowań gradientu i sekundowych momentów gradientów, które są określone przez wyrażenia:

$$m_w^{(t+1)} \leftarrow \beta_1 m_w^{(t)} + (1-\beta_1)\nabla_w L^{(t)}, \; v_w^{(t+1)} \leftarrow \beta_2 v_w^{(t)} + (1-\beta_2)(\nabla_w L^{(t)})^2,$$

gdzie w to parametry; $L^{(t)}$ to funkcja straty; t to indeks bieżącej iteracji, β_1 i β_2 to odpowiednio współczynniki zapomnienia dla gradientów i momenty sekundowe gradientów. Obliczone momenty są korygowane zgodnie z następującymi wzorami:

$$\hat{m}_w = m_w^{(t+1)} / \left(1-\beta_1^{(t+1)}\right), \; \hat{v}_w = v_w^{(t+1)} / \left(1-\beta_2^{(t+1)}\right)$$

Wagi przeliczane są według wzoru:

$$w^{(t+1)} \leftarrow w^{(t)} - \eta \hat{m}_w / \left(\sqrt{\hat{v}_w} + \delta\right)$$

gdzie η to rozmiar kroku lub współczynnik uczenia; δ jest małe, aby zapobiec dzieleniu przez zero.

4 Tworzenie zbioru danych treningowych i trenowanie sieci neuronowej

Postanowiono wykorzystać tryb uczenia nadzorowanego do trenowania opracowanej sieci neuronowej. W tym celu 90% obrazów powinno zostać umieszczonych w zbiorze danych treningowych, a 10% w zbiorze danych testowych. Kod programu sekwencyjnie odczytuje i przygotowuje do treningu obrazy z tablicy wejściowej. Pierwsze 90% przetworzonych obrazów należy do zbioru treningowego, a reszta do zbioru testowego. Ponieważ liczba obrazów źródłowych jest niewielka (mniej niż 10 000 dla najczęściej występującej klasy), augmentacja danych jest najlepszym sposobem na zwiększenie funkcjonalności w trenowaniu modeli uczących się. Następujące transformacje afiniczne tworzą nowe klatki:

• obrót klatki oryginalnej w pionie (obrót o 180 stopni),

• zmniejszenie i zwiększenie jasności klatki (maksymalnie o 70%),

• obrót o 2 stopnie w losowym kierunku względem środka klatki,

• powiększenie obrazu (maksymalnie o 10%).

Takie rozszerzone obrazy są generowane tylko dla próbki treningowej.

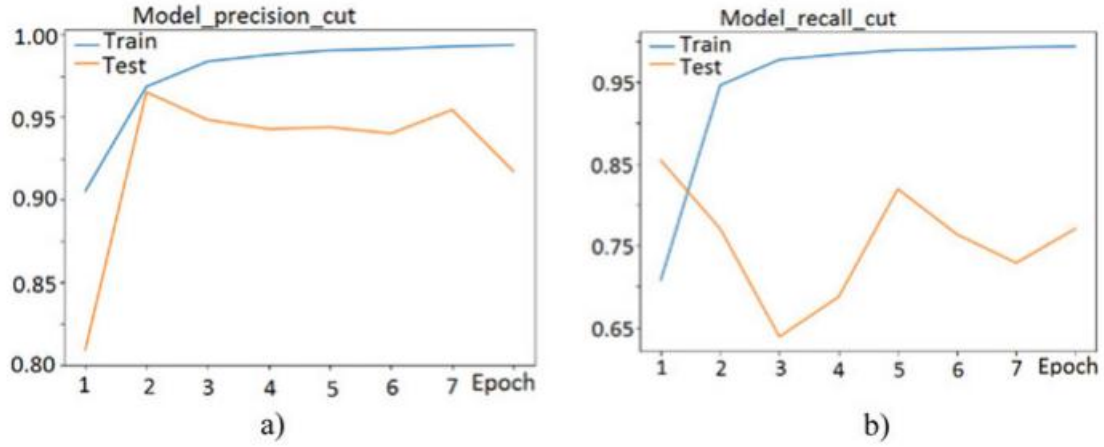
Oszacowanie strat przeprowadza się z wykorzystaniem entropii krzyżowej kategorialnej. Aby zwiększyć tempo uczenia się, wybrano wartość 10–5. Proces uczenia monitoruje zmianę dokładności klasyfikacji i stratę w stosunku do liczby kroków. Maksymalny okres uczenia sieci neuronowej został ograniczony do 100 epok. Jeśli w ciągu pięciu kroków dokładność klasyfikacji modelu poprawi się o mniej niż 0,001, trening sieci zostaje przerwany. Do treningu sieci neuronowej zastosowano metodę transferu uczenia się . Polega ona na wykorzystaniu oryginalnej sekcji splotowej wstępnie wytrenowanej sieci i rozwiniętej sekcji klasyfikatora. Wagi sekcji splotu są pobierane z VGG16 i „zamrożone" (pozostają niezmienione podczas treningu). Wówczas proces uczenia nie wpływa na sekcję splotu. Wręcz przeciwnie, wagi własnego klasyfikatora są dostosowywane podczas treningu sieci. Dokładność klasyfikatora po treningu jest oceniana za pomocą dwóch metryk: precyzji i rozpoznawania. Metryka precyzji lub dodatnia wartość predykcyjna (PPV) pokazuje ułamek odpowiednich połączeń szynowych sklasyfikowanych jako należące do danej klasy:

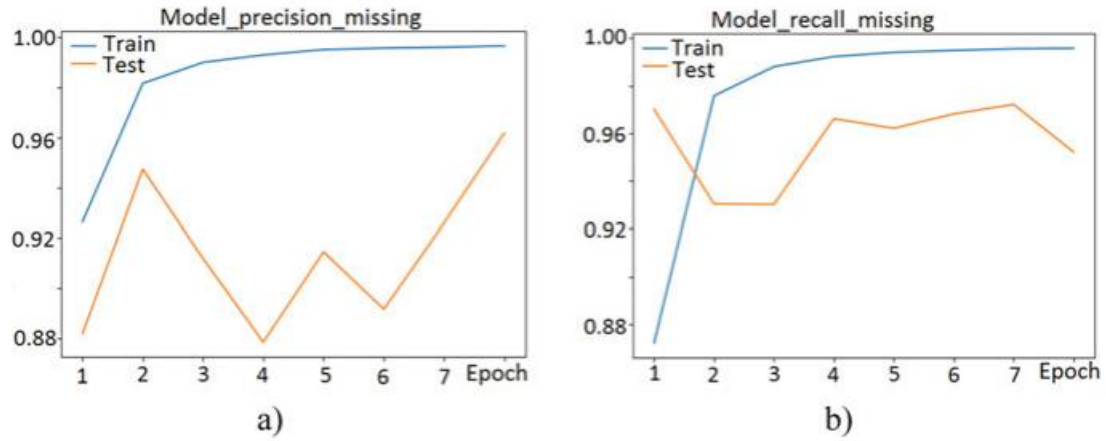
$$PPV_K = TP_K/(TP_K + FP_K),$$

gdzie K = 1, 8 to numer klasy złącza szynowego, TPK to prawdziwie pozytywne prognozy, FPK to fałszywie pozytywne, tj. wyniki testów błędnie wskazujące na przynależność złącza szynowego do klasy K. Metryka Recall lub współczynnik prawdziwie pozytywnych wyników (TPR) pokazuje zdolność klasyfikatora do wykrywania danej klasy złączy szynowych w zestawie wszystkich analizowanych obrazów:

$$TPR_K = TP_K/(TP_K + FN_K)$$

gdzie FNK jest wynikiem fałszywie negatywnym, tj. wynikami testów, które błędnie klasyfikują złącza szynowe jako nienależące do klasy K. Wyniki trenowania klasyfikatora w sieci VGG16 dla klas „przeciętych" i „brakujących" przedstawiono na rysunkach 6 i 7.





Dla rozpatrywanego zbioru danych trenowanie zakończyło się po 8 epokach. Czasami sensowne jest ocenienie obu metryk za pomocą zagregowanego kryterium jakości w postaci wyniku F1, który jest średnią harmoniczną dwóch wartości: precyzji i czułości:

$$F_1 = 2\frac{PPV \times TPR}{PPV + TPR}$$

Wyniki treningu sieci neuronowej dla wszystkich klas złączy szynowych przedstawiono w tabeli 2.

Metrics	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8
PPV	0.96	0.98	0.91	0.98	0.93	0.95	1.00	0.99
TPR	0.98	0.90	1.00	1.00	0.78	0.95	0.98	0.95
F_1 score	0.97	0.94	0.95	0.99	0.85	0.95	0.99	0.97

Największym problemem jest rozdzielenie klas „połączeń" i „przecięć". Powodem tego jest brak równowagi w zbiorze danych źródłowych. Obrazów połączeń z wadą „przecięć" jest pięć razy mniej niż ramek ze sprawnymi połączeniami szynowymi typu „połączeń". Utworzyliśmy dodatkowe obrazy „przecięć" zgodnie z powyższą metodą i otrzymaliśmy 7000 obrazów. Przeprowadzono ponowne trenowanie sieci neuronowej VGG16, uzyskując wyniki dla klasy „przecięć": PPV = 0,97; TPR = 0,93 i wynik F1 = 0,95. Tabela 3 zawiera ogólne dane dotyczące metryk dokładności opracowanego klasyfikatora opartego na głębokiej sieci neuronowej.

Metric	Accuracy	Loss	Precision	Recall
Mean	0.963	0.199	0.964	0.962

Dzięki temu opracowana sieć neuronowa może skutecznie klasyfikować stany styków szynowych z dokładnością 96%.

5 Wykorzystanie wizualizacji akcentu i klasyfikatora styków szynowych do szkolenia operatorów wagonów pomiarowych

Nowoczesne technologie rozszerzonej rzeczywistości (AR) są wykorzystywane w przemysłowych zastosowaniach systemów cyberfizycznych do budowy interaktywnych interfejsów użytkownika. Główną zaletą interfejsów AR jest ich zdolność do umieszczania obiektów wirtualnych w tej samej przestrzeni i czasie co obiekty rzeczywiste. Technologia AR umożliwia tworzenie interaktywnych i kontekstowych interfejsów użytkownika, które powinny zapewniać widzenie komputerowe i rozpoznawanie obiektów. Aby rozwiązać problem diagnostyki technicznej torów kolejowych i innych elementów infrastruktury kolejowej z wykorzystaniem sieci neuronowych, potrzebne są wysokiej jakości i reprezentatywne próbki obrazów do uczenia maszynowego. W tym celu

konieczne jest przeszkolenie operatorów wagonów pomiarowych, którzy tworzą próbki treningowe, identyfikując anomalie na uzyskanych obrazach. Wymaga to wstępnego przeszkolenia personelu na typowych przykładach. Zaproponowano utworzenie zbioru obrazów szyn z wykorzystaniem rzeczywistości rozszerzonej, umożliwienie operatorom wyszukiwanie wadliwych złączy szynowych i elementów złącznych, a następnie wykorzystanie uzyskanych danych jako zbiorów danych treningowych i testowych dla sieci neuronowej. Podczas treningu operator wózka pomiarowego może być przeciążony dużą ilością informacji, a niektóre obiekty rzeczywiste wymagające jego uwagi mogą nakładać się na obiekty wirtualne w jednej scenie. Aby rozwiązać ten problem i poprawić użyteczność oraz wydajność urządzeń rozszerzonej rzeczywistości (AR) w szkoleniu operatorów w zakresie klasyfikacji złączy szynowych, zbudowano adaptacyjny interfejs użytkownika. Identyfikacja elementów toru i ich klasyfikacja są realizowane za pomocą sztucznej sieci neuronowej.

6 Wnioski

Przedstawiono głęboką sieć neuronową VGG16, zaprojektowaną do klasyfikacji złączy szynowych i identyfikacji usterek. Sieć neuronowa działa jako część kompleksu diagnostycznego toru kolejowego, który przetwarza obrazy szyn podczas ruchu wózka pomiarowego. Zaproponowano zmodyfikowaną architekturę wstępnie wytrenowanej głębokiej sieci w celu poprawy efektywności podejmowania decyzji, zwiększając dokładność klasyfikacji złączy szynowych. Opracowana metoda uczenia sieci neuronowej opiera się na równoważeniu zbiorów danych oraz metodzie transferu uczenia. Badania eksperymentalne wykazały możliwość osiągnięcia dokładności klasyfikacji na poziomie co najmniej 96%. Dalszy rozwój inteligentnego systemu diagnostycznego ma być realizowany z wykorzystaniem architektur głębokich sieci splotowych o dużej liczbie warstw. Przykładowo, wskazane jest zbudowanie architektury RetinaNet ze szkieletem zbudowanym na sieci neuronowej, takiej jak ResNet lub EfficientNet.

Informacyjne wsparcie utrzymania ruchu i napraw systemu cyberfizycznego o profilu budowy maszyn

Zagadnienia zarządzania utrzymaniem ruchu i naprawą systemów cyberfizycznych o profilu budowy maszyn rozpatrywane są na przykładzie maszyn sterowanych numerycznie (CNC) w oparciu o teorię budowy systemów eksperckich. Utrzymanie ruchu maszyn CNC obejmuje zespół działań organizacyjnych i technicznych zapewniających utrzymanie ich parametrów wyjściowych na zadanym poziomie przez cały okres eksploatacji; przegląd i kontrolę stanu technicznego maszyny i układu sterowania: czyszczenie, wymianę chłodziwa, mycie i smarowanie mechanizmów maszyny i elementów układu sterowania, wymianę zużytych części, podzespołów, uszkodzonych elementów i jednostek sterujących maszyny; kontrolę napędów hydraulicznych i elektrycznych. Autorzy postawili i rozwiązali problem optymalnego doboru działań podczas utrzymania ruchu i napraw złożonego wyposażenia maszyn CNC. Zaproponowano informacyjno-logiczny model wspomagania decyzji podczas jego realizacji, który pozwala, w zależności od aktualnego stanu podzespołów maszyny i warunków technicznych wytwarzania części, znaleźć optymalną opcję utrzymania ruchu i regulacji centrów obróbczych CNC. Reguły produkcyjne zawarte w modelu są skonstruowane zgodnie z typem: jeżeli... (warunki są spełnione), to... (realizacja konsekwencji). Autorzy opracowali oprogramowanie implementujące rozwiązanie zadania, które zostało przetestowane podczas konserwacji i naprawy pionowego centrum obróbczego frezarskiego z CNCChallenger MCV-2418.

Słowa kluczowe: System cyberfizyczny · Podejmowanie decyzji · Model informacyjno-logiczny · Konserwacja i naprawa · Pionowe centrum obróbcze frezarskie z CNC Challenger MCV-2418

1 Wstęp

Wraz z rozwojem nauki i techniki, technologia wytwarzania produktów na urządzeniach cyberfizycznych o profilu budowy maszyn, w tym maszyn sterowanych numerycznie (CNC), stała się jedną z niezastąpionych i ważnych technologii w przemyśle wytwórczym. Maszyny CNC są elementami nowoczesnego sprzętu produkcyjnego, a ich niezawodność bezpośrednio wpływa na jakość produktów. Utrzymanie tego typu urządzeń odgrywa strategiczną rolę w funkcjonowaniu przedsiębiorstw budowy maszyn i może stanowić znaczną część kosztów cyklu życia produktu. Wdrożenie efektywnego systemu zarządzania utrzymaniem ruchu pozwala zrozumieć jego znaczenie w skali globalnej poprzez integrację wiedzy i informacji uzyskanych od różnych uczestników (np. operatorów, zespołów utrzymania ruchu, producentów maszyn CNC) i źródeł (np. automatycznych alarmów, raportów grup utrzymania ruchu, kierownictwa). Autorzy wielu prac rozważali różne podejścia do tworzenia zautomatyzowanych systemów cyberfizycznych dla złożonych systemów utrzymania ruchu urządzeń. W związku z tym autorzy zaproponowali podejście wspomagające podejmowanie decyzji w zakresie zautomatyzowanego planowania utrzymania ruchu urządzeń CNC w ośrodkach badawczych. Podstawowa struktura systemu zarządzania utrzymaniem ruchu opiera się na: klasyfikacji prac konserwacyjnych; Podział tych prac pomiędzy różne działy serwisowe zgodnie z wymaganą wiedzą i bezpieczeństwem; planowanie interwencji zapobiegawczych; definiowanie opcji checklist i dostosowanych reakcji w celu konsolidacji prac konserwacyjnych; szkolenie zespołu utrzymania ruchu. Główne rezultaty to: śledzenie działań konserwacyjnych w celu zapewnienia, że zadania konserwacyjne są wykonywane zgodnie z wymaganiami producentów maszyn CNC, a także angażowanie operatorów maszyn w konserwację ich sprzętu w celu poprawy wydajności i niezawodności aktywów. W rzeczywistości podejście to zapewnia metodologię dla tych firm, które po raz pierwszy rozważają możliwość usprawnienia planowania konserwacji, umożliwiając użytkownikowi końcowemu otrzymanie podstawowych usług konserwacyjnych sprzętu, a także rozpoczęcie gromadzenia danych konserwacyjnych. Jak zauważają autorzy , nierozsądna strategia konserwacji zwiększa koszty jej wdrożenia i może znacznie obniżyć wydajność maszyn CNC. Dlatego, aby uzyskać naukową i rozsądną strategię konserwacji na poziomie systemu, konieczne jest uwzględnienie nie tylko stanu zużycia jednostek sprzętu, ale także ich wzajemnego wpływu na inne jednostki maszyn CNC. Zużycie sprzętu wpływa na jakość wytwarzanych produktów, ponieważ liczba usterek rośnie wraz ze wzrostem zużycia . Autorzy proponują strategię utrzymania ruchu grupowego maszyn CNC, uwzględniając zależności ekonomiczne, strukturalne i stochastyczne między jednostkami krytycznymi, oraz optymalizują strategię utrzymania ruchu grupowego. Model utrzymania ruchu grupowego maszyn CNC składa się z czterech podmodeli: podmodeli zużycia jednostek, podmodeli decyzji utrzymania ruchu grupowego, podmodeli procesu utrzymania ruchu oraz podmodeli kosztów utrzymania ruchu. Wykorzystując model utrzymania ruchu grupowego maszyn CNC, można określić czas i cel jego wdrożenia w zależności od stopnia zużycia jednostek. Następnie można obliczyć koszty poszczególnych usług technicznych. W modelu utrzymania ruchu grupowego zależność ekonomiczna i strukturalna między komponentami są określane przez koszty, natomiast zależność stochastyczna jest określana przez wskaźnik awaryjności. W tym przypadku metoda Monte Carlo jest wykorzystywana do symulacji procesu eksploatacji maszyny, a długoterminowe koszty utrzymania ruchu maszyn CNC są obliczane według określonego progu wskaźnika awaryjności.

Na koniec, algorytm genetyczny jest wykorzystywany do optymalizacji wartości progowych wskaźnika awaryjności urządzeń podczas konserwacji zapobiegawczej i konserwacji grupowej. Ze względu na brak danych monitorujących i niejasne zrozumienie procesu zużycia podzespołów maszyn CNC zaproponowano metodę oceny stanu i podejmowania decyzji dotyczących konserwacji maszyn CNC o kilku stanach, opartą na częściowo obserwowalnych łańcuchach procesów decyzyjnych Markowa. Z jednej strony, biorąc pod uwagę niedoskonałą informację operacyjną o maszynach CNC oraz wpływ różnych działań konserwacyjnych na proces zużycia maszyn, zbudowano dynamiczny model oceny ich stanu roboczego. Z drugiej strony opracowano model decyzyjny dotyczący konserwacji maszyn CNC w celu minimalizacji całkowitych kosztów strat (w tym kosztów konserwacji i potencjalnych kosztów awarii). Podejmowanie decyzji dotyczących utrzymania ruchu w dużych systemach produkcyjnych jest złożonym procesem, ponieważ wymaga integracji różnych informacji . W praktyce popularna jest polityka ograniczonej kontroli. Utrzymanie ruchu jest przeprowadzane, gdy stan zużycia maszyny CNC osiągnie wartość progową. Autorzy, opracowując strukturę opartą na modelach łańcuchów Markowa z czasem dyskretnym, ocenili wydajność systemu w ramach polityki ograniczonej kontroli w systemach produkcyjnych składających się z maszyn CNC o kilku stanach. W zaawansowanych systemach produkcyjnych maszyny CNC stanowią ważne wyposażenie do wytwarzania komponentów produktów o wysokiej precyzji, natomiast z punktu widzenia utrzymania ruchu są one traktowane jako „produkty" dostarczane przez producentów maszyn CNC. Przedstawiono nowy system planowania utrzymania ruchu oparty na wiedzy, który ma ułatwić wymianę informacji i wiedzy między wszystkimi interesariuszami, w tym producentami obrabiarek, użytkownikami (systemy produkcyjne), dostawcami usług utrzymania ruchu i dostawcami części (do obrabiarek), w ramach rozwijającego się modelu biznesowego „Produkt-Usługa". Aby poprawić efektywność planowania utrzymania ruchu, wdrożono zasady uzasadnienia oparte na analizie przypadków. Z punktu widzenia cyklu życia produktu, drogie i złożone maszyny CNC są zarówno urządzeniami, jak i produktami, których niezawodność wpływa nie tylko na jakość wytwarzanych części, ale także na zyski dostawców urządzeń. Efektywna konserwacja i serwis odgrywają ważną rolę w poprawie niezawodności urządzeń i wydajności produkcji. Ponowne wykorzystanie wiedzy historycznej to skuteczny sposób na rozwiązywanie nowych problemów i poprawę efektywności konserwacji. Autorzy przyjęli dwupoziomową ontologię opartą na rozumowaniu i przykładach z życia wziętych, aby przedstawić wiedzę ekspercką za pomocą wyszukiwania zorientowanego na adaptację. Na podstawie podobieństwa semantycznego i korelacji wybrano najbardziej adaptacyjny przypadek historyczny do rozwiązania obecnego nowego problemu. Analiza literatury wykazała, że obecnie do przeprowadzania złożonej konserwacji technicznej i napraw urządzeń CNC stosuje się kilka podejść do automatyzacji procesów decyzyjnych, w zależności od specyficznych warunków produkcji wyrobów maszynowych. W związku z tym autorzy niniejszego rozdziału proponują podejście do zarządzania konserwacją maszyn CNC oparte na teorii systemów eksperckich , które ilustruje przykład pionowych centrów obróbczych frezarskich. Rodzaj urządzenia służy do wykonywania następujących operacji:

- obróbka powierzchni frezem promieniowym;
- frezowanie rowków, występów, gwintowanie itp.;
- wiercenie otworów przelotowych i nieprzelotowych;
- rozstawianie;
- obróbka frezami wiertarskimi o zadanym rozmiarze.

Należy zauważyć, że niezależnie od doskonałości wyposażenia obrabiarki, operacje jego regulacji i kontroli parametrów technologicznych będą zawsze obecne. Wynika to ze zużycia zespołów roboczych i różnic w technologiach wytwarzania części. Dlatego specjaliści ds. utrzymania ruchu potrzebują dogłębnej wiedzy nie tylko z zakresu konstrukcji, teorii procesów produkcyjnych w tokarkach, ale także umiejętności ich regulacji w zależności od właściwości i stanu obrabianego materiału, stanu technicznego, zużycia części, zespołów i mechanizmów.

2 Materiały i metody

2.1 Systemy utrzymania ruchu pionowych centrów obróbkowych

Utrzymanie ruchu maszyn CNC obejmuje zespół działań organizacyjnych i technicznych zapewniających utrzymanie parametrów wyjściowych na zadanym poziomie przez cały okres eksploatacji; przegląd i kontrolę stanu technicznego maszyny i układu sterowania: czyszczenie, wymianę chłodziwa, mycie i smarowanie mechanizmów maszyny i elementów układu sterowania, wymianę zużytych części, podzespołów mechanizmów oraz niesprawnych elementów i układów sterowania maszyny; przegląd napędów hydraulicznych i elektrycznych. Każde wyposażenie obrabiarki charakteryzuje się pogorszeniem parametrów technicznych podczas eksploatacji, co wyraża się w naturalnej zmianie geometrii. Oznacza to, że części maszyny, poddane oddziaływaniom mechanicznym i erozyjnym, z czasem zmieniają swoje wymiary. W rezultacie dochodzi do naruszenia równoległości konstrukcji, co oczywiście wpływa na sztywność całej maszyny i prowadzi do jej awarii.

Przyczyny wypadków i awarii pionowych centrów obróbczych CNC:

1. Wady konstrukcyjne:

a. wady konstrukcyjne zespołów i ich mocowania; niewystarczająca sztywność i odporność na wibracje, wady układu smarowania; brak urządzeń zabezpieczających itp.

b. wady konstrukcyjne części (nieprawidłowe przypisanie tolerancji, metod obróbki cieplnej, wymiarów i kształtu geometrycznego itp.); Wady produkcyjne:

2. Wady produkcyjne:

a. wady metalowe (wady jakościowe materiału, odlewu i detali);

b. wady montażowe maszyny (słaba jakość dopasowania współpracujących części; słaba jakość okablowania układów pneumatycznego i hydraulicznego maszyny; wycieki olejów smarowych; obecność ciał obcych i wiórów);

3. Niewłaściwa konserwacja maszyny:

a. konserwacja sprzętu w stanie brudnym (wnikanie wilgoci, pyłu i wiórów między trące części maszyny);

b. Brak lub słaba jakość smarowania (brak oleju w zbiornikach maszyny; awaria pompy lub jej napędu; zatkany przewód olejowy; użycie zanieczyszczonego oleju lub oleju o niewłaściwej lepkości);

4. Nieprawidłowa regulacja:

a. słaba jakość regulacji i regulacji (stosowanie posuwów, głębokości skrawania i prędkości powyżej obliczonych;

b. stosowanie stępionych narzędzi skrawających).

Do wykonania wszystkich operacji frezowania na maszynach CNC niezbędne jest ich staranne przygotowanie i regulacja. Wysoka jakość wykonania tych prac w optymalnym czasie ma pozytywny wpływ na gotowy element.

2.2 Zadanie optymalnego doboru środków podczas konserwacji i napraw maszyn CNC

W sformalizowanej formie sformułowanie problemu optymalnego wyboru środków konserwacji i napraw maszyn CNC, w zależności od ich aktualnego stanu, przedstawia się następująco: należy znaleźć taką listę środków, aby w warunkach operatora:

$$Tx_i^{\min} \le Tx_i \le Tx_i^{\max}, \quad i = 1, 2, \ldots N, \qquad (1)$$

który stanowi model matematyczny wspomagania decyzji podczas konserwacji i napraw maszyn CNC, obejmujący informacje o ich częściach i zespołach, aktualnym stanie, metodach rozwiązywania problemów, charakterystyce procesów konserwacji i napraw, spełnione są następujące postanowienia:

$$k_{opt} = \arg\min Q(k), \quad k \in K. \qquad (2)$$

Kryterium optymalności Q przedstawia się jako sumę względnych strat kryteriów przyjętych z pewnymi wartościami ich ważności: szacowanych kosztów realizacji procesów utrzymania i napraw maszyn CNC, technologiczności i bezpieczeństwa procesów:

$$Q(k) = \sum_{j=1}^{3} (\rho_j \ \varpi_j(k)), \qquad (3)$$

gdzie: ρ_1, ρ_2, ρ_3 — współczynniki wagowe ustalone przez ekspertów.

$$\rho = \left\{ \rho_j : \rho_j \ge 0, \quad j = 1, 2, 3; \ \sum_{j=1}^{3} \rho_j = 1 \right\}. \qquad (4)$$

Tutaj: Tx_i, $Tx^{\min}{}_i$, $Tx^{\max}{}_i$ odpowiednio są wartościami bieżącymi i granicznymi i-tej charakterystyki maszyny CNC, N — liczba charakterystyk; $\rho_j \varpi_j(k)$ — ważone straty dla każdego j-tego kryterium; $\rho_j \varpi_j(k) = \varpi_j(F_j(k))$, j = 1, ...3, k ∈ K — zależności funkcyjne, które przekształcają odpowiednie funkcje $F_j(k)$ do postaci bezwymiarowej; $F_1(k)$ — szacowane koszty wdrożenia procesów konserwacji i napraw maszyn CNC; $F_2(k)$, $F_3(k)$ — odpowiednio wartości produkcyjności i bezpieczeństwa procesów ich wdrożenia. Dla funkcji $F_1(k)$ będziemy szukać minimum dla funkcji $F_2(k)$, $F_3(k)$ — maksimum. Zależności funkcyjne, które przekształcają funkcje $F_j(k)$ do postaci bezwymiarowej to:

$$\varpi_1(k) = \frac{F_1(k) - F_1^0}{F_1^{\max} - F_1^0}, \quad \varpi_l(k) = \frac{F_l^0(k) - F_l(k)}{F_l^0 - F_l^{\min}}, \ l = 2, 3. \qquad (5)$$

Tutaj: F^{max}_1 — wartość maksymalna funkcji $F_1(k)$, dla której szukane jest minimum; F^{min}_2 , F^{min}_3 — wartość minimalna funkcji $F_2(k)$, $F_3(k)$, dla której szukane jest maksimum; F^0_1 , F^02 , F^0_3 — wartości optymalne odpowiednich $F_j(k)$, j = 1, ... 3. Przy rozwiązywaniu zadania (1–5) wartości współczynników ustala się na ρ j , j = 1, ... 3, co spełnia zależność (4) i odzwierciedla względną ważność kryteriów Fj(k). Dla każdej k-tej wygenerowanej listy czynności szacunkowy koszt $F_1(k)$ stanowi sumę kosztów poszczególnych zdarzeń, $F_2(k)$ i $F_3(k)$ obliczanych odpowiednio jako wartość możliwości produkcyjnej pomnożona przez wartość bezpieczeństwa procesów poszczególnych zdarzeń zawartych w k-tej wersji wygenerowanej listy.

3 Wyniki

Aby znaleźć optymalny wariant listy działań konserwacyjnych i naprawczych maszyn CNC, spełniający wymagania prawne, opracowano model informacyjno-logiczny wspomagający podejmowanie decyzji, który zawiera wiele reguł produkcyjnych . Jako przykład podano szereg reguł, za pomocą których można utworzyć listę procedur konserwacji i napraw maszyn CNC. Reguły produkcyjne zawarte w modelu są konstruowane zgodnie z typem: jeśli... (warunki są spełnione), to... (realizacja konsekwencji). Obecnie baza danych zawiera ponad 250 reguł, których konkretny typ rozpatrzono w poniższych przykładach dotyczących pionowych centrów obróbczych z CNC, w szczególności maszyny Challenger MCV-2418 .

Reguły są gromadzone przez ekspertów i autorów podczas kontaktów z operatorami-regulatorami pionowych centrów obróbczych z CNC.

1. Reguły doboru działań eliminujących wady podczas pracy na pionowym centrum obróbczym z CNCChallenger MCV-2418:

a. Jeśli „wada przy rozwiercaniu otworów” = „otwór stał się owalny”, to „bicie trzpienia” = „wymiana trzpienia”;

b. Jeśli „wada przy rozwiercaniu otworów” = „otwór stał się owalny”, to „nieprawidłowe tryby skrawania” = „korekta pras tnących”;

c. Jeśli „wada przy rozwiercaniu otworów” = „otwór stał się owalny”, to „bicie wrzeciona” = „wymiana łożysk i szyjek”;

d. Jeśli „wada przy rozwiercaniu otworów” = „obrobiona powierzchnia okazała się nierówna”, to „frez został wyszczerbiony” = „sprawdź frez, w razie potrzeby przetnij ponownie, wymień na nowy”;

e. Jeśli „wada przy rozwiercaniu otworów” = „obrobiona powierzchnia okazała się nierówna”, to „duży posuw frezu” = „sprawdź poprawność doboru posuwu, przeprowadź obróbkę wykańczającą”;

f. Jeśli „wada podczas cięcia końców i występów” = „nieprostopadłe ustawienie końca lub występu do osi elementu”, to „nieprawidłowy montaż frezu” = „sprawdź poprawność montażu frezu”;

g. Jeśli „wada frezowania” = „frezowana powierzchnia ma metalowy pływak”, to „frez jest stępiony” = „naostrz frez”;

h. Jeśli „wada frezowania” = „frezowana powierzchnia ma metalowy pływak”, to „frez jest stępiony” = „wymień frez”;

i. jeśli „wada frezowania” = „frezowana powierzchnia ma nieprawidłowe wymiary”, to „nieprawidłowe powiązanie zer w układzie współrzędnych” = „sprawdź powiązanie, w razie potrzeby połącz ponownie”;

j. jeśli „wada frezowania” = „frezowana powierzchnia ma nieprawidłowe wymiary”,

to „nieprawidłowe powiązanie narzędzia” = „sprawdź powiązanie, w razie potrzeby połącz ponownie”;

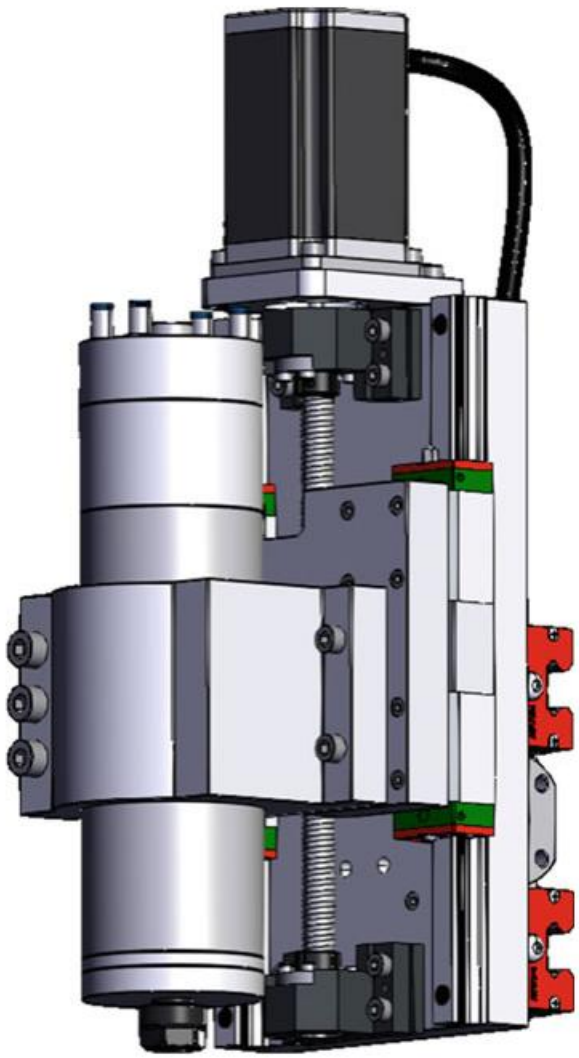
k. jeśli „wada frezowania” = „frezowana powierzchnia ma nieprawidłowe wymiary”, to „nieprawidłowe powiązanie narzędzia” = „sprawdź powiązanie narzędzia”;

l. jeśli „wada frezowania” = „frezowana powierzchnia ma nieprawidłowe wymiary”, to „nieprawidłowa wartość korektora narzędzia” = „sprawdź korektor narzędzia”;

m. jeśli „wada frezowania” = „frezowana powierzchnia ma nieprawidłowe wymiary”, to „sprawdź „podstawę” pod kątem obecności wiórów” = „wydmuchaj „podstawę” z wiórów”.

