

Koneksjonizm i sieci neuronowe stały się podstawą sztucznej inteligencji i nauk kognitywnych. Obecnie konferencje na temat sieci neuronowych z perspektywy sztucznej inteligencji (lub inteligencji obliczeniowej, jak niektórzy by to ujęli) odbywają się regularnie i zwykle cieszą się dość liczną frekwencją (np. Międzynarodowe Wspólne Konferencje na temat Sieci Neuronowych). Na najważniejszych konferencjach kognitywistyki prace oparte na modelach koneksjonistycznych zajmują zwykle główne miejsce. Na wielu konferencjach i czasopismach inżynierskich prace z wykorzystaniem modeli sieci neuronowych są na porządku dziennym. Ich popularność i atrakcyjność osiągnęły w pewnym sensie stabilny stan. Innymi słowy, stały się integralną częścią badań i eksploracji inteligencji i poznania. W tej części najpierw dokonam krótkiego przeglądu historii modeli koneksjonistycznych, identyfikując główne idee i główne obszary zastosowań, a następnie przejdę do zajęcia się kwestią przetwarzania symbolicznego w modelach koneksjonistycznych; na koniec rozszerzę dyskusję na hybrydowe modele koneksjonizmu, które uwzględniają zarówno koneksjonistyczne, jak i symboliczne metody przetwarzania.

Rewolucja koneksjonistyczna lat 80

Modele koneksjonistyczne – czyli modele składające się z sieci prostych jednostek przetwarzających, połączonych ze sobą za pomocą różnego rodzaju wzorców łączności – powróciły w latach 80. XX wieku jako główny paradygmat nauk o kognitywistyce i sztucznej inteligencji po okresie uśpienia w późnych latach 60. i 70. Ta koneksjonistyczna rewolucja w latach 80. przyniosła ze sobą nowe paradygmaty, nowe podejścia, nowe idee i nowe techniki, a także nowe emocje i nowe kontrowersje. Podekscytowanie i kontrowersje już w dużej mierze ucichły, ale wiele pomysłów i technik koneksjonizmu pozostało, które stały się integralną częścią zestawu narzędzi sztucznej inteligencji i nauk kognitywnych.

Przegląd koneksjonizmu

Ogólnie rzecz biorąc, koneksjonizm to sposób uchwycenia i zrozumienia mechanizmów i procesów poznania poprzez budowanie modeli z wykorzystaniem sieci prostych, neuronowych elementów (jednostek) przetwarzających, z których każda wykonuje proste obliczenia numeryczne. Modele koneksjonistyczne zostały zastosowane do wielu różnych zadań: na przykład postrzegania obiektów i zdarzeń; wymawianie tekstów angielskich; przechowywanie i odzyskiwanie (odpowiednio kontekstowo) informacji z pamięci; tworzenie i rozumienie języka; uczenie się umiejętności; rozumowanie i tak dalej. Modele koneksjonistyczne (sieci neuronowe) opierają się na założeniu, że poznanie powstaje w wyniku interakcji dużej liczby prostych elementów lub jednostek przetwarzających (tj. „neuronów”). Podstawową ideą jest to, że mózg składa się z ogromnej liczby takich jednostek i że razem są one zdolne do niezwykle złożonego przetwarzania poznawczego (takiego jak percepcja, język, kontrola motoryczna itd.). Chociaż koneksjoniści często twierdzą, że uchwycili w swoich modelach zasady biologicznych procesów neuronalnych, jednostki (węzły) w modelach koneksjonistycznych rzadko jednak odpowiadają poszczególnym neuronom biologicznym. Często było to konieczne ze względu na praktyczne względy obliczeniowe lub wynikało z braku pełnej wiedzy na temat biologii procesów neuronowych. W modelu koneksjonistycznym „reprezentacja” jest często wzorcem aktywacji zbioru jednostek przetwarzających w modelu (choć istnieje wiele odmian i alternatyw dla tego; więcej później). Przetwarzanie odbywa się poprzez propagację aktywacji pomiędzy jednostkami przetwarzającymi (węzłami) poprzez wzajemne połączenia między nimi. Tym, co pośredniczy w propagacji aktywacji, są numeryczne „wagi” połączeń pomiędzy parami jednostek przetwarzających. Uczenie się odbywa się poprzez (zwykle stopniową) zmianę „wag” połączeń w zależności od aktywności w sieci. Uczenie się czasami odbywa się w oparciu o sygnały „błędu” dostarczane ze źródeł zewnętrznych, w postaci albo sygnału sukcesu/porażki (Sun et al. 2005; Sutton i Barto 1998), albo sygnału wskazującego stopień niedopasowania pomiędzy rzeczywistym wynikiem z

sieci oraz pożądany wynik lub cel. Oto kilka szczególnie interesujących (i potencjalnie użytecznych) właściwości modeli koneksjonistycznych, wymienionych w książce Rumelharta i in. z 1986 r., a także w innych odpowiednich publikacjach:

- * Równoległość (to znaczy fakt, że w większości tych modeli węzły mogą jednocześnie wykonywać obliczenia, a łącza mogą jednocześnie propagować informacje).’
- * Adaptacyjność (czyli wbudowana w większość tych modeli zdolność uczenia się, np. poprzez modyfikację wag na łączach łączących węzły).’
- * Płynna degradacja (to znaczy możliwość uniknięcia katastrofalnych awarii w obliczu błędów w przetwarzaniu lub wprowadzaniu danych).’
- * Automatyczne uzupełnianie (nowych lub znanych) wzorców (jako nieodłączna właściwość wielu z tych modeli).’
- * Spontaniczna generalizacja (czyli umiejętność spontanicznego zastosowania istniejącej wiedzy w różnych sytuacjach).’
- * Solidność (to znaczy wrodzona odporność na uszkodzenia wynikająca z pięciu poprzednich właściwości, wynikająca z konstrukcji takich modeli).’
- * Adresowalność treści (czyli możliwość wyszukiwania informacji w oparciu o częściową znajomość jej treści).’
- * Optymalizacja i spełnianie ograniczeń (to znaczy zdolność do znalezienia optymalnych lub prawie optymalnych rozwiązań, które spełniają wiele „miękkich” ograniczeń, jako nieodłączna właściwość niektórych z tych modeli).

