

Wstęp

Uczenie się za pomocą systemów sztucznej inteligencji – co zazwyczaj nazywam uczeniem maszynowym – ma bogatą historię, a w ciągu ostatnich dwudziestu lat dziedzina ta przeżyła coś w rodzaju renesansu. Uczenie maszynowe składa się głównie z różnorodnego zestawu algorytmów i technik, które zostały zastosowane do problemów z szerokiego zakresu dziedzin. Jakikolwiek przegląd metod i zastosowań będzie nieuchronnie niekompletny, przynajmniej na poziomie konkretnych algorytmów i technik. Istnieje wiele doskonałych wstępów do formalnych i statystycznych szczegółów algorytmów i technik uczenia maszynowego dostępnych gdzie indziej. Niniejsza część skupia się na uczeniu maszynowym jako ogólnym sposobie „myślenia o świecie” i zapewnia ogólną charakterystykę głównych celów uczenia maszynowego. Istnieje wiele filozoficznych obaw dotyczących uczenia maszynowego, ale po bliższym zbadaniu nie zawsze jest jasne, czy zastrzeżenia rzeczywiście przemawiają konkretnie przeciwko uczeniu maszynowemu. Wiele z nich wydaje się być raczej ukierunkowanych na uczenie maszynowe jako konkretną instancję jakiegoś bardziej ogólnego zjawiska lub procesu. Jedną z ogólnych zasad płynących z tej części jest taka, że uczenie maszynowe jest pod wieloma względami mniej niezwykłe i osobliwe, niż się czasem uważa.

Trzy szerokie klasy wnioskowania

Na bardzo wysokim poziomie można wyróżnić trzy różne, niekoniecznie wyczerpujące, strategie wnioskowania: analogiczną, dziedzinową i strukturalną. Jako przykład ogólności tej taksonomii można podać, że zarówno logika dedukcyjna, jak i indukcyjna są rodzajami wnioskowania strukturalnego. Wnioskowanie przez analogię ma na celu odwzorowanie jakiejś sytuacji lub problemu na najistotniejszych przykładach historycznych, dobrze znanych lub osobistych; następnie wyciąga się wnioski, korzystając z analogicznego mapowania, aby przełożyć wyniki historyczne na obecny problem. Problemy i rezultaty nie muszą być duże ani znaczące: jeśli ktoś ma wcześniejsze doświadczenia z włącznikami światła i żarówkami, może wyciągnąć analogiczne wnioski, aby podjąć decyzję o tym, jak włączyć światło, wchodząc do nowego pokoju. Jeśli analogie są odpowiednie, wówczas wnioskowanie przez analogie może wspierać wnioski dotyczące bardzo rzadkich sytuacji lub zjawisk lub wnioski z bardzo ograniczonych danych. Wiarygodność wnioskowania przez analogie jest jednak w dużym stopniu zależna od analogicznych przypadków i odwzorowań, a niewiele wiadomo na temat poszukiwania takich przypadków lub opracowywania odpowiednich odwzorowań. Jeśli chodzi o tematykę tego rozdziału, analogiczne wnioskowanie rzadko jest przeprowadzane przez maszynę; zazwyczaj w myślach wyciągamy analogiczne wnioski. Wnioskowanie specyficzne dla domeny wykorzystuje techniki specjalnie dostosowane do wiedzy o konkretnych problemach, środowiskach i reakcjach występujących w domenie. Używając wyspecjalizowanych algorytmów i ograniczeń, często można wyciągnąć dość rozbudowane wnioski, nawet przy ograniczonej ilości danych. Metody specyficzne dla danej dziedziny można jednak opracowywać i stosować jedynie po uzyskaniu znacznej wcześniejszej wiedzy w tej dziedzinie, co może wykluczać powszechne stosowanie takich metod. Uczenie maszynowe i wnioskowanie specyficzne dla domeny można często rozumieć jako część konkretnej domeny, a nie jako wyróżniającą się i nowatorską strategię wnioskowania. Ponadto, ponieważ każda technika musi wykorzystywać pewne informacje specyficzne dla domeny (np. możliwe wartości zmiennej), nie jest jasne, czy można narysować ostrą linię, aby dokładnie rozgraniczyć metody „specyficzne dla domeny”, chociaż istnieje wiele oczywistych wniosków algorytmów, które mają zastosowanie tylko w bardzo specyficznych sytuacjach. Wnioskowanie strukturalne wykorzystuje (względnie) algorytmy ogólnodomenowe, których powodzenie zależy od wewnętrznej struktury danych, a nie cech treści semantycznej danych. Oznacza to, że wnioskowanie strukturalne koncentruje się na związkach między zmiennymi, obiektami lub predykatami, a nie na jakichkolwiek ich nieodłącznych właściwościach. Ten typ wnioskowania ma z konieczności charakter ogólny, ponieważ