2. Zasady doboru środków zaradczych w przypadku awarii pionowego centrum obróbczego Challenger MCV-2418:

a. W przypadku awarii wrzeciona :



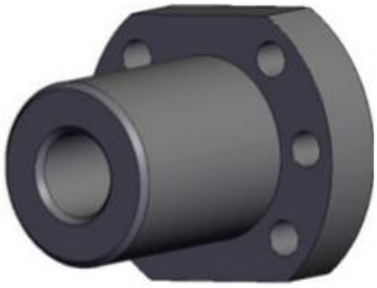
(1) jeśli „awaria wrzeciona” = „zwiększone bicie promieniowe”, to krzywizna wału roboczego” = „korekta wału wrzeciona”;

(2) jeśli „awaria wrzeciona” = „zwiększone bicie promieniowe”, to „niedostateczne dokręcenie łożysk, luzy” = „regulacja łożysk, dokręcenie nakrętki mocującej łożysko”;

(3) jeśli „awaria wrzeciona” = „hałas w górnym łożysku”, to „diagnostyka łożysk” = „wyczyszczenie łożyska z zanieczyszczeń poprzez wymianę smaru lub łożyska”.

b. w przypadku awarii pary śrub kulowych (śruba kulowa) :

a)



b)

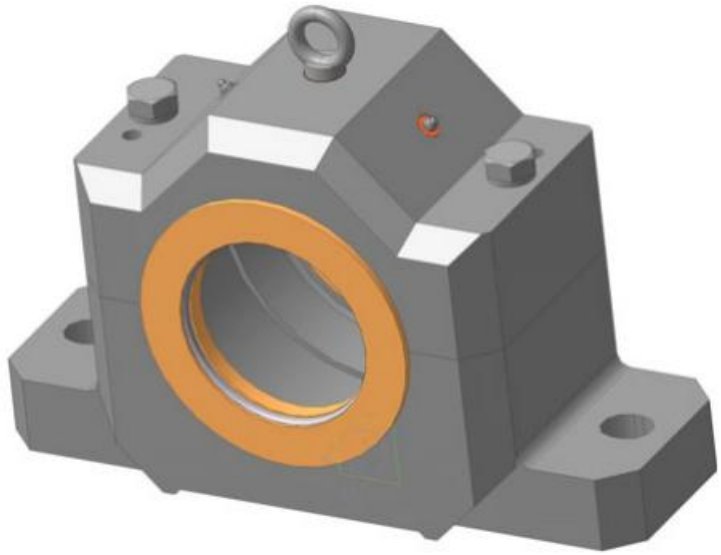
(1) jeśli „awaria śruby” = „nieznaczne pogorszenie układu pozycjonowania”, to „zużycie śruby” = „kompensacja osi maszyny”;

(2) jeśli „awaria śruby kulowej” = „poważne pogorszenie układu pozycjonowania”, to „zużycie śruby kulowej” = „wymiana śruby kulowej”;

(3) jeśli „usterka nakrętki śruby” = „nieznaczne pogorszenie stanu układu pozycjonowania”, to „zużycie łożyska kulkowego” = „kompensacja osi maszyny”;

(4) jeśli „usterka nakrętki śruby” = „poważne pogorszenie stanu układu pozycjonowania”, to „diagnostyka łożysk kulkowych” = „czyszczenie łożysk kulkowych z wymianą środka smarującego lub wymiana łożysk kulkowych”;

c. w przypadku usterek uniwersalnego podparcia mocującego :



(1) jeśli „usterka podparcia” = „pogorszenie stanu układu pozycjonowania”, to „zużycie łożyska” = „wymiana łożyska”;

(2) jeśli „usterka podparcia” = „zwiększone bicie promieniowe”, to „zużycie łożyska” = „wymiana łożyska”;

(3) jeśli „usterka podparcia” = „niewspółosiowość osi wałów”, to „zużycie łożyska” = „wymiana łożyska”.

Informacje o charakterystyce każdego zdarzenia podano w tabeli .

Rule number	Manufacturability (scores from 1 to 10)	Estimated cost of the event (RUB)	Replacement safety (scores from 1 to 10, where 10 is complete safety, 1 is a serious injury to a worker)
…	…	…	.
24	8	32,000	8
25	8	15,000	7
26	7.6	30,000	8
27	10	11,000	9.8
28	4	55,000	6
29	10	11,000	9.8
30	4	24,000	7
31	5	27,000	7
32	5	27,000	7
…	…	…	…

4. Dyskusja

W wyniku wdrożenia modelu wspomagania decyzji w zakresie wyboru optymalnych opcji konserwacji i napraw maszyn CNC możliwe jest sformułowanie szeregu miar o różnej skuteczności. Obecnie rozmiar generowanego zbioru zdarzeń nie przekracza tysiąca, dlatego poszukiwanie optymalnej opcji odbywa się metodą przeszukiwania kompletnego. Autorzy opracowali pakiet oprogramowania implementujący rozwiązanie problemów (1–5) i przetestowali go podczas konserwacji i naprawy pionowego centrum obróbczego z frezarką Challenger MCV-2418 CNC.

5. Wnioski

W wyniku przeprowadzonych badań opracowano podejście do zarządzania utrzymaniem ruchu i naprawami maszyn sterowanych numerycznie (CNC), oparte na wykorzystaniu teorii budowy systemów ekspertowych. Autorzy postawili i rozwiązali problem optymalnego doboru środków podczas złożonych prac konserwacyjnych i naprawczych urządzeń CNC. Autorzy przeanalizowali zagadnienia budowy informacyjno-logicznego modelu wspomagania decyzji w trakcie jego realizacji, który pozwala, w zależności od aktualnego stanu podzespołów maszyny i warunków technicznych wytwarzania części, znaleźć najlepszą opcję utrzymania ruchu i uruchomienia pionowych centrów obróbczych frezarskich z CNC. Zatwierdzenie opracowanego pakietu oprogramowania realizującego rozwiązanie problemu (1–5) podczas utrzymania ruchu i napraw systemów cyberfizycznych na przykładzie pionowego centrum obróbczego frezarskiego z CNC Challenger MCV-2418 wykazało wysoką skuteczność proponowanego podejścia.

Model dynamiki systemu do zarządzania produkcją szkła float

Na podstawie analizy mapy celów strategicznych wieloetapowej produkcji wysokiej jakości szkła float określono strukturę akumulatorów modelu dynamiki systemu. Opracowano układ równań różniczkowych zwyczajnych dynamiki systemu wieloetapowej produkcji wysokiej jakości szkła płaskiego. Opracowano strukturę modelu dynamiki systemu produkcji szkła płaskiego. Wyznaczono równania łączące akumulatory i wartości stałych. Przedstawiono model dynamiki systemu funkcjonowania produkcji szkła płaskiego w środowisku AnyLogic. Przeprowadzono eksperyment obliczeniowy w celu symulacji funkcjonowania produkcji szkła float. W środowisku

symulacyjnym AnyLogic wykorzystano metodę Eulera do przeprowadzenia eksperymentów z modelem dynamiki systemu wieloetapowej produkcji wysokiej jakości szkła float w celu rozwiązania układu równań różniczkowych. Symulację przeprowadzono w okresie 10 lat i 3650 dni. Wykresy wzrostu współczynników wad szkła w zależności od sezonowych składowych spadku jakości surowców i zużycia głównych urządzeń, dynamiki napełnienia magazynu szkła standardowego i magazynu szkła Jumbo w tonach, z uwzględnieniem początkowych pozostałości i sezonowości popytu na te rodzaje szkła, kosztów utrzymania urządzeń w rublach, kosztów gazu, energii elektrycznej i surowców.

Słowa kluczowe: modelowanie symulacyjne · mapa celów strategicznych · dynamika systemu · AnyLogic · produkcja szkła float

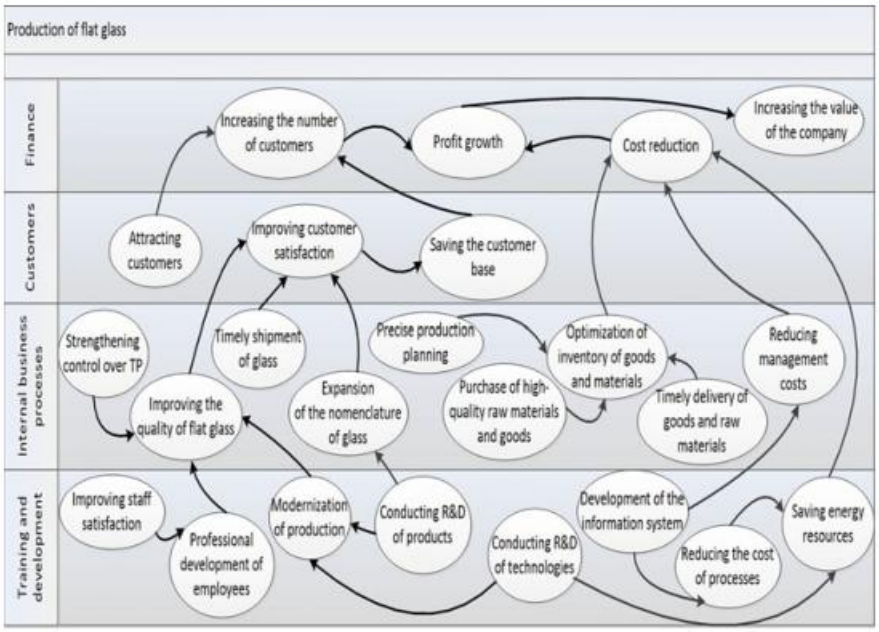
1 Wstęp

Współczesny poziom automatyzacji ciągłych procesów technologicznych pozwala mówić o istnieniu wystarczająco efektywnych systemów cyberfizycznych przedsiębiorstw przemysłowych o ciągłym charakterze produkcji . Do zarządzania produkcją w takich przedsiębiorstwach stosuje się systemy zarządzania jakością (SZJ) zgodne z normami serii ISO 9000. Zautomatyzowane SZJ oparte na ARIS, Business Studio, ELMA itp. zapewniają monitorowanie i prognozowanie wydajności operacyjnej za pomocą metod statystycznych, co pozwala na realizację mapy celów strategicznych (SGM) przedsiębiorstwa. Systemy te należą do systemów BPM (Business Process Management) . Aby zrealizować koncepcję Przemysłu 4.0 , przedsiębiorstwa muszą rozwijać cyfrowe bliźniaki, oprócz systemów BPM, które umożliwiają symulację procesów produkcyjnych w czasie rzeczywistym i podejmowanie świadomych decyzji dotyczących zarządzania produkcją, aby sprostać indywidualnym wymaganiom klientów . Tradycyjnie optymalną równowagę między sprzecznościami „Szybko-Tanie-Jakość" w postaci formułowania zrównoważonych wskaźników określają eksperci. W złożonych systemach zastosowanie modelowania matematycznego może znacznie poprawić jakość decyzji podejmowanych przez ekspertów. Wieloetapowa produkcja szkła float metodą float obejmuje szereg złożonych i wzajemnie powiązanych reakcji chemicznego utleniania-redukcji oraz fizycznych procesów wymiany ciepła. Składa się ona z następujących etapów: przygotowania wsadu, topienia szkła, formowania wstęgi szklanej na stopionej cynie, wyżarzania wstęgi szklanej, kontroli defektów, cięcia wstęgi szklanej na tafle o zadanym rozmiarze oraz pakowania powstałych tafli. Na jakość szkła float istotny wpływ ma proces technologiczny przygotowania mieszanki wsadowej, w którym konieczne jest kontrolowanie procesu formowania mieszanki proszków o różnych właściwościach . Początkowy i końcowy etap produkcji szkła float: przygotowanie wsadu, cięcie wstęgi szklanej na tafle i ich pakowanie należą do dyskretnych procesów technologicznych (TP). Pozostałe etapy realizowane są w ciągłych TP. Zatem wieloetapowa produkcja szkła płaskiego ma charakter dyskretno-ciągły. Analizę takiej produkcji przeprowadzono zgodnie z etapami analizy systemowej . Do analizy funkcjonowania tak złożonego układu nieliniowego niezbędne jest modelowanie matematyczne . Zastosowanie metody symulacji „Dynamika Systemu" daje całościowy obraz funkcjonowania obiektu sterowania, koncentrując się na trendach globalnych, a nie na pojedynczych zdarzeniach. Model dynamiki systemu jest definiowany przez układ równań różniczkowych opisujących interakcję jednostek pamięci masowej, przepływy między jednostkami pamięci masowej oraz zmienne dynamiczne. Środowisko modelowania symulacyjnego AnyLogic, oparte na podstawowych typach modeli: „dynamika systemu", „zdarzenie dyskretne" i „agent", pozwala budować hierarchiczne modele złożonych systemów. Model dynamiki systemu wieloetapowej produkcji wysokiej jakości szkła float w środowisku AnyLogic zapewnia możliwość optymalizacji kosztów działań i określenia wartości parametrów modelu w celu minimalizacji lub

maksymalizacji wartości docelowej funkcjonalności. Celem pracy jest modelowanie matematyczne procesu produkcji wielonumerycznego szkła float z wykorzystaniem metody symulacji „Dynamika systemowa" oraz eksperymentu obliczeniowego w celu zbadania wpływu na koszt produkcji szkła float następujących parametrów: jakości surowców, stanu głównych urządzeń technologicznych, wielkości magazynów szkła o rozmiarach standardowych i „Jumbo".

2. Opracowanie modelu matematycznego

Analiza funkcjonowania przedsiębiorstw zajmujących się produkcją szkła float pozwala określić główne cele zarządzania zautomatyzowaną, wieloetapową produkcją wysokiej jakości szkła float. Powiązania między tymi celami przedstawiono na SGM,



opracowanym w Business Studio. Na podstawie analizy celów perspektywy „Wewnętrzne Procesy Biznesowe" wieloetapowej produkcji wysokiej jakości szkła float w SGM, zdefiniowano zapasy modelu dynamiki systemu: S_1 – „Finanse przedsiębiorstwa", S_2 – „Produkcja wsadu", S_3 – „Wytop szkła", S_4 – „Szkło formowane", S_5 – „Magazyn szkła standardowego", S_6 – „Magazyn szkła „Jumbo", S_7 – „Stan normalny urządzeń", S_8 – „Stan awaryjny urządzeń". Wartości początkowe, jednostki i oznaczenia zapasów w modelu dynamiki systemu podano w tabeli.

No	Stock (S)	Designations in model	Initial values	Units
1	Enterprisefinances	Finance	10,000,000	rubles
2	Charge production	Charge	700	ton
3	Glass melting	Glass_furnace	700	ton
4	Molded glass	Glass_	700	ton
5	Standard glass warehouse	Glass_warehouse_st	40,000	ton
6	"Jumbo" glass warehouse	Glass_warehouse_jumbo	10,000	ton
7	Normal state of equipment	Normal_state	100	%
8	Emergency state of equipment	Emergency_state	0	%

Zmianę stanów zapasów w dynamice układu opisują równania różniczkowe zwyczajne. Podstawowy układ równań dynamiki układu wieloetapowej produkcji wysokiej jakości szkła float ma postać:

$$\begin{cases} dS_1/dt = r_1 - r_2 \\ dS_2/dt = r_3 + r_7 - r_4 \\ dS_3/dt = r_4 - r_5 - r_6 \\ dS_4/dt = r_6 - r_7 - r_8 - r_9 \\ dS_5/dt = r_8 - r_{10} \\ dS_6/dt = r_9 - r_{11} \\ dS_7/dt = r_{13} - r_{12} \\ dS_8/dt = r_{12} - r_{13} \end{cases}, \quad (1)$$

gdzie r_1 — przepływ „Dochód", r_2 — przepływ „Zużycie", r_3 — przepływ „Objętość dostaw surowców", r_4 — przepływ „Załadunek wsadu", r_5 — przepływ „Odpady surowców", r_6 — przepływ „Wyprodukowane szkło", r_7 — przepływ „Kontrola jakości", r_8 — przepływ „Standardowy plan produkcji szkła", r_9 — przepływ „Plan produkcji szkła „Jumbo", r_{10} — przepływ „Dostawa szkła standardowego", r_{11} — przepływ „Dostawa szkła „Jumbo", r_{12} — przepływ „Zużycie sprzętu", r_{13} — przepływ „Konserwacja sprzętu". Zapasy wiążą przepływy. Ich oznaczenia w modelu dynamiki systemu i jednostki miary podano w Tabeli 2.

No	Flow (r)	Designations in model	Units
1	Income	Income	rubles
2	Consumption	Consumption	rubles
3	Supply volume of raw materials	Supply_of_raw_ materials	ton
4	Charge loading	Charge_loading	ton
5	Waste of raw materials	Waste	ton
6	Produced glass	Flow	ton
7	Quality control	Quality_control	%
8	Standard glass production plan	Plan_st	ton
9	"Jumbo" glass production plan	Plan_jumbo	ton
10	Standard glass shipment	Shipment_st	ton
11	"Jumbo" glass shipment	Shipment_jumbo	ton
12	Equipment wear	Equipment_wear	%
13	Equipment maintenance	Maintenance	%

Wartości przepływu: „Dochód", „Zużycie", „Objętość dostaw surowców", „Obciążenie wsadu", „Odpady surowców", „Wyprodukowane szkło", „Kontrola jakości", „Plan produkcji szkła standardowego", „Plan produkcji szkła „Jumbo", „Dostawa szkła standardowego", „Dostawa szkła „Jumbo", „Zużycie sprzętu", „Konserwacja sprzętu" są określane na podstawie następujących zależności:

$$r_1 = v_1 + v_2, \quad (2)$$

$$r_2 = v_3 + v_4 + v_5 + v_6, \quad (3)$$

$$r_3 = v_7 + r_5, \quad (4)$$

$$r_4 = S_2, \quad (5)$$

$$r_5 = S_3 * c_1, \quad (6)$$

$$r_6 = S_3 - r_5, \quad (7)$$

$$r_7 = S_4 * v_8, \quad (8)$$

$$r_8 = v_9, \quad (9)$$

$$r_9 = v_{10}, \quad (10)$$

$$r_{10} = v_{11}, \quad (11)$$

$$r_{11} = v_{12}, \quad (12)$$

$$r_{12} = v_{10} * c_2, \quad (13)$$

$$r_{13} = delay(r_{12}), \quad (14)$$

gdzie v_1 — zmienna „Koszt na tonę szkła standardowego”, v_2 — zmienna „Koszt na tonę szkła „Jumbo””, v_3 — zmienna „Koszt energii elektrycznej na dzień”, v_4 — zmienna „Koszt gazu ziemnego na dzień”, v_5 — zmienna „Koszt surowców na dzień”, v_6 — zmienna „Koszt konserwacji sprzętu na dzień”, v_7 — zmienna „Objętość surowców”, v_8 — zmienna „Wada szkła”, v_9 — zmienna „Plan produkcji szkła standardowego”, v_{10} — zmienna „Plan produkcji szkła „Jumbo”, v_{11} — zmienna „Zamawianie szkła standardowego”, v_{12} — zmienna „Zamawianie szkła „Jumbo””, v_{13} — zmienna „Zużycie sprzętu”, c_1 — stała „Procent odpadów surowców”, c_2 — stała „Współczynnik zużycia sprzętu”, delay() — funkcja opóźnienia otwarcia przepływu na okres. Tabela 3 przedstawia oznaczenia zmiennych dynamicznych w modelu dynamiki układów oraz ich jednostki miary.

No	Dynamic variable (v)	Designations in model	Units
1	Cost per ton of standard glass	Cost_st	rubles
2	Cost per ton of “Jumbo” glass	Cost_jumbo	rubles
3	Cost of electricity per day	Cost_enrgy_d	rubles
4	Cost of natural gas per day	Cost_gas_d	rubles
5	Cost of raw materials per day	Cost_raw_dd	rubles
6	Cost of equipment maintenance per day	Maintenance_costs	rubles
7	Volume of raw materials	Rav_volume	ton
8	Glass defect	Defect_glass	%
9	Standard glass production plan	Plan_glass_st	ton
10	“Jumbo” glass production plan	Plan_glass_jumbo	ton
11	Standard glass ordering	Order_st	ton
12	“Jumbo” glass ordering	Order_jumbo	ton
13	Equipment wear	Maintenance	%
14	Electricity for glass production per day	Energy_volume	MW
15	Gas volume for glass production per day	Gas_volume	$м^3$
16	Raw material quality factor	Quality_raw1	%
17	Salary per day	Salary	rubles
18	Tax per day	Tax	rubles

Wartości zmiennych dynamicznych: „Koszt tony szkła standardowego”, „Koszt tony szkła „Jumbo”, „Dzienny koszt energii elektrycznej”, „Dzienny koszt gazu ziemnego”, „Dzienny koszt surowców”, „Dzienny koszt konserwacji urządzeń”, „Objętość surowców”, „Wada szkła”, „Plan produkcji szkła standardowego”, „Plan produkcji szkła „Jumbo”, „Zamówienie szkła standardowego”, „Zamówienie szkła „Jumbo”, „Zużycie urządzeń” są określane na podstawie następujących zależności:

$$v_1 = v_{11} * c_3, \tag{15}$$

$$v_2 = v_{12} * c_4, \tag{16}$$

$$v_3 = v_{14} * 24/700 * c_6, \tag{17}$$

$$v_4 = v_{15} * c_7, \tag{18}$$

$$v_5 = v_7 * v_{16} * c_5, \tag{19}$$

$$v_6 = S_8 * r_{13} * c_{10}, \tag{20}$$

$$v_7 = r_8 + r_9, \tag{21}$$

$$v_8 = S_8/100 + (1 - v_{16}) + c_{11}, \tag{22}$$

$$v_9 = S_4 * v_{11}/(v_{11} + v_{12} + r_7), \tag{23}$$

$$v_{10} = S_4 * v_{12}/(v_{11} + v_{12} + r_7), \tag{24}$$

$$v_{11} = Demayd_st(t), \tag{25}$$

$$v_{12} = Demayd_jambo(t), \tag{26}$$

$$v_{13} = Maintence_F(t), \tag{27}$$

$$v_{14} = S_4 * c_9, \tag{28}$$

$$v_{15} = c_8 * 24/700 * S_2, \tag{29}$$

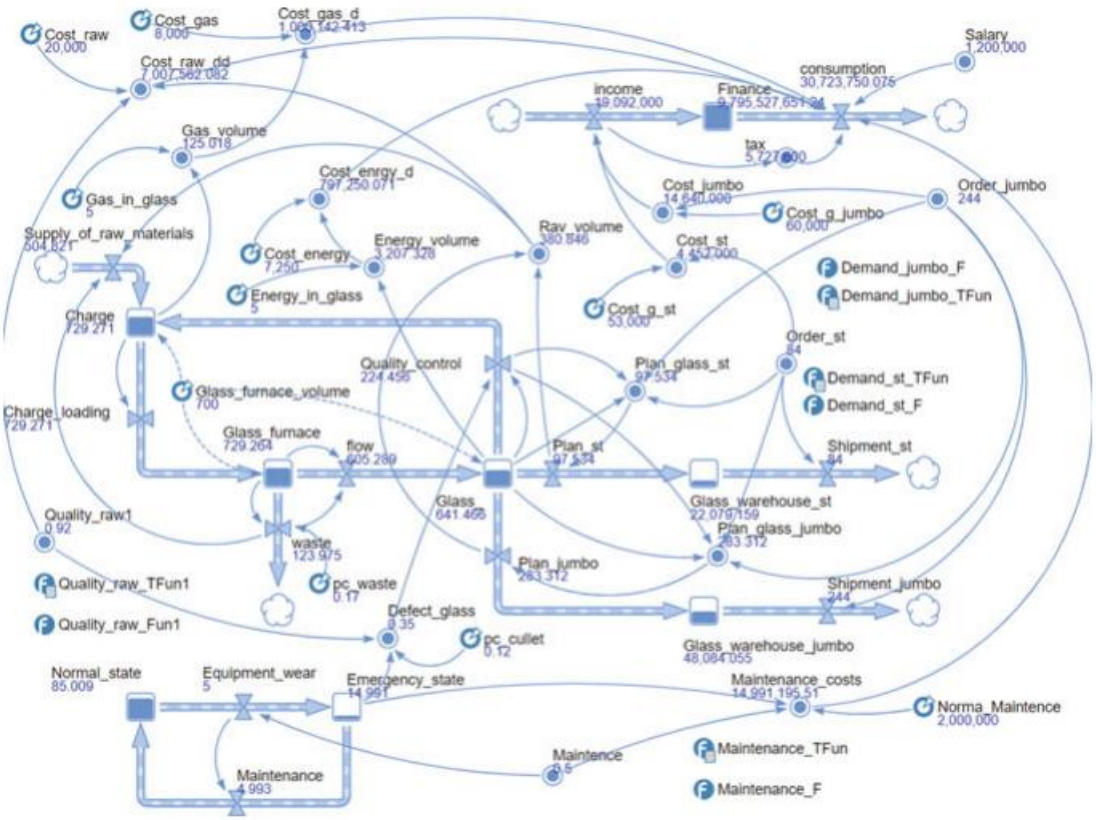
$$v_{16} = Quality_raw1_F(t), \tag{30}$$

$$v_{17} = c_{13} * Salary_F(t), \tag{31}$$

$$v_{18} = r_1 * c_{14}, \tag{32}$$

gdzie c_3—stała „Cena szkła standardowego za tonę”, c_4—stała „Cena szkła „Jumbo” za tonę”, c_5—stała „Cena surowca za tonę”, c_6—stała „Cena/mocy elektrycznej za megawat”, c_7—stała „Cena gazu za 1000 metrów sześciennych”, c_8—stała „Zużycie gazu na godzinę na 700 ton”, c_9—stała „Zużycie energii elektrycznej na godzinę na 700 ton”, c_{10}—stała „Czas naprawy urządzeń”, c_{11}—stała „Procent stłuczki szklanej w wsadzie”, c_{12}—stała „Objętość pieca szklarskiego”, c_{13}—stała „Średnie wynagrodzenie pracowników”, c_{14}—stała „Współczynnik odliczeń podatkowych”, Demand_st—funkcja uzyskiwania popytu na szkło standardowe, Demand_jumbo—funkcja uzyskiwania popytu na Szkło „Jumbo”, Quality_raw1_F — funkcja zmiany jakości surowca, Maintenance_F — funkcja zużycia maszyn, Salary_F — funkcja zmiany wynagrodzenia pracowników. Na podstawie układu równań różniczkowych (1), równań przepływów (2–14) oraz

wyrażeń dla zmiennych dynamicznych (15–32) opracowano strukturę modelu matematycznego dynamiki układu wieloetapowej produkcji wysokiej jakości szkła float

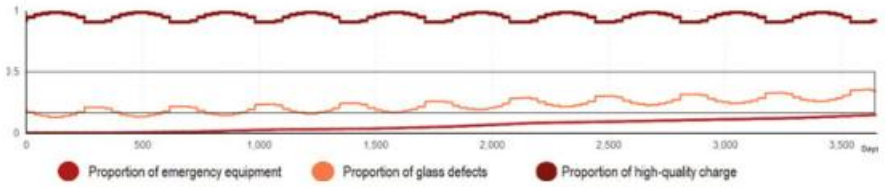


w oprogramowaniu symulacyjnym AnyLogic. Tabela 4 przedstawia oznaczenia stałych w modelu dynamiki układu, ich wartości i jednostki miary.

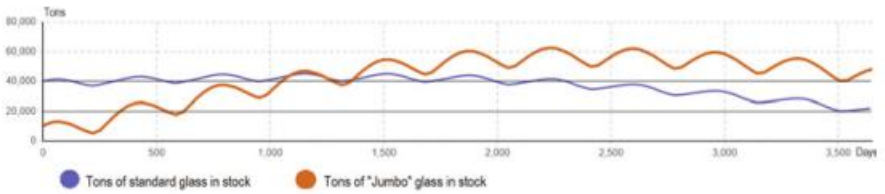
No	Constant (c)	Designations in model	Values	Units
1	Percentage of raw material waste	Pc_waste	18	%
2	Coefficient of equipment wear	EE	10	coefficient
3	Price of standard glass per ton	Cost_g_st	53,000	rubles
4	Price of "Jumbo" glass per ton	Cost_g_jumbo	60,000	rubles
5	Price of raw material per ton	Cost_raw	20,000	rubles
6	Price of electricity per megawatt	Cost_energy	7250	rubles
7	Price of gas per1000 cubic meters	Cost_gas	8000	rubles
8	Gas consumption per hour per 700 tons	Gas_in_glass	5	1000 $м^3$
9	Electricity consumption per hour per 700 tons	Energy_in_glass	5	MW
10	Equipment repair rate	Norma_Maintence	2,000,000	rubles
11	Percentage of cullet glass in the charge	Pc_cullet	12	%
12	Volume of glass furnace	Glass_furnace_vol	700	ton
13	Average salary of employees	Salary_c	1,200,000	rubles
14	Tax deductions coefficient	Tax_c	0.3	Coefficient

3 Modelowanie matematyczne

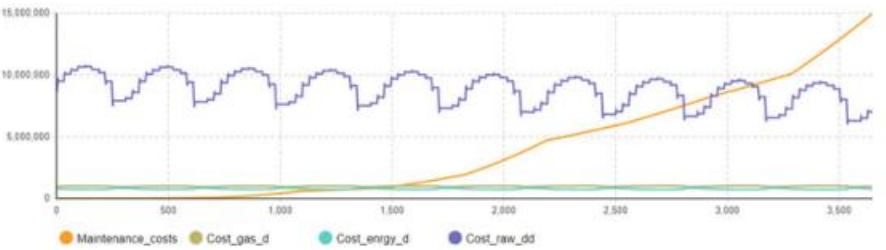
W oprogramowaniu AnyLogic do przeprowadzania eksperymentów z modelem dynamiki układu wieloetapowej produkcji wysokiej jakości szkła float, metoda Eulera została wykorzystana do rozwiązania układu równań różniczkowych (1). Metoda Newtona została wykorzystana do rozwiązania równań algebraicznych (2–32). Wybraną jednostką czasu modelu jest dzień. Symulację przeprowadzono w przedziale 10 lat – 3650 dni. Symulacje numeryczne przeprowadzono dla tych parametrów symulacji, wartości początkowych z tabeli 1 oraz wartości stałych z tabeli 4. Rysunek 3 przedstawia wykresy wzrostu współczynników defektów szkła w zależności od sezonowych składowych spadku jakości surowców i zużycia głównych urządzeń.



Skład chemiczny surowców nie jest stabilny i zmienia się w zależności od partii. Konserwacja urządzeń oparta na stanie technicznym, elastyczna kontrola receptur wsadu i program wsadowy mogą obniżyć koszty produkcji szkła float i zmniejszyć wskaźnik braków o 14%. Rysunek 4 przedstawia wykresy dynamiki zapełnienia magazynu szkła standardowego i magazynu szkła „Jumbo" w tonach, z uwzględnieniem sald początkowych i sezonowości popytu na te rodzaje szkła.



Zastosowanie automatycznej identyfikacji wad szkła float, opartej na wykorzystaniu sztucznych sieci neuronowych i transformacji falkowych, pozwala na zmniejszenie liczby wad niezidentyfikowanych, a dzięki optymalnemu rozkrojowi na zwiększenie produkcji szkła standardowego i szkła „Jumbo". Rysunek 5 przedstawia wykresy kosztów utrzymania urządzeń, gazu, energii elektrycznej i surowców w rublach.



Kontrola i zarządzanie stanem urządzeń w oparciu o „konserwację opartą na stanie" pozwala na obniżenie kosztów utrzymania urządzeń oraz odsetka braków szkła float.