Bardziej szczegółowe omówienie właściwości sieci neuronowych można znaleźć w literaturze koneksjonistycznej. Historycznie rzecz biorąc, koneksjonizm wyrósł z kilku odrębnych wątków badań. Niektóre z wczesnych badań dotyczyły sieci jednostek binarnych. Inne wczesne badania skupiały się na systemach adaptacyjnych. Prace nad programowaniem dynamicznym i uczeniem się przez wzmacnianie również przyczyniły się do powstania sieci neuronowych. W szczególności McCulloch i Pitts w latach czterdziestych XX wieku badali proste sieci „neuronów” progowych binarnych pod kątem operacji logicznych. Donald Hebb opracował teorię poznania dotyczącą składania komórek i zaproponował w szczególności pogląd, że mogą to być specyficzne zmiany synaptyczne leżą u podstaw psychologicznych zasad uczenia się. Frank Rosenblatt w latach pięćdziesiątych XX wieku sformułował zasady uczenia się dla sieci neuronowych w celu kojarzenia dowolnych wzorców poprzez dostosowywanie wag (tj. idei perceptronu). Bernard Widrow zaproponował modele adaptacyjnych układów liniowych. Sieci neuronowe odrodziły się w latach 80. XX wieku w wyniku negatywnej reakcji na dominujące wówczas podejście symboliczne w sztucznej inteligencji i kognitywistyce oraz jego postrzeganą porażkę. Ten ruch z lat 80. XX wieku był znany jako „rewolucja koneksjonistyczna”. Opierając się na modelach, które eliminują scentralizowaną manipulację sekwencyjną symbolami i często obejmują przetwarzanie rozproszone (w architekturach masowo równoległych), koneksjonizm był często określany w tamtych czasach jako „równoległe przetwarzanie rozproszone”. Z tego punktu widzenia do poznania należy podchodzić bardziej w kategoriach mechanizmów spełniania ograniczeń, rozpoznawania wzorców i dostosowywania wagi, a nie wyrażnej manipulacji symbolami.

Uczenie się koneksjonistyczne

W tym miejscu warto dokonać krótkiego przeglądu głównych paradygmatów uczenia się w sieciach neuronowych. Przyjrzymy się w szczególności następującym rodzajom uczenia się: uczeniu się pod

nadzorem, uczeniu się bez nadzoru i uczeniu się przez wzmacnianie. Choć tego typu uczenie można przeprowadzić także w innych typach modeli, są one szczególnie widoczne w sieciach neuronowych. Algorytmy nadzorowanego uczenia się wymagają sygnału sprzężenia zwrotnego (ze źródeł zewnętrznych) dla każdego węzła wyjściowego sieci, aby proces uczenia mógł przebiegać. Typowy przykład takiego uczenia się ma miejsce w trójwarstwowej sieci neuronowej ze sprzężeniem zwrotnym (gdzie sieć ze sprzężeniem zwrotnym oznacza sieć bez połączeń ze sprzężeniem zwrotnym między warstwami i bez połączeń bocznych w warstwie). W takiej sieci wzorce wejściowe prezentowane są w pierwszej warstwie (warstwa wejściowa), a każda kolejna warstwa (warstwa ukryta i warstwa wyjściowa) jest kolejno aktualizowana (poprzez propagację aktywacji), w wyniku czego na końcu powstaje wzór wyjściowy. Ten wzorzec wyjściowy jest porównywany z pożądanym wzorcem wyjściowym i w ten sposób obliczany jest sygnał błędny. Sygnał błędny jest propagowany wstecz przez sieć w celu obliczenia aktualizacji wag pomiędzy warstwami i zmniejszenia błędny. Słynny algorytm uczenia się z propagacją wsteczną jest jedną z klas takich algorytmów uczenia się, który jest szeroko stosowany we wszystkich obszarach kognitywistyki i sztucznej inteligencji (Rumelhart i in. 1986). W tym względzie uzyskano interesujące wyniki teoretyczne dotyczące obliczeń sieci neuronowych. Wykazano, że warstwowe sieci neuronowe z wyprzedzeniem są uniwersalnymi aproksymatorami, to znaczy mogą reprezentować zasadniczo dowolną funkcję. Wykazano, że rekurencyjne sieci neuronowe (tj. sieci z połączeniami ze sprzężeniem zwrotnym) są odpowiednikami Turinga (tj. odpowiednikami maszyny Turinga – ogólnego modelu obliczeń) i są w stanie reprezentować dużą klasę nieliniowych układów dynamicznych. Uzyskano różnorodne wyniki dotyczące uczenia się nadzorowanego w sieciach neuronowych. Na przykład w przypadku klasyfikacji wykazano, że algorytmy uczenia się sieci neuronowych zbiegają się z tylnymi prawdopodobieństwami klas. Algorytmy te zostały wykorzystane zarówno w modelowaniu poznawczym, jak i zastosowaniach praktycznych. Bayesowskie metody statystyczne zostały wykorzystane do analizy uczenia się nadzorowanego i projektowania nowych algorytmów uczenia się. Z drugiej strony uczenie się bez nadzoru nie wymaga sygnałów błędnych ani rozróżnienia między węzłami wejściowymi, ukrytymi (wewnętrznymi) i wyjściowymi. Ma zatem zastosowanie w wielu sytuacjach, gdzie nie ma możliwości uczenia się pod nadzorem. Na przykład samoorganizujące się sieci, będące formą uczenia się bez nadzoru, były szeroko badane i wykorzystywane w modelowaniu poznawczym, a także w zastosowaniach przemysłowych. Polegają one na samoorganizacji węzłów i łączy w odpowiedzi na dane z sieci środowiska. Innym podejściem do uczenia się bez nadzoru jest określenie modelu sposobu, w jaki środowisko generuje dane. Uczenie się bez nadzoru staje się zatem statystycznym problemem znalezienia najlepszego modelu pasującego do danych. Uczenie się przez wzmacnianie plasuje się gdzieś pomiędzy uczeniem się pod nadzorem i bez nadzoru: nie wymaga dokładnego sygnału błędny, a jedynie wskazania, czy prąd wyjściowy jest dobry, czy zły (lub jak dobry jest, czy zły), zwykle w postaci liczby. Dlatego ma szersze zastosowanie niż uczenie się pod nadzorem i często jest bardziej przydatne niż uczenie się bez nadzoru (ze względu na dostępność informacji zwrotnej). Uczenie się przez wzmacnianie stosowano w modelowaniu uczenia się zwierząt i uczenia się umiejętności ludzkich, a także w inżynierii sterowania i innych zastosowaniach przemysłowych. Uzyskano wiele wyników teoretycznych dotyczących różnych algorytmów uczenia się przez wzmacnianie. Istnieją również modele sieci neuronowych oparte na teorii prawdopodobieństwa. Wykazano matematycznie, że wiele modeli uczenia się sieci neuronowych jest powiązanych z obliczeniami probabilistycznymi (tj. Odpowiadającymi teorii prawdopodobieństwa). Zapewnia to w pewnym sensie matematycznie rygorystyczną podstawę do uczenia się sieci neuronowych. W szczególności warto zauważyć, że niektóre algorytmy uczenia się bez nadzoru badane w literaturze dotyczącej sieci neuronowych można wyprowadzić z sieci Bayesa