takie metody są wyraźnie zaprojektowane tak, aby nie wykorzystywać żadnych informacji dziedzinowych z wyjątkiem informacji „strukturalnych” o obiektach wnioskania (np. liczby wartości zmiennych, tego, czy przestrzenne położenie obiektów jest istotne, i tak NA). Zaleta wnioskania strukturalnego jest oczywista: metody można zastosować w dowolnej dziedzinie, w której występują odpowiednie cechy strukturalne i które można odkryć na podstawie danych. Metody te nie ograniczają się zatem do dziedzin, w których posiadamy znaczną wiedzę wcześniej, ani nie musimy mieć żadnego znaczącego doświadczenia z sytuacjami tego typu. Wady takich wnioskowań są również oczywiste: nie można wnioskować o mechanizmach specyficznych dla domeny (ponieważ informacje specyficzne dla domeny są wykluczone), a wnioskowanie na podstawie małych zbiorów danych może być dość trudne. Wnioskowanie strukturalne jest podstawą wielu i prawdopodobnie większości maszyn ramy i metody uczenia się, w tym wiele dobrze znanych, takie jak różne formy regresji, algorytmy uczenia się sieci neuronowych, takie jak propagacja wsteczna, oraz algorytmy uczenia się przyczynowego wykorzystujące sieci Bayesa. We wszystkich tych metodach algorytm działa poprzez wyodrębnianie i wykorzystywanie zależności strukturalnych między zmiennymi bez względu na znaczenie lub dziedzinę zmiennych. Na przykład, jeśli dokonujesz klasyfikacji za pomocą sztucznej sieci neuronowej, możesz otrzymać zbiór danych zawierający pomiary różnych cech widżetów, a także pewną kategorię docelową. Algorytm uczenia się sieci neuronowej (np. standardowa propagacja wsteczna) wykorzystuje następnie jedynie prawidłowości statystyczne w zbiorze danych, aby poznać odpowiednią strukturę między zmiennymi, którą można następnie wykorzystać do przewidzenia kategorii docelowej (np. „funkcjonalna” vs. „wadliwa”). ”) dla przyszłych widżetów. Dokładne „znaczenie” zmiennych nie ma znaczenia dla algorytmu uczenia się. W przypadku wszystkich tych metod nie trzeba dużo wiedzieć o dziedzinie podstawowej, aby je zastosować, chociaż informacje specyficzne dla domeny (np. zmienna X przyjmuje wartość przed zmienną Y) można zazwyczaj uwzględnić na różne sposoby. Ogólność metod uczenia maszynowego częściowo wyjaśnia ich popularność w stosunkowo nowych dziedzinach nauki, takich jak bioinformatyka, w których istnieje znaczna niepewność co do tego, jakie modele i metody są odpowiednie. Istnieje naturalny podział metod wnioskowania strukturalnego na metody logiczne i statystyczne. Metody logiczne zazwyczaj mają na celu modelowanie struktury w kategoriach relacji dedukcyjnych, być może uzupełnionych różnymi reprezentacjami braku precyzyjnej wiedzy o sytuacji. Metody często wykorzystują różne typy logiki modalnej, aby pomóc w reprezentowaniu i wnioskowaniu niepewnych możliwości. Najczęstszym zastosowaniem logicznych metod uczenia maszynowego jest wnioskowanie na podstawie wcześniejszej wiedzy, która koduje informacje strukturalne dotyczące konkretnej dziedziny. W tym rozdziale skupimy się bardziej na metodach statystycznych, które wykorzystują większe ilości danych do wnioskowania o zależnościach strukturalnych. Większość z tych metod wykorzystuje dane do określenia, które zmienne są istotne informacyjnie dla jakich innych, a następnie wykorzystuje brak takich powiązań informacyjnych do opracowania prostych, ale dokładnych modeli o znacznej mocy predykcyjnej. Istnieje oczywista różnica pomiędzy algorytmami uczącymi się a wyuczonym modelem. Na przykład pewna konkretna sieć neuronowa (z wagami połączeń itp.) jest modelem wyuczonym; propagacja wsteczna to algorytm, za pomocą którego uczymy się modelu. Algorytmy uczenia maszynowego są rodzajem wnioskowania strukturalnego, ponieważ uczenie się nie ma wewnętrznego odniesienia do badanej dziedziny; w szczególności algorytm uczenia nie wykorzystuje (istotnej) informacji semantycznej o zmiennych. Ta obserwacja pozostawia otwartą kwestię, czy wyuczony model rzeczywiście ma interesującą treść semantyczną. Do tego pytania powrócimy w dalszej części tego rozdziału. W międzyczasie jednak, myśląc o tych procesach, należy pamiętać o rozróżnieniu pomiędzy algorytmem uczenia się a wyuczonym modelem.