4. Wnioski

Tu rozpatrzono zagadnienia analizy systemowej produkcji szkła float. W oparciu o cele strategicznej mapy celów przedsiębiorstwa opracowano układ równań różniczkowych dla funkcjonowania wieloetapowej produkcji wysokiej jakości szkła float. Zdefiniowano równania

łączące zapasy i wartości stałych. W oprogramowaniu do modelowania symulacyjnego AnyLogic opracowano strukturę modelu dynamiki systemu. Przeprowadzono eksperyment obliczeniowy dotyczący modelowania funkcjonowania produkcji szkła float. Wyniki modelowania symulacyjnego pokazują, że w celu obniżenia kosztów produkcji wysokiej jakości szkła float konieczne jest: elastyczne sterowanie produkcją wsadu w celu korygowania zmian jakości surowca, inteligentne sterowanie stanem głównych urządzeń procesowych, automatyczna identyfikacja defektów oraz optymalizacja rozkroju tafli szkła float.

Analiza interwałowa bezpieczeństwa zasobów informatycznych i telekomunikacyjnych infrastruktury krytycznej

Celem badań jest nowe podejście metodologiczne do rozwiązania problemu analizy interwałowej bezpieczeństwa zasobów informatycznych i telekomunikacyjnych infrastruktury krytycznej. Podejście to stanowi jedną z możliwości praktycznego zastosowania teorii średnich interwałowych (obliczeń interwałowych). Przeprowadzono analizę specyfiki tego podejścia, które decydują o jego przydatności i przydatności do oceny wskaźników bezpieczeństwa tak złożonych systemów technicznych w danym przedziale czasowym. Rozważono teoretyczne aspekty konstrukcji algorytmów obliczania średnich interwałowych poziomów bezpieczeństwa zasobów informatycznych i telekomunikacyjnych infrastruktury krytycznej, a także cechy obliczania górnej i dolnej średniej wartości poziomów bezpieczeństwa. Zaproponowano sekwencję obliczeń i wyrażeń analitycznych dla takich obliczeń na przykładzie konkretnego wskaźnika bezpieczeństwa. Jednocześnie analiza przedziałowa nie charakteryzuje się dużą złożonością matematyczną i obliczeniową, ale pozwala na uzyskanie szacunków przedziałowych bezpieczeństwa zasobów, które są odpowiednie do zadań kontroli i zarządzania, oszczędzają zasoby obliczeniowe i ostatecznie przyczyniają się do zwiększenia niezawodności kontroli bezpieczeństwa nowoczesnych infrastruktur krytycznych.

Słowa kluczowe: analiza przedziałowa · bezpieczeństwo · metoda · średnie przedziałowe · wskaźnik · zasoby informacyjne i telekomunikacyjne · infrastruktura krytyczna

1 Wstęp

Tworzenie i doskonalenie infrastruktury krytycznej dowolnego kraju jest trudne, a często wręcz niemożliwe, bez jednoczesnego i, w niektórych przypadkach, zaawansowanego rozwoju systemów i kompleksów informatycznych i telekomunikacyjnych, które stanowią kluczowe elementy takich infrastruktur i stanowią materialną podstawę terminowej, niezawodnej i bezpiecznej wymiany danych między ich użytkownikami. Przez infrastrukturę krytyczną (IK) rozumie się infrastrukturę ekonomiczną, przemysłową, energetyczną, informacyjną i/lub telekomunikacyjną, której przerwanie lub zakłócenie prowadzi do sytuacji nadzwyczajnej lub poważnych negatywnych konsekwencji na długi okres czasu dla obronności, bezpieczeństwa, stosunków międzynarodowych, gospodarki, innych sfer gospodarki kraju lub dla życia ludności zamieszkującej dane terytorium. Jednocześnie zasoby informacyjne IK obejmują cały zbiór danych zorganizowanych w celu skutecznego pozyskiwania wiarygodnych informacji w celu zapewnienia sprawnego funkcjonowania IK. Jest to zbiór pojedynczych dokumentów i zbiorów dokumentów, a także wielu dokumentów i zbiorów dokumentów w podsystemach informacyjnych

CI – bibliotekach, archiwach, funduszach, bankach danych i innych podsystemach informacyjnych. Zasoby telekomunikacyjne CI mogą obejmować wszystkie numery abonentów, adresy IP i adresy domen dostępne w sieciach, systemach i kompleksach telekomunikacyjnych, liczbę i przepustowość łączy telekomunikacyjnych (przewodowych, optycznych, radiowych i satelitarnych), kanały i ścieżki transmisji informacji, routery, stacje komutacyjne i węzły, a także zasób częstotliwości radiowych. Jednocześnie na szczególną uwagę zasługują zadania analizy bezpieczeństwa zasobów informacyjnych i telekomunikacyjnych (I&TR) współczesnych CI. Stanowią one kluczowe zagadnienie w teorii i praktyce zarządzania ochroną danych w CI i zapewnienia ich bezpiecznego funkcjonowania. Jednocześnie istniejące trudności w tworzeniu bezpiecznych systemów I&TR (Infrastruktury Informatycznej i Zarządzania Zasobami Ludzkimi), złożoność rozwiązania problemu wielokryterialnej analizy ich bezpieczeństwa wynikają między innymi z braku uniwersalnych metod uwzględniających szereg właściwości i cech podsystemów ochrony elementów tych zasobów, niekompletności i heterogeniczności informacji początkowych na temat jakości ochrony informacji w systemie I&TR (Infrastruktury Informatycznej i Zarządzania Zasobami Ludzkimi) jako całości. Dlatego opracowanie nowych metod obliczania i wielokryterialnej analizy interwałowej bezpieczeństwa I&TR nowoczesnych systemów IK z niekompletną informacją początkową jest ważnym i pilnym zadaniem.

2 Prace pokrewne

Wiele współczesnych badań naukowych poświęconych jest rozwojowi metod analizy bezpieczeństwa informacji i informacji (ITR) złożonych zarządzanych systemów informacyjnych oraz organizacyjno-technicznych [9–23]. W pracach tych formułowane i szczegółowo opisywane są różne podejścia do analizy bezpieczeństwa. Jednocześnie praktyczne zastosowanie tych podejść do analizy porównawczej poziomu bezpieczeństwa różnych zasobów CI oraz do opracowania obiecujących kierunków poprawy ochrony informacji w CI staje się problematyczne. W szczególności wynika to z uwzględnienia procesów przejściowych zachodzących w działających CI, wielokryterialnego charakteru współczesnych wymagań dotyczących bezpieczeństwa I&T CI oraz zarządzania procesem zapewniania bezpieczeństwa informacji, które determinują formułowanie nie tylko wektora, ale także dynamicznego zadania analizy bezpieczeństwa zasobów infrastruktur tej klasy, którego podejścia do rozwiązania nie zostały zbadane w ramach istniejących metod. Zaproponowano przeprowadzenie analizy bezpieczeństwa z uwzględnieniem algorytmów zarządzania ryzykiem, ale zalety tej metody są rekompensowane przez duże koszty obliczeniowe wymagane do jej praktycznego wdrożenia. Ponadto, poszczególne techniki analizy bezpieczeństwa w zdecydowanej większości nie uwzględniają parametrów organów i procesów zarządzania ochroną I&TR CI przy opracowywaniu systemów wskaźników bezpieczeństwa (SI) zasobów tej klasy. W przypadku, gdy taki problem został jednak rozwiązany, na przykład analizowano wartości bezwzględne SI, ale nie uwzględniano odchyleń tych SI od wymagań dla nich, co jest istotne dla analizy bezpieczeństwa zasobów tak złożonych infrastruktur. Analiza prac pokazuje, że gromadzenie dużej ilości heterogenicznych danych o stanie SI dla złożonych systemów technicznych, przewidziane w istniejących metodach, prowadzi do dużych nakładów czasowych na zbieranie statystyk, co negatywnie wpływa na wydajność analizy. To z kolei wpływa na czas reakcji podsystemu ochrony I&TR, wydłużając jego „czas reakcji" na zagrożenia realizowane przez intruza i wprowadzane działania kontrolne, skoncentrowane na uruchomieniu określonych środków ochrony I&TR CI. Analiza prac w tym zakresie zasługuje na szczególną uwagę pod kątem dokładności analizy bezpieczeństwa. I tak na przykład w istniejących metodach tradycyjne kryterium dokładności – wariancja błędu estymacji – jest równa wariancji a priori samego badanego procesu bezpieczeństwa I&TR, co nie może być akceptowalne dla nowoczesnych systemów sterowania o

wysokiej precyzji. Sztuczne sieci neuronowe służą do zwiększenia wiarygodności analizy SI. Umożliwiają one uwzględnienie jednego ważnego aspektu niepewności danych początkowych – niekompletności i niespójności informacji pomiarowej (obserwowanej) o wartościach SI, ale niestety nie są w stanie uwzględnić rozmycia takiej informacji i jej „zaszumienia". Z drugiej strony, badania w zakresie analizy bezpieczeństwa złożonych systemów w warunkach niepewności, choć już prowadzone, nie miały charakteru uniwersalnego. Poświęcone były głównie bieżącej ocenie wskaźników bezpieczeństwa informacji w ramach jednego rodzaju niepewności (np. rozmycia). Zatem analiza odpowiednich prac pozwala mówić o obiektywnej potrzebie rozwoju istniejących metod analizy bezpieczeństwa CI I&TR w przypadku uwzględnienia współczesnych wymagań podsystemu zarządzania bezpieczeństwem CI, wymagań urzędników (audytorów, administratorów) związanych z czasem trwania cykli monitorowania bezpieczeństwa i zarządzania bezpieczeństwem systemów tej klasy. Innymi słowy, istnieje obiektywna potrzeba uzyskania nie punktowych, lecz przedziałowych oszacowań SI. Analiza bezpieczeństwa może i powinna być przeprowadzana z określoną częstotliwością, powinna uwzględniać wszystkie rodzaje niepewności w danych początkowych. Podstawą matematyczną i metodologiczną takiej analizy mogą być algorytmy obliczania średnich przedziałowych SI dla CI I&TR. Pozwoli to zaoszczędzić zasoby obliczeniowe, zwiększyć niezawodność i adekwatność analizy.

3 Część teoretyczna — Aspekty konstruowania algorytmów obliczania średnich interwałowych wartości poziomów bezpieczeństwa

Bezwładność cykli kontroli bezpieczeństwa I&TR oraz procedur zarządzania bezpieczeństwem informacji krążących i przechowywanych w CI podkreśla potrzebę uzyskiwania nie punktowych (natychmiastowych), lecz interwałowych ocen bezpieczeństwa. Istotnym elementem naukowej i praktycznej nowości tych badań jest, naszym zdaniem, fakt, że wyróżniają się one od dotychczasowych prac obecnością złożonej, dualnej, zarówno „poziomej", jak i „pionowej" analizy interwałowej. Innymi słowy, ocena SI dla I&TR CI jest przeprowadzana zarówno w przedziałach czasowych („poziomo"), jak i wyznaczane są średnie interwałowe (dolna i górna granica) wartości SI („pionowo"). Jednocześnie ocena przedziałów czasowych („pozioma") jest zapewniona dzięki zastosowaniu modeli matematycznych opartych na ciągłych łańcuchach Markowa, gdy zmiana stanów dyskretnych (wartości) SI zachodzi w losowych momentach, ale jest szacowana w z góry określonym przedziale czasowym. Analiza przedziałowa SI („pionowa") wynika z zastosowania metod teorii średnich przedziałowych, które pozwalają na obliczenie górnych i dolnych wartości średnich (poziomów) konkretnego SI dla I&TR CI w tym przedziale czasowym. Racjonalnym narzędziem metodologicznym i analitycznym do rozwiązywania takich problemów mogą być, naszym zdaniem, algorytmy obliczania przedziałów (estymacja przedziałowa, obliczanie średnich przedziałowych), które stanowią jedną z kluczowych metod matematycznych w teorii średnich przedziałowych. Algorytmy obliczania średnich przedziałowych (ACIA) są w stanie matematycznie poprawnie łączyć metody opisu różnych rodzajów niepewności, traktując je jako przypadki szczególne. Za pomocą ACIA możliwe jest przetwarzanie heterogenicznych niekompletnych, rozmytych i sprzecznych informacji. ACIA umożliwia wykonywanie obliczeń interwałowych i uzyskanie ostatecznej oceny interwałowej wskaźników bezpieczeństwa dla CI I&TR, a fizyczne znaczenie i działanie ACIA przedstawia się następująco. Dane początkowe dla działania ACIA to z góry określone przedziały czasowe Δt, w których przeprowadzana będzie analiza interwałowa wskaźników bezpieczeństwa dla CI I&TR. Przedziały te Δt mogą wynosić od jednej do 20 minut, w zależności od czasu trwania cyklu sterowania (bezwładności sterowania) zabezpieczenia dla CI I&TR. Kolejnym krokiem w pracy ACIA jest określenie wskaźników SI dla CI I&TR dla każdego stanu bezpieczeństwa $\psi(\Theta n)$ na zbiorze wszystkich zidentyfikowanych wartości SI, poprzez analizę statystyk dla czasu Δt

chwilowych (punktowych, krokowych) wartości wskaźników SI. Jednocześnie wartości liczbowe parametrów charakteryzujących różne aspekty bezpieczeństwa informacji mogą służyć jako wskaźniki SI dla CI I&TR: dostępność, poufność, integralność. Na przykład czas $I_{hcp}(\Delta t)$ złamania i złamania haseł dla dostępu do I&TR może być SI dla poufności, czas $I_{spr}(\Delta t)$ bezpiecznego zapewnienia I&TR użytkownikom i podsystemom CI – dla dostępności, a czas $I_{ra}(\Delta t)$ przywrócenia dostępu użytkowników i podsystemów CI do I&TR po zhakowaniu i złamaniu haseł – dla integralności. Kolejnym krokiem w pracy ACIA jest obliczenie górnej i dolnej średniej wartości (poziomów) określonego SI dla I&TR CI w przedziale czasowym Δt. Obliczenia przeprowadza się na podstawie analizy statystycznej parametrów bezpieczeństwa mierzonych i obserwowanych w przedziale Δt, wykorzystując wyrażenia matematyczne leżące u podstaw metod teorii średnich przedziałowych . Na przykład, dla konkretnego SI charakteryzującego integralność zasobów – czasu przywrócenia dostępu do I&TR dla użytkowników i podsystemów CI po zhakowaniu i złamaniu haseł – obliczane są dokładne wartości górnego $I_{ra}(\Delta t)$ i dolnego I ra(Δt) jego średniego poziomu obserwowanego w przedziale czasowym Δt. Analiza, na przykład, czasu odzyskiwania dostępu do I&TR dla użytkowników i podsystemów CI po zhakowaniu i złamaniu haseł, z wykorzystaniem ACIA, charakteryzuje się szeregiem osobliwości. W szczególności zakłada się, że SI początkowo ma N możliwych stanów (wartości): {Θ1, Θ2, ..., Θn, ..., ΘN}. W tym przypadku SI dla każdego stanu ψ(Θn) w przedziale czasowym Δt jest określana na podstawie zbioru wszystkich możliwych stanów tego SI dla CI I&TR.

4. Część metodologiczna (określanie górnych i dolnych średnich wartości poziomów bezpieczeństwa) i dyskusja

Poszukiwanie górnych i dolnych średnich wartości SI_y dla I&TR CI w określonym przedziale czasowym podlega również zasadom obliczania średnich interwałowych: kolejnym krokiem w pracy ACIA jest, w naszym przykładzie, obliczenie dla określonego SI $I_{ra}(\Delta t)$, który charakteryzuje integralność zasobów telekomunikacyjnych (dla czasu odzyskiwania dostępu użytkowników i podsystemów CI do I&TR po złamaniu i skompromitowaniu haseł), w przedziale czasowym Δt, wykorzystując zidentyfikowane stany (wartości):

$$
\begin{aligned}
I_{ra}(\Delta t) &= \sum_{n=1}^{N} \psi(\Theta_n)\ p_n(\Delta t) \\
&= (\psi(\Theta 1) p1(\Delta t)) + (\psi(\Theta 2) p2(\Delta t)) + \ldots + (\psi(\Theta n) pn(\Delta t)) \\
&\quad + \ldots + (\psi(\Theta N) pN(\Delta t)), \qquad (1)
\end{aligned}
$$

gdzie ψ(Θn) jest wskaźnikiem charakteryzującym n-ty (n = 1, ..., N) stan w przedziale czasowym Δt; pn(Δt) to prawdopodobieństwo, że konkretny SI $I_{ra}(\Delta t)$ dla CI I&TR znajduje się w stanie (ma zidentyfikowaną wartość) Θn w przedziale czasowym Δt. Nawet jeśli prawdopodobieństwa pn(Δt) dla tego konkretnego SI $I_{ra}(\Delta t)$ nie są zdefiniowane (nieznane), zawsze istnieje możliwość określenia a priori na podstawie analizy eksperckiej górnej granicy I ra(Δt) i dolnej granicy $I_{ra}(\Delta t)$ średniego poziomu bezpieczeństwa, w naszym przypadku — pod względem czasu odzyskiwania dostępu do I&TR dla użytkowników i podsystemów CI po włamaniu i złamaniu haseł, obserwowanego w przedziale czasowym Δt. W tym przypadku SI ψ dla CI I&TR będzie traktowany jako pewna „cecha" (zidentyfikowany parametr bezpieczeństwa), a funkcje $\underline{I}_{ra}(\Delta t)$ i $\underline{I}_{ra}(\Delta t)$ jako górna i dolna średnia tej cechy. Z punktu widzenia ACIA można to zapisać matematycznie jako

$$
\{\underline{I}_{ra}(\Delta t) = \underline{R}(\psi); \underline{I}_{ra}(\Delta t) = \underline{R}(\psi), \qquad (2)
$$

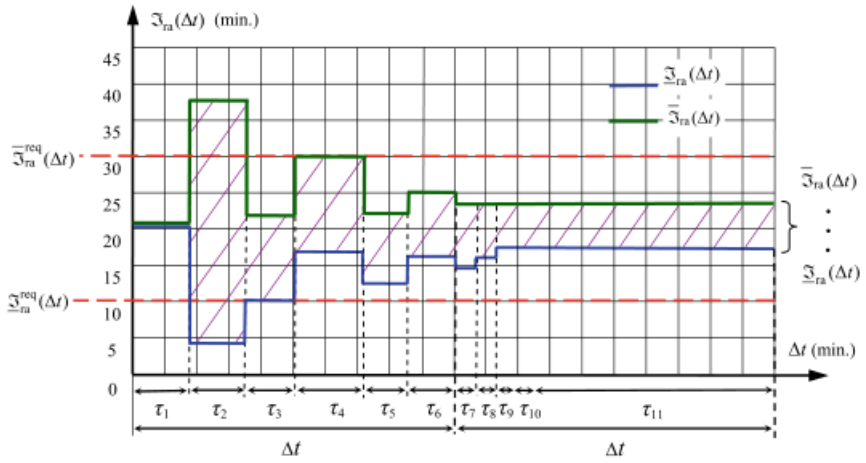
gdzie $\underline{R}(\psi)$ i $\underline{R}(\psi)$ to średnie przedziałowe górne i dolne. Jednocześnie algorytmy obliczania średnich przedziałowych przewidują, że w przypadku całkowitego braku jakichkolwiek informacji a priori o wartościach SI, górna granica $\underline{I}_{ra}(\Delta t)$ i dolna granica $\underline{I}_{ra}(\Delta t)$ średniego poziomu ochrony są równe odpowiednio jeden i zero:

$$\{\underline{I}_{ra}(\Delta t) = 1; \underline{I}_{ra}(\Delta t) = 0. \qquad (3)$$

Biorąc pod uwagę fakt, że w ramach ACIA przyjęto nazywać dowolną funkcję numeryczną f(y) „znakiem" zdarzenia losowego (zmiennej losowej) (co więcej, y ∈ Y, a Y jest przestrzenią elementarnych wyników wartości SI), w naszym przypadku „znakiem" jest poziom (wartość) konkretnego SI dla I&TR CI w przedziale czasowym Δt. Wówczas koncepcja „średnich przedziałowych" odpowiednio charakteryzuje górny i dolny średni poziom bezpieczeństwa dla I&TR CI w przedziale czasowym Δt, a dla konkretnego SI, na przykład czas $I_{ra}(\Delta t)$, jaki zajmuje użytkownikom i podsystemom CI uzyskanie dostępu do I&TR po zhakowaniu i złamaniu haseł, możliwe jest określenie jego „dokładnej" średniej $R_{ra}(f)$:

$$R_{ra}(f) = \int_Y f(y) dF(y), \qquad (4)$$

gdzie F(y) jest funkcją rozkładu zmiennej losowej — wartości SI, a średnia przedziałowa naszego konkretnego „znaku" to przedział $[\underline{R}_{ra}(f) \dots \underline{R}_{ra}(f)]$, który, bez uwzględnienia pojęcia „znaku", można zapisać jako $[\underline{I}_{ra}(\Delta t) \dots \underline{I}_{ra}(\Delta t)]$. Przykład zastosowania teorii średnich przedziałowych w celu określenia średniej w przedziale $[\underline{I}_{ra}(\Delta t) \dots \underline{I}_{ra}(\Delta t)]$ dla czasu przywrócenia dostępu po zhakowaniu i złamaniu haseł pokazano na rysunek .



Rysunek ilustruje, że przy nieznanej a priori funkcji rozkładu F(y) można mówić jedynie o „granicach" $\underline{F}(y) \le F(y) \le \underline{F}(y)$ dla wszystkich wartości y ∈ Y. Jednocześnie, w „dokładnych" granicach

$$\underline{R}_{ra}(f) \le R_{ra}(f) \le \underline{R}_{ra}(f), \qquad (5)$$

charakteryzujące średnie górnego przedziału $\underline{I}_{ra}(\Delta t)$ i dolnego przedziału $\underline{I}_{ra}(\Delta t)$,

$$\underline{I}_{ra}(\Delta t) \le I_{ra}(\Delta t) \le \underline{I}_{ra}(\Delta t), \qquad (6)$$

„Dokładna" średnia jest znajdowana dla konkretnego SI dla CI I&TR, w naszym przypadku dla czasu $I_{ra}(\Delta t)$ przywrócenia dostępu użytkowników i podsystemów CI do I&TR po zhakowaniu i skompromitowaniu haseł. Analiza wykresów na rysunku pozwala zauważyć, że biorąc pod uwagę początkowy, początkowy czas przywracania dostępu użytkowników $I_{ra}(\Delta t) \approx 20\,\min$ (podany jako przykład), stan może przyjmować wartości od 0 do 45 min, a jego oszacowanie osiąga stan stabilny już w ósmym kroku (τ8) działania i w drugim kroku (przedział czasu Δt) analizy przedziałowej. Wykresy na rysunku, oprócz obserwacji zmiany szacowanych wartości interwału $I_{ra}(\Delta t)$ (tj. jego przejścia z jednego stanu do drugiego w losowych odstępach czasu τ), pozwalają uzyskać odpowiedź na pytanie istotne, wręcz fundamentalne dla analizy bezpieczeństwa: czy uzyskane wartości dolnego interwału $\underline{I}_{ra}(\Delta t)$ i górnego $\underline{I}_{ra}(\Delta t)$ analizowanego SI dla CI I&TR (sektor zacieniowany) mieszczą się w granicach dopuszczalnych, wymaganych średnich wartości dolnego $\underline{I}_{ra}^{req}(\Delta t)$ i górnego $\underline{I}_{ra}^{req}(\Delta t)$ określonych przez politykę bezpieczeństwa lub administratora bezpieczeństwa CI. Aby określić inne wskaźniki bezpieczeństwa, na przykład w ramach określania poziomu poufności – dla czasu $I_{hcp}(\Delta t)$ włamania i złamania haseł dostępu do I&TR; W ramach określania poziomu dostępności – dla czasu $I_{spr}(\Delta t)$ bezpiecznego zapewnienia I&TR użytkownikom i podsystemom CI w przedziale czasu Δt, zadanie znalezienia ich średnich interwałowych jest rozwiązywane w podobny sposób.

5 Wnioski

Rozważana metoda pokazuje, że istnieje praktyczna możliwość analizy bezpieczeństwa I&TR CI, uwzględniając współczesne wymagania podsystemów zarządzania bezpieczeństwem CI, wymagania urzędników (audytorów, administratorów bezpieczeństwa) związane z bezwładnością procesów decyzyjnych, z czasem trwania cykli monitorowania bezpieczeństwa i zarządzania bezpieczeństwem systemów tej klasy. Podejście to opiera się na algorytmach obliczania średnich interwałowych i umożliwia uzyskanie nie punktowych (chwilowych), lecz interwałowych estymacji SI, podczas gdy analiza bezpieczeństwa jest przeprowadzana z zadaną częstotliwością i uwzględnia niepewność danych początkowych. Proponowane podejście metodologiczne nie charakteryzuje się dużą złożonością matematyczną i obliczeniową, ale pozwala na uzyskanie szacunków przedziałowych bezpieczeństwa zasobów CI, adekwatnych do zadań kontroli i zarządzania. Uwzględnia to niepewność obserwowanych i kontrolowanych parametrów bezpieczeństwa i oszczędza zasoby obliczeniowe, co ostatecznie prowadzi do wzrostu niezawodności analizy bezpieczeństwa współczesnych infrastruktur krytycznych. Praktyczne zastosowanie proponowanego podejścia do obliczeń przedziałowych jest możliwe zarówno w ramach prac badawczych, jak i w systemach automatycznej kontroli bezpieczeństwa zasobów złożonych zarządzanych infrastruktur krytycznych. Kierunkiem dalszych badań może być rozwój metod analizy bezpieczeństwa łączących obliczenia przedziałowe z tradycyjnymi algorytmami filtrowania.

Przydzielanie praw dostępu do aplikacji w korporacyjnej sieci mobilnej: Rozwój oprogramowania

Tu przedstawiono badanie i implementację metody przydzielania praw dostępu do aplikacji w korporacyjnej sieci mobilnej. Przeanalizowano istniejące rozwiązania w zakresie przydzielania praw dostępu do aplikacji i usług intranetu, uzasadniono potrzebę stworzenia metody

zapewniającej bezpieczeństwo informacji podczas dostępu do sieci korporacyjnych o zróżnicowanych wymaganiach co do poziomu bezpieczeństwa oraz opisano postęp i wyniki prac rozwojowych.

Słowa kluczowe: Prawa dostępu · Zróżnicowanie uprawnień · Aplikacja · Korporacyjna sieć mobilna · System informatyczny · Serwer WWW

1 Wstęp

Podczas tworzenia pojedynczej przestrzeni informacyjnej dużej firmy jednym z najtrudniejszych problemów jest zarządzanie prawami dostępu do aplikacji i usług intranetu. Każdy serwer WWW lub serwer aplikacji ma własne mechanizmy kontroli uprawnień. W niektórych przypadkach są to po prostu pliki konfiguracyjne edytowane za pomocą edytora tekstu, w innych – rozbudowany interfejs graficzny, a w jeszcze innych – wbudowane narzędzia systemu operacyjnego. Występuje, gdy konfiguracja systemu różnicowania uprawnień jest umieszczana w kodzie źródłowym aplikacji, tj. każda zmiana ustawień pociąga za sobą restrukturyzację modułów oprogramowania . Podczas pierwszego wprowadzania lub usuwania konta użytkownika z systemu, podczas publikowania lub usuwania zasobu, administrator musi wprowadzić zmiany w każdym z systemów, co prowadzi do błędów, nieprawidłowych lub nieaktualnych ustawień. W prostych przypadkach, gdy zestaw aplikacji jest niewielki, zadanie scentralizowanej kontroli można rozwiązać, wykorzystując własne rozwiązania lub mechanizmy wbudowane w serwery WWW lub serwery aplikacji. Na przykład magazyny informacji o użytkownikach można skonsolidować za pomocą oprogramowania metakatalogowego lub aplikacji natywnych w jeden centralny katalog główny, po czym serwery WWW i aplikacji (z dodatkowymi narzędziami) mogą umieszczać informacje o kontroli dostępu w tym samym katalogu . Jednak gdy liczba aplikacji rośnie, a system staje się bardziej złożony, koszt wdrożenia takiego rozwiązania może być równy kosztowi migracji systemów na nową, zunifikowaną platformę. Technologie mobilne są obecne w biznesie od dawna, a pracodawcy na całym świecie wykorzystują je do celów zawodowych. Pojawiające się problemy w obszarach bezpieczeństwa i niezawodności przechowywania danych muszą być stale rozwiązywane. Najpoważniejszym ryzykiem dla każdej firmy jest oczywiście zwiększone prawdopodobieństwo wycieku informacji biznesowych, co z kolei prowadzi do różnych strat – finansowych, technicznych i wizerunkowych. W ramach mobilności korporacyjnej należy wprowadzić surowe zasady bezpieczeństwa, z których co najmniej dwie muszą być absolutnie obowiązkowe: zarządzanie urządzeniami mobilnymi (MDM – Mobile Device Management), które zapewnia ich konfigurację zgodnie z określonymi zasadami, a także rozdzielenie danych osobowych i służbowych na urządzenia mobilne za pomocą odpowiednich aplikacji korporacyjnych przeznaczonych do obsługi poczty elektronicznej i plików oraz zintegrowanych z korporacyjną usługą katalogową. Główne zagrożenia dla firm wiążą się właśnie z niewystarczalnością stosowanych środków ochrony i rozdzieleniem praw dostępu do aplikacji. Znaczenie tej pracy jest oczywiste, ponieważ zagrożenia są szczególnie wysokie w przypadku zapewniania mobilnego dostępu do informacji biznesowych kadrze kierowniczej firmy. Im wyższy status i pozycja danej osoby, tym wyższa wartość poufnych informacji, z których korzysta w swojej pracy. Dostęp mobilny bezpośrednio wpływa na jakość zarządzania firmą, a ostatecznie na wydajność i efektywność jej pracy. Dlatego ważne jest jasne zdefiniowanie listy potencjalnych zagrożeń i na tej podstawie zbudowanie polityki korzystania z osobistych urządzeń mobilnych z rozdzieleniem poziomów dostępu. Celem pracy jest opracowanie metody przydzielania uprawnień dostępu do aplikacji w sieciach korporacyjnych o zróżnicowanych wymaganiach dotyczących poziomu bezpieczeństwa, która uwzględni specyfikę działań różnych użytkowników i przydzieli dostęp do aplikacji niezbędnych do pracy. Na podstawie analizy trendów i perspektyw

rozwoju nowoczesnych korporacyjnych sieci mobilnych ujawniono sprzeczność między wymaganiami bezpieczeństwa informacji przy dostępie do chronionych usług i informacji za pomocą uniwersalnych urządzeń mobilnych a możliwościami technicznymi systemów bezpieczeństwa informacji, które pozwalają zapewnić bezpieczeństwo informacji, gdy taki dostęp odbywa się w sieciach korporacyjnych o zróżnicowanych wymaganiach bezpieczeństwa.

2 Analiza przydzielania praw dostępu w sieciach korporacyjnych

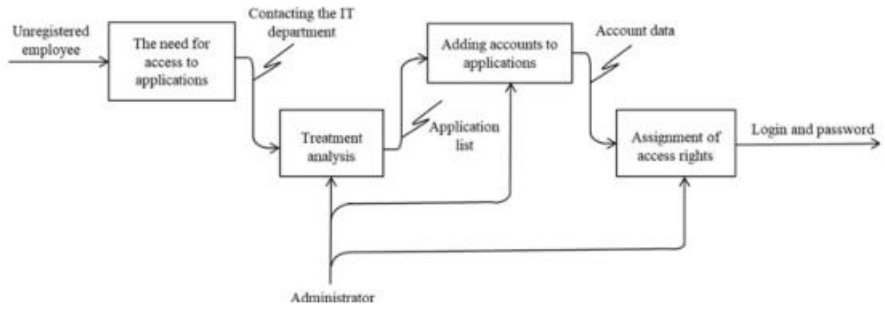
Współczesne warunki organizacji działalności większości firm wymagają technologii informatycznych, aby utrzymać konkurencyjność, a tym samym stworzyć warunki do powszechnego wykorzystania zaawansowanych technologicznie rozwiązań [9, 10]. Dział Informatyki zajmuje się wsparciem i rozwojem infrastruktury IT przedsiębiorstwa, a jego pracownicy zazwyczaj zajmują się następującymi czynnościami:

• wsparciem serwerów zewnętrznych i wewnętrznych firmy;

• administracją systemami;

• wsparciem użytkowników.