Reprezentacje koneksjonistyczne

Ogólnie rzecz biorąc, istnieją dwie ogólne kategorie reprezentacji koneksjonistycznych: (1) reprezentacja lokalistyczna, w której każdy węzeł reprezentuje indywidualną koncepcję; oraz (2) reprezentacja rozproszona, w której każde pojęcie jest reprezentowane przez wzorzec aktywacji w zestawie węzłów, z których każdy może nie być interpretowalny (przenikliwy poznawczo).¹ Jednakże poza tymi dwiema przybliżonymi kategoriami istnieje całe spektrum dostępne różne techniki reprezentacji:

* W pełni lokalistyczna reprezentacja. Jak wspomniano wcześniej, charakteryzuje się to reprezentowaniem każdej koncepcji za pomocą indywidualnego, dedykowanego węzła w sieci. Innymi słowy, jest to jeden węzeł dla jednego pojęcia (tj. istnieje mapowanie jeden do jednego pomiędzy węzłami i pojęciami).

* Rozproszona reprezentacja lokalna. Zamiast jednego węzła dla jednego pojęcia, możliwe jest użycie zbioru węzłów dla jednego pojęcia, każdego z nich który wykonuje to samo przetwarzanie, więc wybicie jednego węzła spowoduje, że nie wpływają drastycznie na wydajność systemu. Ten zestaw węzłów ma na celu reprezentowanie tylko tej konkretnej koncepcji i żadnej z nich węzły te uczestniczą w reprezentacji wszelkich innych koncepcji. Przedstawienie to można uznać za odmianę określenia lokalisty reprezentacji. Jest to zbiór węzłów dla jednego pojęcia (tzn. mapowanie jeden do jednego pomiędzy zbiorami węzłów i pojęć).

* Lokalnie rozproszona reprezentacja. Ta forma reprezentacji charakteryzuje się podzieleniem przestrzeni reprezentacji (zbiór wszystkich dostępnych węzłów) na kilka podprzestrzeni (tj. podzbiorów węzłów), a dla każdej z tych podprzestrzeni stosowana jest reprezentacja rozproszona. Oznacza to, że zbiór pojęć jest reprezentowany w sposób rozproszony w podprzestrzeni reprezentacyjnej. Ale poza tą konkretną podprzestrzenią reprezentacyjną to, czy węzeł jest aktywny, czy nie, nie ma znaczenia dla reprezentacji pojęć w tym zbiorze. Reprezentacja jest zlokalizowana w odniesieniu do zbioru pojęć, ale jest rozproszona w obrębie zbioru.² Można ją nazwać zbiorem węzłów dla zbioru pojęć (tj. istnieje odwzorowanie jeden do jednego pomiędzy zbiorami węzłów i zbiorami pojęć).³

* W pełni rozproszona reprezentacja. Każdy węzeł uczestniczy w reprezentacji wszystkich zaangażowanych koncepcji, a każde pojęcie jest reprezentowane przez wszystkie węzły używane w sieci.³ Można to nazwać wszystkimi węzłami dla wszystkich koncepcji.⁴

* Inne typy reprezentacji rozproszonej. Oprócz tego istnieją inne typy reprezentacji rozproszonej, które nie pasują do żadnego z powyższych opisów.

Każda z tych technik reprezentacji ma swoją własną charakterystykę oraz zalety i wady, dlatego każda jest odpowiednia w określonych sytuacjach.

Koneksjonizm i poznanie

Następnie dokonano krótkiego przeglądu różnych obszarów modelowania koneksjonistycznego w odniesieniu do rozumienia poznania, w zgodzie z wspomnianymi wcześniej użytecznymi właściwościami modeli koneksjonistycznych. Podczas gdy neuronauka zajmuje się zjawiskami na poziomach często nie wyższych niż sieci komórkowe i lokalne, modele poznawcze muszą uwzględniać ludzkie zachowania na wyższych poziomach (np. badane w psychologii). Dlatego uzasadnione jest pewne uproszczenie. W ten sposób, chociaż wiarygodność biologiczna jest w pewnym stopniu zagrożona, możemy jednak założyć, że uproszczone modele mogą mieć to samo ogólne podejście do obliczeń, co biologiczne systemy neuronowe. Ze względu na swoje charakterystyczne cechy modele koneksjonistyczne zaoferowały nowe teorie poznawcze i wygenerowały wyjaśnienia, które często radykalnie różnią się od wyjaśnień poprzednich teorii. Pokróćce omówię kilka przykładów.

Pamięć. W modelach koneksjonistycznych pamięć jest to często konstruktywny proces obejmujący interakcję prostych jednostek przetwarzających (węzłów). Na przykład przywołanie może być procesem konstruowania wzorca aktywacji na zestawie jednostek, który jest podobny do pewnego wzorca, którego doświadczyliśmy wcześniej. Często podlega wpływom wynikającym z interakcji jednostek, które mogą uzupełnić brakujące szczegóły lub skorygować niedokładności. Jedną z przydatnych właściwości modeli koneksjonistycznych wyszkolonych na przykład za pomocą propagacji wstecznej jest to, że mogą dowiedzieć się, jakiej podstawy w ramach reprezentacji rozproszonej użyć do wewnętrznego reprezentowania pojęć, tak aby procesy oparte na podobieństwie (np. Uogólnienie) mogły opierać się na odpowiednich cechach. Dotychczas modele koneksjonistyczne stosowano do zagadnień pamięci semantycznej, pamięci epizodycznej, uczenia się pojęć, kategoryzacji i tak dalej.

Uczenie się ukryte i jawne. Pokazano modele koneksjonistyczne aby były odpowiednie do zajęcia się psychologicznym rozróżnieniem pomiędzy uczeniem się jawnym i ukrytym, opracowane na podstawie empirycznych badań psychologicznych. Jak wykazali Sun i inni rozróżnienie to można uchwycić poprzez zastosowanie reprezentacji lokalnej i reprezentacji rozproszonej w sieciach koneksjonistycznych. W przedstawieniach lokalnych (lub symbolicznych) każda jednostka jest łatwa do interpretacji i ma jasne znaczenie pojęciowe. Cecha ta oddaje właściwość wiedzy jawnej, która jest bardziej dostępna i łatwiejsza w manipulowaniu. Natomiast jednostki reprezentacyjne w reprezentacji rozproszonej są razem zdolne do wykonywania zadań, ale generalnie nie mają indywidualnego znaczenia. Sun i inni dokładnie pokazali, w jaki sposób efekty interakcji uczenia się ukrytego i jawnego można ująć w modelach opartych na tym rozróżnieniu. Podobnie Cleeremans i McClelland (1991), a także inni, zajęli się modelowaniem danych psychologicznych dotyczących czysto utajonego uczenia się.