Przybliżona taksonomia uczenia maszynowego

Założmy, że mamy zbiór danych D : zbiór punktów danych, z których każdy zawiera pomiary wartości zmiennych V dla konkretnej osoby lub jednostki. Ze zbiorem danych może wystąpić wiele komplikacji: wartości zmiennych mogą zostać przypisane lub wywnioskowane; może nie być oczywiste, jak określić zmienne lub osoby; punkty danych mogą nie być niezależne (np. jeśli istnieją dane szeregów czasowych); mogą istnieć niezmiernie czynniki, które wpływają na zmienne w zbiorze danych; i tak dalej. Jeśli chodzi o opracowanie przybliżonej taksonomii metod uczenia maszynowego, te subtelności są w dużej mierze nieistotne. Biorąc to pod uwagę, zasadniczo wszystkie metody uczenia maszynowego zakładają, że sytuacja jest „dobrze przedstawiona” na różne sposoby, na przykład przy użyciu dobrze zdefiniowanych zmiennych. Na najbardziej ogólnym poziomie algorytmy uczenia maszynowego można podzielić na dwie klasy (z małym środkiem) w zależności od tego, czy algorytm wymaga specyfikacji zmiennej docelowej w zbiorze danych. Algorytmy nadzorowanego uczenia się zakładają, że pewna zmienna X jest wyznaczona jako cel przewidywania, wyjaśniania lub wnioskowania oraz że wartości X w zbiorze danych stanowią „podstawowe” wartości uczenia się. Oznacza to, że algorytmy uczenia się nadzorowanego wykorzystują znane wartości X do określenia, czego należy się nauczyć. Najpopularniejszy typ algorytmu nadzorowanego uczenia się ma na celu opracowanie modelu klasyfikacji lub kategoryzacji: mając informacje o różnych osobach i kategoriach, do których one należą, algorytm tworzy wyuczony model, który można wykorzystać do przewidywania przynależności do kategorii nowych osób. Można na przykład przewidzieć, które gadżety produkowane w fabryce najprawdopodobniej zawiodą. Jeśli dysponuje się danymi na temat wydajności wielu różnych widżetów, a także pomiarami innych istotnych funkcji, można użyć algorytmu uczenia maszynowego, aby nauczyć się modelu, który będzie przewidywał wydajność przyszłych widżetów. W odpowiednich warunkach algorytmy klasyfikacji mogą generować modele, które pozwalają na dokonywanie nowych, uzasadnionych uogólnień na temat grup w oparciu o relacje między cechami. Algorytmy klasyfikacji można również wykorzystać do rozpoznawania lub identyfikacji poprzez klasyfikację w „kategorie” składającą się z dokładnie jednego elementu. Przykłady algorytmów uczenia się nadzorowanego obejmują algorytmy uczenia się dla sztucznych sieci neuronowych, drzew decyzyjnych i maszyn wektorów nośnych; liczne formy regresji; i większość metod uczenia się przez wzmacnianie. Algorytmy uczenia się bez nadzoru nie wyróżniają żadnych konkretnych zmiennych jako celu ani celu, a zatem mają na celu zapewnienie ogólnej charakterystyki pełnego zbioru danych. Prawdopodobnie najpowszechniejszym zastosowaniem uczenia się bez nadzoru są algorytmy grupowania: rozdzielanie różnych osób na „naturalne” grupy według tej lub innej metryki. Algorytmy te czasami rysują stosunkowo arbitralne linie między jednostkami, ale mogą być dość skuteczne w odkrywaniu grup, gdy one faktycznie istnieją. Można na przykład zmierzyć postawy ludzi w odniesieniu do różnych kwestii politycznych, a następnie ustalić, czy istnieją naturalne grupy, które można zdefiniować na podstawie tych przekonań. Dane wyjściowe algorytmu grupowania mogą w pewnych warunkach służyć później jako zmienna docelowa dla algorytmu uczenia się nadzorowanego. Zwykle dość trudno jest zweryfikować wyniki lub model algorytmu uczenia się bez nadzoru, właśnie dlatego, że zwykle nie ma „podstawowej prawdy”, z którą można porównać wydajność modelu. Rzadko istnieje jeden poprawny, prawdziwy sposób grupowania różnych jednostek, są tylko lepsze i gorsze sposoby. Różne metody rozkładu prawdopodobieństwa lub szacowania gęstości również wchodzi w zakres uczenia się bez nadzoru. Standardowy pogląd na wyuczone modele jest taki, że ich treść semantyczna jest całkowicie statystyczna: powiązania między zmiennymi dostarczają informacji, które można wykorzystać do przewidywania, ale uważa się, że żadna inna treść semantyczna – na przykład struktura przyczynowa – nie może być przypisana tym modelom. Jednak w ostatnich latach nastąpił znaczny wzrost zainteresowania algorytmami uczenia maszynowego, które unikają stosowania założeń specyficznych dla domeny, ale tworzą wyuczone modele o bogatej zawartości semantycznej. W szczególności wyuczone modele można wykorzystać do przewidywania przyszłego zachowania lub cech pod wpływem interwencji lub manipulacji spoza systemu. Te algorytmy uczenia maszynowego są zazwyczaj