Przyjrzyjmy się bliżej wsparciu użytkowników. Specjalista ds. wsparcia użytkowników jest odpowiedzialny za instalację, przydzielanie uprawnień oraz diagnostykę oprogramowania i sprzętu; zapewnia wsparcie techniczne i doradztwo użytkownikom końcowym; jest odpowiedzialny za organizację napraw sprzętu komputerowego; zapewnia dostępność materiałów eksploatacyjnych do komputerów i sprzętu biurowego; doradza użytkownikom w kwestiach technicznych. Najbardziej czasochłonny proces wiąże się z terminowym przydzielaniem praw dostępu do aplikacji sieci korporacyjnej, ponieważ wymaga to zrozumienia funkcji i zadań wykonywanych przez użytkownika zgodnie z opisami stanowisk. W związku z tym następuje znaczne rozszerzenie obowiązków specjalisty ds. wsparcia użytkowników . Przy tradycyjnym podejściu do organizacji dostępu, pracownicy, którzy często korzystają z kilku aplikacji, muszą nie tylko otrzymać konto dla każdej z nich, ale także wielokrotnie przechodzić procedurę autoryzacji . Rezultatem jest dyskomfort w pracy. Dodatkowo hasła są niekiedy zapisywane na „kartce" leżącej obok komputera, co już samo w sobie stanowi słaby punkt w systemie bezpieczeństwa; do działu wsparcia bardzo często zgłaszają się użytkownicy, którzy zapomnieli jednego ze swoich haseł. Schemat modelu pracy działu IT przedstawiono na rysunku



Obecnie działalność specjalisty ds. wsparcia użytkowników sprowadza się do otrzymania wniosku o uprawnienia dostępu do każdej aplikacji z osobna. W kolejnym etapie wniosek jest analizowany i określana jest potrzeba utworzenia konta, które nadaje użytkownikowi login i hasło

do przeprowadzenia procedury autoryzacji w aplikacji . W przypadku wykrycia takiej sytuacji tworzone jest konto, które nadaje użytkownikowi uprawnienia dostępu. Procedura ta jest akceptowalna dla małych firm, w których liczba pracowników i aplikacji jest ograniczona do kilkudziesięciu. W firmach zatrudniających więcej pracowników niż ta liczba użytkowników, specjalista ds. wsparcia może mieć szereg czasochłonnych zadań:

• w przypadku zwolnienia pracownika lub przeniesienia go z jednego działu do innego zachodzi konieczność zmiany dostępu do niektórych aplikacji;

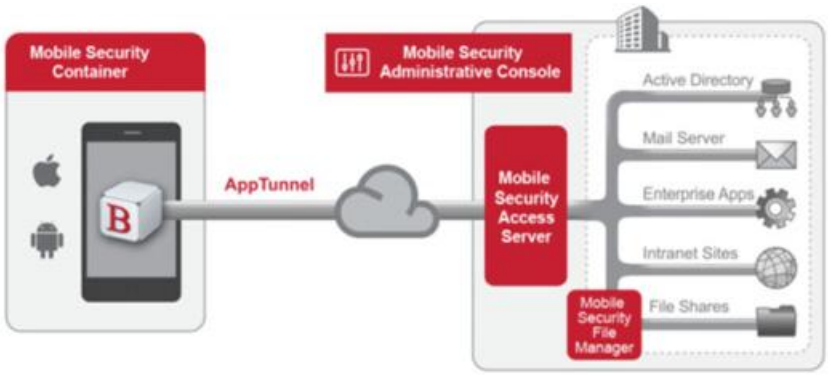
• wraz ze wzrostem liczebności personelu firmy, proces zakładania kont i autoryzacji zajmuje większą część czasu pracy, co uniemożliwia wykonywanie innych obowiązków na stanowisku.

3 Przegląd i analiza istniejących rozwiązań

Obecnie na rynku dostępnych jest wiele różnych systemów ograniczających dostęp do aplikacji korporacyjnych. Poniżej znajduje się przegląd niektórych systemów, który daje obraz stanu tej dziedziny.

3.1 Oracle Mobile Security Suite

Nowy pakiet, w połączeniu z rozwiązaniami do zarządzania tożsamością i dostępem, zapewnia zintegrowaną platformę Oracle, która umożliwia organizacjom zarządzanie dostępem do wszystkich aplikacji na wszystkich typach urządzeń, w tym gadżetach mobilnych, komputerach stacjonarnych i laptopach. Rozwiązanie spełnia nowoczesne koncepcje COPE (urządzenie należy do firmy, jest używane osobiście przez pracownika, w tym do celów osobistych) lub BYOD (urządzenie należy do pracownika i jest używane osobiście, w tym do celów służbowych). Strukturę kompleksu przedstawiono na rysunku 2.



Nowy pakiet, choć dostępny jako samodzielne rozwiązanie, jest ściśle zintegrowany z Oracle Mobile Suite w zakresie tworzenia, integracji i wdrażania aplikacji. Jedną z głównych funkcji Oracle Mobile Security Suite jest zarządzanie bezpieczeństwem urządzeń mobilnych. Wśród narzędzi do zarządzania i kontroli:

• Tryb geofencingu ograniczający dostęp lub funkcjonalność w zależności od lokalizacji;

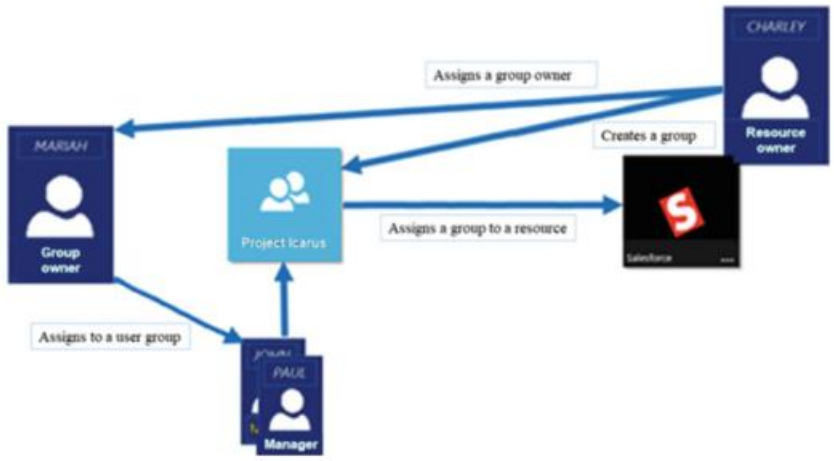
• Zarządzanie zasadami korzystania z aplikacji, w tym ograniczanie funkcji kopiowania/wklejania/drukowania w celu zapobiegania wyciekom danych;

• bezpieczne tunele sieciowe dla każdej aplikacji, zapewniające ochronę informacji bez użycia połączeń VPN poprzez wirtualne sieci prywatne na poziomie urządzeń, które wyczerpują baterie;

• szyfrowanie danych lokalnych z obsługą możliwości zdalnego usunięcia informacji firmowych i aplikacji z osobistego urządzenia pracownika po jego zwolnieniu.

3.2 MS Azure Active Directory

Usługa Azure Active Directory (Azure AD) umożliwia zarządzanie aplikacjami w chmurze, aplikacjami lokalnymi i zasobami za pośrednictwem grup firmowych. Zasoby mogą być specyficzne dla katalogu, takie jak uprawnienia do zarządzania obiektami za pośrednictwem ról w katalogu, lub nie, takie jak aplikacje SaaS, usługi Azure, witryny SharePoint i zasoby lokalne. Azure AD umożliwia udzielanie dostępu do zasobów firmowych poprzez nadanie praw dostępu pojedynczemu użytkownikowi lub grupie Azure AD. Korzystając z grup, właściciel zasobu (lub właściciel katalogu Azure AD) może przypisać zestaw uprawnień dostępu wszystkim członkom grupy zamiast udzielać praw indywidualnie. Właściciel zasobu lub katalogu może również przyznać uprawnienia do zarządzania listą członków innej osobie, takiej jak kierownik działu lub administrator pomocy technicznej, umożliwiając jej dodawanie i usuwanie członków w razie potrzeby. Strukturę usługi Azure AD przedstawiono na rysunku 3.



Istnieją cztery sposoby przypisywania użytkownikom praw dostępu do zasobu:

• bezpośrednie mianowanie. Właściciel zasobu bezpośrednio przypisuje użytkownika do zasobu.

• przypisanie do grupy. Właściciel zasobu przypisuje grupę usługi Azure AD do zasobu, która automatycznie przyznaje wszystkim jej członkom dostęp do zasobu. Członkostwem w grupie zarządza zarówno właściciel grupy, jak i właściciel zasobu, co pozwala każdemu z nich dodawać i usuwać członków grupy.

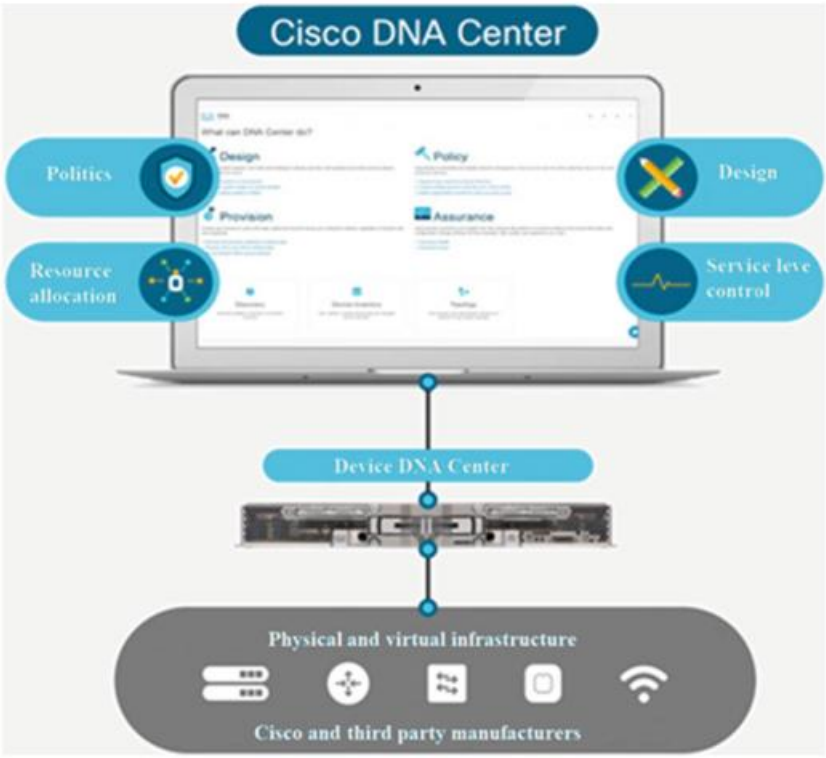
• Przypisanie oparte na regułach. Właściciel zasobu tworzy grupę i używa reguły, aby określić, którzy użytkownicy są przypisani do danego zasobu. Reguła opiera się na atrybutach przypisanych poszczególnym użytkownikom. Właściciel zasobu zarządza regułą, aby określić, które atrybuty i wartości są wymagane do udzielenia dostępu do zasobu.

• Przypisanie ze źródła zewnętrznego. Dostęp jest zapewniany ze źródła zewnętrznego, takiego jak katalog lokalny lub aplikacja SaaS. W takiej sytuacji właściciel zasobu przypisuje grupę do zapewnienia dostępu do zasobu, a członkowie grupy są zarządzani przez źródło zewnętrzne.

3.3 Cisco DNA Center

Kompleksowy system zarządzania siecią wspierający rozwój i innowacje w firmie. Obecnie sieć odgrywa ważniejszą rolę w biznesie niż kiedykolwiek wcześniej. System automatyzuje wdrażanie

infrastruktury, zarządza jej cyklem życia i łączy ją z siecią, a także utrzymuje jakość i bezpieczeństwo aplikacji, pozwalając personelowi IT skupić się na projektach sieciowych o strategicznym znaczeniu dla firmy. Dzięki Cisco DNA Center nie musisz już poświęcać zbyt wiele czasu na udostępnianie zasobów sieciowych i wykonywanie czasochłonnych zadań związanych z rozwiązywaniem problemów. Funkcje takie jak automatyczne łączenie urządzeń z siecią i zarządzanie obrazami oprogramowania pozwalają na instalację i aktualizację urządzeń w ciągu kilku minut, a nie godzin. Dzięki gotowemu do użycia urządzeniu Cisco można łatwo podłączyć nowe biura zdalne do sieci. DNA Center Assurance przekształca każdy punkt w sieci w czujnik, który generuje ciągły strumień danych telemetrycznych dotyczących wydajności aplikacji i łączności użytkowników w czasie rzeczywistym. Dodatkowo dostępne są dane automatycznego śledzenia ścieżki oraz funkcja wspomaganego odzyskiwania. Wszystko to pozwala rozwiązać problemy z siecią w ciągu kilku minut, zanim doprowadzą one do poważnych konsekwencji. Dzięki integracji z rozwiązaniami bezpieczeństwa Cisco Stealthwatch zagrożenia mogą być wykrywane i usuwane nawet w ruchu szyfrowanym. DNA Center oferuje również otwartą, rozszerzalną platformę, która obsługuje szeroką gamę zewnętrznych aplikacji i systemów do udostępniania danych i analiz dzięki wbudowanym funkcjom. DNA Center to jedyny scentralizowany system zarządzania siecią, który łączy wszystkie te funkcje. DNA Center oferuje jeden panel sterowania dla wszystkich podstawowych funkcji sieci. Platforma ta umożliwia działom IT szybsze i bardziej efektywne reagowanie na zmiany i stawianie czoła pojawiającym się wyzwaniom. DNA Center to system zarządzania siecią, kontroler rdzenia i platforma analityczna, na których zbudowana jest sieć oparta na intencjach Cisco. Oprócz zarządzania urządzeniami i konfiguracją, DNA Center oferuje pakiet rozwiązań programowych, które zapewniają kompleksową platformę zarządzania siecią; programowalny kontroler sieciowy do automatyzacji urządzeń i usług wirtualnych; mechanizm kontroli gwarantujący gwarantowaną jakość usług dla wszystkich użytkowników. Oprogramowanie DNA Center działa na urządzeniu DNA Center i kontroluje wszystkie urządzenia Cisco, zarówno fizyczne, jak i wirtualne (w sieci i poza nią). Menu główne DNA Center zawiera cztery sekcje ogólne. Sieć została zaprojektowana z wykorzystaniem wizualnych map fizycznych i topologii logicznych. Funkcjonalność platformy umożliwia importowanie istniejących map, obrazów i topologii bezpośrednio z modułu Cisco Prime Infrastructure and Application Policy Infrastructure Controller Enterprise Module (APIC-EM). Ułatwia to i przyspiesza przeprowadzanie aktualizacji. Wykrywanie urządzeń odbywa się automatycznie. W tym celu można użyć protokołu Cisco Discovery Protocol (CDP) lub po prostu wprowadzić zakres adresów IP. DNA Center tłumaczy informacje zebrane w regułach na określone konfiguracje sieci i urządzeń, zgodnie z różnymi typami, kompilacjami, modelami, systemami operacyjnymi, rolami i limitami zasobów urządzeń sieciowych. Za pomocą DNA Center można tworzyć sieci wirtualne, reguły kontroli dostępu, reguły kopiowania ruchu i reguły aplikacji. Po utworzeniu reguł w DNA Center, ich konfiguracja jest bardzo prosta dzięki funkcji „przeciągnij i upuść". Każdej kategorii tożsamości (użytkownicy, urządzenia, aplikacje itp.) w rejestrze DNA Center przypisana jest polityka, która zawsze będzie powiązana z tą tożsamością. Proces ten jest w pełni zautomatyzowany. Nowym urządzeniom dodawanym do sieci przypisana jest polityka oparta na ich tożsamościach. To znacznie ułatwia konfigurację biura zdalnego. DNA Center Assurance to kompleksowe rozwiązanie, które zapewnia wyższy i bardziej spójny poziom usług, aby sprostać rosnącym potrzebom biznesowym. DNA Center Assurance to coś więcej niż tylko monitorowanie sieci i rozwiązywanie problemów. Proaktywne reagowanie i predykcyjne działanie sieci pomagają zapewnić wysoką wydajność klientów, aplikacji i usług. W rezultacie osiągana jest spójność usług, a proces optymalizacji sieci staje się proaktywny. Jednocześnie mniej czasu poświęca się na rozwiązywanie problemów. Strukturę Cisco DNA Center przedstawiono na rysunku 4.



3.4 G Suite Google

Funkcje zarządzania urządzeniami mobilnymi Google służą do administrowania, zabezpieczania i śledzenia firmowych urządzeń mobilnych. Umożliwiają one kontrolowanie różnych urządzeń, w tym telefonów, tabletów i smartwatchy. Pracownicy organizacji mogą korzystać z zasobów organizacji na urządzeniach osobistych lub firmowych. Dostępne są dwa poziomy zarządzania: podstawowy i zaawansowany. Dostępne opcje różnią się w zależności od używanego trybu. Niektóre funkcje zaawansowane, takie jak audyt urządzeń i ustawianie zasad, nie są dostępne we wszystkich wersjach. W podstawowym trybie kontroli można wykonywać następujące czynności:

• ustawiać reguły wymagające skonfigurowania blokady ekranu lub hasła w celu ochrony danych firmowych na urządzeniach.

• usuwać dane firmowe ze zgubionych lub skradzionych urządzeń;

• udzielać użytkownikom dostępu do rekomendowanych aplikacji firmowych na urządzenia z systemem Android.

• publikować i dystrybuować prywatne aplikacje;

• przeglądać listę urządzeń z dostępem do danych firmowych w Konsoli administracyjnej Google.

Jeśli potrzebujesz więcej funkcji do kontrolowania urządzeń z dostępem do danych firmowych, możesz skorzystać z zaawansowanego trybu zarządzania, który umożliwia:

• ustanowienie reguły wymagającej używania silniejszych haseł;

• zapewnić użytkownikom dostęp do aplikacji firmowych w Katalogu Zalecanego Oprogramowania (dla Androida i iOS);

• korzystać z profili służbowych na urządzeniach z Androidem, aby oddzielić aplikacje osobiste od firmowych;

• zablokować dostęp do niektórych ustawień i funkcji urządzenia. Na przykład można zabronić

połączenia się z sieciami komórkowymi lub przez Wi-Fi, tworzenia zrzutów ekranu itp.;

• monitorować zgodność z polityką i otrzymywać raporty dotyczące użytkowników, urządzeń i wersji systemu operacyjnego.

3.5 Analiza porównawcza systemów kontroli dostępu do aplikacji korporacyjnych

Analizę porównawczą systemów ograniczania praw dostępu do aplikacji korporacyjnych przedstawiono w tabeli.

Function	Analyzed systems			
	Oracle Mobile Security Suite	MS Azure Active Directory (Azure AD)	Cisco DNA Center	G Suite Google
Free version	–	–	–	+
Convenient interface	+	+	+	+
Open source	–	–	–	–
Support for data encryption	+	+	+	+
Additional plugins	+	–	+	–
Multi-language support	+	+	+/–	+

Po przeanalizowaniu gotowych rozwiązań zdecydowano się na samodzielne opracowanie modułu, bez angażowania płatnego oprogramowania. Wynika to z faktu, że istniejące rozwiązania mają albo ograniczoną funkcjonalność, która nie wystarcza do zaimplementowania wszystkich funkcji modułu, albo są zbędne, tzn. środki finansowe wydane na zakup gotowego rozwiązania pozostaną w pełni niewykorzystane. Ponadto, przy zakupie oprogramowania konieczne jest przeznaczenie środków na szkolenie personelu w zakresie przygotowania, konfiguracji i utrzymania stabilnego działania zakupionego produktu. Ponadto, w związku z sankcjami nakładanymi przez firmy zagraniczne na wykorzystywanie wyników działalności intelektualnej oraz w celu zwiększenia bezpieczeństwa dostępu do aplikacji w sieciach korporacyjnych, konieczne jest opracowanie unikalnej metody przydzielania uprawnień z różnymi wymaganiami dotyczącymi poziomu bezpieczeństwa.

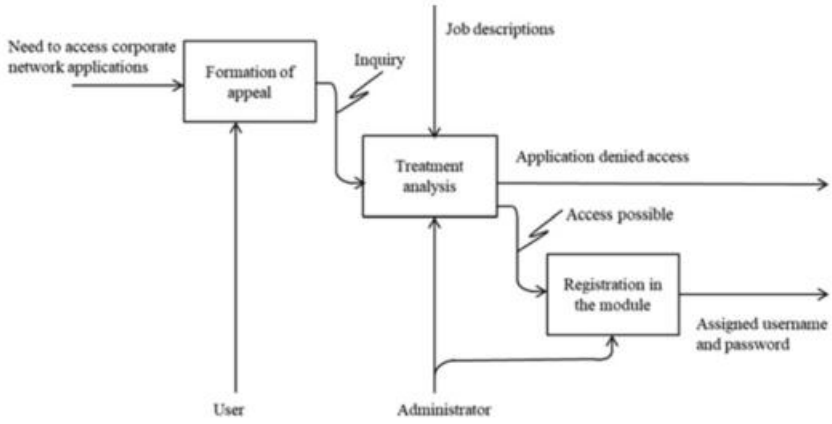
4 Projektowanie i rozwój metody przydzielania praw dostępu aplikacjom w korporacyjnej sieci mobilnej

W ramach rozwiązywanego problemu zastosowano system wielodostępny, który działa na całej flocie komputerowej organizacji. W związku z tym rozważane są systemy DBMS typu klient-serwer. System DBMS typu klient-serwer umożliwia klientowi i serwerowi wymianę minimalnej wymaganej ilości informacji. W tym przypadku główne obciążenie obliczeniowe spoczywa na serwerze. Klient może wykonywać funkcje wstępnego przetwarzania przed przesłaniem informacji do serwera, ale jego głównym zadaniem jest organizacja dostępu użytkowników do serwera . W większości przypadków system DBMS typu klient-serwer jest znacznie mniej wymagający pod względem przepustowości sieci komputerowej niż system DBMS typu serwer plików, zwłaszcza

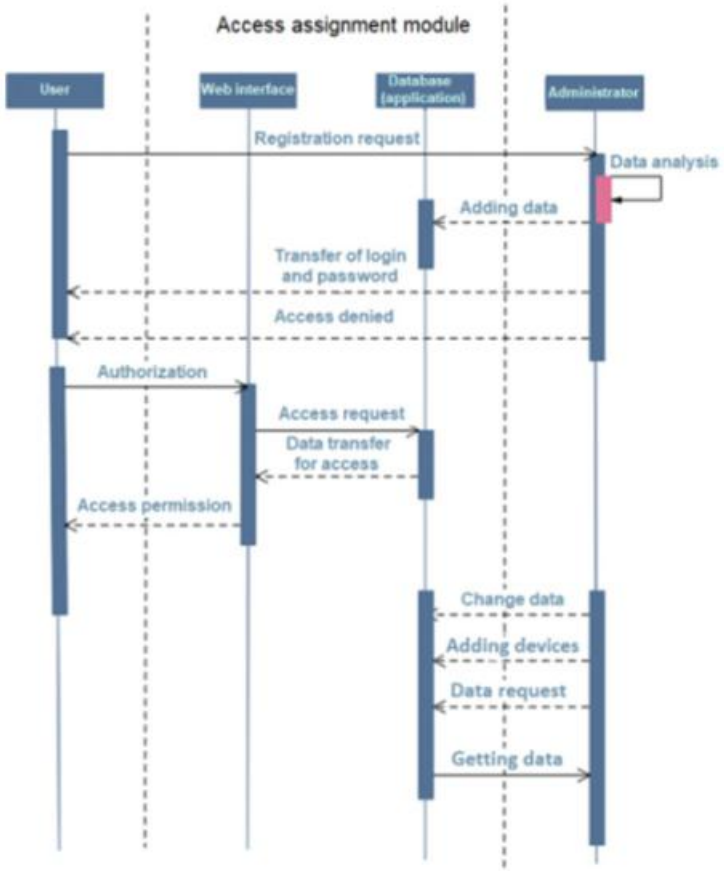
podczas wykonywania operacji wyszukiwania w bazie danych według parametrów określonych przez użytkownika, ponieważ w celu przeprowadzenia wyszukiwania nie ma potrzeby odbierania całej tablicy danych od klienta: klient przekazuje parametry żądania do serwera, a serwer przeszukuje lokalną bazę danych w poszukiwaniu otrzymanego żądania. Wynik zapytania, który zazwyczaj jest o kilka rzędów wielkości mniejszy niż cała tablica danych, jest zwracany klientowi, który wyświetla wynik użytkownikowi. Jako bazę danych wybrano MS Access. Jako język programowania i środowisko programistyczne wybrano Embarcadero RAD Studio.

4.1 Algorytm implementacji metody systemu informatycznego do przydzielania praw dostępu aplikacjom korporacyjnej sieci komórkowej

Budowa modelu systemu informatycznego rozpoczyna się od opisu funkcjonowania systemu jako całości. Rozważmy procesy zachodzące podczas interakcji w sieciach korporacyjnych. Aby uzyskać prawa dostępu, użytkownik musi skontaktować się bezpośrednio z administratorem, który po przeanalizowaniu opisu stanowiska i zakresu obowiązków użytkownika doda jego dane do modułu lub odrzuci je, jeśli praca z aplikacjami korporacyjnej sieci komórkowej nie jest wymagana. Aby zalogować się do sieci, użytkownik wprowadza swój login i hasło w interfejsie webowym modułu, po czym otrzymuje listę dostępnych aplikacji. Administrator może dodawać, usuwać, edytować i przeglądać bazę danych. Wymagania funkcjonalne opracowanej metody obejmują możliwość autoryzacji użytkownika, umożliwienie administratorowi zarządzania prawami dostępu użytkowników do różnych aplikacji, zarządzanie pamięcią masową aplikacji, rejestrowanie operacji użytkownika i prowadzenie ewidencji. Ze schematu przedstawionego na rysunku 5 wynika, że proces przydzielania praw dostępu powinien sprowadzać się do analizy wniosków użytkowników i rejestrowania ich za pomocą modułu przydzielania praw dostępu.



Na rysunku 6 przedstawiono algorytm metody przyjmowania i przetwarzania wniosków o przydzielenie praw dostępu do aplikacji w korporacyjnej sieci mobilnej.



Proponowana metoda reguluje kolejność procesu ubiegania się o dostęp do aplikacji. Użytkownik kontaktuje się bezpośrednio z administratorem, który analizuje zakres obowiązków pracownika i albo przyznaje prawa dostępu do wymaganych aplikacji w korporacyjnej sieci mobilnej, albo odmawia ich przyjęcia. Aby zalogować się do sieci, należy wprowadzić nazwę użytkownika i hasło, po czym pracownik uzyskuje dostęp do niezbędnych aplikacji. Pracą zarządza administrator, który może edytować bazę danych.

4.2 Koncepcja implementacji programowej metody przydzielania praw dostępu aplikacjom w korporacyjnej sieci mobilnej

Podczas implementacji programowej proponowanej metody przydzielania praw dostępu aplikacjom należy wziąć pod uwagę, że system musi zapewniać następujące funkcje:

• autoryzacja: podczas korzystania z modułu użytkownicy (administrator, pracownik) muszą dokonać autoryzacji w interfejsie użytkownika;

• zarządzanie użytkownikami: administrator utrzymuje, dodaje i edytuje informacje o pracownikach (login, hasło itp.);

• zarządzanie repozytorium aplikacji firmowych: administrator utrzymuje, dodaje i edytuje listę aplikacji w repozytorium;

• zarządzanie aplikacjami mobilnymi: przypisywanie uprawnień dostępu do aplikacji pracownikom firmy, danych o urządzeniach przechowywanych w bazie danych;

• prowadzenie ewidencji: rejestrowanie transakcji wykonanych przez użytkowników.

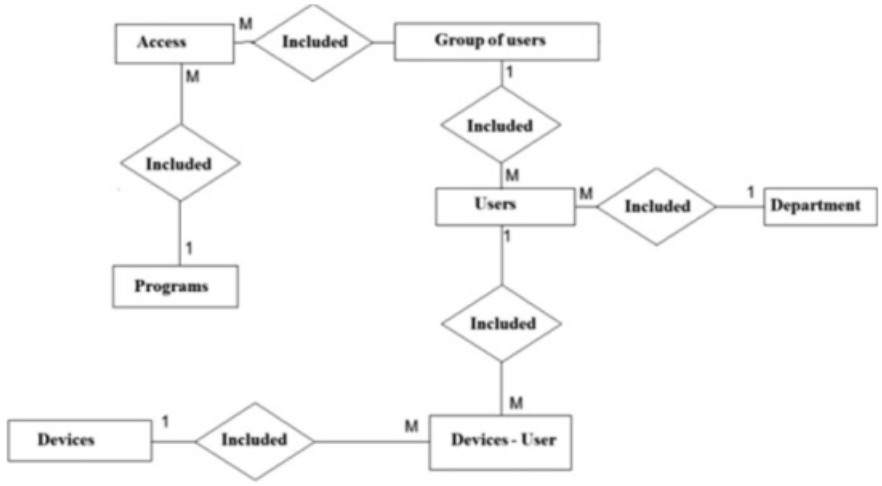
Dane wejściowe powinny zawierać informacje wprowadzone do bazy danych przez administratora:

• informacje o użytkowniku — wszystkie niezbędne informacje o użytkowniku urządzeń mobilnych i aplikacji (login, hasło itp.);

• urządzenie mobilne — wszystkie niezbędne informacje o urządzeniu mobilnym (nazwa, identyfikator, IMEI, model, ...);

• lista aplikacji i wszystkie niezbędne informacje o aplikacji (nazwa, opis, wersja, ...).

Na rysunku 7 przedstawiono model danych.



Podczas analizy obszaru tematycznego zidentyfikowano następujące encje:

• encja „Urządzenia” (kod, IMEI, nazwa urządzenia, adres IP, wersja systemu operacyjnego, data rejestracji);

• encja „Dział” (kod, nazwa działu);

• encja „Użytkownicy” (kod, nazwisko, imię, login, hasło, kod grupy, adres e-mail, komentarz, kod działu);

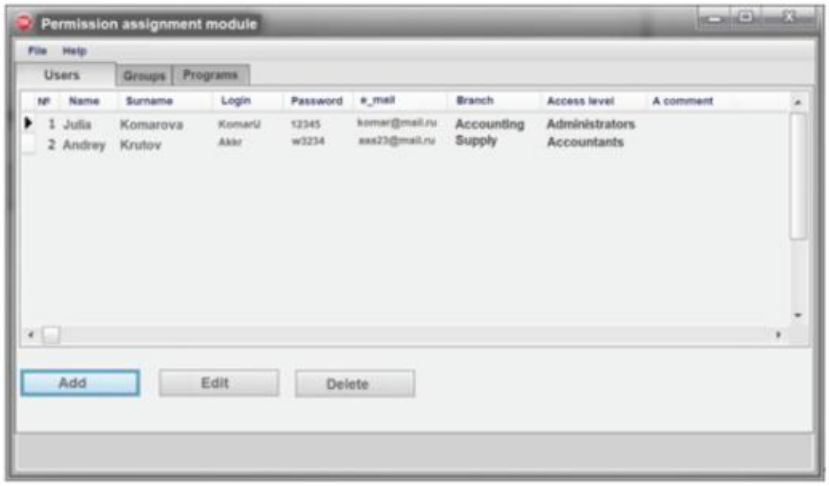
• encja „Grupy użytkowników” (kod, nazwa grupy, komentarz);

• istota „Programu” (kod, nazwa programu, link do programu, wersja programu, data dodania, opis);

• encja „Urządzenia-użytkownicy” (kod, kod urządzenia, kod użytkownika);

• encja „Dostęp” (Kod, Kod grupy, Kod programu).

Główne okno programu składa się z następujących elementów:



• zakładka do przeglądania informacji o użytkownikach;

• zakładka do przeglądania informacji o grupach użytkowników;

• zakładka do przeglądania informacji o programach.

Menu główne (rys. 8) umożliwia dodawanie nowych danych o użytkownikach, grupach roboczych, programach i używanych aplikacjach, a także daje możliwość edycji wprowadzonych danych.

5. Wnioski

W trakcie realizacji projektu przeanalizowano działania pracowników działu w zakresie przydzielania praw dostępu do aplikacji; zdefiniowano dane wejściowe, stałe i wyjściowe; rozważono funkcje i przeznaczenie poszczególnych komponentów projektowanego modułu; przeprowadzono analizę istniejących rozwiązań, przegląd istniejących systemów i wybór strategii projektowania modułu, w wyniku czego podjęto decyzję o samodzielnym opracowaniu tego modułu, bez angażowania płatnego oprogramowania; dokonano wyboru systemu zarządzania bazą danych (DBMS) i środowiska programistycznego interfejsu użytkownika; zbadano i uzasadniono podstawy decyzji projektowych, określono optymalny system zarządzania bazą danych (DBMS) – MS Access – oraz środowisko programistyczne – Embarcadero Delphi 10.2. Opisano krok po kroku projekt bazy danych, zbudowano model przepływów informacji, rozważono fizyczny schemat interakcji poszczególnych części modułu, na podstawie którego tworzona jest baza danych w wybranym systemie zarządzania bazą danych. Opracowano interfejs użytkownika z formularzami wyświetlającymi informacje przechowywane w bazie danych. Formularze umożliwiają wprowadzanie i wyprowadzanie danych, ich edycję, przetwarzanie i przechowywanie danych o użytkownikach firmowej sieci komórkowej (pracownikach firmy) oraz aplikacjach mobilnych, a następnie nadawanie pracownikom uprawnień dostępu do aplikacji.

Skalowanie sieci i przechwytywanie kluczy za pomocą połączonych systemów zestawów

Omówiono zintegrowane podejście do wykrywania bloków zawierających niezłamany klucz w bezprzewodowych sieciach sensorowych o strukturze układów kombinatorycznych oraz skalowania tych sieci. Polega ono na wykorzystaniu zarówno pojedynczych bloków układów

kombinatorycznych, jak i zbiorów bloków zawierających określony element. Takie zbiory bloków nazywane są blokami dualnymi. Jeśli elementy interpretowane są jako identyfikatory kluczy, wówczas znajomość składu bloku dualnego, w połączeniu z informacjami o rozmieszczeniu bloków, pozwala na planowanie operacji przechwytywania bloków z kluczem interesującym intruza. Z drugiej strony, zaproponowano budowę sieci o strukturze układu kombinatorycznego, składającej się z całkowicie lub częściowo połączonych podsieci odpowiadających blokom i posiadających połączenia odpowiadające blokom dualnym, zapewniając w ten sposób etap skalowania. Powstała sieć ma strukturę układu kombinatorycznego, która wraz z blokami zawiera bloki dualne; jej węzły odpowiadają parom bloków i blokom dualnym. W rozdziale podano przykłady i oszacowania skalowania podczas konwersji projektu kombinatorycznego na projekt kombinatoryczny łączony.