Pamięć ukryta i jawna. Modele koneksjonistyczne zajęły się również rozróżnieniem między pamięcią jawną i ukrytą. (Pamięć ukryta odnosi się do skutku doświadczenia bez wyraźnego odniesienia do tego wcześniejszego doświadczenia – czasami nawet bez żadnego świadomego przypomnienia tego doświadczenia.) Pamięć jawna niedawnych doświadczeń może być zaburzona u niektórych pacjentów z uszkodzeniem mózgu, którzy mimo to wykazują dobrą pamięć ukrytą, sugerując, że do tworzenia nowych wyraźnych wspomnień może być wymagany specjalny układ mózgowy. Modele koneksjonistyczne zazwyczaj uwzględniają pamięć ukrytą opartą na połączeniach między dużą pulą jednostek w sieci o rozproszonych reprezentacjach.

Język. Modele koneksjonistyczne sugerują realną alternatywę dla idei, że procesy poznawcze w języku muszą być reprezentowane jako system reguł. Koneksjoniści stworzyli modele fleksji morfologicznej, konwersji pisowni na dźwięk, przetwarzania i rozumienia zdań oraz szeregu innych aspektów, które wyjaśniają wiele ważnych zjawisk psycholingwistycznych, często pomijanych przez teorie oparte na regułach. Jednym z problemów jest wrażliwość wzorców językowych na częstotliwość i spójność. W podejściu opartym na regułach nie uwzględniono faktu, że wyjątki nie są arbitralne. Na przykład wyjątki od czasu przeszłego regularnego czasowników angielskich występują w klastrach o wspólnych cechach fonologicznych. Model koneksjonistyczny pokazał, że taki model, który nauczył się wag połączeń w celu wygenerowania czasu przeszłego słowa z czasu teraźniejszego, może uchwycić szereg psychologicznych aspektów nabywania czasu przeszłego. W modelach koneksjonistycznych przetwarzanie języka jest często procesem zaspokajania ograniczeń podlegającym ograniczeniom semantycznym, kontekstowym, syntaktycznym i innym. Ponadto w rekurencyjnych sieciach neuronowych prowadzono prace koneksjonistyczne nad uczeniem się struktur gramatycznych zdań.

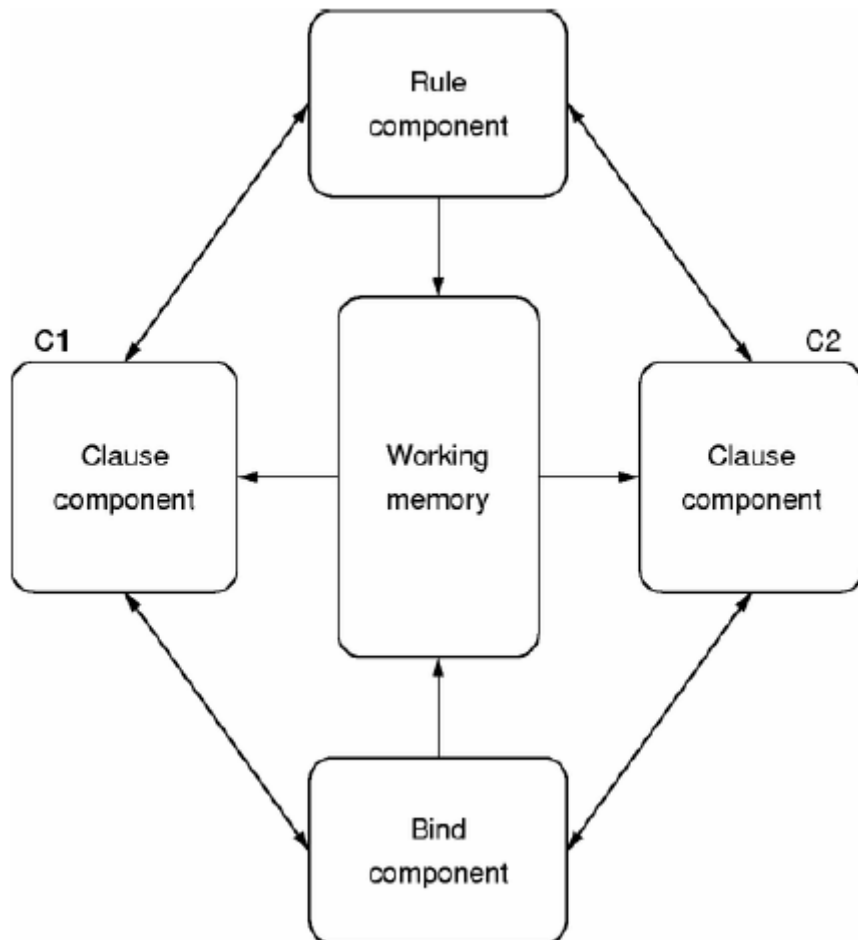
Rozumowanie. Jeśli chodzi o poznanie wyższego poziomu, takie jak rozumowanie i rozwiązywanie problemów, modele koneksjonistyczne (czystych form) pod wieloma względami są niewystarczające. Jednakże wykonano pewne istotne prace w tym obszarze. Na przykład poczyniono prace w obszarze rozumowania przez analogię. Dokonano również innych prac w obszarach rozumowania opartego na

regułach, a nawet rozumowania logicznego. W tym względzie badacze często wybierają modele „hybrydowe”. Modele hybrydowe (które zostaną omówione później) często w taki czy inny sposób wykorzystują tradycyjne metody symboliczne. Na przykład mogą przypisywać jednostki i połączenia za pomocą algorytmów symbolicznych, a następnie przeprowadzać procesy spełniania ograniczeń lub uczenie się przez propagację wsteczną w sposób koneksjonistyczny. Mogą też implementować rozumowanie symboliczne, używając bezpośrednio wag połączeń.

Wiązanie. Badacze często postrzegają tzw. „problem wiążący” jako podstawowy problem w koneksjonistycznych modelach rozumowania, a także w innych obszarach kognitywistyki. „Wiązanie” odnosi się do połączenia wielu dowolnych elementów podczas przetwarzania lub reprezentacji, włączając na przykład przypisanie dowolnego elementu do slotu w opisie strukturalnym (np. przypisanie „CPU-213” do slotu „CPU” w ustrukturyzowany opis „komputera”, który obejmuje gniazda procesora, pamięci, dysku itd.). Aby rozwiązać ten problem, zaproponowano kilka (częściowych) rozwiązań. Jednak opracowanie sieci neuronowych, które uczą się tworzyć własne wiązania, zarówno w znanych, jak i nieznanych domenach, pozostaje trudnym problemem.