metodami uczenia się bez nadzoru, chociaż często chce się poznać strukturę przyczynową, aby wpłynąć na jakąś konkretną zmienną lub spowodować zmianę. Można się zastanawiać, jak możliwe jest takie wnioskowanie przyczynowe, biorąc pod uwagę całkowicie standardową maksymę w nauce i filozofii, że „korelacja nie jest przyczynowością”. Metody uczenia maszynowego służące do odkrywania przyczyn muszą przyjmować założenia o pewnej treści przyczynowej, ale zazwyczaj wykorzystują jedynie założenia ogólne dla danej dziedziny dotyczące sposobów powiązania przyczynowości i korelacji. Na przykład szeroko dyskutowane założenie przyczynowe Markowa stwierdza, że zmienna nie dostarcza informacji o swoich nieskutkach, jeśli znane są już wartości bezpośrednich przyczyn zmiennej. Założenie to ma treść przyczynową, lecz na bardzo wysokim poziomie ogólności. Ten typ uczenia się przyczynowego jest trudniejszy niż uczenie się czysto statystyczne (np. grupowanie, klasyfikacja, estymacja gęstości, przybliżanie funkcji), w którym po prostu próbuje się znaleźć powiązania informacyjne między zmiennymi. Z wyjątkiem bardzo nietypowych okoliczności, zbiór związków przyczynowych pomiędzy niektórymi cechami będzie ścisłym podzbiorem zbioru relacji informacyjnych. Oznacza to, że (prawie wszystkie) relacje przyczynowe są relacjami informacyjnymi, ale nie wszystkie relacje informacyjne są przyczynowe. Biorąc pod uwagę tę asymetrię w możliwościach uczenia się różnych typów modeli, można mieć nadzieję, że modele statystyczne wystarczą do wszystkich interesujących zastosowań. Informacje czysto statystyczne nie są jednak wystarczające, aby przewidzieć, kiedy system ulegnie zmianie, czy to z powodu działań lub polityki danej osoby, czy może z powodu załamania się struktury przyczynowej na różne sposoby. Informacje przyczynowe są potrzebne do przewidywania prawdopodobnych skutków większości interwencji, decyzji politycznych lub innych zewnętrznych zmian w systemie. Czasami musimy stawić czoła trudniejszemu problemowi uczenia się. Algorytmy uczenia maszynowego muszą równoważyć trzy czynniki: (1) złożoność wyuczonego modelu, która zapewnia większą dokładność w reprezentowaniu wejściowego zbioru danych; (2) możliwość uogólnienia wyuczonego modelu na nowe dane, co umożliwi wykorzystanie modelu w nowych kontekstach; oraz (3) wykonalność obliczeniowa uczenia się i korzystania z modelu, co jest warunkiem koniecznym, aby algorytmy miały wartość praktyczną. Złożoność świata będzie czasami większa niż złożoność modeli praktycznych w konkretnej sytuacji, co sugeruje, że żaden pojedynczy model nie będzie wystarczający. Można mieć nadzieję, że istnieją „podproblemy”, które nie są bardziej złożone niż dostępne modele. Jeśli tak jest, to z dużym prawdopodobieństwem można dowiedzieć się więcej o świecie poprzez (1) określenie zakresu podproblemów; (2) nauczenie się odpowiedniego modelu (lub modeli) dla każdego podproblemu; a następnie (3) integrowanie wyników modelu w sposób zasadniczy. Tę trzyetapową propozycję realizują różne techniki „metalearningu”. Jako ilustrację rozważmy przypadek wzmacniania dla prostej klasyfikacji binarnej: na przykład, czy widget jest uszkodzony, czy nie. Zamiast próbować nauczyć się całego modelu w jednym kroku, algorytm wzmacniający najpierw uczy się prostego modelu klasyfikacji, który działa dość dobrze, choć zazwyczaj nie tak dobrze, jak chce lub wymaga użytkownik. Następnie system wyodrębnia wszystkie przypadki, dla których ten prosty model dokonał błędnej prognozy, i uczy się drugiego modelu klasyfikacji właśnie dla tych przypadków. Wyniki tych dwóch modeli można integrować na różne sposoby, aby uzyskać klasyfikator dla wszystkich przypadków. Ten ujednoczony klasyfikator będzie błędnie prognozował dla innych przypadków, więc można nauczyć się trzeciego klasyfikatora dla tych błędnie sklasyfikowanych przypadków, zintegrować nowy klasyfikator z ujednoczonym i iterować.¹ W ten sposób wzmacnianie buduje ujednoczony klasyfikator składający się z liczby „słabych” klasyfikatorów, z których każdy koncentruje się na dokładnej klasyfikacji zredukowanego podzbioru danych. Modele hierarchiczne, takie jak mieszaniny ekspertów (np. Jordan i Jacobs 1994), funkcjonują podobnie. Żaden przegląd uczenia maszynowego nie byłby kompletny bez omówienia uczenia się bayesowskiego. Algorytm uczenia się Bayesa wymaga określenia (prawdopodobnie nieskończonego) zbioru możliwych hipotez lub modeli, a także rozkładu prawdopodobieństwa – „wcześniejszego rozkładu prawdopodobieństwa” – względem tych hipotez. Po