Słowa kluczowe: System zbiorów · Projekt kombinatoryczny · Projekt kombinatoryczny łączony · Struktura sieci kombinatorycznej · Podsieć wewnętrzna · Podsieć zewnętrzna · Selektywne przechwytywanie węzłów · Skalowanie sieci

1 Wstęp

W erze Przemysłu 4.0 pojawił się i rozwija Internet Rzeczy, oparty na powszechnym stosowaniu bezpiecznych struktur sieci czujników bezprzewodowych (WSN). Istnieje ugruntowana praktyka i technologia budowy sieci WSN, w szczególności dla „inteligentnego" domu, w oparciu o standard IEEE 802.15.4 , protokół routingu w ZigBee . Jednocześnie wykorzystywane są zasoby, które są ograniczone zarówno pod względem zużycia energii, jak i ilości wymaganej pamięci . Bezpieczeństwo takich systemów można zapewnić stosując symetryczne systemy szyfrowania oparte na wstępnej dystrybucji kluczy w węzłach sieci . Zarządzanie kluczami sieci sensorowej obejmuje następujące kroki: wstępna dystrybucja klucza, odnajdywanie klucza współdzielonego oraz określanie ścieżki dostarczania klucza (ustalanie ścieżki-klucza). Obecnie algorytmy tych etapów są realizowane z wykorzystaniem pewnych typów kombinatorycznych bloków projektowych, które są specyficznymi systemami zbiorów zdefiniowanymi przez zbiór elementów i zbiór pewnych jego podzbiorów (bloków) . Równolegle, również z wykorzystaniem systemów zbiorów, badane są struktury sieciowe. Kombinatoryczne bloki projektowe jako modele matematyczne sieci komputerowych są implementowane. Tu rozważamy problemy wykrywania bloków zawierających nieskompromitowany klucz w bezprzewodowej sieci sensorowej o kombinatorycznej strukturze projektowej oraz skalowania tych sieci. Proponujemy jednolite podejście do ich rozwiązania, wykorzystując zarówno kombinatoryczne bloki projektowe, jak i zbiory bloków zawierające wspólny element. Zbiory te nazywane są blokami dualnymi. Podczas selektywnego przechwytywania węzłów na każdym etapie atakujący zna zestaw K skompromitowanych kluczy i zestaw T numerów kluczy, które nie zostały jeszcze skompromitowane . Aby skompromitować klucz o tym numerze należącym do T, atakujący powinien przechwycić węzeł zawierający ten klucz. W ten sposób może przechwycić dowolny odpowiedni węzeł z bloku dualnego odpowiadającego temu numerowi klucza. Biorąc pod uwagę szereg znanych podejść do skalowania sieci poprzez modyfikację znanych projektów kombinatorycznych, proponujemy uzupełnienie takich struktur strukturami opartymi na blokach dualnych i ocenę powstałej struktury w porównaniu z oryginalną.

2 Niektóre projekty kombinatoryczne stosowane w WSN

W tej sekcji przypomnimy podstawowe koncepcje projektów kombinatorycznych i rozważymy przykłady projektów kombinatorycznych oraz ich znane modyfikacje proponowane w celu skalowania WSN. System zbiorów (X, B) jest określony przez skończony zbiór X elementów i zbiór

B jego podzbiorów. Odmianami systemów zbiorów obecnie używanymi do dystrybucji przedkluczowej są zrównoważone projekty bloków niekompletnych (BIBD) z parametrami (v, b, r, k), gdzie v to liczba elementów, b to liczba bloków, r to liczba bloków zawierających dowolny dany element, k to liczba elementów dowolnego bloku. Dodatkowo, piąty parametr λ jest używany jako liczba bloków, z których każdy zawiera daną parę elementów. Jeśli v = b, BIBD jest symetryczny z parametrami (v, k) i dodatkowym parametrem λ. W skrócie SBIBD. SBIBD o parametrach ($n^2 + n + 1, n + 1, 1$) nazywa się płaszczyzną rzutową (lub geometrią) rzędu n . Oznacza się ją jako PG(2, n). Znane są algorytmy konstrukcji PG(2, n) dla dowolnej potęgi liczby pierwszej n. BIBD ($n^2, n^2 + n, n, 1$) nazywa się płaszczyzną afiniczną rzędu n; jest to resztkowa BIBD uzyskana z PG(2, n) przez usunięcie dowolnego bloku i usunięcie elementów tego bloku z pozostałych bloków. Zatem z każdego PG(2, n) można uzyskać $n^2 + n + 1$ różnych płaszczyzn afinicznych. Suma wszystkich płaszczyzn afinicznych odpowiadających danym PG(2, n) tworzy zbiór bloków o zmodyfikowanej konstrukcji.

Przykład 1. Rozważmy geometrię rzutową cykliczną PG(2, 2) = (X, B) , gdzie

X = {0, 1, 2, 3, 4, 5, 6},

B = {{0, 1, 5}, {1, 2, 6}, {2, 3, 0}, {3, 4, 1}, {4, 5, 2}, {5, 6, 3}, {6, 0, 4}}.

Istnieje 21 bloków zmodyfikowanego projektu resztkowego MRPG(2, 2), odpowiadających PG(2, 2) na tym samym zbiorze X

B = {{0, 1}, {0, 2}, {0, 3}, {0, 4}, {0, 5}, {0, 6}, {1, 2},

{1, 3}, {1, 4}, {1, 5}, {1, 6}, {2, 3}, {2, 4}, {2, 5},

{2, 6}, {3, 4}, {3, 5}, {3, 6}, {4, 5}, {4, 6}, {5, 6}}.

Liczba bloków MRPG(2, n) jest zdefiniowana wzorem

$|B| = (n^2 + n + 1)(n + 1)$

(zamiast $(n^2 + n + 1)$ bloków PG(2, n)).

Następnie rozważmy poprzeczny projekt kombinatoryczny. Liniowy poprzeczny projekt kombinatoryczny TD(k, n) jest zbudowany z elementów k parami rozłącznych zbiorów w_i zawierających po n elementów każdy i składa się z n^2 zbiorów $Y_j, j = \overline{(1, n^2)}$, z których każdy zawiera k elementów. Ponadto każdy zbiór Y_j ma dokładnie jeden element z każdego zbioru w_i, i = 0, ..., k − 1, a jeśli j ≠ s, to Y j i Y s mają co najwyżej jeden wspólny element. Zatem, kombinatoryczny projekt TD(k, n) jest zdefiniowany na zbiorze X o mocy kn. Jeśli n jest liczbą pierwszą lub potęgą liczby pierwszej, to TD(k, n) istnieje. Następnie, w interpretacji algebraicznej, zbiór X jest zbiorem par (y, x), $x \in X_1$, gdzie X_1 jest zbiorem k elementów ciała F_n, $y \in F_n$; bloki liniowego poprzecznego kombinatorycznego projektu TD(k, n) są określone przez parę elementów $(a, b) \in GF(n^2)$:

$B_{a,b} = \{(ax + b, x) : x \in X_1\}$.

Do każdego bloku z TD(k, n) przypisuje się zbiór niepustych podzbiorów o niższej kardynalności uzyskany z innych bloków poprzez ekstrakcję tego bloku. Łącząc powstałe zbiory, otrzymujemy resztkowy poprzeczny kombinatoryczny projekt blokowy RTD(k, n), odpowiadający TD(k, n).

Przykład 2 Rozważmy przekrojowy projekt kombinatoryczny TD(3, 3) = (X, B), gdzie

X = {00, 01, 02, 10, 11, 12, 20, 21, 22};

B = {{00, 10, 20}, {01, 11, 22}, {02, 12, 21}, {00, 11, 21}, {01, 12, 20},

{02, 10, 22}, {00, 12, 22}, {01, 10, 21}, {02, 11, 20}}.

Zbiór bloków odpowiadającego im resztkowego poprzecznego projektu bloków kombinatorycznych RTD(3, 3) (używane są tylko bloki o tej samej randze) jest następujący:

B = {{11, 21}, {01, 12}, {02, 22}, {12, 22}, {01, 21}, {02, 11}, {00, 21}, {12, 20}, {02, 10},

{00, 12}, {10, 21}, {02, 20}, {10, 20}, {01, 22}, {02, 12}, {01, 10}, {00, 10}, {11, 22},

{02, 21}, {00, 22}, {00, 20}, {01, 11}, {12, 21}, {11, 20}, {01, 20}, {00, 11}, {10, 22}}.

Liczba bloków tak zdefiniowanego resztkowego poprzecznego kombinatorycznego projektu blokowego RTD(k, n) odpowiadająca TD(k, n) jest równa n^3 zamiast n^2 bloków początkowego TD(k, n). Należy zauważyć, że użycie kluczy geometrii rzutowej do wstępnej dystrybucji kluczy jest ograniczone przez fakt, że liczba elementów w bloku jest większa niż pierwiastek kwadratowy liczby węzłów. Jest to powód preferowania stosowania poprzecznych kombinatorycznych projektów blokowych. Aby zwiększyć odporność na kompromis kluczy, sieć jest podzielona na komórki, w których używane są różne pule kluczy. Jednocześnie uzyskuje się zmniejszenie liczby kluczy w węzłach. Geometria rzutowa jest wykorzystywana do dystrybucji kluczy w komórkach oraz do transwersalnych projektów bloków kombinatorycznych do komunikacji między komórkami.

3 Połączone projekty kombinatoryczne

Układ zbiorów (X, B) jest powiązany z połączonym układem zbiorów (X, B, D, M), w którym rozważane są dwa kolejne zbiory. Zbiór D bloków dualności jest w relacji jeden do jednego ze zbiorem X. Bloki dualności to zbiory bloków zawierające element odpowiadający danemu blokowi dualności. Zbiór główny M to zbiór wszystkich odrębnych par formy (bloku, elementu bloku).

Przykład 3 Połączona geometria rzutowa CPG(2, 2) = (X, B, D, M) odpowiadający PG(2, 2) jest zdefiniowany przez zbiory

X = {0, 1, 2, 3, 4, 5, 6};

B = {{0, 1, 5}, {1, 2, 6}, {2, 3, 0}, {3, 4, 1}, {4, 5, 2}, {5, 6, 3}, {6, 0, 4}}.

D = {{{0, 1, 5}, {2, 3, 0}, {6, 0, 4}}, {{0, 1, 5}, {1, 2, 6}, {3, 4, 1}},

{{1, 2, 6}, {2, 3, 0}, {4, 5, 2}}, {{2, 3, 0}, {3, 4, 1}, {5, 6, 3}},

{{3, 4, 1}, {4, 5, 2}, {6, 0, 4}}, {{0, 1, 5}, {4, 5, 2}, {5, 6, 3}},

{{1, 2, 6}, {5, 6, 3}, {6, 0, 4}}}.

M = {({0, 1, 5}, 0), ({0, 1, 5}, 1), ({0, 1, 5}, 5) ,

({1, 2, 6}, 1), ({1, 2, 6}, 2), ({1, 2, 6}, 6),

({3, 4, 1}, 3), ({3, 4, 1}, 4), ({3, 4, 1}, 1),

({4, 5, 2}, 4), ({4, 5, 2}, 5), ({4, 5, 2}, 2),

({5, 6, 3}, 5), ({5, 6, 3}, 6), ({5, 6, 3}, 3),

({6, 0, 4}, 6), ({6, 0, 4}, 0), ({6, 0, 4}, 4)}.

Zbiór D bloków dualności o kombinowanym zmodyfikowanym projekcie resztkowym CMRPG(2, 2)b odpowiadający zmodyfikowanemu projektowi resztkowemu MRPG(2, 2) z przykładu 1

D = {{0, 1}, {0, 2}, {0, 3}, {0, 4}, {0, 5}, {0, 6}}, {{0, 1}, {1, 2}, {1, 3}, {1, 4}, {1, 5}, {1, 6}},

{{0, 2}, {1, 2}, {2, 3}, {2, 4}, {2, 5}, {2, 6}}, {{0, 3}, {1, 3}, {2, 3}, {3, 4}, {3, 5}, {3, 6}},

{{0, 4}, {1, 4}, {2, 4}, {3, 4}, {4, 5}, {4, 6}}, {{0, 5}, {1, 5}, {2, 5}, {3, 5}, {4, 5}, {5, 6}},

{{0, 6}, {1, 6}, {2, 6}, {3, 6}, {4, 6}, {5, 6}}.

Główny zestaw tego połączonego zmodyfikowanego projektu resztowego CMRPG(2, 2) odpowiadający MRPG(2, 2) jest następujący:

M = {({0, 1}, 0), ({0, 1}, 1), ({0, 2}, 0), ({0, 2}, 2), ({0, 3}, 0) ,

({0, 3}, 3), ({0, 4}, 0), ({0, 4}, 4), ({0, 5}, 0), ({0, 5}, 5), ({0, 6}, 0),

({0, 6}, 6), ({1, 2}, 1), ({1, 2}, 2), ({1, 3}, 1), ({1, 3}, 3),

({1, 4}, 1), ({1, 4}, 4), ({1, 5}, 1), ({1, 5}, 5), ({1, 6}, 1), ({1, 6}, 6),

({2, 3}, 2), ({2, 3}, 3), ({2, 4}, 2), ({2, 4}, 4), ({2, 5}, 2), ({2, 5}, 5),

({2, 6}, 2), ({2, 6}, 6), ({3, 4}, 3), ({3, 4}, 4), ({3, 5}, 3), ({3, 5}, 5),

({3, 6}, 3), ({3, 6}, 6), ({4, 5}, 4), ({4, 5}, 5), ({4, 6}, 4),

({4, 6}, 6), ({5, 6}, 5), ({5, 6}, 6)}.

(2)

Zastosowanie do głównego zbioru zmodyfikowanego projektu resztowego CMRPG(2, n) daje poprawny wzór

$|M| = (n^2 + n + 1)(n^2 + n)$.

Przykład 4. Połączony, poprzeczny projekt kombinatoryczny CTD(3, 3) = (X, B, D, M) jest zdefiniowany przez zbiory

X = {00, 01, 02, 10, 11, 12, 20, 21, 22};

B = {{00, 10, 20}, {01, 11, 22}, {02, 12, 21}, {00, 11, 21}, {01, 12, 20},

{02, 10, 22}, {00, 12, 22}, {01, 10, 21}, {02, 11, 20}};

D = {{{00, 10, 20}, {00, 12, 22}}, {{01, 11, 22}, {01, 12, 20}},

{{02, 12, 21}, {02, 10, 22}}, {{00, 10, 20}, {02, 10, 22}},

{{01, 11, 22}, {02, 11, 20}}, {{02, 12, 21}, {00, 12, 22}},

{{00, 10, 20}, {02, 11, 20}}, {{02, 12, 21}, {01, 10, 21}},

{{02, 10, 22}, {00, 12, 22}}}.

M = {({00, 10, 20}, 00), ({00, 10, 20}, 10}), ({00, 10, 20}, 20), . . . ,

({02, 11, 20}, 02), ({02, 11, 20}, 11), ({02, 11, 20}, 20)}.

(3)

Zestaw 9 bloków dualności połączonego projektu bloku kombinatorycznego resztkowego poprzecznego CRTD(3, 3) odpowiadającego projektowi bloku kombinatorycznego resztkowego poprzecznego RTD(3, 3) z przykładu 2 to

D = {{{00, 22}, {00, 21}, {00, 12}, {00, 10}, {00, 20}, {00, 11}},

{{01, 12}, {01, 21}, {01, 22}, {01, 10}, {01, 11}, {01, 20}},

{{02, 22}, {02, 11}, {02, 10}, {02, 20}, {02, 12}, {02, 21}},

{{02, 10}, {10, 21}, {10, 20}, {01, 10}, {00, 10}, {10, 22}},

{{11, 21}, {02, 11}, {11, 22}, {01, 11}, {11, 20}, {00, 11}},

{{01, 12}, {12, 22}, {12, 20}, {00, 12}, {02, 12}, {12, 21}},

{{11, 21}, {01, 21}, {00, 21}, {10, 21}, {02, 21}, {12, 21}},

{{02, 22}, {12, 22}, {01, 22}, {11, 22}, {00, 22}, {10, 22}}}.

Główny zestaw tego połączonego resztkowego poprzecznego kombinatorycznego projektu blokowego CRTD(3, 3) składa się z 54 par

M = {({11, 21}, 11), ({11, 21}, 21), . . . , ({10, 22}, 10), ({10, 22}, 22)}.

Zastosowany do głównego zestawu połączonego zmodyfikowanego resztkowego poprzecznego kombinatorycznego projektu odpowiadającego poprzecznemu kombinatorycznemu projektowi TD(2, k, n), istnieje poprawny wzór

$|M| = kn^3$.

4 Przechwytywanie niezłamanego klucza

Aby zaimplementować atak polegający na przechwyceniu niezłamanego klucza, opisany we wstępie, konieczna jest znajomość składu bloków dualnych. Osiąga się to za pomocą algorytmów numerowania bloków, obliczania bloku o danym numerze oraz obliczania zbioru numerów bloków tworzących blok dualny odpowiadający danemu elementowi (kluczowi). Algorytmy te, zastosowane do geometrii rzutowych, tripletów Steinera oraz liniowych i kwadratowych układów kombinatorycznych poprzecznych. Przykłady bloków dualnych, które można wykorzystać do implementacji tego ataku.

5 Skalowanie w oparciu o połączone układy kombinatoryczne

W części 2 zaimplementowano znaną technikę skalowania opartą na wykorzystaniu resztkowych układów bloków kombinatorycznych. Pokażmy, że możliwe jest znaczne zwiększenie skalowania w oparciu o połączone układy kombinatoryczne. Zbudujemy sieć na zbiorze węzłów odpowiadających elementom głównego zbioru połączonego projektu bloków. Wówczas liczba węzłów będzie k razy większa niż liczba bloków w oryginalnym projekcie bloków, gdzie k jest liczbą elementów bloku. Strukturę sieci konstruujemy jako kompozycję w pełni lub częściowo połączonych sieci dwóch typów. Podsieci pierwszego typu, nazywajmy je sieciami wewnętrznymi, budowane są na zbiorze węzłów reprezentowanych przez pary o tych samych pierwszych elementach, a podsieci drugiego typu (zewnętrzne) budowane są na zbiorach węzłów o tych samych drugich elementach. Te sieci wewnętrzne odpowiadają blokom oryginalnego projektu, a podsieci zewnętrzne odpowiadają jego blokom dualności.

Przykład 5. Na zbiorze głównym (1) połączonej geometrii rzutowej CPG(2, 2) można zbudować sieć z 7 podsieciami wewnętrznymi o strukturze K3 (liczby kursywą).

{({0, 1, 5}, 0), ({0, 1, 5}, 1), ({0, 1, 5}, 5)},

{({1, 2, 6}, 1), ({1, 2, 6}, 2), ({1, 2, 6}, 6)},

{({2, 3, 0}, 2), ({2, 3, 0}, 3), ({2, 3, 0}, 0)},

{({3, 4, 1}, 3), ({3, 4, 1}, 4), ({3, 4, 1}, 1)},

{({4, 5, 2}, 4), ({4, 5, 2}, 5), ({4, 5, 2}, 2)},

{({5, 6, 3}, 5), ({5, 6, 3}, 6), ({5, 6, 3}, 3)},

{({6, 0, 4}, 6), ({6, 0, 4}, 0), ({6, 0, 4}, 4)}.

i 7 zewnętrznych podsieci o tej samej strukturze

{({0, 1, 5}, 0), ({2, 3, 0}, 0), ({6, 0, 4}, 0)},

{({0, 1, 5}, 1), ({1, 2, 6}, 1), ({3, 4, 1}, 1)},

{({1, 2, 6}, 2), ({2, 3, 0}, 2), ({4, 5, 2}, 2)},

{({2, 3, 0}, 3), ({3, 4, 1}, 3), ({5, 6, 3}, 3)},

{({3, 4, 1}, 4), ({4, 5, 2}, 4), ({6, 0, 4}, 4)},

{({0, 1, 5}, 5), ({4, 5, 2}, 5), ({5, 6, 3}, 5)},

{({1, 2, 6}, 6), ({5, 6, 3}, 6), ({6, 0, 4}, 6)}.

Przykład 6 Na głównym zestawie (2) łączonego zmodyfikowanego projektu resztkowego CMRPG(2,2) z przykładu 3 można uzyskać 21 wewnętrznych podsieci o strukturze K_2

{({0, 1}, 0), ({0, 1}, 1)}, {({0, 2}, 0), ({0, 2}, 2)},

{({0, 3}, 0), ({0, 3}, 3)}, {({0, 4}, 0), ({0, 4}, 4)},

{({0, 5}, 0), ({0, 5}, 5)}, {({0, 6}, 0), ({0, 6}, 6)},

{({1, 2}, 1), ({1, 2}, 2)}, {({1, 3}, 1), ({1, 3}, 3)},

{({1, 4}, 1), ({1, 4}, 4)}, {({1, 5}, 1), ({1, 5}, 5)},

{({1, 6}, 1), ({1, 6}, 6)}, {({2, 3}, 2), ({2, 3}, 3)},

{({2, 4}, 2), ({2, 4}, 4)}, {({2, 5}, 2), ({2, 5}, 5)},

{({2, 6}, 2), ({2, 6}, 6)}, {({3, 4}, 3), ({3, 4}, 4)},

{({3, 5}, 3), ({3, 5}, 5)}, {({3, 6}, 3), ({3, 6}, 6)},

{({4, 5}, 4), ({4, 5}, 5)}, {({4, 6}, 4), ({4, 6}, 6)},

{({5, 6}, 5), ({5, 6}, 6)}.

i 7 zewnętrznych podsieci struktury K_6

{({0, 1}, 0), ({0, 2}, 0), ({0, 3}, 0), ({0, 4}, 0), ({0, 5}, 0), ({0, 6}, 0)},

{({0, 1}, 1), ({1, 2}, 1), ({1, 3}, 1), ({1, 4}, 1), ({1, 5}, 1), ({1, 6}, 1)},

{({0, 2}, 2), ({1, 2}, 2), ({2, 3}, 2), ({2, 4}, 2), ({2, 5}, 2), ({2, 6}, 2)},

{({0, 3}, 3), ({1, 3}, 3), ({2, 3}, 3), ({3, 4}, 3), ({3, 5}, 3), ({3, 6}, 3)},

{({0, 4}, 4), ({1, 4}, 4), ({2, 4}, 4), ({3, 4}, 4), ({4, 5}, 4), ({4, 6}, 4)},

{({0, 5}, 5), ({1, 5}, 5), ({2, 5}, 5), ({3, 5}, 5), ({4, 5}, 5), ({5, 6}, 5)},

{({0, 6}, 6), ({1, 6}, 6), ({2, 6}, 6), ({3, 6}, 6), ({4, 6}, 6), ({4, 6}, 6)}.

Przykład 7 Na głównym zestawie (3) kombinowanego poprzecznego projektu kombinatorycznego CTD(3, 3) z przykładu 4 można uzyskać 9 wewnętrznych podsieci o strukturze K_3 {({00, 10, 20}, 00), ({00, 10, 20}, 10), ({00, 10, 20}, 20)},

. . . ,

{({02, 11, 20}, 02), ({02, 11, 20}, 11), ({02, 11, 20}, 20)}

oraz 9 zewnętrznych podsieci o tej samej strukturze

{({00, 10, 20}, 00), ({02, 10, 22}, 00), ({00, 12, 22}, 00)},

. . . ,

{({01, 11, 22}, 22), ({02, 11, 20}22,), ({00, 12, 22}, 22)}.

Przykład 8 Na głównym zbiorze (4) kombinowanego zmodyfikowanego resztkowego poprzecznego projektu kombinatorycznego CRTD(3, 3) z przykładu 4 można uzyskać 27 wewnętrznych podsieci o strukturze K_2

{({11, 21}, 11), ({11, 21}, 21)}, {({01, 12}, 01), ({01, 12}, 12)}, . . .

. . . , {({00, 11}, 00), ({00, 11}, 11)}, {{10, 22}, 10), ({10, 22}, 22)}

i 9 zewnętrznych podsieci struktury K_6

{({00, 21}, 00), ({00, 12}, 00), ({00, 10}, 00), ({00, 22}, 00), ({00, 20}, 00), ({00, 11}, 00)},

{({01, 12}, 01), ({01, 21}, 01), ({01, 22}, 01), ({01, 10}, 01), ({01, 11}, 01), ({01, 20}, 01)},

{({02, 22}, 02), ({02, 11}, 02), ({02, 10}, 02), ({02, 20}, 02), ({02, 12}, 02), ({02, 21}, 02)},

{({02, 10}, 10), ({10, 21}, 10), ({10, 20}, 10), ({01, 10}, 10), ({00, 10}, 10), ({10, 22}, 10)},

{({11, 21}, 11), ({02, 11}, 11), ({11, 22}, 11), ({01, 11}, 11), ({11, 20}, 11), ({00, 11}, 11)},

{({01, 12}, 12), ({12, 22}, 12), ({12, 20}, 12), ({00, 12}, 12), ({02, 12}, 12), ({12, 21}, 12)},

{({12, 20}, 20), ({02, 20}, 20), ({10, 20}, 20), ({00, 20}, 20), ({11, 20}, 20), ({01, 20}, 20)},

{({11, 21}, 21), ({01, 21}, 21), ({00, 21}, 21), ({10, 21}, 21), ({02, 21}, 21), ({12, 21}, 21)},

{({02, 22}, 22), ({12, 22}, 22), ({01, 22}, 22), ({11, 22}, 22), ({00, 22}, 22), ({10, 22}, 22)}.

Struktury sieciowe opisane w tej sekcji mają ważne właściwości:

– zbiory węzłów podsieci wewnętrznych tworzą podział zbioru wszystkich węzłów na podzbiory o tej samej kardynalności;

– podobnie zbiory węzłów podsieci zewnętrznych również tworzą podział zbioru wszystkich węzłów na podzbiory o tej samej kardynalności;

– każdy element sieci jest elementem pewnej podsieci wewnętrznej i pewnej podsieci zewnętrznej.

Z perspektywy wstępnej dystrybucji kluczy, właściwości te umożliwiają różnym pulom kluczy i różnym modelom kombinatorycznym dystrybucję kluczy w podsieciach. Na przykład, mniejsze podsieci mogą korzystać z płaszczyzn rzutowych, a wyższe z poprzecznych projektów kombinatorycznych . Każdy węzeł używa dwóch zestawów kluczy do komunikacji w swoich odpowiednich podsieciach wewnętrznych i zewnętrznych. Pozwala to zmniejszyć ilość pamięci wymaganej do przechowywania kluczy, a także zapewnia ochronę przed ich kompromitacją. Na tym etapie skalowania węzły korzystają z pul kluczy o tej samej pojemności, a pule kluczy są chronione przed wzajemną kompromitacją.

6 Wnioski

Tu rozważamy dwa ważne problemy praktyczne związane z przechwytywaniem nieskompromitowanego klucza i skalowaniem sieci komputerowej, której struktura odpowiada projektowi kombinatorycznemu. W pierwszym z nich wyznaczany jest zbiór węzłów zawierający klucz interesujący najeźdźcę. W drugim wykorzystywane są w pełni lub częściowo połączone podsieci, których skład jest określony przez bloki konstrukcji kombinatorycznej. Połączenia wewnętrzne uzupełniane są połączeniami zewnętrznymi komputerów odpowiadającymi tym samym elementom bloków. Algorytmicznie rozwiązanie osiąga się w oparciu o koncepcję połączonej konstrukcji bloków kombinatorycznych, gdy zestaw bloków dualnych jest używany wraz z zestawem bloków. Umożliwiło to, poprzez skalowanie zbiorów elementów dwóch znanych schematów kombinatorycznych, zbudowanie nowych, struktur sieciowych o większej skali, które mają właściwości charakterystyczne dla praktyki wstępnej dystrybucji kluczy. Uzyskaliśmy również liczbę węzłów w takich strukturach sieciowych i ich podsieciach odpowiadających blokom i blokom dualnym.

Podsystem wideokonferencji bezpiecznego internetu bezserwerowego (TheOoL.Net)

W artykule omówiono warunki rozwoju prywatnego internetu bezserwerowego z wbudowanym, superbezpiecznym podsystemem komunikacji wideo i audio, umożliwiającym prowadzenie konferencji z udziałem trzech lub więcej uczestników. Przedstawiono krótki opis zasad leżących u podstaw prywatnej architektury internetu bezserwerowego, która oprócz protokołu przesyłania danych nakładkowych obejmuje hosting bezserwerowy, przetwarzanie w chmurze, bezserwerową wyszukiwarkę oraz zintegrowany system płatności. Przedstawiono uproszczony model graficzny architektury i działania bezserwerowego, superbezpiecznego podsystemu komunikacji wideo i audio. Omówiono podstawowe tryby jego działania oraz metody zapewnienia bezpieczeństwa i prywatności rozmów i korespondencji.

Słowa kluczowe Synteza parametryczna · Układ automatycznego sterowania · Regulator PID · Metoda modalna · Bieguny · Charakterystyka przejściowa

1 Wstęp

Ostatnie procesy łączenia wirtualnej przestrzeni internetu, odzwierciedlającej przestrzeń mentalną jednostek i społeczeństwa , nie tylko z globalnym systemem finansowym i systemem

zarządzania społecznego, ale także poprzez IoT (Internet Rzeczy) i IoE (Internet Wszystkiego), z systemami sterowania obiektami technicznymi w świecie fizycznym, prowadzą do łączenia przestrzeni wirtualnej i rzeczywistej . Było to szczególnie widoczne od 2020 roku dzięki pandemii COVID-19 i globalnemu lockdownowi, które znacznie przyspieszyły proces tworzenia zdalnych miejsc pracy i ich automatyzacji . W rezultacie nastąpił gwałtowny wzrost zapotrzebowania na usługi komunikacji internetowej i bezpieczny zdalny dostęp do korporacyjnych systemów informatycznych, a także znacznie wzrosło zapotrzebowanie na bezpieczeństwo informacji. W rzeczywistości rynek bezpieczeństwa informacji rośnie nieprzerwanie od co najmniej dwudziestu lat, średniorocznie o 7%, co stanowiło ekonomiczne uzasadnienie rozwoju bezpiecznego internetu bezserwerowego (projekt TheOoL) i bezserwerowego bezpiecznego systemu wideokonferencji jako jego integralnej części. Technicznymi warunkami wstępnymi dla projektu były, po pierwsze, inherentne podatności sieci opartych na stosie protokołów TCP/IP, które początkowo były otwarte i zdecentralizowane i które oferują szeroki wachlarz wbudowanych możliwości dla web intelligence na trzeciej do siódmej warstwie modelu OSI/ISO implementacji i pozyskiwania informacji o systemach operacyjnych i aplikacjach działających na określonych węzłach sieciowych i ich sprzęcie w celu identyfikacji potencjalnie podatnych komponentów. Po drugie, ich niespójność z architekturą większości współczesnych systemów przechowywania i przetwarzania danych, które są projektowane i konstruowane jako systemy scentralizowane .Oczywiste jest, że posiadanie informacji o składzie programowym i sprzętowym węzłów sieciowych oraz topologii sieci pociąga za sobą znaczne ryzyko dla bezpieczeństwa informacji , a zagrożenia te materializują się z zadziwiającą regularnością . Kilkadziesiąt firm i nieformalnych zespołów programistycznych pracuje nad rozwiązaniem problemów bezpieczeństwa sieci w ramach koncepcji Web 2.0 . Naszym zdaniem twórcy projektu „niewidzialnego internetu" I2P [24] opracowali najlepsze rozwiązanie dla rozwoju prywatnej, nakładkowej sieci ruchu internetowego, która pomimo szeregu znanych luk zapewnia wysoki poziom bezpieczeństwa ruchu internetowego, w przeciwieństwie do jej bezpośredniego konkurenta – projektu TOR , którego luki pozwalają rozwiązać problem naruszenia bezpieczeństwa połączenia przy użyciu ograniczonych zasobów . Pomimo powszechnej atrakcyjności pracy zdalnej, która jest raczej niemożliwa bez wideokonferencji, praktycznie nie ma publicznych systemów do prywatnej komunikacji wideo. Istniejące usługi komercyjne są projektowane w sposób jednolity, z wykorzystaniem scentralizowanej architektury z pełnym deszyfrowaniem ruchu po stronie serwera , co samo w sobie stanowi lukę z punktu widzenia klienta, zwłaszcza w przypadku poufnych rozmów. W efekcie właściciele najpopularniejszych systemów, w tym ZOOM, Skype itp., a także ich pracownicy mają pełny dostęp do rozmów swoich klientów. Uważamy, że drugim źródłem zagrożenia dla prywatności klientów w tego typu systemach wideokonferencyjnych są scentralizowane usługi generowania i zarządzania używanymi przez nich kluczami kryptograficznymi i certyfikatami. Problem polega na tym, że oprócz klientów uczestniczących w rozmowach, właściciele i zespoły konserwacyjne takich usług kryptograficznych, w różnym stopniu, posiadają certyfikaty i klucze niezbędne do deszyfrowania ruchu wideokonferencyjnego, co znacznie zwiększa prawdopodobieństwo przechwycenia takiej rozmowy przez odpowiednich właścicieli lub personel konserwacyjny. Ogólnie rzecz biorąc, nasza analiza obecnie dostępnych rozwiązań dla sieci prywatnych ujawniła szereg wspólnych problemów: Powolny dostęp do treści w porównaniu z „otwartym internetem"; brak zaufanego hostingu prywatnego, co oznacza, że dla ich chronionych zasobów konieczne jest korzystanie z serwerów i łączy danych o niższej przepustowości; brak zaufanych protokołów do konfigurowania katalogów wyszukiwania, co skutkuje korzystaniem ze scentralizowanych usług, w których rejestracja wiąże się z ujawnieniem informacji o topologii sieci i lokalizacji geograficznej serwerów; problem bezpieczeństwa komunikacji audio i wideo nie został rozwiązany w momencie pisania tego artykułu. Naszym

zdaniem rozwiązanie powyższych problemów leży w przejściu z paradygmatu Web 2.0 do paradygmatu Web 3/Web 3.0 , tj. przejściu od zorientowanej topologicznie, scentralizowanej architektury usług do jej całkowitej decentralizacji. W tym paradygmacie wdrożenie tego systemu komunikacji wideo i audio stanowi osobne wyzwanie, ponieważ wymaga podsystemu o wysokiej wydajności i bezpieczeństwie dla przetwarzania rozproszonego, złożonej i efektywnej czasowo kryptografii, a także zaufanego systemu kierowania i routingu ruchu. Jednak wdrożony podsystem umożliwiłby ograniczenie dostępu do kluczy kryptograficznych i certyfikatów dla uczestników poszczególnych wideokonferencji, tym samym znacznie zmniejszając prawdopodobieństwo przechwycenia rozmów poprzez wyeliminowanie wszystkich innych stron z obsługi takiej komunikacji wideo i audio, z wyjątkiem bezpośrednich uczestników rozmów. Zastosowanie rozproszonego systemu obliczeniowego umożliwia dalsze zwiększenie prywatności rozmów dzięki planowanej migracji komponentów zdecentralizowanego serwera wideo do losowo wybranych węzłów sieciowych.