Przykłady koneksjonistycznego przetwarzania symbolicznego

Biorąc pod uwagę znaczenie symboli w procesach poznawczych, modele koneksjonistyczne muszą być w stanie uchwycić symbole i przetwarzanie symboliczne, zwłaszcza wyszukiwanie i reprezentację wiedzy. Podejmowano wiele prób umożliwienia modelom koneksjonistycznym przeprowadzania przetwarzania symbolicznego. Zaproponowano różne schematy, które łącznie można nazwać koneksjonistycznym implementacjonizmem. Obejmują one modele wiązania zmiennych, implementacje systemów produkcyjnych i logiki pierwszego rzędu, a następnie włączenie logiki modalnej i logiki rozmytej, wszystkie oparte na typowych modelach sieci koneksjonistycznych, takich jak wielowarstwowe sieci ze sprzężeniem zwrotnym z uczeniem się z propagacją wsteczną. Techniki stosowane do realizacji przetwarzania symbolicznego w takich sieciach znacznie się różnią w zależności od modelu. Wczesnym przykładem koneksjonistycznego implementacjonizmu jest rozproszony system produkcji koneksjonistycznej Touretzky’ego i Hintona (1988), który zaimplementował system produkcyjny przy użyciu modeli koneksjonistycznych. W DCPS istniała pamięć robocza, w której przechowywano fakty; istniały dwa składniki klauzuli, z których każdy służył do dopasowania jednego z dwóch warunków reguły (gdzie każda reguła była ograniczona do dwóch warunków); istniał także komponent reguły, który służył do wykonania akcji reguły dopasowującej zmieniającą pamięć roboczą; ponadto zastosowano komponent wiązania w celu wymuszenia ograniczeń dotyczących zmiennych, które mogą istnieć w regule. Każdy z tych komponentów został zaimplementowany jako sieć koneksjonistyczna. Patrz rysunek 5.1. Ogólnie rzecz biorąc, był to złożony modułowy system koneksjonistyczny, zaprojektowany specjalnie do realizacji ograniczonego systemu produkcyjnego, jako demonstracja możliwości implementacji złożonych systemów symbolicznych w klasycznych modelach koneksjonistycznych z reprezentacją rozproszoną.



Oprócz tego podejścia zaproponowano i zbadano wiele innych metod i technik wdrażania rozumowania opartego na regułach i przetwarzania symbolicznego w sieciach neuronowych, w tym na przykład rekurencyjna pamięć autoasocjacyjna (RAAM), reprezentacja holograficzna i iloczyn tensorowy reprezentacji .

Emergentne przetwarzanie symboliczne w modelach koneksjonistycznych

Zarówno od strony koneksjonistycznej, jak i symbolistycznej niektórzy badacze uważają, że poznanie wysokiego poziomu, zwłaszcza to, które jest rozciągnięte w czasie lub obejmuje wyraźne rozumowanie werbalne, często można lepiej uchwycić poprzez zastosowanie bardziej symbolicznych ram (więcej na ten temat w następnym Sekcja). Jednak wielu koneksjonistów wierzy, słusznie lub nie, że procesy koneksjonistyczne leżą u podstaw wszystkich aspektów ludzkiego poznania. Wierzą, że ludzkie rozumowanie i rozwiązywanie problemów często wynika z wglądu lub intuicji lub bezpośrednio z percepcji, a zatem do uchwycenia wszystkich ich subtelności może być potrzebne podejście koneksjonistyczne, a nie symboliczne. Na przykład Elman (1990) pokazał, jak prawidłowości językowe mogą powstać bez wyraźnej reprezentacji reguł językowych i jak uczenie się może prowadzić do odkrycia wewnętrznych reprezentacji, które mogą uchwycić struktury językowe na podstawie współwystępowania słów. W prostych sieciach rekurencyjnych Elmana (SRN) w kroku czasu t do sieci został zaprezentowany sygnał wejściowy, który spowodował wzór aktywacji w warstwie ukrytej i wyjściowej. W kroku czasu $t + 1$ do sieci zaprezentowano kolejne wejście w sekwencji, a kopia aktywacji jednostek ukrytych w kroku czasu t jest również przesyłana z powrotem do jednostek ukrytych. Dlatego

też każdy wkład przekazywany do SRN był przetwarzany w kontekście tego, co miało miejsce wcześniej. W swoich eksperymentach Elman wytrenował SRN w przewidywaniu następnego słowa w zdaniu. Podczas uczenia wyniki sieci zbliżyły się do prawdopodobieństw przejścia między słowami w zdaniach. Na przykład po pierwszym rzeczowniku jednostki czasownika będą bardziej aktywne jako możliwe następne słowo, a czasowniki, które są zwykle kojarzone z tym konkretnym rzeczownikiem, będą bardziej aktywne niż te, które tego nie robią. Elman zbadał strukturę wewnętrznych reprezentacji SRN i stwierdził, że reprezentacje wewnętrzne są wrażliwe na różnice syntaktyczne, a także na szereg rozróżnień semantycznych. W ten sposób SRN była w stanie opracować reprezentacje bytów, które różniły się w zależności od kontekstu użycia, w przeciwieństwie do tradycyjnych reprezentacji symbolicznych, które zachowywały swoją tożsamość niezależnie od kombinacji, w jakie zostały umieszczone. Podsumowując, praca Elmana pokazała, jak proste sieci mogą uczyć się prawidłowości statystycznych w sekwencjach czasowych i w jaki sposób mogą wystarczyć do wytworzenia wielu zachowań, które lingwiści przypisują regułom gramatycznym. Jako inny przykład Miikkulainen (1993) pokazał, co zestaw połączonych ze sobą wielowarstwowych sieci wyprzedzających z algorytmem uczenia się z propagacją wsteczną może osiągnąć w zakresie rozumienia języka naturalnego. Zamiast po prostu wdrażać skrypty i ramki, jego modele pokazały, w jaki sposób te symboliczne reprezentacje (skrypty i ramki) mogą wyłonić się w wyniku uczenia się metodą propagacji wstecznej na podstawie danych szkoleniowych. Ponieważ reprezentacje te wyłoniły się z sieci neuronowych w wyniku uczenia się, można je było uogólnić na nowe sytuacje (przynajmniej w pewnym stopniu). Praca ta rozszerzyła zatem zakres reprezentacji symbolicznej w modelach koneksjonistycznych.