otrzymaniu danych algorytm uczący wykorzystuje regułę Bayesa do określenia prawidłowego (za pomocą rachunku prawdopodobieństwa) rozkładu prawdopodobieństwa w oparciu o hipotezy na podstawie tych danych. Rozumowanie bayesowskie wychwytuje intuicję, że przekonania po obserwacji pewnych danych należy wyrażać na podstawie prawdopodobieństwa każdego możliwego wyjaśnienia na podstawie tych danych. Wyrażona w języku potocznym reguła Bayesa stwierdza: Prawdopodobieństwo hipotezy po zaobserwowaniu pewnych danych $[P(H | D)]$ jest równe (1) wcześniejszemu prawdopodobieństwu hipotezy $[P(H)]$, pomnożonemu przez (2) prawdopodobieństwo zobaczenia takich danych, gdyby hipoteza rzeczywiście była prawdziwa $[P(D | H)]$ podzielone przez (3) prawdopodobieństwo zobaczenia tych danych w pierwszej kolejności $[P(D)]$. Pomysł, że uczenie się bayesowskie jest racjonalne, ma długą historię filozoficzną. Bayesjanizm miał jednak przez wiele lat stosunkowo niewielkie znaczenie praktyczne z powodu prostego faktu: z wyjątkiem przykładów zabawek obliczenia wymagane do uczenia się metodą bayesowską szybko stają się zbyt trudne do wykonania analitycznego lub ręcznego. Rozwój nowoczesnych komputerów cyfrowych umożliwił przeprowadzenie lub przybliżenie uczenia się bayesowskiego w bardziej realistycznych sytuacjach, w związku z czym bayesyzm ponownie pojawił się jako dominujący temat w uczeniu się. Wiele algorytmów uczenia maszynowego można postrzegać jako implementujące lub przybliżające uczenie się bayesowskie przy różnych założeniach lub ograniczeniach dotyczących przestrzeni hipotez, wcześniejszego rozkładu prawdopodobieństwa, funkcji wiarygodności i tak dalej.