2 Zdecentralizowany Internet według TheOoL

Ponieważ prywatny system komunikacji wideo i audio („SoPVAC") nie jest samodzielnym produktem, lecz podsystemem w ekosystemie prywatnego zdecentralizowanego Internetu TheOoL , trudno opisać ten system bez krótkiego omówienia architektury i funkcjonalności ekosystemu TheOoL. TheOoL to wariant implementacji Internetu w paradygmacie Web3 z elementami Web 3.0 i koncentruje się przede wszystkim na zapewnieniu prywatności członkom sieci, co oznacza możliwość całkowitego ukrycia cyfrowego śladu lub wygenerowania fałszywego śladu pod pełną kontrolą użytkownika, całkowite i wyłączne posiadanie przez użytkownika wszystkich kluczy kryptograficznych oraz wyłączne prawo użytkownika do zarządzania wszelkimi danymi generowanymi w trakcie jego aktywności. W sieci TheOoL nie ma właścicieli serwerów ani właścicieli usług; Jest to całkowicie bezserwerowy, rozproszony system peer-to-peer zarządzany przez algorytmy, które implementują procesy obsługi i dostarczania zasobów sieciowych w formie zdecentralizowanej autonomicznej organizacji, czyli DAO . Z technicznego punktu widzenia TheOoL to rozproszona geograficznie, zaawansowana sieć kolejek automatycznych. Na poziomie logicznym jest to chmura danych i obliczeń oparta na prywatnym i bezpiecznym zdecentralizowanym ekosystemie („PSDE") , oparta na następujących zasadach:

• Bezpośrednia interakcja użytkownika z informacjami o użytkowniku za pośrednictwem pliku karty informacyjnej z kolejnością w hierarchii.

• Solidność informacji o użytkowniku i niemożność ich rozpowszechniania, modyfikacji lub usunięcia bez zgody właściciela informacji lub osób przez niego upoważnionych.

• Dostępność informacji o użytkowniku za pośrednictwem zautomatyzowanego, zdecentralizowanego pliku karty informacyjnej.

• Kontrolowalność informacji o użytkowniku, co oznacza całkowitą kontrolę nad cyfrowym śladem użytkownika po jego stronie i brak możliwości pozyskania danych przez kogokolwiek innego.

• Przechowywanie informacji o użytkowniku „wszędzie i nigdzie". Oznacza to maksymalną decentralizację i fragmentację między węzłami wykonującymi dane użytkownika a wykonywanymi algorytmami.

• Abstrahowanie użytkowników od metadanych, informacji technicznych, topologii lub procesów operacyjnych sieci.

• Ukrywanie metadanych i informacji technicznych przed użytkownikami.

Oprogramowanie TN TheOoL jest takie samo, a jego znaczenie w danym momencie jest automatycznie definiowane przez algorytmy PSDE w zależności od wymagań sieci. TN może jednocześnie pełnić kilka ról (Tabela 1). Oprogramowanie TN TheOoL jest takie samo, a jego znaczenie w danym momencie jest automatycznie definiowane przez algorytmy PSDE w zależności od wymagań sieci. TN może jednocześnie pełnić kilka ról. Architektura oprogramowania węzła PSDE obejmuje cztery (4) warstwy. Komponenty są rozmieszczone w węzłach systemu zgodnie z mapą lokalizacji komponentów programowych czterowarstwowych architektur PSDE. Komponenty programowe są rozproszone, umożliwiając obsługę funkcji użytkownika PSDE oraz sterowanie i zarządzanie sprzętem TN po stronie ich właścicieli. Zapewniono realizację głównych założeń rozwoju PSDE, a mianowicie całkowite ukrycie metadanych i topologii przechowywania danych oraz w pełni zautomatyzowaną konserwację oprogramowania PSDE i obsługiwanych danych. Silnikiem bezserwerowego systemu bezpieczeństwa i zarządzania PSDE jest rozproszony rejestr TheOoL (blockchain). Zawiera on dwie powiązane ze sobą rozproszone bazy danych: deskryptor bezpieczeństwa bazy danych dla bezserwerowego silnika bezpieczeństwa danych, który zapewnia kontrolę nad dostępem do danych; bazę danych PS oraz bazę danych dla inteligentnych kontraktów — skryptów, które zarządzają interakcją użytkownika z PSDE. Każde zadanie otrzymywane przez PSDE od użytkowników jest inteligentnym kontraktem (specjalnym skryptem), który definiuje role i nagrody dla określonych węzłów PSDE w zakresie tego zadania. Zgodnie z tym skryptem główne zadanie użytkownika jest podzielone na zestaw mniejszych zadań procesowych („zadanie techniczne”), którym przypisano wykonawcę węzła i nagrody za wykonanie każdego zadania. W świecie rzeczywistym nagrody są przyznawane właścicielom odpowiednich węzłów sprzętowych, którzy nie mają dostępu do procesu wykonania, a także do wszelkich informacji o zadaniu. Sieć MN obsługuje rozproszoną bazę danych rejestru. Sieć ta zawiera nie tylko bloki i skrypty inteligentnych kontraktów, ale jest również odpowiedzialna za zarządzanie adresami danych i ich konwersję z przestrzeni adresowej pewnego podsystemu do przestrzeni adresowej innego podsystemu. W rzeczywistości adresy danych w podsystemach architektury czwartej warstwy różnią się od ich prezentacji w podsystemie architektury trzeciej warstwy. Ma to na celu usprawnienie implementacji szóstej i siódmej zasady PSDE. Sieć MN pełni tutaj rolę bezpiecznego transkrybenta, który zapewnia maksymalną abstrakcję i ukrycie topologii przechowywania i przetwarzania danych przed dowolnym członkiem sieci. Do pewnego stopnia MN reprezentuje punkty podatności z punktu widzenia możliwości naruszenia wyżej wymienionych zasad. Dlatego przypisanie dowolnej sieci TN funkcji MN odbywa się w sposób losowy; MN migruje w nieskończoność do sieci TN; Dane na MN są szyfrowane, podczas gdy same mapowania adresów są przechowywane na różnych MN, tj. szyfrowane łącze SDSE i odpowiadające mu szyfrowane łącze DT&SP są przechowywane na różnych MN, a te MN są, z dużym prawdopodobieństwem, własnością różnych osób lub podmiotów w świecie rzeczywistym. Sieć DLCN została skonfigurowana w podobny sposób i z tego samego powodu; w tej sieci wykonywane są algorytmy konsensusu odpowiedzialne za zapewnienie legalności modyfikacji danych w rozproszonych rejestrach. Aby zastosować i zmienić deskryptory bezpieczeństwa, używamy bezpiecznego i niezawodnego algorytmu konsensusu Proof Of Work (PoW), który jest modyfikowany w celu wyeliminowania konkurencyjnego wzrostu złożoności obliczeniowej, co jest typowe dla innych sieci PoW i co eliminuje nadmierne zużycie energii wymagane do obsługi konsensusu w sieci TheOoL. Jedna z implementacji algorytmu Proof of Stake (PoS) [39, 40] została wybrana jako algorytm konsensusu dla rozproszonego rejestru inteligentnych kontraktów w płatnościach. Nie należy jednak rozumieć, że w sieci TheOoL stosowane są dwa oddzielne algorytmy konsensusu.

W rzeczywistości używamy hybrydy PoW i PoS, która działa jako nierozłączny mechanizm zapewniający podwójną walidację najważniejszych wpisów w rozproszonym rejestrze, a mianowicie wpisów dotyczących płatności i zmian w deskryptorach bezpieczeństwa. Metadane techniczne (wpisy dotyczące fizycznej lokalizacji części danych lub węzłów wykonawczych zadań w ramach przetwarzania rozproszonego) są obsługiwane wyłącznie przez PoW. Podwójna walidacja jest wymagana w celu znacznego zmniejszenia prawdopodobieństwa powodzenia ataków na rozproszony rejestr . Szyfrowanie i generowane komputerowo podpisy w rozproszonym rejestrze TheOoL są wykonywane przy użyciu sprawdzonych (na przykład w kryptografii sieci Bitcoin) krzywych eliptycznych (ECC) , w tym krzywej eliptycznej SECP521R1.

3. Bezserwerowy system wideokonferencji

Aby zaimplementować SoPVAC , używamy komponentów 1.1 i 2.2 w sieci AN oraz komponentów 3.1, 3.3 i 4.3 w sieci TN. W sieci AN (UI) interfejs użytkownika SoPVAC jest uruchamiany i współpracuje z grupą serwerów wirtualnych – routerów SoPVAC, które są automatycznie tworzone i uruchamiane w sieci TN PSDE dla każdej konkretnej wideokonferencji w ramach jednego konkretnego inteligentnego kontraktu. Sieć TN można przedstawić jako wiele węzłów:

$$S_t \equiv S_m \cup S_d \cup S_s \cup S_{ss} \equiv \{X_1 \dots X_i \dots X_\xi\}, X_i = \{a, b, c\}, \quad i = \overline{1 \dots \xi} \quad (1)$$

gdzie S_m — mnogość MN, S_d — mnogość DLCN, S_s — mnogość SCN, S_{ss} — mnogość SSN, ξ — całkowita liczba TN w PSDE, X_i — mnogość parametrów określonego TN , które są niezbędne do działania SoPVAC i które obejmują a — liczbę białych spacji w kolejce, b — indeks wydajności węzła i c — dostępną przestrzeń w podsystemie dyskowym węzła. Zdecentralizowane aplikacje lub DApps odpowiedzialne za działanie systemu wyszukiwania („SDSE") i systemu przechowywania danych („DT&SP") zapewniają przesyłanie na AN listy kontaktów i bieżącej wersji lokalnej aplikacji SoPVAC („AN UI"), która współpracuje poprzez UWFC (Tabela 2) z wirtualnym routerem wideokonferencji i SoPVAC DApp („CDApp"). W celu obsługi SoPVAC utrzymywane są 2 indeksy dla każdego użytkownika — lista subskrybentów w SDSE: A — lista abstrakcyjnych subskrybentów (AA) zarejestrowanych jako użytkownicy SoPVAC; A_o — lista AA o statusie „online" i „dostępny". Zdefiniujmy zbiory A i A_o:

$$A = \{a_1 \dots .a_n\}, A_o = \{a_1 \dots a_k\}, a_i = \{id_i, t_i, Q_i^s, t_i^{max}\} \quad (2)$$

gdzie a_i — abonent na liście przypisany przez czas życia t_i, t^{max}_i — maksymalny czas przechowywania, ustalony klucz publiczny ECC Q^s_i oraz unikalny identyfikator id_i, a zakres zmian indeksu i jest zdefiniowany przez przynależność do jednej lub drugiej listy:

$$i = \overline{1 \dots n} | a_i \in A, i = \overline{1 \dots k} | a_i \in A_0$$

Węzeł interfejsu użytkownika zawiera wielu subskrybentów zarejestrowanych w węźle, których definiujemy jako:

$$U = \{u_1 \dots u_r\}, u_j = \{id_j, d_j^s, Q_j^s\}, Q_j^s = d_j^s \cdot G, j = \overline{1 \dots r} \quad (3)$$

gdzie id_j — unikalny identyfikator abonenta, d^s_j — stały klucz prywatny abonenta, $Q^s{}_j$ — stały klucz publiczny abonenta, G — punkt zakotwiczenia krzywej eliptycznej wspólny dla sieci PSDE, r — liczba abonentów obsługiwanych przez konkretny interfejs użytkownika sieci AN. Za zarządzanie pierwszą listą odpowiadają tylko interfejsy SDSE i AN. Dla każdego abonenta wykonywana jest następująca funkcjonalność:

$$\left\{t_j = 0 | id_j \in A\right\}, \left\{f : \mathrm{U} \to \left\{a = \left\{id_j, t_j = 0, Q_j^s, t_j^{max}\right\} \in A\right\} | id_j \notin A\right\}, j = \overline{1 \ldots r}$$

w mnogościach (2, 3). Mówiąc inaczej, po uruchomieniu interfejsu użytkownika AN węzeł sprawdza dostępność wpisów o u j w mnogości A zarejestrowanej w węźle. Jeśli subskrybenci są zarejestrowani, interfejs użytkownika AN wysyła komunikat do SDSE o wyczyszczeniu czasu od ostatniego połączenia. W przeciwnym razie w indeksach SDSE uruchamiana jest procedura rejestracji nowego AA. Podczas rejestracji użytkownik ustawia maksymalną wartość limitu czasu połączenia. Ten parametr definiuje okres, w którym SDSE przechowuje informacje o zarejestrowanym AA i powiązane z nim dane w pamięci masowej w chmurze. Po osiągnięciu tej wartości wszystkie informacje są automatycznie usuwane, a kolejne połączenia AA są nawiązywane jako nowe. Ponieważ druga lista wymaga stałego monitorowania aktywności subskrybentów, jest ona zarządzana — oprócz SDSE i interfejsu użytkownika AN — przez procesy obliczeniowe uruchamiane w ramach CDApp. Interfejs użytkownika (UI) aktywowany przez użytkownika łączy się ze znanymi mu modułami CDApp w sieci MN, a następnie aktualizuje listę dostępnych sieci MN i powiadamia moduł CDApp w sieci MN o zmianie statusu AA zarejestrowanych w nim na „online" i „available", określając grupy użytkowników, które będą miały dostęp do tych informacji. CDApp synchronizuje informacje o aktywacji subskrybenta za pośrednictwem indeksów SDSE, które są publicznie dostępne w odpowiednich grupach lub w publicznym indeksie SDSE, jeśli subskrybent musi komunikować się z szerokim gronem nieznanych mu kontrahentów. Wpis w SDSE pozostaje aktywny przez 2 minuty, po czym SDSE automatycznie go usuwa, jeśli jego czas ważności nie zostanie przedłużony, a status abstrakcyjnych subskrybentów zarejestrowanych w odpowiednim interfejsie użytkownika sieci AN zostaje przekonwertowany na „offline". Następnie CDApp przypisuje dwa monitorujące sieci MN (CMN) z (1) do węzła z interfejsem użytkownika sieci AN, których zadaniem jest przedłużenie czasu ważności wpisu subskrybenta w SDSE:

$$f : S_m \to \left\{\{CMN_1, CMN_2\} | \max_{a,b} S_m\right\}$$

Wpis dotyczący numeru CMN przypisanego abonentowi jest również wprowadzany do SDSE. Jest to konieczne, ponieważ to właśnie numer CMN zapewnia komunikację abonenta z innymi członkami sieci SoPVAC. Dla każdego n abonenta w zbiorach (2, 3) realizowana jest następująca funkcjonalność:

$$f : U \to \left\{a = \left\{id_j, t_j = 0, Q_j^s, t_j^{max} = 2, \{CMN_1, CMN_2\}\right\} \in A_o\right\} | id_j \notin A_o,$$
$$j = \overline{1 \ldots k}$$

CMN uzyskuje oznaczenie abonenta raz na minutę i, jeśli zostanie nawiązane połączenie z odpowiednim interfejsem użytkownika AN, wydłuża czas życia wpisu o aktywności abstrakcyjnych abonentów w indeksach SDSE. Ponadto CMN w żadnym wypadku nie odbiera ani nie przetwarza strumieni audio i wideo ani żadnych wiadomości tekstowych komunikatora. Cele CMN obejmują nawiązanie połączenia początkowego, inicjalizację połączenia oraz inicjalizację wirtualnych routerów wideokonferencji. Chociaż CMN, z konieczności, posiada informacje o identyfikatorze klienta na warstwie transportowej modelu ISO/OSI, nie przewiduje możliwości przejęcia kontroli nad użytkownikiem i deanonimizacji, ponieważ sieć TheOoL wykorzystuje I2P jako warstwę transportową. CMN w tym przypadku nie zna identyfikatora I2P klienta, lecz router tunelu identyfikatorów I2P [24], który jest najbliższy CMN. Router ten jest zazwyczaj węzłem sieci I2P, który po pierwsze, niekoniecznie jest węzłem TheOoL, a po drugie, jest automatycznie zastępowany co 10 minut (wartość domyślna). W związku z tym numery CMN pozostają

niezmienione dla abonenta tak długo, jak długo są aktywne, z wyjątkiem sytuacji, gdy działanie CMN zostanie zakończone. W takim przypadku możliwe są dwa następujące scenariusze:

1. Dostępny jest co najmniej jeden (1) numer CMN. W takim przypadku PSDE automatycznie przydziela nowy numer CMN w miejsce uszkodzonego.

2. Wszystkie numery CMN są wyłączone. W takim przypadku klient interfejsu użytkownika AN inicjuje algorytm połączenia początkowego, odbiera nową parę numerów CMN i generuje nowy wpis o dostępności.

Aby utrzymać dostępność SoPVAC, wystarczy mieć tylko jeden numer CMN. Do systemu wprowadzany jest dodatkowy w celu zwiększenia jego odporności na błędy. Po uruchomieniu interfejsu użytkownika AN i przydzieleniu numeru CMN użytkownik jest gotowy do wysyłania wiadomości tekstowych do użytkowników za pośrednictwem lokalnego komunikatora w trybie p2p z listy dopuszczonych w indeksie SDSE w statusie „dostępny", a użytkownik może wykonywać z nimi połączenia głosowe i wideo. To powiedziawszy, żadne dane połączeń ani konwersacji nie mogą być zapisywane, nawet lokalnie, a także nie ma możliwości opóźnionej komunikacji, na przykład dla użytkownika wysyłającego wiadomości tekstowe w trybie „offline". Implementacja protokołu RTP (Real-time Transport Protocol) w trybie p2p jest odpowiedzialna za przetwarzanie multimediów w SoPVAC. Protokół ten jest szeroko stosowany do przesyłania plików multimedialnych w trybie „online"; jest dobrze udokumentowany i nie posiada niezbędnych funkcji dla aplikacji w SoPVAC. Dla porównania, procedura inicjalizacji wideokonferencji, przesyłania danych i przetwarzania danych powinna zostać opisana bardziej szczegółowo. Aby rozpocząć wideokonferencję, abonent, który ją inicjuje, wybiera innych uczestników wideokonferencji z listy dostępnych abonentów i wysyła im żądania połączenia („żądanie uzgadniania"). SoPVAC wykonuje procedurę sprawdzania dostępności abonentów za pośrednictwem przydzielonego im numeru CMN. Następnie abonenci z potwierdzonym statusem aktywnym otrzymują żądanie uzgadniania. Jeśli zaakceptują żądanie, wykonywana jest procedura generowania i przesyłania odpowiednich kluczy szyfrowania sesji przy użyciu protokołu punkt-punkt opartego na algorytmie Diffiego-Hellmana (uzgadnianie). Protokół ECC jest używany jako protokół asymetryczny wraz z generowaniem efemerycznych par kluczy do obliczania sekretów i par kluczy stanu stacjonarnego zarejestrowanych dla AA w SDSE, aby zapewnić działanie algorytmu podpisu cyfrowego. Do szyfrowania ruchu, za pomocą odpowiedniego klucza szyfrowania sesji, stosujemy Advanced Encryption Standard (AES) . Wybraliśmy ten standard ze względu na jego wysoką odporność kryptograficzną i jednocześnie niską złożoność obliczeniową algorytmu, co jest szczególnie ważne podczas przetwarzania strumieni multimedialnych w trybie online. W rzeczywistości oznacza to, że AES działa wystarczająco szybko, aby wykluczyć jakiekolwiek znaczące opóźnienia w transmisji danych audio i wideo, dzięki wykonywaniu procedur szyfrowania/deszyfrowania. Mimo to w momencie pisania tego artykułu znany jest tylko jeden algorytm implementacji ataku AES – tzw. atak algebraiczny opisany przez Nielsa Fergusona w 2001 roku. Atak ten jednak nigdy nie został faktycznie przeprowadzony. Jego nieważność została udowodniona na późniejszym etapie [48]. Inne ataki są związane z konkretnymi implementacjami AES. Są one w większości złożone i trudno je wdrożyć w praktyce. Krotka 10 kluczy sesji jest generowana podczas uzgadniania dla każdego uczestnika wideokonferencji; klucze te są używane pojedynczo podczas sesji komunikacyjnej do szyfrowania AES ruchu multimedialnego i wiadomości tekstowych wysyłanych przez komunikator. Krotki kluczy każdego uczestnika, zaszyfrowane wspólnym sekretem ECC, są przekazywane pozostałym uczestnikom wideokonferencji. Interfejs użytkownika gospodarza wideokonferencji losowo definiuje numer

pozycji klucza szyfrującego w krotce, który jest wspólny dla wszystkich uczestników i jest im przekazywany. W ciągu 5 minut uczestnicy używają klucza do szyfrowania swojego ruchu wychodzącego, który znajduje się na obliczonej w ten sposób pozycji w wygenerowanej przez nich krotce kluczy. Po upływie tego czasu interfejs użytkownika gospodarza wideokonferencji generuje i przesyła nowy numer pozycji oraz przekazuje go pozostałym uczestnikom wideokonferencji. W sieci OoL przesyłane są zatem strumienie danych zaszyfrowane kluczami, które są różne dla każdego uczestnika wideokonferencji, a klucze te są losowo wymieniane co 5 minut. Interfejs użytkownika gospodarza wideokonferencji losowo definiuje numer pozycji klucza szyfrującego w krotce, która jest wspólna dla wszystkich uczestników i jest im udostępniana. W ciągu 5 minut uczestnicy używają klucza do szyfrowania swojego ruchu wychodzącego, który znajduje się na obliczonej w ten sposób pozycji w wygenerowanej przez nich krotce kluczy. Po upływie tego czasu interfejs użytkownika gospodarza wideokonferencji generuje i przesyła nowy numer pozycji oraz udostępnia go innym uczestnikom wideokonferencji. W sieci OoL przesyłane są zatem strumienie danych zaszyfrowane kluczami, które są różne dla każdego uczestnika wideokonferencji, a klucze te są losowo wymieniane co 5 minut. W przypadku dwóch (2) abonentów wideokonferencji, SoPVAC nie tworzy żadnych encji bezpośrednio powiązanych z wideokonferencją. Transmisja danych odbywa się w trybie p2p między węzłami uczestników konwersacji, którzy przechowują dwie (2) krotki kluczy AES i utrzymują otwarty jeden przychodzący zaszyfrowany strumień danych i jeden wychodzący zaszyfrowany strumień danych za pośrednictwem sieci I2P. Zasadniczo ważne jest, aby były to różne strumienie przesyłane przez różne tunele I2P. W przypadku n > 2 uczestników wideokonferencji SoPVAC przewiduje 3 tryby pracy. W pierwszym trybie (rys. 4a) host i wszyscy uczestnicy nawiązują między sobą bezpośrednie połączenia p2p. Ta opcja dla wideokonferencji z wieloma użytkownikami wymaga, aby przepustowość połączenia sieciowego każdego interfejsu użytkownika AN wynosiła 2n (n połączeń przychodzących i wychodzących), co implikuje dodatkowe obciążenie węzłów abonenckich, które w tym schemacie niezależnie wysyłają ruch do sieci uczestników. Oznacza to wysokie obciążenie sieci, opłaty za użytkowanie I2P (za bezpieczeństwo warstwy transportowej), co przenosi funkcję routera multicast na węzeł kliencki. Do kluczowych zalet tego schematu należą niski koszt, ponieważ nie są wykorzystywane dodatkowe zasoby w chmurze PSDE, oraz zwiększone bezpieczeństwo, ponieważ nie ma strażników w transmisji i przetwarzaniu danych. W pierwszym trybie SoPVAC może być dość wydajny, nawet w przypadku przeciętnego połączenia sieciowego, gdy użytkownicy uciekają się do konferencji głosowych. Szerokopasmowe połączenie sieciowe jest wymagane do wydajnego przetwarzania strumienia wideo. Aby ograniczyć ruch, SoPVAC umożliwia działanie z wykorzystaniem wirtualnych routerów wideokonferencji (VRC). W drugim trybie host interfejsu użytkownika AN adresuje CDApp za pośrednictwem CMN, a w sieci SCN (Tabela 1) tworzony jest wirtualny router wideokonferencji (VRC). Jego główną funkcją jest zbieranie strumieni danych od uczestników wideokonferencji i przekierowywanie ich do odbiorników. W tym trybie magnetowid nie posiada kluczy szyfrujących; jedynie odbiera zaszyfrowane strumienie danych od uczestników wideokonferencji i przesyła je zgodnie z przeznaczeniem. Dzięki VCR liczba jednocześnie otwartych strumieni multimedialnych w węźle abonenckim zostaje zredukowana do n + 1 (n przychodzących i 1 wychodzących), co oznacza, że w tym trybie VCR działa jako router multicast. W związku z tym wymagania dotyczące szybkości połączenia sieciowego AN UI są zmniejszone o połowę, dzięki nieznacznemu obniżeniu poziomu prywatności konwersacji. Mamy tu węzeł pośredniczący, który realizuje proces VCR, a cały ruch wideokonferencyjny przechodzi przez ten węzeł. Potencjalnie skompromitowany właściciel węzła pośredniczącego jest w stanie przechwycić ruch; jednak nie jest w stanie go odczytać, ponieważ nie posiada odpowiednich kluczy szyfrujących. I wreszcie, trzeci tryb działania oznacza, że VCR, oprócz funkcji routingu, przetwarza przychodzący ruch multimedialny dla uczestników wideokonferencji. Oznacza to, że

VCR otrzymuje klucze szyfrujące na równi z pozostałymi uczestnikami wideokonferencji. Para kluczy ECC używana przez VCR do szyfrowania kluczy AES to para właściciela SCN i to właśnie tutaj uruchamiany jest proces oprogramowania VCR. Proces rozpowszechniania klucza AES jest modyfikowany kilkakrotnie. VCR generuje tylko jeden (1) klucz AES. Każdy klient przesyła tylko jeden (1) (aktywny) klucz z krotki wygenerowanej przez VCR. Krotki z kluczami AES nie są przesyłane między klientami. VCR zbiera strumienie multimedialne od uczestników wideokonferencji, generuje z nich pojedynczy strumień, szyfruje wynikowy strumień za pomocą swojego klucza, a następnie udostępnia go wszystkim uczestnikom wideokonferencji. Dzięki temu magnetowid uzyskuje pełny dostęp do rozmowy bez zmniejszania obciążenia interfejsu użytkownika AN do dwóch (2) strumieni multimedialnych, zmniejszając tym samym wymagania dotyczące połączenia sieciowego i sprzętu węzła abonenckiego, dzięki znacznemu obniżeniu poziomu prywatności rozmowy. Kolejną wadą trybów pracy magnetowidu jest ogólne pogorszenie odporności na błędy wideokonferencji SoPVAC, ponieważ awaria magnetowidu automatycznie kończy wideokonferencję dla wszystkich jej klientów. Aby rozwiązać ten problem, a także zminimalizować ryzyko przechwycenia rozmowy związanej z każdą wideokonferencją, w SoPVAC jednocześnie przypisane są trzy (3) magnetowidy (podstawowy, zapasowy i rezerwowy). Działają one w sieci „pływających serwerów wirtualnych" („FVS"). Łącznie PSDE przypisuje „z (1) listy 10 SCN dla działania magnetowidu, znany już FVS $f : S_s \rightarrow \left\{ V = \{v_1, \ldots, v_{10}\} | \max_{a,b} S_s \right\}$. Początkowo dwa serwery są wybierane losowo spośród nich random(V) = $\{v^r_1, v^r_2\}$. Podstawowy magnetowid zostaje podniesiony do v^r_1, który otrzymuje pierwsze klucze i rozpoczyna obsługę wideokonferencji. Zapasowy magnetowid zostaje podniesiony do v^r_2, który otrzymuje pierwsze klucze i — 5 minut później — drugie klucze szyfrujące dla wideokonferencji. v^r_2 nie obsługuje wideokonferencji przez pierwsze 5 minut; otrzymuje wszystkie informacje wideokonferencyjne z v^r_1. Interfejs użytkownika utrzymuje „lekki kontakt" z zapasowym magnetowidem za pośrednictwem jego CMN, a w przypadku rozłączenia podstawowego magnetowidu natychmiast przełącza się na niego. CDApp rozpoczyna monitorowanie dostępności pozostałych FVS. Te osiem (8) SCN generuje trzeci $V^r_3 = V \setminus \{v^r_1, v^r_2\}$, który zastępuje rozproszony magnetowid. Pasywne procesy magnetowidu są wykonywane w węzłach V^r_3; otrzymują listę CMN uczestników wideokonferencji i natychmiast przechodzą w tryb gotowości. W przypadku rozłączenia jednego z FVS, CDApp żąda od chmury PSDE nowego. Po upływie 5 minut FVS zaczynają „pływać". Zapasowy magnetowid staje się podstawowym i zaczyna obsługiwać wideokonferencję z nowymi kluczami szyfrującymi. Nowy zapasowy magnetowid jest przydzielany przez FVS, a nowy SCN jest przypisywany do zapasowego magnetowidu. Następnie procedura zmiany serwera i klucza szyfrującego jest powtarzana co 5 minut. W ten sposób skompromitowany właściciel SCN jest w stanie przechwycić nie więcej niż 5 minut wideokonferencji. W przypadku drugiego trybu pracy może zostać przechwycony zaszyfrowany ruch, a liczba pakietów danych w tym ruchu jest zdecydowanie niewystarczająca do odszyfrowania [46]. W przypadku trzeciego trybu działania, czas zmiany kluczy i serwerów może zostać skrócony w zależności od potrzeb organizatora wideokonferencji, co niewątpliwie dodatkowo zmniejsza ryzyko potencjalnego podsłuchu, ale znacznie zwiększa koszt wideokonferencji. Dlatego w trzecim trybie działania nie jest wskazane omawianie żadnych istotnych informacji. Trzeci tryb działania został wprowadzony w SoPVAC wyłącznie w celu umożliwienia masowych wydarzeń komunikacyjnych, na przykład wykładów, seminariów czy webinariów o otwartym dostępie. Z punktu widzenia bezpieczeństwa jedyną zaletą korzystania z SoPVAC w trzecim trybie działania jest całkowita anonimowość wszystkich

uczestników rozmowy, oczywiście pod warunkiem, że twarze lub inne dane identyfikacyjne tych uczestników nie zostaną uchwycone na kamerze. SoPVAC nie rejestruje żadnych informacji o wideokonferencji ani o jej organizacji i nie pozwala żadnemu z jej uczestników rejestrować takich informacji, z wyjątkiem transakcji płatności za zasoby wykorzystywane w rozproszonym rejestrze TheOoL. Po zakończeniu wideokonferencji wszystkie procesy oprogramowania SCN zostają zakończone bez rejestrowania, natychmiast po dokonaniu płatności (co jest procesem w pełni zautomatyzowanym).