Hybrydowe modele koneksjonistyczne

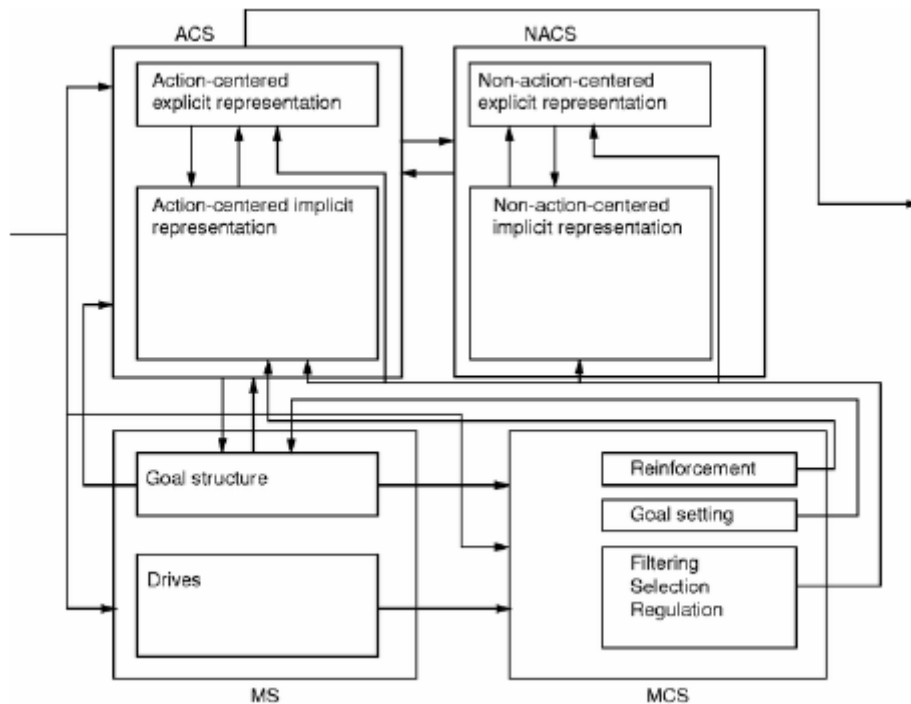
Idea modelu hybrydowego

W przeciwieństwie do koneksjonistycznego implementacjonizmu, hybrydowe modele koneksjonizmu można uznać za syntezę modeli koneksjonistycznych i tradycyjnych modeli symbolicznych. Celem takich modeli jest zatem odsunięcie nas od starej debaty na temat koneksjonizmu kontra symboliki w stronę nowej i produktywnej syntezy. W wyniku łączenia różnorodnych reprezentacji i procesów, symbolicznych lub koneksjonistycznych, stają się one bardziej wyraziste, potężniejsze, często wydajniejsze, a przez to bardziej przydatne zarówno w modelowaniu poznawczym, jak i zastosowaniach praktycznych (przemysłowych). Podstawowe uzasadnienie modeli hybrydowych można zwięźle podsumować jako „używanie odpowiedniego narzędzia do właściwego zadania”. Mówiąc dokładniej, powinno być oczywiste, że procesy poznawcze nie są jednorodne; prawdopodobnie stosuje się szeroką gamę reprezentacji i procesów, pełniących różne role i służących różnym celom. Niektóre procesy i reprezentacje poznawcze najlepiej uchwycą modele symboliczne, inne modele koneksjonistyczne. Istnieje zatem potrzeba pluralizmu w modelowaniu ludzkiego poznania, co w naturalny sposób prowadzi do rozwoju hybrydowych modeli koneksjonistycznych, zapewniających niezbędne narzędzia obliczeniowe i ramy pojęciowe. Na przykład, aby uchwycić pełen zakres ludzkich możliwości uczenia się poprzez umiejętności, architektura poznawcza musi uwzględniać zarówno wiedzę ukrytą, jak i jawną. Architekturę obejmującą oba procesy można wdrożyć obliczeniowo poprzez kombinację modeli symbolicznych (które wychwytyują wiedzę jawną) i modeli koneksjonistycznych (które wychwytyują wiedzę ukrytą). Rozwój inteligentnych systemów do różnych zastosowań praktycznych może również zyskać na odpowiednim połączeniu różnych technik, ponieważ obecnie żadna pojedyncza technika nie jest w stanie wykonać wszystkiego wydajnie i skutecznie. Obszernie argumentowano za względną przewagą modeli koneksjonistycznych nad modelami symbolicznymi. Zalety modeli koneksjonistycznych obejmują ogromną równoległość, możliwości uczenia się i tolerancję na błędy, jak wymieniono wcześniej. Zalety modeli symbolicznych obejmują wyraźną reprezentację i przetwarzanie, łatwość określania etapów przetwarzania symbolicznego i

wynikającą z tego precyzję przetwarzania. Mając na uwadze te względne zalety, stosunkowo łatwo można uzasadnić połączenie modeli koneksjonistycznych i symbolicznych. Niektóre istniejące dychotomie poznawcze są w tym względzie bardzo istotne. Na podstawie danych empirycznych psychologowie zaproponowali szereg dychotomii. Te dychotomie obejmują uczenie się ukryte i jawne, pamięć ukrytą i jawną, przetwarzanie automatyczne i kontrolowane, uczenie się przypadkowe i zamierzone i tak dalej. Istnieje również dobrze znane rozróżnienie wiedzy proceduralnej i deklaratywnej. Dowody na te dychotomie znajdują się w danych eksperymentalnych, które wyjaśniają różne dysocjacje i różnice w działaniu w różnych warunkach. Chociaż nie ma konsensusu co do szczegółów dychotomii, istnieje konsensus co do jakościowej różnicy między różnymi typami poznania. Co więcej, większość badaczy wierzy w konieczność uwzględnienia obu stron dychotomii, gdyż każda z nich pełni inną funkcję i jest przez to niezbędna. Wokół nich zbudowano architektury poznawcze, obejmujące zarówno techniki koneksjonistyczne, jak i symboliczne niektóre z tych dychotomii. W nawiązaniu do powyższych dychotomii Smolensky (1988) zaproponował bardziej abstrakcyjne rozróżnienie przetwarzania pojęciowego i subkonceptualnego i powiązał to rozróżnienie z modelem koneksjonistycznym i symbolicznym. Przetwarzanie pojęciowe obejmuje wiedzę, która posiada następujące cechy: (1) publiczny dostęp, (2) niezawodność i (3) formalność. To właśnie ich zdaniem uchwycą modele symboliczne. Z drugiej strony istnieją inne rodzaje zdolności poznawczych, takie jak umiejętności i intuicja, które nie dają się wyrazić w formach językowych i nie odpowiadają powyższym kryteriom. Zdaniem Smoleńskiego i wielu innych badaczy próby modelowania takich zdolności w kategoriach symbolicznych były daremne i należy je postrzegać jako znajdujące się na innym poziomie poznania – poziomie subpojęciowym. Poziom ten lepiej radzą sobie modele koneksjonistyczne, które przewyższają pewne poważne problemy, jakie napotykają modele symboliczne w modelowaniu przetwarzania subkonceptyjnego. Zatem połączenie tych dwóch typów modeli może prowadzić do znacznych korzyści w uchwyceniu pełnego zakresu zdolności poznawczych. Idee te stanowią podstawę do budowania hybrydowych modeli koneksjonistyczno-symbolicznych.