Zakres i ograniczenia uczenia maszynowego

Podobnie jak w przypadku uczenia się przez ludzi, wartość uczenia maszynowego jest mniejsza w wynikach, a większa w sposobie, w jaki wyniki mogą zostać wykorzystane do przyszłych zadań: przewidywania, planowania, klasyfikacji, rozpoznawania i tak dalej. Jako społeczność wiemy, jak wiele zrobić dzięki uczeniu maszynowemu. Uczenie maszynowe stanowi dużą część współczesnej informatyki i istnieje wiele różnych algorytmów i technik, które nadają się do szerokiego zakresu warunków. Do grupowania, klasyfikacji i uczenia się przyczynowego istnieją oczywiście algorytmy dla prostych sytuacji: zbiory danych zawierające wszystkie istotne zmienne, czyste pomiary, proste zależności (np. liniowe) i brak brakujących punktów danych. Istnieją jednak również algorytmy odporne na zmiany we wszystkich tych wymiarach: zaszumione dane, niezmierzone zmienne, złożone relacje, brakujące dane, błąd w doborze próby i tak dalej. Istnieje wiele historii sukcesu każdego z tych algorytmów w zastosowaniach w świecie rzeczywistym. Istnieją również algorytmy do obsługi danych szeregów czasowych, a w szczególności do prowadzenia monitorowania systemów i wykrywania uszkodzeń. Istnieją dość potężne algorytmy klasyfikacji tekstu i obrazów, które są wysoce wyspecjalizowane do tych celów (choć zazwyczaj nadal nie dorównują ludzkim możliwościom pod względem dokładności). Fuzja informacji – integracja informacji z wielu różnych źródeł – stała się ostatnio centralnym elementem uczenia maszynowego w świecie rzeczywistym. Jednocześnie znane są teoretyczne ograniczenia uczenia maszynowego, a wiele z nich odzwierciedla ograniczenia uczenia się przez człowieka. Na przykład, jeśli dane są zbyt zaszumione – jeśli są w zasadzie losowe – wówczas uczenie się będzie prawie niemożliwe. Algorytmy uczenia maszynowego wykorzystują wnioskowanie strukturalne, więc jeśli w danych nie ma wzorców, nie można niczego wywnioskować. Uczenie się wymaga również pewnych różnic w świecie, czy to między jednostkami, czy między czasami, czy między miejscami. Algorytmy uczenia maszynowego nie mogą dowiedzieć się niczego o funkcji o stałej wartości, ponieważ nie ma się czego uczyć: stała funkcja jest zawsze taka sama. I chociaż w niektórych sytuacjach nauka jest zdecydowanie łatwiejsza niż w innych, w najgorszym przypadku nauka jest prawie zawsze trudna. Mówiąc dokładniej, zasadniczo wszystkie interesujące problemy związane z uczeniem maszynowym są na tyle trudne, że w najgorszym przypadku wymagają (jak sądzimy) algorytmu z wykładniczo wieloma krokami obliczeniowymi. Bardziej interesującym ograniczeniem uczenia maszynowego jest niemożność, nawet w pozornie łatwych warunkach, wywnioskowania cech