4 Podsystem wiadomości tekstowych SoPVAC

W niektórych przypadkach może zaistnieć konieczność wysłania wiadomości do użytkownika, który jest aktualnie offline. W tym przypadku SoPVAC posiada wbudowany podsystem wiadomości tekstowych. Aby przesłać wiadomości do użytkownika offline lub nagrać wiadomość do późniejszego odczytu , repozytorium komunikatora jest wykorzystywane w rozproszonym magazynie danych osobowych, do którego dostęp jest realizowany za pośrednictwem DT&SP. W tym przypadku istotne jest, kto wysyła wiadomość do magazynu – nadawca czy odbiorca – ponieważ definiuje to, kto ustala parametry przechowywania wiadomości i uiszcza opłaty za przechowywanie, a także kto jest właścicielem takich wiadomości w repozytorium danych, co decyduje o prawie do dysponowania informacjami, na przykład o udostępnianiu ich osobom trzecim. Kluczowe parametry pamięci masowej obejmują czas przechowywania, wymaganą szybkość dostępu do danych, współczynnik redundancji (określa on liczbę kopii zapasowych, która domyślnie wynosi dwie (2)) oraz minimalną liczbę równoległych wątków odczytu, która zależy od liczby potencjalnych czytelników i jest definiowana przez właściciela informacji. Właściciel ustala parametry pamięci masowej dla wszystkich wiadomości komunikatora (domyślnie dotyczą one wszystkich wiadomości), a także może ustalić pewne reguły dla konkretnej konwersacji. Domyślnie nadawca jest właścicielem wiadomości w SoPVAC; istnieje jednak opcja, w której prawa własności do wiadomości mogą zostać przeniesione na innego uczestnika konwersacji, jeśli nadawca nie chce płacić za utrzymanie konwersacji w repozytorium SoPVAC. Aby aktywować ten tryb, wymagana jest jednak uprzednia zgoda uczestników konwersacji. Rozważono metodę parametrycznej syntezy regulatora PID układu sterowania ze sprzężeniem zwrotnym za pomocą obiektów wyższego rzędu (powyżej drugiego), uwzględniającą rozmieszczenie złożonych biegunów sprzężonych. Metoda została zaimplementowana w środowisku programowym MathCAD. Wydajność metody wykazała dobre wyniki odpowiadające zadaniu projektowemu na przykładzie modelu ACS za pomocą obiektu czwartego rzędu. Proces przejściowy przebiegał bez nadmiernej korekty. Dla każdej konwersacji tworzone jest specjalne repozytorium w DT&SP i odpowiadający mu indeks w SDSE. Przechowuje ono metadane wiadomości w celu ich pobrania z DT&SP, a także link do deskryptorów bezpieczeństwa grupy biorącej udział w konwersacji. Dzięki temu każdy członek grupy może umieścić swoją wiadomość w konwersacji, niezależnie od dostępności pozostałych uczestników. Po uruchomieniu interfejs użytkownika AN skanuje indeksy grup z nim powiązanych, odbiera i odczytuje wiadomości. Skanowanie indeksów odbywa się z częstotliwością zdefiniowaną przez użytkownika przez cały czas działania interfejsu użytkownika AN. Szyfrowanie AES zapewnia ochronę wiadomości przed nieautoryzowanym odczytem. Transfer klucza AES odbywa się w oparciu o protokół Diffiego-Hellmana z wykorzystaniem krzywych eliptycznych. Szyfrowanie i transmisja wiadomości odbywa się w następującej kolejności:

1. W zaszyfrowanej wiadomości generowany jest losowy klucz AES. Generowana jest efemeryczna para kluczy ECC. Tworzona jest tabela sekretów w celu przesłania klucza z punktu 1

do uczestników konwersacji. Dla każdego uczestnika konwersacji wykonywane są następujące czynności:

(a) Pobieramy klucz publiczny ECC w stanie ustalonym z listy abonentów SoPVAC.

(b) Obliczamy wspólny sekret dla klucza prywatnego pary efemerycznej (punkt 2) i klucza publicznego w stanie ustalonym odbiorcy (punkt 3.a).

(c) Szyfrujemy klucz AES za pomocą protokołu uzyskanego w punkcie 1. Wspólny sekret jest używany jako klucz szyfrujący (punkt 3.b).

(d) Umieszczamy zaszyfrowany klucz i stacjonarny klucz publiczny odbiorcy w tabeli.

Umieszczamy zaszyfrowaną wiadomość, efemeryczny klucz publiczny pary oraz tabelę sekretów w odpowiednim repozytorium. Aby odczytać wiadomość, odbiorca musi wybrać z tabeli sekretów wpis zawierający jego stacjonarny klucz publiczny i AES odpowiadający jego zaszyfrowanemu kluczowi. Odbiorca musi następnie obliczyć wspólny sekret, stosując efemeryczny klucz publiczny pary i klucz prywatny ze swoich stacjonarnych par kluczy ECC, a następnie odszyfrować klucz AES. Następnie odbiorca może odszyfrować wiadomość.

5. Wnioski

Podsumowując, proponowany system komunikacji prywatnej zapewnia wysoki poziom prywatności dzięki zdecentralizowanej architekturze i wyeliminowaniu właścicieli infrastruktury serwerowej z SoPVAC, a tym samym eliminuje potencjalne zagrożenie naruszenia prywatności przez tych właścicieli z powodu ich zawodności lub błędów. Nawet w trybach operacyjnych wykorzystujących tymczasowe zasoby scentralizowane (VCR), SoPVAC nie pozwala ich właścicielom na niezawodne przechwytywanie rozmów, ponieważ żaden z właścicieli węzłów technicznych PSDE nie wie i nie może wiedzieć, który z węzłów i kiedy otrzymuje proces oprogramowania VCR do wykonania. Przeszkodą jest z jednej strony losowa zasada wyboru węzłów do VFS i ich nieustanne losowe zmiany, a z drugiej strony, a nawet w jeszcze większym stopniu, sama natura organizacyjna prywatnego internetu The OoL, który przyjmuje formę DAO. Innymi słowy, jest to otwarta organizacja peer-to-peer, w której interakcja jest regulowana przez algorytmy TheOoL. Każdy uczestnik może dołączyć do tej organizacji w dowolnym momencie. Zarówno subskrybenci, jak i właściciele węzłów technicznych automatycznie stają się uczestnikami rynku zasobów obliczeniowych w sieci TheOoL i są finansowo zachęcani do oferowania serwerów sprzętowych w chmurze PSDE. Ci (nazwijmy ich dostawcami) członkowie sieci mogą pochodzić z dowolnego miejsca geograficznego, mieć dowolną narodowość i wyznanie. Ich węzły mogą być wybierane do wykonywania dowolnych zadań w ramach PSDE na równych prawach z serwerami sprzętowymi potencjalnych atakujących. W rzeczywistości, im więcej dostawców działa w sieci TheOoL, tym mniejsza szansa, że którykolwiek z nich przechwyci konkretną konwersację. A nawet jeśli uda im się przechwycić część konwersacji, atakujący muszą, po pierwsze, przełamać kryptograficzną ochronę danych, a po drugie, muszą być w stanie zidentyfikować uczestników konwersacji, co jest skutecznie uniemożliwione dzięki abstrakcji użytkownika od infrastruktury technicznej PSDE i anonimizacji użytkowników. To znacznie komplikuje (prawie uniemożliwia) identyfikację przechwyconego ruchu i zestawianie go z konkretnymi członkami sieci TheOoL w świecie rzeczywistym.

Utworzenie optymalnego zestawu środków ochronnych do zarządzania ryzykiem bezpieczeństwa informacji w przemysłowych systemach sterowania przedsiębiorstw zajmujących się wydobyciem gazu

W niniejszym artykule opisano elementy uogólnionej struktury nowoczesnego zautomatyzowanego systemu sterowania procesami oraz poziomy sterowania procesami. Opracowano algorytm oceny ryzyka bezpieczeństwa informacji w zautomatyzowanych systemach sterowania procesami oraz tworzenia optymalnego zestawu środków ochronnych dla ich przetwarzania z wykorzystaniem podejścia metaheurystycznego. Podano podstawowe funkcje algorytmu genetycznego. Wskazano na konieczność poszukiwania optymalnej konfiguracji środków ochronnych, która poprawi efektywność bezpieczeństwa informacji w zautomatyzowanych systemach sterowania procesami w warunkach oddziaływania destrukcyjnego. Przedstawiono sformalizowane sformułowanie problemu optymalizacyjnego. Przeprowadzono testy proponowanego oprogramowania dla przedsiębiorstwa zajmującego się wydobyciem gazu Achimgaz. Potwierdzono efektywność algorytmu optymalizacyjnego i złożoność programów go implementujących. Określono preferowane obszary dla parametrów algorytmu genetycznego: prawdopodobieństwa crossing-over i mutacji, rodzaj crossing-over oraz objętość populacji początkowej.

Słowa kluczowe Zautomatyzowany system sterowania · Bezpieczeństwo informacji · Narzędzia bezpieczeństwa informacji · Przedsiębiorstwo wydobywcze gazu · Ocena ryzyka · Środki zarządzania ryzykiem · Algorytm genetyczny

1 Wstęp

Współcześnie systemy informatyczne stanowią istotną część organizacji. Organizacje te tracą również wiele przewag konkurencyjnych z powodu utraty bezpieczeństwa. Istnieje wiele prac badawczych i norm z zakresu zarządzania ryzykiem bezpieczeństwa (ISRM), w tym NIST 800–30 i ISO/IEC 27005. Jednak niewiele prac badawczych poświęcono ograniczaniu ryzyka bezpieczeństwa informacji. Chociaż normy wyjaśniają ogólne zasady i zalecenia, nie zawierają szczegółów dotyczących wdrażania ISRM. Radzenie sobie z ryzykiem bezpieczeństwa informacji w warunkach ograniczeń finansowych wymaga dużego wysiłku. Zarządzanie ryzykiem informacji w zautomatyzowanym systemie sterowania procesami (APCS) stanowi ważny obszar bezpieczeństwa informacji. Konieczność oceny i przetwarzania ryzyka informacji w wielopoziomowej strukturze hierarchicznej wymaga opracowania odpowiedniego algorytmu. W branżach, w których ochrona systemów zautomatyzowanych jest szczególnie krytyczna, na przykład w przypadku zautomatyzowanych systemów sterowania działających w obszarze kompleksów paliwowo-energetycznych, możliwe są znaczne straty materialne spowodowane awariami kosztownego sprzętu oraz negatywne konsekwencje ekonomiczne, środowiskowe lub humanitarne, w tym ofiary w ludziach. Zapewnienie gwarantowanego poziomu bezpieczeństwa informacji w systemach APCS wymaga doskonalenia metod oceny i przetwarzania ryzyka związanego z bezpieczeństwem informacji (SI). Dlatego też prowadzenie badań teoretycznych i rozwiązywanie problemów praktycznych w zakresie oceny i przetwarzania ryzyka związanego z SI i APCS jest pilnym zadaniem. Jednym z zadań w budowie systemu bezpieczeństwa informacji w systemach APCS jest opracowanie zestawu środków ochronnych. Jednocześnie konieczne jest uwzględnienie nie tylko wymagań bezpieczeństwa, ale także kosztów wdrożenia i utrzymania systemu bezpieczeństwa informacji. Zazwyczaj przy opracowywaniu zestawu zabezpieczeń stosuje się metodę opartą na podejściu eksperckim, które sprowadza się do preferencji ekspertów. W niniejszym artykule zaproponowano zastosowanie algorytmu genetycznego do doboru i optymalizacji konfiguracji systemu bezpieczeństwa informacji dla zautomatyzowanych

systemów sterowania procesami. Algorytmy genetyczne stanowią szczególny przypadek metod ewolucyjnych służących do heurystycznego rozwiązywania problemów optymalizacji i syntezy strukturalnej. Algorytmy genetyczne opierają się na znajdowaniu najlepszych rozwiązań poprzez dziedziczenie i ulepszanie użytecznych właściwości zbioru obiektów danej aplikacji w procesie symulacji ich ewolucji.

2. Sformułowanie problemu

Celem badania jest udoskonalenie istniejących metod i technik oceny ryzyka bezpieczeństwa informacji w systemach sterowania przemysłowego (ICS) oraz opracowanie algorytmu opartego na podejściu metaheurystycznym, który pozwala na wybór zestawu optymalnych środków w celu zmniejszenia ryzyka wynikającego z rzeczywistych zagrożeń dla bezpieczeństwa informacji w systemach sterowania przemysłowego (ICS). Celem badania jest zwiększenie bezpieczeństwa przemysłowych systemów sterowania i zapewnienie niezbędnego poziomu ochrony informacji z wykorzystaniem podejścia opartego na ryzyku. Osiągnięcie tego celu jest bezpośrednio związane z rozwiązaniem następujących problemów naukowych i technicznych:

1. przeprowadzenie analizy typowej struktury systemu sterowania przemysłowego (APC), jego elementów i zasobów;

2. opracować schemat APCS opisujący różne typy aktywów narażonych na zagrożenia bezpieczeństwa, podkreślając cechy interakcji aktywów w rozwiązywaniu problemu zarządzania ryzykiem bezpieczeństwa informacji;

3. w oparciu o istniejącą metodę opisaną w źródle [1], opracować algorytm optymalizacyjny z wykorzystaniem schematu APCS opisującego różne typy aktywów i ich relacje;

4. utworzyć zestaw środków ochronnych, określić główne aktywa zautomatyzowanego systemu sterowania procesami, podatności i zagrożenia dla bezpieczeństwa informacji zautomatyzowanego systemu sterowania procesami. Jednocześnie należy uwzględnić wymagania bezpieczeństwa, a także koszty wdrożenia i utrzymania systemu bezpieczeństwa informacji;

5. stworzyć narzędzie programowe implementujące algorytm optymalizacyjny do oceny ryzyka bezpieczeństwa informacji i utworzenia optymalnego zestawu środków ochronnych do postępowania z ryzykiem.

3 Analiza typowej struktury przemysłowych systemów sterowania

Zautomatyzowany system sterowania procesami – system sterowania człowiek-maszyna, który zapewnia automatyczne gromadzenie i przetwarzanie informacji niezbędnych do optymalizacji sterowania obiektem technologicznym zgodnie z przyjętym kryterium. Obiekt sterowania technologicznego (TCO) to zespół urządzeń technologicznych i jest on na nim zaimplementowany zgodnie z odpowiednimi instrukcjami lub regulacjami procesu technologicznego produkcji. TCO i APCS tworzą zautomatyzowany kompleks technologiczny. APCS powinien zapewniać trzy cechy procesu technologicznego, podobne do trzech cech informacji, które ma zapewniać SI – są to poufność, integralność i dostępność. W przypadku zautomatyzowanych systemów sterowania procesami – wydajność, niezawodność i bezpieczeństwo. Wydajność to redukcja kosztów zasobów i energii niezbędnych do wytworzenia produktu . Niezawodność – zapewnienie stałości jakości produktu, w tym w warunkach awarii urządzeń technologicznych, zmniejszenie

częstotliwości awarii [3]. Bezpieczeństwo – zwiększenie żywotności urządzeń technologicznych poprzez dobór optymalnych trybów pracy, ograniczenie udziału człowieka w procesie technologicznym . Zarządzanie obiektem technologicznym odbywa się na różnych poziomach zarządzania, z których każdy ma swoje własne zadania i możliwości . Typowy system sterowania procesem składa się z trzech poziomów: dolnego, środkowego i górnego. Dolny poziom zautomatyzowanego systemu sterowania procesem składa się z urządzeń obiektowych, w tym czujników i elementów wykonawczych. Ich działanie polega na pozyskiwaniu parametrów procesu, przetwarzaniu ich do odpowiedniej postaci w celu dalszego przesłania na wyższy poziom (funkcje czujników), a także na odbieraniu sygnałów sterujących i wykonywaniu odpowiednich czynności (funkcje elementów wykonawczych). Urządzenia obiektowe są połączone z jednym z modułów podsystemu wejścia/wyjścia (I/O) za pomocą kabli elektrycznych. Podsystem I/O składa się ze sprzętowych modułów I/O. Moduły różnią się rodzajem sygnału elektrycznego, którego używają do komunikacji z urządzeniami obiektowymi, oraz kierunkiem transmisji sygnału. Jeśli do modułu podłączony jest czujnik, to moduł wprowadza sygnał do systemu i nazywany jest modułem wejściowym; jeśli podłączony jest element wykonawczy, to moduł wyprowadza akcję sterującą z systemu i nazywany jest modułem wyjściowym. Na poziomie pośrednim znajdują się programowalne sterowniki logiczne (PLC) oraz urządzenia komunikacji z obiektem (OCD). OCD to zestaw modułów, które zapewniają interfejs z czujnikami i elementami wykonawczymi oraz umożliwiają odbiór, przetwarzanie i generowanie sygnałów różnego typu w szerokim zakresie napięcia, prądu, mocy i czasu trwania impulsu. PLC to urządzenie mikroprocesorowe przeznaczone do sterowania procesami przemysłowymi. Zasada działania PLC polega na zbieraniu sygnałów z czujników i przetwarzaniu ich z generowaniem sygnałów sterujących do elementów wykonawczych. Przetwarzanie to odbywa się zgodnie z wbudowanymi algorytmami sterowania, za pomocą których PLC i OCD sterują elementami wykonawczymi niższego rzędu i analizują informacje z czujników. Na najwyższym poziomie znajduje się system gromadzenia danych i operacyjnego sterowania dyspozytorskiego (ang. SCADA – Supervisory Control and Data Acquisition) lub rozproszony system sterowania (DCS). System ten to oprogramowanie konfigurowane i instalowane na komputerach dyspozytorskich. Na tym poziomie odbywa się sterowanie produkcją wyrobów. Proces ten obejmuje gromadzenie danych pochodzących z zakładów produkcyjnych, ich akumulację, przetwarzanie i wydawanie poleceń sterujących do niższych etapów. Wyższy poziom systemu składa się z serwerów wejścia/wyjścia, serwera automatycznego systemu sterowania dyspozytorskiego (ADCS) oraz stanowiska operatora. Na tym poziomie zaangażowany jest człowiek, czyli operator. Wykonuje on lokalne sterowanie urządzeniami technologicznymi za pośrednictwem interfejsu człowiek-maszyna. Często na wyższych poziomach architektury wykorzystywane są sieci IP i Ethernet. Komunikacja między różnymi poziomami odbywa się za pomocą węzłów komutacyjnych, a dostęp do czujników odbywa się za pośrednictwem magistrali polowych z wykorzystaniem specjalnych protokołów, takich jak Profibus.

4. Sformułowanie problemu oceny ryzyka bezpieczeństwa informacji i sformułowania optymalnego zestawu środków ochrony ich przetwarzania

Aby skutecznie rozwiązać problem przetwarzania niedopuszczalnych ryzyk bezpieczeństwa informacji w warunkach ograniczeń finansowych, konieczne jest określenie metod optymalizacji, które najlepiej odpowiadają specyfice rozwiązywanego problemu. W tym celu przeprowadzono analizę istniejących metod matematycznych stosowanych w podobnych problemach. Problem optymalizacyjny: konieczne jest znalezienie takiego optymalnego zestawu środków ochrony przed ryzykiem dla danych parametrów wejściowych, który zapewni minimum funkcji celu – całkowitego ryzyka resztkowego przy wdrażaniu środków redukcji ryzyka.

$$\sum_{i=1}^{N_T} r_{\lambda}^{i} \rightarrow \min$$

Należy przy tym uwzględnić ograniczenia finansowe Cmax dotyczące wdrażania środków zarządzania ryzykiem. Koszt wdrożenia środków zarządzania ryzykiem s^{j}_{c} nie może przekroczyć limitów finansowych C_{max} ustalonych przez przedsiębiorstwo.

$$\sum_{j=1}^{N_{RTM}} s_{c}^{j} \leq C_{\max}$$

Zatem r^{i}_{λ} — ryzyko resztkowe przy podejmowaniu działań ochronnych; i — liczba zagrożeń; N_T — liczba zagrożeń; s^{j}_{c} — koszt wdrożenia działań ochronnych; N_{RTM} — liczba działań mających na celu ograniczenie ryzyka (RTM); C_{max} — limity finansowe ustalone przez przedsiębiorstwo na wdrożenie działań ochronnych. W problemie optymalizacyjnym podejmuje się decyzję o wyeliminowaniu grupy zagrożeń w celu osiągnięcia minimalnej ogólnej oceny ryzyka resztkowego. Kolejny warunek uwzględnia ograniczenia przedsiębiorstwa dotyczące kosztów finansowych wdrożonych środków ochronnych. Zakłada się, że pojawienie się zagrożeń jest zdarzeniem niezależnym. Ponadto zakłada się, że po zneutralizowaniu wybranego zagrożenia w procesie wdrażania wybranego zestawu RTM nie pojawią się żadne nowe zagrożenia. Ostatni aspekt jest dość silnym założeniem. Po jego usunięciu rozwiązanie problemu optymalizacyjnego staje się na ogół dość skomplikowane. Najbardziej preferowane jest sformułowanie problemu wyznaczania RTM w oparciu o obliczenie minimalnego ryzyka resztkowego w chwili bieżącej. Zadanie sformułowania optymalnych środków ochronnych związanych z redukcją ryzyka, przy ustalonych ograniczeniach kosztów finansowych, można rozwiązać poprzez enumerację wszystkich proponowanych rozwiązań, czyli enumerację wyczerpującą. Główną zaletą podejścia z enumeracją wyczerpującą jest łatwość implementacji, wyszukiwanie funkcji celu dużej klasy oraz obliczenie optimum globalnego. Główną wadą tej metody jest czas potrzebny na znalezienie rozwiązania, czyli złożoność czasowa – 2^{NRTM}. Dlatego metoda siłowa jest stosowana tylko dla niewielkiej liczby zabezpieczeń. Wraz ze wzrostem liczby RTM problem nie może być rozwiązany tą metodą w optymalnym czasie. Aby pozbyć się tej wady, stosuje się bardziej złożone metody optymalizacji. W rozpatrywanym problemie zastosowano kryterium Walda (kryterium „maxmin”), a występują ograniczenia finansowe, dlatego proponuje się jego rozwiązanie metodą algorytmu genetycznego.

5 Analiza ryzyka

W dziedzinie technologii informatycznych i bezpieczeństwa informacji istnieje następująca definicja ryzyka: ryzyko bezpieczeństwa informacji to możliwość, że zagrożenie będzie w stanie wykorzystać podatność zasobu lub grupy zasobów i tym samym spowodować szkodę organizacji. Ryzyko jest rozpatrywane w kontekście cennego zasobu (aktywa). Aktywem jest wszystko, co ma wartość dla organizacji. Może to być sprzęt, finanse, procesy, zasoby ludzkie i informacyjne. Pojęcie ryzyka jest często mylone z pojęciem zagrożenia. Zagrożenie to możliwość wystąpienia niekorzystnego zdarzenia, które, z pewnym stopniem destrukcyjności, uszkadza mechanizmy obronne lub bezpośrednio wpływa na cenny zasób, prowadząc do negatywnych konsekwencji. Konsekwencje mogą być wyrażone w postaci szkód finansowych, materialnych, moralnych, reputacyjnych lub innych [6]. Szkodę w realizacji zagrożenia określa się jako wartość:

$$y_{W}^{i} = 3_{i}^{T} + \Pi_{i}^{T} + \tau_{i}^{T}$$

Tutaj 3^{T}_{i} są wskaźnikami wpływu zagrożenia t^{i} na koszty zbioru aktywów $A^{i} \cdot \Pi^{T}_{i}$ — wskaźniki wpływu zagrożenia t^{i} na stratę zbioru aktywów A^{i}, mierzone w procentach, τ^{T}_{i} — czas zakończenia lub zakłócenia działania grupy aktywów A^{i}, gdy zagrożenie t^{i} zostanie zrealizowane, mierzony w minutach. Wszystko to jest określane przez grupę ekspertów w oparciu o metodę Delphi. Po ocenie możliwych szkód w realizacji zagrożenia t^{i}, właściciel ryzyka musi ocenić możliwość realizacji tego zagrożenia. W tym przypadku stosowana jest metoda ekspercka drzew zdarzeń. Zaletą analizy drzew zdarzeń (ETA) w porównaniu z wieloma innymi metodami analizy ryzyka jest jej zdolność do modelowania sekwencji i interakcji różnych czynników ochrony, które towarzyszą wystąpieniu zdarzenia wyzwalającego. W ten sposób system i jego interakcje ze wszystkimi czynnikami ochrony w rozwoju niekorzystnego scenariusza zostają przedstawione wizualnie, co przyczynia się do dalszej oceny ryzyka. Za pomocą ETA ocenia się możliwość neutralizacji zagrożenia na każdym z istniejących poziomów zabezpieczenia i na podstawie uzyskanych wyników określa się możliwość zrealizowania zagrożenia t^{i} jako:

$$p^{i} = (1 - p^{i}_{1}) \times (1 - p^{i}_{2}) \times (1 - p^{i}_{3}) \times (1 - p^{i}_{4}) \times (1 - p^{i}_{5}) \times (1 - p^{i}_{6}) \times (1 - p^{i}_{7})$$

Mówimy o poziomach ochrony, stosując zasadę „ochrony dogłębnej” . Zgodnie z proponowaną zasadą, wszystkie środki łagodzące ryzyko dzielą się na poziomy ochrony według sfer wpływu: środki administracyjne (kary finansowe za naruszenie ustalonych zasad bezpieczeństwa informacji itp.), bezpieczeństwo fizyczne (stworzenie środków zapobiegających nieautoryzowanemu dostępowi do obiektu), obwód (zapora sieciowa filtrująca ruch z sieci zewnętrznej), sieć wewnętrzna (monitorowanie urządzeń wewnątrz sieci itp.), urządzenie (stosowanie ochrony antywirusowej itp.), aplikacja (monitorowanie luk w zabezpieczeniach itp.), informacja (szyfrowanie danych, tworzenie kopii zapasowych itp.). Zatem każdy środek łagodzący ryzyko można przypisać do jednego z poziomów ochrony. S_{Adm} – środki administracyjne, S_{Ph} – bezpieczeństwo fizyczne, S_{Per} – obwód, S_{IN} – sieć wewnętrzna, S_{Host} – urządzenie, S_{App} – aplikacja lub S_{Data} – dane. Oprócz oceny szkód wynikających z implementacji zagrożenia i możliwości jego implementacji, konieczne jest określenie oceny podatności. Systemy oceny podatności specjalizujące się w przemysłowych systemach sterowania obecnie nie istnieją . Wykorzystuje się wiele różnych metod oceny podatności, takich jak US-CERT i Skala Analizy Krytycznych Podatności Instytutu SANS. Jednakże, Wspólny System Punktacji Podatności (CVSS) jest dominującą techniką dla większości baz danych podatności. VSS określa wartość istotności implementacji konkretnej podatności na podstawie wartości jej charakterystyk. Wartości wektora podatności vi i jego wynik zostaną obliczone za pomocą wzoru:

$$\begin{aligned} \mathrm{CVSS}(v^{i}) &= \mathrm{CVSS{:}3.1}/\mathrm{AV} \\ &\quad : X/\mathrm{AC} : X/\mathrm{PR} \\ &\quad : X/\mathrm{UI} : X/S \\ &\quad : X/C : X/I \\ &\quad : X/A : X, \end{aligned}$$

gdzie X jest wynikiem metryki, CVSS (v^{i}) jest ilościowym wynikiem podatności v_{i}, mieszczącym się w przedziale od 0 do 10. Wielkość wektora podatności CVSS dla konkretnego cennego zasobu

można obliczyć, biorąc pod uwagę maksymalne kryteria wszystkich podatności powiązanych z zasobem:

$$\begin{aligned}\mathrm{CVS}(a^j) &= \mathrm{CVSS{:}3.1}/\mathrm{AV} \\ &\quad : \max X/\mathrm{AC} : \max X/\mathrm{PR} \\ &\quad : \max X/\mathrm{UI} : \max X/S \\ &\quad : \max X/C : \max X/I \\ &\quad : \max X/A : \max X\end{aligned}$$

6 Opis metody zarządzania ryzykiem i bezpieczeństwem informacji w zautomatyzowanym systemie sterowania procesami

Zarządzanie ryzykiem w zautomatyzowanym systemie sterowania procesami (IS APCS) opiera się na normie GOST ISO/IEC 27005. Zgodnie z tą normą proces zarządzania ryzykiem bezpieczeństwa informacji obejmuje etapy ustalania kontekstu, oceny, postępowania, akceptacji, monitorowania i ponownej oceny ryzyka bezpieczeństwa informacji. Procesy oceny i postępowania z ryzykiem mogą być przeprowadzane iteracyjnie. Podejście iteracyjne pozwala na zrównoważone inwestowanie czasu i wysiłku w dobór mechanizmów kontroli i zarządzania, zapewniając jednocześnie odpowiednią ocenę ryzyka wysokiego poziomu. Podczas ustalania kontekstu, zgodnie z jego modelem, określa się relacje między aktywami na różnych poziomach zautomatyzowanego systemu. Kolejnym krokiem jest identyfikacja zagrożeń i podatności na zagrożenia dla zidentyfikowanych aktywów i relacji. Analiza ryzyka obejmuje ocenę szkód wynikających z realizacji zagrożenia, ocenę możliwości realizacji zagrożenia oraz określenie środków postępowania z ryzykiem. Na tym etapie określa się również ograniczenia finansowe. Ocenę ryzyka przeprowadza się dla każdego indywidualnego zagrożenia zgodnie ze wzorem:

$R = P \cdot V \cdot D,$

gdzie P to prawdopodobieństwo realizacji zagrożenia, V to ocena podatności, D to szkody w przypadku ataku. Poszukiwany jest optymalny zestaw środków ochronnych w celu zmniejszenia ryzyka związanego z systemami informatycznymi opartymi na APCS, poprzez zastosowanie algorytmu genetycznego na etapie przetwarzania ryzyka. Następnie, tworzony jest optymalny zestaw środków ochronnych do postępowania z ryzykiem, który spełnia określone ograniczenia finansowe na etapie akceptacji ryzyka. Poszczególne zasoby APCS mogą być modernizowane. Ponadto, informacje o nowych lukach w zabezpieczeniach komputerowych w istniejącym oprogramowaniu są publikowane w źródłach publicznych. Wszystko to prowadzi do konieczności monitorowania i ponownej oceny ryzyka systemu w całym cyklu życia. Zatem metodę zarządzania ryzykiem APCS można przedstawić w postaci poniższego schematu blokowego. Metodę tę zaproponował Andrey Rimsha w swojej rozprawie doktorskiej . W oparciu o istniejącą metodę opracowano zoptymalizowany algorytm genetyczny, różniący się od algorytmu zaproponowanego przez A.S. Rimshę, opartego na programowaniu dynamicznym. Zmodyfikowano etapy oceny, przetwarzania i akceptacji ryzyka. W tym przypadku, na etapie przetwarzania i akceptacji ryzyka, wykorzystano algorytm genetyczny do poszukiwania optymalnego zestawu środków ochronnych w celu ograniczenia ryzyka związanego z SI APCS.

7. Konstrukcja algorytmu optymalizacyjnego do rozwiązania problemu optymalnego zestawu środków ochronnych

Ze względu na ogólnie duży wymiar problemu, a także konieczność uwzględnienia ograniczeń zasobów i różnorodności produktów, rozwiązywany problem należy do tzw. problemów NP-

zupełnych w dziedzinie optymalizacji dyskretnej. W związku z tym jego rozwiązanie wymaga dużych nakładów obliczeniowych. Celowe jest rozwiązanie rozważanego problemu formowania zestawu środków ochronnych w oparciu o metody optymalizacji. W tym przypadku, dla problemów o stosunkowo małym wymiarze, wskazane jest zastosowanie metody przeszukiwania wyczerpującego, która gwarantuje wyznaczenie ekstremum globalnego. Przy rozwiązywaniu problemu formowania zestawu środków ochronnych o stosunkowo dużym wymiarze konieczne jest zastosowanie ewolucyjnych metod populacyjnych , do których należy algorytm genetyczny. Poniżej znajduje się opis algorytmu genetycznego użytego do rozwiązania problemu . Należy zauważyć, że algorytmy genetyczne mogą implementować różne kryteria docelowe (funkcje dopasowania) i ograniczenia. Ta właściwość jest bardzo przydatna do tworzenia optymalnego zestawu środków ochronnych w warunkach ograniczeń finansowych. W rozważanym problemie optymalnego zestawu środków ochronnych konieczne jest wprowadzenie odpowiednich definicji i założeń dotyczących stosowania algorytmów genetycznych. Zatem każde możliwe rozwiązanie problemu, tj. zestaw środków ochronnych, jest możliwe. Istotną zaletą algorytmów genetycznych jest możliwość wykorzystania kilku punktów dla rozważanej przestrzeni poszukiwań, co umożliwia paralelizację poszukiwań. W tym przypadku algorytm genetyczny wykorzystuje informację zawartą w całym zestawie dopuszczalnych rozwiązań [19–22]. Główne koncepcje algorytmu genetycznego jako metody rozwiązywania problemu optymalizacji zestawu środków ochronnych są następujące .

Gen. Reprezentuje jednostkę tzw. informacji dziedzicznej, tj. jest integralną częścią chromosomu, służącą do wewnętrznej reprezentacji alternatywnych rozwiązań. W niniejszej pracy gen jest środkiem ochronnym.

Chromosom. Jest to uporządkowana sekwencja genów w postaci zakodowanej struktury danych, która determinuje decyzję. Chromosom będzie zestawem mechanizmów obronnych zakodowanych jako liczba binarna. Każdy bit równy 1 oznacza, że w konfiguracji uwzględniono środek ochronny o odpowiadającym mu numerze. Zakres zmiany kodu:

$G = (d_0 d_1 \ldots d_L)_2 = (0 \ldots 2L)_{10}$,

gdzie L to liczba dostępnych zabezpieczeń, di to informacja, czy zabezpieczenie jest zawarte w zestawie.