Przykłady modeli hybrydowych

Przykładem hybrydowych modeli koneksjonistyczno-symbolicznych jest CLARION, który składa się z dwóch poziomów: poziomu symbolicznego i poziomu koneksjonizmu. Obydwa poziomy działają raczej niezależnie, ale ich wyniki są połączone. Poziom koneksjonistyczny składa się z sieci neuronowych, które działają poprzez rozprzestrzenianie aktywacji i uczą się w oparciu o metody uczenia się przez wzmacnianie. Poziom symboliczny działa według symbolicznych zasad. Dzięki integracji wyników obu typów procesów model był w stanie uchwycić różnorodne dane dotyczące uczenia się, rozumowania i rozumowania człowieka. W szczególności CLARION to architektura integracyjna złożona z szeregu odrębnych podsystemów, z podwójną reprezentacyjną strukturą w każdym podsystemie. Jego podsystemy obejmują podsystem skoncentrowany na działaniu (ACS), podsystem nieskoncentrowany na działaniu (NACS), podsystem motywacyjny (MS) i podsystem metakognitywny (MCS). Rolą podsystemu skoncentrowanego na działaniu jest kontrolowanie działań, niezależnie od tego, czy dotyczą one zewnętrznych ruchów fizycznych, czy wewnętrznych operacji umysłowych. Rolą podsystemu nieskoncentrowanego na działaniu jest utrzymywanie wiedzy ogólnej. Rolą podsystemu motywacyjnego jest zapewnienie podstawowych motywacji do percepcji, działania i poznania, w zakresie zapewniania impulsu i informacji zwrotnej. Rolą podsystemu metapoznawczego jest monitorowanie i modyfikowanie działania pozostałych podsystemów. Każdy z tych oddziałujących na siebie podsystemów składa się z dwóch poziomów reprezentacji. W każdym podsystemie poziom najwyższy (symboliczny) koduje wiedzę jawną, a poziom dolny (koneksjonistyczny) koduje wiedzę ukrytą. Rozróżnienie wiedzy ukrytej i jawnej oparto na danych psychologicznych.



Podsystem skoncentrowany na działaniu jest centralną częścią CLARION. W nim proces podejmowania decyzji o działaniu jest zasadniczo następujący: Obserwując aktualny stan świata, dwa poziomy procesów (ukryte i jawne) podejmują odrębne decyzje zgodnie z własną wiedzą, a ich wyniki są „zintegrowane”. W ten sposób dokonuje się ostatecznego wyboru akcji, a następnie akcja jest wykonywana. Akcja w jakiś sposób zmienia świat. Porównując zmieniony stan świata ze stanem poprzednim, system uczy się (np. zgodnie z uczeniem przez wzmacnianie). Następnie cykl się powtarza. Na najniższym (koneksjonistycznym) poziomie rozwija się wiedza ukryta (procedury reaktywne). Procedury reaktywne opracowane w wyniku uczenia się przez wzmacnianie mogą wykazywać zachowania sekwencyjne bez wyraźnego (symbolicznego) planowania. Na najwyższym (symbolicznym) poziomie podsystemu skoncentrowanego na działaniu wiedza jawna jest ujmowana w postaci reguł symbolicznych. Istnieje wiele sposobów uczenia się wiedzy jawnej, w tym niezależne uczenie się poprzez testowanie hipotez i „uczenie się oddolne”. CLARION był używany do różnych celów, w tym na przykład do zrozumienia uczenia się umiejętności ludzkich. Za pomocą CLARION symulowano i wyjaśniano wiele dobrze znanych zadań związanych z uczeniem się umiejętności, począwszy od prostych umiejętności reaktywnych po złożone umiejętności poznawcze. Ponadto w CLARION zamodelowano i wyjaśniono wiele zadań związanych z rozumowaniem, zadaniami metakognitywnymi, zadaniami motywacyjnymi i zadaniami związanymi z interakcjami społecznymi. Uwzględniając różne dane psychologiczne, CLARION zapewnia szczegółowe wyjaśnienia danych ludzkich, które rzucają nowe światło na zjawiska poznawcze. Na przykład, rozliczając zadania związane z uczeniem się umiejętności, CLARION przypisał pewne różnice w wydajności różnym sposobom interakcji między procesami ukrytymi i jawnymi (odpowiednio na dwóch poziomach). Za pomocą tego prostego pojęcia firma CLARION wyjaśniła szeroką gamę danych dotyczących uczenia się umiejętności ludzkich, które nie były wcześniej wyjaśniane w jednolity sposób. Oprócz powyższej pracy istnieje kilka innych podejść, które na różne sposoby łączą metody koneksjonistyczne i symboliczne.

Zagadnienia dotyczące modeli hybrydowych

Przyjmując i rozwijając hybrydowe modele koneksjonistyczno-symboliczne, należy odpowiedzieć na wiele pytań, aby wypracować podejście oparte na zasadach. Najważniejsze z nich to:

* Jakie są względne zalety i wady poszczególnych podejść do opracowywania modeli hybrydowych?

* Na ile poznawczo wiarygodne jest każde z tych podejść?

Mówiąc dokładniej, pojawia się szereg pytań dotyczących architektur modeli hybrydowych, a także uczenia się w tych modelach. Po pierwsze, modele hybrydowe obejmują zwykle wiele różnych typów procesów i reprezentacji, a co za tym idzie, wiele heterogenicznych mechanizmów oddziałujących na siebie w złożony sposób. Konieczne jest rozważenie sposobów strukturyzowania tych różnych komponentów, czyli innymi słowy, rozważenie architektur. Niektóre problemy związane z architekturą obejmują:

* Jak zdecydować, czy reprezentacja określonej części architektury powinna mieć charakter symboliczny, lokalistyczny czy rozproszony?’

* Jakie są właściwe i oparte na zasadach sposoby przezwyciężenia prawdopodobnej heterogeniczności modeli hybrydowych?’