struktury przyczynowej lub informacyjnej jednostki na podstawie pomiarów na poziomie grupy. Załóżmy, że chcemy dowiedzieć się czegoś o jednostce (np. jak wykształcenie wpływa na późniejsze dochody?), ale mierzymy cechy tylko na poziomie grupy (np. średnie w różnych grupach wykształcenia, dochody i inne istotne zmienne). Załóżmy dalej, że każda jednostka ma dokładnie ten sam typ podstawowych relacji (choć niekoniecznie te same wartości), a cechy na poziomie grupy są prostymi, deterministycznymi funkcjami cech na poziomie jednostki (np. Wartość średnia lub całkowita). Nawet przy tych silnych założeniach upraszczających istnieje wiele interesujących przypadków, w których relacje informacyjne między cechami na poziomie grupy nie są takie same, jak relacje między odpowiadającymi im cechami na poziomie indywidualnym. Oznacza to, że wyuczony model cech na poziomie grupy niekoniecznie jest taki sam, jak model jednostki, nawet jeśli każda osoba ma ten sam model. Możliwość ta stwarza poważne wyzwanie metodologiczne w zakresie wykorzystania uczenia maszynowego w dziedzinach, w których skupiają się przede wszystkim jednostki, ale gromadzenie danych odbywa się głównie w przypadku grup (np. części ekonomii, inne nauki społeczne i bioinformatyka).

Filozoficzne wyzwania stojące przed uczeniem maszynowym

Uczenie maszynowe jest głównym obszarem badań w informatyce i statystyce, a wiele, a być może prawie wszystkie, najważniejsze problemy uczenia maszynowego mają charakter obliczeniowy i algorytmiczny (np. „czego można się nauczyć w określonych warunkach?” lub „czy ten algorytm może działać szybciej?”), a nie koniecznie filozoficzne. Nawet koncepcje uczenia maszynowego, które mogą wydawać się filozoficzne, często okazują się mniej filozoficzne, niż mogłoby się wydawać. Jako jeden przykład rozważmy tak zwane twierdzenia „nie ma darmowego lunchu”, które czasami są potocznie określane jako: „Algorytmy odnoszą sukcesy tylko wtedy, gdy są „dostrojone” do swoich potrzeb. domena; nie ma uniwersalnych algorytmów uczenia się.” To sformułowanie sugeruje różne argumenty filozoficzne, ale wszystko opiera się na niezrozumieniu faktycznych twierdzeń. Na przykład można pokusić się o stwierdzenie, że uczenie maszynowe jest bezcelowe, ponieważ można by pomyśleć, że twierdzenia o zakazie darmowego lunchu implikują, że właściwy wybór algorytmu wymaga znajomości podstawowej prawdy, co wyeliminowałoby potrzebę uczenia się. Ten sugerowany argument nie rozumie sensu, w jakim żaden algorytm nie ma przewagi nad innymi. Twierdzenia o zakazie darmowego lunchu są pod wieloma względami po prostu precyzyjnym stwierdzeniem starożytnej sceptycznej obserwacji, że każda przyszłość jest zgodna z przeszłością. Jeśli jakkolwiek przyszłość jest możliwa, biorąc pod uwagę przeszłe obserwacje, to żaden algorytm uczenia się nie ma żadnej przewagi nad innymi. Ale teraz można wyraźnie zobaczyć, że prawie każde ograniczenie przestrzeni możliwości wystarczy, aby pokonać twierdzenia o zakazie darmowego lunchu; na przykład pojedyncze założenie o słabej regularności może wystarczyć do zdefiniowania „dziedziny” i wybrania uprzywilejowanej klasy lepszych algorytmów. Z pewnością nie trzeba a priori znać rzeczywistej, leżącej u podstaw prawdy. Oczywiście, jak w przypadku wielu złych argumentów, w tym sugerowanym zarzucie jest ziarno prawdy. Każda interesująca metoda uczenia maszynowego przyjmuje założenia dotyczące natury świata, a algorytmy mogą łatwo zawieść, jeśli te założenia okażą się fałszywe. Ważną częścią uczenia maszynowego jest sprawdzenie, czy założenia algorytmu rzeczywiście są spełnione, przynajmniej w przybliżeniu (np. poprzez sprawdzenie, czy rozkład danych jest w przybliżeniu gaussowski). Takich testów często brakuje zarówno w praktyce, jak i w retoryce uczenia maszynowego. Jeśli odpowiednie założenia są fałszywe, należy zastosować inne metody, które nie przyjmują tych założeń, nawet jeśli te inne metody będą zazwyczaj odpowiednio słabsze. Błędem jest myślenie o uczeniu maszynowym jako o „czarnej skrzynce”, która po prostu przyjmuje dane jako dane wejściowe i zwraca prawdę. Praktyka uczenia maszynowego jest natomiast znacznie bliższa wykorzystaniu statystyki w nauce – jako narzędzia do dokładniejszego badania struktury danych. Do konkretnego zadania należy wybrać odpowiednie narzędzie (tj. algorytm uczenia maszynowego), a narzędzi można używać z różnym