Populacja Jest to zbiór osobników — potencjalnych rozwiązań (optymalny zestaw środków ochronnych i ocena ryzyka resztkowego), które są reprezentowane przez chromosomy. Populacja składa się z osobników o różnych chromosomach. Jej wielkość jest ograniczona przez K, maksymalną liczbę osobników w populacji. Każdy przypadek jest definiowany przez zbiór: X = {G, C, R}, gdzie G to kod genetyczny, C to koszt środków ochronnych, R to całkowite ryzyko, uwzględniające podjęte środki ochronne. Rozważmy główne operatory algorytmu genetycznego.

Krzyżowanie Funkcją krzyżowania jest generowanie nowych potomków. Dla każdej pary tworzony jest nowy egzemplarz, który dziedziczy cechy rodziców. W pracy wykorzystano krzyżowanie wielopunktowe. W krzyżowaniu wielopunktowym losowo wybiera się k punktów odcięcia $dot_i \in$

$\{1, 2, \ldots, N_{gen}\}$, $i = \underline{1, k,}\ N_{gene}$ to liczba genów u osobnika. Następnie punkty odcięcia są

sortowane od najmniejszych do największych. Części chromosomów ograniczone punktami odcięcia są wymieniane między sobą. Ostatecznie powstają dwa nowe potomstwo. Część osobnika z pierwszym genem do pierwszego punktu odcięcia nie bierze udziału w wymianie. Krzyżowanie wielopunktowe wymaga wprowadzenia kilku zmiennych, czyli punktów odcięcia. Aby

uzyskać potomstwo, należy wybrać osobniki o największym dopasowaniu. Jeśli crossover wielopunktowy ma dużą, parzystą liczbę punktów przecięcia, to wygodnie jest przedstawić go dla chromosomów jako pierścienie. W tym przypadku chromosom wygląda jak zamknięty pierścień, a wybór punktów przecięcia następuje z równym prawdopodobieństwem na całym okręgu. Cztery punkty (granice wymiany) crossovera są wyznaczane losowo. Crossover generuje dziecko poprzez wybranie trasy od jednego rodzica i zachowanie pozycji i kolejności strażników od drugiego rodzica. Na przykład, dla rodziców G_1 = (0 0 1 1 | 0 0 | 1 1 1 | 00 | 0) i G_2 = (1 0 1 0 | 1 1 | 0 1 1 | 0 1 | 1) tworzymy dziecko w następujący sposób. Pierwszym krokiem jest zamiana wybranych podtras, w wyniku czego otrzymujemy Ch_1 = (X X X X | 1 1 | X X X | 0 1 | X) i Ch_2 = (X X X X | 0 0 | X X X | 0 0 | X), gdzie „X" jest pustą pozycją. Następnie w miejsce „X" wstawiamy numery środków ochronnych od pierwotnych rodziców, dla których nie ma konfliktów (nie tworzy się żaden cykl). W rezultacie otrzymujemy dwoje dzieci: Ch_1 = (0 0 1 1 | 1 1 | 1 1 1 | 0 1 | 0) i Ch_2 = (1 0 1 0 | 0 0 | 0 1 1 | 0 0 | 1). Mutacja Jest to losowa zmiana w jednej lub kilku pozycjach na chromosomie. Zgodnie ze schematem AG, operator mutacji jest wykonywany z zadanym prawdopodobieństwem Pm. Mutacja pozwala na utrzymanie różnorodności genetycznej osobników, dzięki czemu populacja nie ulega degeneracji, a chromosomy nie upodabniają się do siebie. Funkcją mutacji jest zmienność kodu genetycznego. Losowo odwrócone dwie cyfry binarne w chromosomie. Inwersja bitowa została użyta jako operator mutacji dla rozwiązywanego problemu. Reprodukcja Jest to proces, w którym chromosomy są kopiowane do populacji pośredniej w celu dalszej „propagacji" zgodnie z wartościami ich funkcji celu. Zatem jeden lub więcej potomków z lepszymi wynikami funkcji celu ma większe szanse na przejście do następnego pokolenia. Przed zastosowaniem selekcji należy obliczyć koszt środków ochronnych i wielkość ryzyka, biorąc pod uwagę zestaw środków ochronnych dla populacji. Populacja początkowa musi składać się z co najmniej dwóch osobników. W pracy operator reprodukcji jest implementowany zgodnie z regułą ruletki. Jako funkcję dopasowania, funkcja celu jest używana

$$\sum_{i=1}^{N_T} r_\lambda^i \rightarrow \min$$

Algorytm składa się z następujących kroków:

1. wczytanie wszystkich niezbędnych danych z pliku;

2. utworzenie populacji początkowej;

3. obliczenie funkcji dopasowania dla populacji początkowej;

4. wejście do pętli while z warunkiem sprawdzenia liczby populacji o wstępnie określonej maksymalnej dopuszczalnej wartości populacji;

5. zastosowanie operatora reprodukcji;

6. zastosowanie operatora crossing-over dla wybranych osobników;

7. mutacja osobników;

8. wybór najlepszego osobnika na podstawie najniższej wartości funkcji dopasowania.

Program implementujący algorytm genetyczny został napisany w języku programowania C++. Aby zbadać wpływ ustawień algorytmu (wielkość populacji początkowej, prawdopodobieństwo crossovera, prawdopodobieństwo mutacji, kryterium zatrzymania) i typów operatorów (wybór puli rodzicielskiej, wybór par rodzicielskich, typ crossovera, typ mutacji, utworzenie nowej populacji),

wskazane jest zastosowanie problemu niskowymiarowego. W tym celu, za pomocą metody siłowej, wyznacza się ekstremum globalne i można porównać rozwiązanie uzyskane na podstawie algorytmu genetycznego. Aby stwierdzić zależność szybkości konwergencji do wartości optymalnej od ustawień algorytmu genetycznego.

8 Testowe zatwierdzenie programu według danych Przedsiębiorstwa Gazowniczego „Achimgaz"

Najpierw tworzona jest lista aktywów na podstawie schematu typowego zautomatyzowanego systemu sterowania procesami w przedsiębiorstwie gazowniczym. Załóżmy, że kierownictwo przedsiębiorstwa gazowniczego ustaliło ograniczenia dotyczące możliwości zmiany procesu krytycznego i zaangażowało firmy zewnętrzne do ciągłego utrzymania systemu sterowania procesami, a także ograniczenia finansowe dotyczące wdrożenia środków zarządzania ryzykiem w wysokości 7 milionów rubli. Podano koszt wdrożenia dla każdego elementu zestawu S_{RTM}. Obliczone prawdopodobieństwa zagrożenia, szkody w przypadku ataku oraz wielkość podatności dla każdego zagrożenia zostały zapisane w pliku o rozszerzeniu CVS. Łącznie do pliku załadowano 14 środków ochronnych G i ich koszt C, 14 ocen podatności V, 14 ocen szkód w realizacji zagrożeń, 14 ocen prawdopodobieństwa realizacji zagrożenia . Określono koszt wdrożenia dla każdego zabezpieczenia. Wartość ryzyka jest obliczana jako suma iloczynów prawdopodobieństwa realizacji zagrożenia pi przez ocenę podatności vi i szkód w przypadku ataku di. W konfiguracji początkowej R_0 = 421 945,8376. Tabela przedstawia wynik programu opartego na algorytmie genetycznym.

N	Percentage of interbreeding individuals from the total population	Mutation percentage		
		30%	50%	70%
50	30%	69,438.3/6889000	46,475.6/7409000	78,580.6/7763000
	50%	46,475.6/7409000	46,475.6/7409000	73,332.1/7795000
	70%	47,375.3/7921000	46,475.6/7409000	46,475.6/7409000
	90%	46,475.6/7409000	47,375.3/7921000	52,557.9/7831000
N	Percentage of interbreeding individuals from the total population	Mutation percentage		
		30%	50%	70%
200	30%	49,874.2/7909000	46,475.6/7409000	73,332.1/7795000
	50%	46,475.6/7409000	46,475.6/7409000	49,874.2/7909000
	70%	47,375.3/7921000	49,874.2/7909000	48,120.7/7799000
	90%	73,332.1/7795000	78,580.6/7763000	52,557.9/7831000

Komórki tabeli przedstawiają stosunek oceny skuteczności systemu ochrony, z uwzględnieniem zestawu środków ochronnych, do kosztu zestawu środków ochronnych, wraz z odpowiadającym odsetkiem osobników krzyżujących się i elitaryzmu w całej populacji, a także potęgą populacji. Najlepsze rozwiązanie jest zaznaczone na żółto. Najlepsze rozwiązanie jest określane jako najskuteczniejsze, z najniższą oceną ryzyka resztkowego i najbardziej odpowiednim kosztem wdrożenia środków ochronnych w ramach ustalonych ograniczeń. Zmienne wartości parametrów algorytmu genetycznego:

• liczba osobników w populacji N = 50, 200;

• prawdopodobieństwo krzyżowania się powyżej 30, 50, 70, 90%;

• prawdopodobieństwo mutacji 30, 50, 70%;

• crossing-over: wielopunktowy;

• mutacja: inwersja bitowa.

Przy potędze populacji 50, przy 50 i 70% skrzyżowanych osobników z całej populacji, przy 50 i 70% mutacji, uzyskano najodpowiedniejsze wartości optymalnej konfiguracji środków ochronnych. Przy koszcie zestawu środków ochronnych wynoszącym 7 409 000, ryzyko resztkowe po wdrożeniu środków ochronnych wyniesie 46 475,6. Przy potędze populacji 200, najodpowiedniejsze wartości optymalnej konfiguracji środków ochronnych uzyskano przy 50% skrzyżowanych osobników z całej populacji i 50% mutacji. Wraz ze wzrostem liczby osobników obserwuje się wzrost czasu działania algorytmu. Na podstawie przeprowadzonych badań udało się zidentyfikować związek między wzrostem prawdopodobieństwa mutacji a optymalnym rozwiązaniem. Optymalna wartość prawdopodobieństwa mutacji wynosi 0,5. Dla każdego nosiciela, zgodnie z jego kodem genetycznym, wybiera się zestaw środków ochronnych, za pomocą którego zmienia się parametry modelu i oblicza się całkowite ryzyko. Ponadto, dla wszystkich wariantów kodu genetycznego, przeprowadzono eksperyment obliczeniowy w celu określenia najlepszych konfiguracji pod kątem redukcji ryzyka i efektywności kosztowej .

9 Analiza i wnioski z wyników testowania oprogramowania w celu znalezienia optymalnej konfiguracji środków ochronnych

Podczas analizy wyciągnięto następujące wnioski:

1. Należy zauważyć, że w proponowanym algorytmie przeprowadza się większą liczbę obliczeń ryzyka – 2RTM – w porównaniu z wyliczeniem wszystkich wariantów 2MOP. Dla 10 epok przeprowadzono 50 069 obliczeń wartości ryzyka, dla 5 epok – 17 518.

2. Czas działania algorytmu został przeanalizowany dla różnej liczby pomiarów w celu zmniejszenia ryzyka. Dla 5 RTM czas wykonania algorytmu genetycznego wyniósł 0,029 s. Dla 9 RTM czas wykonania algorytmu genetycznego wyniósł 0,036 s. Dla 14 RTM czas wykonania algorytmu genetycznego wyniósł 0,048 s.

3. Przy wysokich wartościach prawdopodobieństwa krzyżowania (P_c = 0,7, P_c = 0,9) i prawdopodobieństwa mutacji (P_m = 0,7) algorytm nie zawsze znajdował rozwiązanie spełniające wymagania dokładności. Można to wyjaśnić faktem, że rozwiązania przedwcześnie zbiegają się do jednego punktu.

4. Najlepsze wyniki uzyskano przy 50% krzyżowaniu, czyli prawdopodobieństwie krzyżowania powyżej 0,5, oraz przy 50% mutacji. Dalsze zwiększanie wartości nie poprawia wydajności, a czasami utrudnia poszukiwanie optymalnego rozwiązania.

5. Najbardziej efektywnym rozwiązaniem pod względem liczby pokoleń i czasu poświęconego na znalezienie rozwiązania jest wykorzystanie populacji 50 osobników. 6. Najbardziej opłacalny system ochrony polega na wprowadzeniu kar pieniężnych za nieprzestrzeganie ustalonych zasad (K = 2000), zainstalowaniu czujników alarmowych w pomieszczeniach ze sprzętem (K = 105 000), zainstalowaniu ramek wykrywaczy metali i introskopu telewizyjnego rentgenowskiego na posterunku wartowniczym (K = 2 630 000), zorganizowaniu monitoringu wizyjnego w pomieszczeniach ze sprzętem (K = 520 000), skonfigurowaniu filtrowania według białej listy adresów MAC (K = 4000), zainstalowaniu dwóch dodatkowych zapór sieciowych (K = 408 000), zainstalowaniu bramek transmisji danych z jednej sieci LAN do drugiej tylko w jednym kierunku (K = 2 508 000), utworzeniu zapasowego serwera AD tylko do odczytu i ograniczeniu dostępu do głównego serwera AD (K = 158 000), ustanowieniu systemu Zabbix do monitorowania stanu

sprzętu i programów (K = 58 000), uzyskaniu uprawnień w systemie za pomocą kodu PIN i Karta dostępu (C = 1 008 000), monitorowanie nowo powstałych luk i ich eliminacja za pomocą oprogramowania antywirusowego (C = 4000), regularne tworzenie kopii zapasowych kationów informacyjnych na serwerze zapasowym (C = 4000). Dzięki temu zestawowi środków ochronnych, całkowite ryzyko resztkowe przy wdrożeniu środków zarządzania ryzykiem zmniejszy się o 46 475,6, a koszt wdrożenia środków wynosi 7409 000 rubli, co mieści się w założonych przez przedsiębiorstwo limitach finansowych.

Koszt wdrożenia RTM wynosi zatem 7409 000 rubli, co spełnia ustalone ograniczenia finansowe.

10 Wnioski

W artykule poruszono ważny obszar bezpieczeństwa informacji – zarządzanie ryzykiem informacyjnym w przemysłowych systemach sterowania. Konieczność oceny i przetwarzania ryzyka informacyjnego w wielopoziomowej strukturze hierarchicznej wymagała opracowania odpowiedniego algorytmu. Przeprowadzono analizę bezpieczeństwa informacji w systemach APCS, poziomów sterowania procesami i głównych funkcji systemów APCS, a także rozważono główne zagrożenia dla bezpieczeństwa informacji w systemach APCS. Celem pracy była poprawa efektywności bezpieczeństwa informacji w zautomatyzowanych systemach sterowania procesami w obliczu destrukcyjnych oddziaływań poprzez zastosowanie podejścia opartego na ryzyku. W oparciu o podejście metaheurystyczne opracowano algorytm oceny ryzyka bezpieczeństwa informacji w zautomatyzowanych systemach sterowania procesami oraz doboru optymalnego zestawu środków ochronnych dla ich przetwarzania. Opracowano narzędzie programistyczne w języku programowania C++ do implementacji algorytmu oceny ryzyka bezpieczeństwa informacji w zautomatyzowanych systemach sterowania procesami oraz doboru optymalnego zestawu środków ochronnych dla ich przetwarzania. Program pobiera dane wejściowe z pliku, a następnie konwertuje otrzymane parametry wejściowe zadania na elementy algorytmu: gen, chromosom, populację, funkcję dopasowania. W procesie badawczym, w celu uzyskania najlepszego wyniku optymalizacji, modyfikowano różne parametry i cechy algorytmu, takie jak liczba iteracji, wielkość populacji, prawdopodobieństwo mutacji i krzyżowania.

Zastosowanie zasad sterowania PID w zadaniach modelowania ruchu pojazdów kołowych wyposażonych w układ zapobiegający blokowaniu kół (ABS)

Stały wzrost prędkości i natężenia ruchu drogowego, a co za tym idzie zmniejszenie odległości między uczestnikami ruchu, nieuchronnie prowadzi do wzrostu liczby wypadków, dlatego też aktualność zagadnienia rozważanego przez autorów rozdziału jest niewątpliwa, lecz z czasem będzie ono coraz bardziej istotne. Poprawa skuteczności układu hamulcowego pojazdów kołowych w rzeczywistych warunkach eksploatacji wymaga optymalizacji metod rozwiązywania modelu matematycznego elementu sterującego układem rozdziału siły hamowania pojazdów kołowych. Nowatorstwo naukowe niniejszej pracy polega na wykorzystaniu elementów logiki rozmytej w algorytmie działania elektronicznej jednostki sterującej układem rozdziału siły hamowania w odniesieniu do zadań hamowania pojazdów kołowych.

Słowa kluczowe Pojazd kołowy · Proces hamowania · Układ zapobiegający blokowaniu kół · Symulacja · Sterowanie · Optymalizacja · Logika rozmyta

1 Wstęp

Rosnąca rola sztucznej inteligencji, fuzja robotyki z przemysłem i biznesem, wdrożenie „Internetu rzeczy" w zautomatyzowanej produkcji doprowadziły do czwartej rewolucji przemysłowej. Ale Przemysł 4.0 to tylko część większego procesu. Dziś japoński rząd mówi o narodzinach społeczeństwa superintelektualnego – Społeczeństwa 5.0. Wykorzystanie w samochodzie różnorodnych czujników, które mogą gromadzić i przetwarzać Big Data, jest jednym z trendów w strategii Społeczeństwa 5.0. Samochód porusza się po mieście i gromadzi dane o otoczeniu, pomagając kierowcy podejmować właściwe decyzje w różnych sytuacjach, a na przykład w trybie hamowania minimalizuje błędy kierowcy, zwłaszcza w momencie kontrolowanego poślizgu kół w trybie hamowania . Koncepcja kontrolowanego poślizgu kół podczas hamowania jest znana od około wieku. Dlatego już w latach 30. XX wieku w podwoziach samolotów stosowano układy zapobiegające blokowaniu kół (ABS). Również koncepcja kontrolowanego poślizgu kół w trybie hamowania jest wykorzystywana w problemach ruchu kolejowego. Jednak koncepcja ta upowszechniła się w naziemnym transporcie kołowym dopiero w ostatniej ćwierci XX wieku. Wprowadzenie układów ABS do seryjnych modeli samochodów wiązało się zarówno z wymogami poprawy bezpieczeństwa ruchu drogowego w obliczu wzrostu ograniczeń prędkości i natężenia ruchu, jak i z możliwością implementacji rozwiązań technicznych, które pojawiły się w okresie szybkiego rozwoju elektroniki i technologii informatycznych . Oprócz układów ABS, kontrolowany poślizg kół jest stosowany w układach sterowania sprzęgłem (Traction Control), stabilizacji kursu walutowego (ESP) itp. Obecnie znane są różne możliwości implementacji zasad teorii sterowania automatycznego w elektronicznych jednostkach układów hamulcowych, z których każda ma swoje zalety, jak i wady. Z drugiej strony, w ciągu ostatnich dwóch stuleci obserwuje się wzrost znaczenia i rozpowszechnienia sterowania automatycznego w różnych systemach technicznych. Począwszy od trzeciej ćwierci XIX wieku, istnieją ścisłe zapisy teorii sterowania automatycznego . Początkowo zasady teorii sterowania automatycznego stosowano do liniowych lub zlinearyzowanych postaci problemów; później opracowano i nadal rozwijane są metody mające zastosowanie do problemów nieliniowych. Ważnym krokiem w rozwoju układu sterowania automatycznego jest określenie zasady sterowania i wybór typu regulatora. Przesłanki rozwiązania problemu parametryzacji wszystkich możliwych regulatorów stabilizujących, odnoszących się do konkretnego obiektu sterowania, zaczęły pojawiać się w pracach badaczy w ostatnim ćwierćwieczu XX wieku i pierwotnie dotyczyły jedynie skończenie wymiarowych liniowych układów stacjonarnych ze sprzężeniem zwrotnym. Następnie prowadzono badania zarówno w odniesieniu do układów niestacjonarnych, jak i nieliniowych oraz układów o nieskończonych wymiarach. Przyjrzyjmy się zagadnieniu modelowania układu przeciwblokującego hamulca ze szczególnym uwzględnieniem układu sterowania automatycznego .

2 Cechy modelowania automatycznego układu sterowania układem ABS

Zadaniem automatycznego układu sterowania układem ABS jest utrzymanie wartości argumentu S, odpowiadającej maksymalnej wartości funkcji ϕ(S), poprzez przeniesienie momentu hamowania na tarczę hamulcową. Należy zauważyć, że dla różnych rodzajów nawierzchni maksimum funkcji ϕ(S) znajduje się w pobliżu wartości argumentu 0,2, podczas gdy maksymalne wartości samych funkcji różnią się znacząco. W najprostszych układach automatycznego sterowania (ACS) ABS zaproponowano wykorzystanie zasad regulacji PID w sterowaniu rozmytym z regulacją opartą na odchyleniu współczynnika poślizgu wzdłużnego kół od wartości docelowej 0,2. Ogólnie rzecz biorąc, układy ACS dla ABS są znane w oparciu o zasady regulacji PID, adaptacyjnej, optymalnej, odpornej, rozmytej, inteligentnej, a także ich różne kombinacje i hybrydy, zarówno jedno-, jak i wielowejściowe. Najprostsze systemy pod względem opisu modelu matematycznego i technicznej implementacji rozwiązania z reguły charakteryzują się najgorszą

adekwatnością w stosunku do rzeczywistych procesów. Rozwiązania o większej adekwatności mogą być znacznie bardziej złożone pod względem modelowania matematycznego i wymagają głębokiej wiedzy od programistów z kilku różnych dziedzin kilku różnych dyscyplin naukowych jednocześnie. Zatem zaprojektowanie automatycznego systemu sterowania układem ABS przy dużej liczbie istniejących rozwiązań pozostaje zadaniem nietrywialnym. Głównym celem układu ABS samochodu jest skrócenie drogi hamowania do minimum przy jednoczesnym zachowaniu sterowalności podczas hamowania . Głównymi danymi wejściowymi do obliczenia kluczowych parametrów są prędkość kątowa (przyspieszenie) koła i prędkość liniowa (przyspieszenie) samochodu. Kluczowymi parametrami, które określają działanie układu ABS, są współczynnik poślizgu wzdłużnego S i współczynnik przyczepności wzdłużnej koła do nawierzchni drogi ϕ. Do określenia prędkości kątowej koła stosuje się głównie impulsowe czujniki magnetyczne. Do określenia prędkości liniowej pojazdu znanych jest kilka metod opartych na przetwarzaniu sygnałów czujnika prędkości wałka wyjściowego skrzyni biegów ; jednoczesnym przetwarzaniu sygnałów z kilku czujników prędkości kątowej kół pojazdu; przetwarzaniu danych z akcelerometru; przetwarzaniu danych z modułu GPS, a także różne kombinacje wskazanych metod. Na podstawie otrzymanych danych oprogramowanie jednostki sterującej układu przeciwblokującego (ECU ABS) rozdziela siły hamowania. Skuteczność układu przeciwblokującego jest determinowana zarówno przez błąd wskazań wskazanych czujników, jak i przez adekwatność procesu przetwarzania danych wejściowych przez urządzenie obliczeniowe . Oznacza to konieczność dążenia do wysokiego poziomu dokładności zgodności przybliżeń z rzeczywistymi wartościami parametrów w modelowaniu i obliczaniu ruchu koła samochodu. Zadowalający wynik pierwotnego przetwarzania sygnałów czujników można obecnie w większości przypadków osiągnąć dzięki różnym implementacjom programowym metody filtru Kalmana . Do dalszego przetwarzania danych można zastosować metody wykorzystujące regulator PID, metody sterowania optymalnego ze spełnieniem warunku stabilności Lyapunowa, metody sterowania nieliniowego itp. Metoda logiki rozmytej, omówiona w tym rozdziale, jest jedną z najbardziej satysfakcjonujących metod modelowania ruchu koła samochodu w trybie hamowania w czasie rzeczywistym. Jako przykład modelu matematycznego układu hamulcowego ze sprzężeniem zwrotnym przedstawiono schemat blokowy algorytmu działania modelu układu hamulcowego z ABS, który zawiera operator do rozwiązywania układu równań różniczkowych . Adekwatność rozwiązań modelu matematycznego do rzeczywistego procesu jest określana, po pierwsze, przez zgodność równania rzeczywistego procesu z fizyką zjawiska, po drugie, przez dokładność opisu matematycznego modułów procesu i po trzecie, przez zgodność wyników eksperymentów obliczeniowych i pełnoskalowych z tymi samymi danymi początkowymi, w ramach przyjętych założeń i błędów obliczeń i pomiarów. Podczas modelowania kontrolowanego procesu do wymagań jednoznaczności, stabilności i adekwatności dodaje się jeszcze jedną rzecz — rzeczywisty skumulowany czas rozwiązywania równań modelu musi wyprzedzać czas modelowanego procesu. Ponadto czas wyprzedzenia nie powinien być krótszy niż czas podjęcia i wykonania decyzji przez odpowiednie urządzenia sterujące.

3. Model matematyczny hamowania kół samochodowych

Modele obejmujące rozwiązywanie równań różniczkowych wymagają czasu obliczeniowego komputera do obliczenia symulowanego procesu. Na przykład rozwiązanie układu nieliniowych równań różniczkowych opisujących ruch koła w trybie hamowania awaryjnego jest stabilne w sensie Lapunowa. Układ równań różniczkowych opisujący proces hamowania kół samochodowych wygląda następująco:

$$\begin{cases} M = f(t), & (1) \\ \dot{\omega} = \frac{1}{j} \cdot [-M + \mu \cdot m \cdot g \cdot r], & (2) \\ \omega = \omega_0 + \int \dot{\omega} dt, & (3) \\ \dot{v} = -\mu \cdot g, & (4) \\ v = v_0 + \int \dot{v} \cdot dt, & (5) \\ \lambda = 1 - \frac{\omega \cdot r}{v}, & (6) \\ \mu = \frac{f_0 \cdot \lambda}{a \cdot \lambda^2 + b \cdot \lambda + c}. & (7) \end{cases}$$

gdzie

t to aktualna wartość czasu;

M moment hamowania przyłożony do koła;

ω_0, v_0 to początkowe wartości prędkości kątowej i liniowej;

$\dot{\omega}, \omega$ to przyspieszenie kątowe i prędkość obrotowa;

$\dot{v}, v$ to przyspieszenie liniowe i prędkość koła;

j, m to moment bezwładności koła i masa na koło;

g to współczynnik grawitacji;

r to promień dynamiczny koła;

μ to współczynnik przyczepności wzdłużnej koła z nawierzchnią drogi;

λ to współczynnik poślizgu wzdłużnego;

a, b, c to współczynniki empiryczne (deskryptory nawierzchni drogi);

f_0 to współczynnik tarcia opona-podłoże przy pełnym zablokowaniu kół.

Wynika to z następujących zjawisk fizycznych:

(1) zależność momentu hamowania od czasu;

(2) zmiana przyspieszenia kątowego koła;

(3) zmiana prędkości obrotowej koła;

(4) zmiana opóźnienia ruchu postępowego masy skupionej w środku koła i tworzącej obciążenie normalne w miejscu styku;

(5) zmiana prędkości liniowej koła;

(6) stosunek prędkości postępowej do prędkości obwodowej bieżnika w środku miejsca styku jest współczynnikiem poślizgu względnego;

(7) zależność współczynnika tarcia opony o nawierzchnię w kierunku wzdłużnym od współczynnika poślizgu względnego i nawierzchni drogi.

Jednak w tym przypadku w rozwiązaniu tego układu równań może wystąpić niestabilność, co jest konsekwencją zastosowanych metod numerycznych, w szczególności wyboru kroku całkowania. Wraz ze zmniejszeniem kroku całkowania wzrasta dokładność obliczeń, ale jednocześnie wydłuża się czas obliczeń. Istniejące metody numeryczne (Eulera, Runge-Kutty, Dormana-Prince'a) zakładają rozwinięcie żądanej funkcji w szereg Taylora lub (Adamsa) wykorzystują drugi wzór

interpolacyjny Newtona. W praktyce zastosowanie metod numerycznych o zadanym stopniu dokładności obliczeń, na przykład Runge-Kutty czwartego rzędu, tylko częściowo rozwiązuje problem braku mocy obliczeniowej . Problem zwiększenia szybkości zliczania przy zachowaniu zadanej dokładności obliczeń można rozwiązać, stosując element ECU o wyższej wydajności, na przykład procesor graficzny ogólnego przeznaczenia, co w konsekwencji doprowadzi do wzrostu ceny ECU. Opracowując model ABS, należy uwzględnić obecnie ekonomicznie nieakceptowalny wysoki koszt wysokowydajnych procesorów graficznych do masowej produkcji jednostek sterujących. Jedną z możliwych alternatywnych metod reprezentacji modelu procesu hamowania kół w czasie rzeczywistym, uwzględniającą przyłożone siły hamowania i umożliwiającą wykluczenie układów równań różniczkowych jako etapu pośredniego obliczeń, jest metoda logiki rozmytej z wykorzystaniem zmiennych lingwistycznych. Zastosowanie zmiennych lingwistycznych stosowanych w logice rozmytej będzie wymagało znacznie mniejszej mocy obliczeniowej modułu ECU.

4 Implementacja modelu ABS dla jednego koła sterowanego regulatorem rozmytym

W części obliczeniowej wyznaczymy parametry hamowania koła samochodowego, stosując metodę numeryczną i celowo mały krok całkowania (aby zapewnić wysoką dokładność uzyskanych wyników), a także regulator rozmyty o tych samych danych początkowych. Następnie konieczne będzie porównanie uzyskanych wyników i wyciągnięcie odpowiednich wniosków. Tabela przedstawia dane początkowe do obliczeń.

Name	Designation	Value
Linear speed value at the initial moment of time	v_0	60 km/h
Linear and angular acceleration at the initial moment of time	$\dot{\omega}_0$, $\dot{v}_0$	0
Wheel moment of inertia	j	0.9 $\mathrm{m} \cdot \mathrm{s}^2$
Mass per wheel	m	350 kg
Gravity factor	g	9.8 $\mathrm{m} \cdot \mathrm{s}^2$
Dynamic radius of the wheel	r	0.3 m
Empirical coefficients characterizing the road surface (dry surface)	a; b; c; f_0	0.342; 0.612; 0.046; 0.7
Braking torque applied to the wheel at the initial moment of time	M_0	0
Angular speed value at the initial moment of time	$\omega_0 = \frac{v_0}{r}$	
The law of braking torque variation	$1000 \cdot \mathrm{t}$	

Przy rozwiązywaniu danego układu numerycznie wykorzystamy ulepszoną metodę Eulera i krok obliczeniowy równy 0,005 s. Jako wersję regulatora, alternatywę dla regulatora PID, wykorzystamy model rozmyty regulatora. Pozostawmy model dla jednego koła z poprzedniego rozdziału bez zmian, zastąpimy jedynie podsystem regulatorem. Do utworzenia regulatora rozmytego wykorzystamy zestaw narzędzi Fuzzy Library Logic Toolbox pakietu MATLAB, do pracy wybierzemy regulator Fuzzy Block Logic Controller. Aby uprościć model, wykorzystaliśmy model z jedną wartością wejściową i trzema wartościami wyjściowymi, które są określone przez zakres wartości wejściowej. Zastosujemy parametr prędkości do sygnału wejściowego, w zależności od jego wartości na wyjściu otrzymamy wartości Ki, Kp i Kd – współczynniki dla elementu PID projektowanego regulatora. Ustawmy reguły i uzyskajmy wizualizację w Simulinku. Teraz komponenty rozmyte i PID są kompatybilne w Simulinku. Aby poprawić adekwatność regulatora, używamy dwóch parametrów wejściowych: wartości prędkości dla elementu rozmytego i wartości odchyłki dla elementu PID . W rezultacie otrzymujemy regulator rozmyty typu PID ze współczynnikami P-, I-, D, które zmieniają się w zależności od trybu prędkości.

5. Wnioski

W odniesieniu do procesu komputerowej symulacji ruchu koła samochodu w trybie hamowania w czasie rzeczywistym, obecnie można zalecić następujące rozwiązania. W niektórych przypadkach wynik zastosowania regulatora rozmytego jest dość trudny do przewidzenia i nie zawsze możliwe jest uzyskanie wymaganej dokładności regulacji. W związku z tym konieczne jest przeprowadzenie wstępnej analizy układu i dokonanie rozsądnego wyboru regulatora. Jak pokazały wyniki badań, tam gdzie zapewniona jest regulacja, preferowane jest zastosowanie klasycznych regulatorów PID, a w odniesieniu do zagadnienia modelowania ruchu samochodu z ABS, zastosowanie adaptacyjnego regulatora rozmytego jest uzasadnione i wystarczające. Podczas opracowywania regulatora rozmytego wprowadzenie dodatkowego parametru wejściowego oprócz wartości odchyłki poprawia jakość modelu. W proponowanym uproszczonym modelu wykorzystano wartość prędkości pojazdu. Preferowane jest jednak wykorzystanie pierwszej lub drugiej pochodnej prędkości pojazdu, prędkości kół lub zmiany odchyłki. Wykorzystanie wartości liczbowych (poprzez wartości pochodnych lub zgodnie z danymi akcelerometru) i znaków przyspieszeń w modelu rozmytym umożliwia rozważenie przewidywania zachowania systemu; Aby poprawić adekwatność modelu, należy uwzględnić opóźnienie czasowe, biorąc pod uwagę działanie hydraulicznego układu hamulcowego i przetwarzanie danych przez procesor. Konieczne jest przewidywanie w czasie rzeczywistym wartości współczynnika poślizgu w następnym punkcie czasowym i wykorzystanie jej do kontrolowania błędu. Aby to zrobić, w modelu, na wyjściu sygnału współczynnika poślizgu, podsystem powinien zostać włączony, biorąc pod uwagę opóźnienie czasowe.