* Jak reprezentacja i uczenie się współdziałają w modelach hybrydowych (ponieważ w takich modelach oba aspekty będą prawdopodobnie bardziej złożone)?’

* W jaki sposób konstruuje się różne części modelu hybrydowego, aby osiągnąć optymalne wyniki (w jakimkolwiek sensie właściwym dla danego zadania)?

Po drugie, chociaż wiadomo, że modele czysto koneksjonistyczne, które stanowią część każdego modelu hybrydowego, wyróżniają się zdolnościami uczenia się, hybrydyzacja utrudnia uczenie się. W pewnym sensie modele hybrydowe dziedziczą trudności w uczeniu się od strony symbolicznej i rezygnują w pewnym stopniu z przewagi, jaką mają pod tym względem modele czysto koneksjonistyczne. Niektóre z zagadnień związanych z nauką obejmują:

* Jaki rodzaj uczenia się można przeprowadzić w każdym typie architektury hybrydowej?’

* W jaki sposób w modelach hybrydowych można się nauczyć złożonych struktur symbolicznych, takich jak reguły, ramki i sieci semantyczne? (Jest to szczególnie problem w przypadku wysoce ustrukturyzowanych modeli hybrydowych, w których uczenie się jest szczególnie trudne.)’

* Jaki powinien być związek między symbolicznymi metodami uczenia się, metodami pozyskiwania/nabywania wiedzy i algorytmami uczenia się sieci neuronowych przy opracowywaniu modeli hybrydowych?’

* W jaki sposób można uformować każdy typ architektury za pomocą różnych kombinacji wyżej wymienionych metod?

Pomimo różnorodności, jaka istnieje w badaniach nad hybrydowymi koneksjonistycznymi modelami symbolicznymi, istnieje wyraźny temat jednoczący: poszukiwanie modeli obliczeniowych, które łączą techniki symboliczne i koneksjonistyczne w celu osiągnięcia syntezy i synergii dwóch pozornie różnych paradygmatów. Różne proponowane metody, modele i architektury odzwierciedlają powszechne przekonanie, że metody koneksjonistyczne i symboliczne można z pożytkiem zintegrować oraz że taka integracja może prowadzić do postępu w rozumieniu poznania i inteligencji.

Modele hybrydowe w modelowaniu poznawczym

Modele hybrydowe, takie jak CLARION, zostały wykorzystane w celu rozwiązania szerokiego zakresu problemów

różnorodne zagadnienia z zakresu kognitywistyki i sztucznej inteligencji, obejmujące uczenie się człowieka, rozumowanie, rozwiązywanie problemów, kreatywność, dynamikę motywacyjną, procesy metapoznawcze, a przede wszystkim ludzką świadomość. Modelowanie i wyjaśnianie ludzkiego uczenia się zostało już poruszone. Jeśli chodzi o zrozumienie ludzkiego rozumowania, wykazano, że CLARION jest w stanie uchwycić zarówno ukryte, jak i jawne rozumowanie oraz ich interakcje (Sun 1994; Sun i Zhang 2006). Dzięki szczegółowemu modelowaniu danych dotyczących ludzkiego rozumowania projekt CLARION zapewnił interpretacje danych dotyczących ludzkiego rozumowania na podstawie interakcji procesów ukrytych i jawnych, co doprowadziło do nowych spostrzeżeń na temat ludzkiego rozumowania, wykraczających poza te, które można uzyskać w wyniku wdrożenia ograniczonej formy systemu produkcyjnego, jak w omówionej wcześniej koneksjonistycznej pracy implementacyjnej. W odniesieniu do zrozumienia ludzkiej świadomości CLARION okazał się przydatny w generowaniu hipotez wyjaśniających. W szczególności wysunięto hipotezę, że różnica reprezentacyjna pomiędzy dwoma poziomami reprezentacji w CLARION ma uchwycić zasadniczą różnicę pomiędzy świadomością i nieświadomością (Sun 1999), która nie mogłaby zostać wymyślona bez dostępności narzędzi technicznych, zarówno symbolicznych, jak i nieświadomych. podejścia koneksjonistyczne. Podobnie architektura ACT-R Johna Andersona została wykorzystana do zrozumienia i modelowania interakcji percepcji i poznania na różne sposoby (

Uwagi końcowe

W tej części omówiono trzy typy modeli koneksjonizmu: koneksjonizm klasyczny, koneksjonalizm przetwarzania symbolicznego (implementacjonizm) i koneksjonalizm hybrydowy. Choć klasyczny koneksjonizm przyniósł ciekawe i nowatorskie pomysły, ogranicza go jego prostota i jednolitość. Przetwarzanie symboliczne w ramach takich modeli koneksjonistycznych zostało szeroko zbadane, ale jak dotąd przyniosło ograniczone wyniki. Dlatego też konieczne wydaje się, przynajmniej na krótką metę, opracowanie hybrydowych modeli koneksjonistycznych, uwzględniających metody symboliczne (i ewentualnie inne, np. metody bayesowskie czy logiki rozmytej). Patrząc w przyszłość tej dziedziny, można dostrzec pewne trendy. Na przykład, chociaż istnieją specyficzne dla danej dziedziny zastosowania do modelowania różnych procesów poznawczych, takich jak przetwarzanie języka naturalnego, rozumowanie i podejmowanie decyzji, pamięć i uczenie się, widzenie itd., mogą również istnieć bardziej integracyjne modele koneksjonistyczne, które przekraczają granice wąskich domeny i funkcjonalności. Inną prawdopodobną tendencją jest to, że modele koneksjonistyczne mogą być w coraz większym stopniu powiązane ze statystycznym (bayesowskim) podejściem do uczenia się i rozumowania, które może zapewniać strategię oparte na zasadach (ale prawdopodobnie wiąże się również z wyższymi kosztami obliczeniowymi). Innym, powiązaniem trendem jest rosnąca hybrydyzacja modeli koneksjonistycznych; bardziej modele koneksjonistyczne mogą zawierać komponenty symboliczne, komponenty logiki rozmytej i inne komponenty, które wykraczają poza sferę klasycznych modeli koneksjonistycznych. I odwrotnie, modele symboliczne mogą również w coraz większym stopniu uwzględniać techniki i podejścia koneksjonistyczne. Wreszcie modele koneksjonistyczne mogą być w coraz większym stopniu powiązane z pracami nad biologicznymi systemami neuronowymi, w tym z rozliczaniem danych z obrazowania mózgu. W szczególności w przyszłości więcej modeli może być inspirowanych biologicznymi sieciami neuronowymi, które są nie tylko wykonalne obliczeniowo, ale także biologicznie realistyczne, co może pomóc w rozwoju tej dziedziny.