stopniem umiejętności (np. interpretując wynik algorytmu na różne sposoby). Można mieć nadzieję na wyrafinowany system, który będzie w stanie przyjąć dane wejściowe, określić najlepszy algorytm dla tego typu danych, a następnie zastosować algorytm, ale taki metauczący się obecnie pozostaje w dużej mierze nadzieją. Jednym z najrzadziej omawianych „założeń” algorytmów uczenia maszynowego jest to, że wszystkie wymagają dostarczenia dobrze określonych zmiennych z precyzyjnymi, być może nieskończonymi, zbiorami wartości. Zmienne nie muszą być numeryczne – mogą należeć do różnych kategorii, takich jak „duży” i „mały” – ale muszą być jasno określone: w pewnym sensie musi istnieć jakiś, być może nieznan, fakt dotyczący „prawdziwej” wartości każdej zmiennej dla każdego punktu danych. Uczenie maszynowe opiera się na wnioskowaniu strukturalnym, dlatego musi istnieć możliwość znalezienia wzorców i struktury w danych. Nie jest jasne, co w ogóle oznacza mówienie o „strukturze” wśród zmiennych, które nie są dobrze zdefiniowane. Zagadnienie to nie stanowi poważnego wyzwania w praktyce, ponieważ zasadniczo zawsze mamy do czynienia ze zbiorami danych wynikającymi z procesów pomiarowych określających zmienne; Różni realści metafizyczni również zazwyczaj nie będą się tym martwić. Jeśli jednak zadać sobie pytanie, czy istnieje jakakolwiek stabilna struktura podstawowa, którą należy zmierzyć, wówczas uczenie maszynowe będzie wydawać się przedsięwzięciem daremnym. Wszystkie obserwacje w tej sekcji rodzą naturalne pytanie: Jeśli uczenie maszynowe jest z grubsza analogiczne do statystyki, więc w jakim sensie jest to „uczenie się”? Bardziej kontrowersyjne sformułowanie brzmiałoby: czy maszyna w ogóle się uczy, czy też uczy się człowiek korzystający z algorytmu? Pytanie „czy się uczy?” ma dwa różne aspekty. Pytanie: Pierwsza dotyczy ogólnego zagadnienia filozoficznego, a druga wskazuje na luki w naszym rozumieniu poznania. Pierwsza obawa została najsłynniej przedstawiona przez Searle’a (1980) w jego chińskim pokoju, ale pojawiła się w wielu różnych formach (np. Harnad 1994). Argument zaczyna się od ogólnego twierdzenia, że obliczenia obejmują jedynie manipulację symbolami, podczas gdy poznanie wymaga czegoś więcej. „Coś więcej” poznania różni się w zależności od autora, ale często jest to pojęcie semantyczne, takie jak szczególna właściwość naszych pojęć lub „podstawa” dla nich. Manipulację symbolami charakteryzuje się zatem pojęciem czysto syntaktycznym: zgodnie z tym argumentem można poprawnie manipulować symbolami wyłącznie poprzez badanie cech fizycznej reprezentacji i bez zrozumienia semantyki lub znaczenia symbolu. Argument ten prowadzi następnie do wniosku, że obliczenia nie mogą być poznaniem, ponieważ temu pierwszemu brakuje jakiegokolwiek treści semantycznej ani oparcia w świecie, podczas gdy drugie z konieczności je posiada. Argument ten jest przedstawiany jako ogólny przeciwko idei „poznania jako obliczenia”, a uczenie maszynowe jest niewątpliwie odpowiednim rodzajem obliczeń. Sukcesy uczenia maszynowego wynikają z wnioskowania strukturalnego; metody te wykorzystują wzorce lub prawidłowości statystyczne w danych i są (względnie) obojętne na semantykę zmiennych wejściowych. Bardziej szczegółowa wersja poprzedniego argumentu prowadziła do wniosku, że „uczenie maszynowe” może być przydatne, ale nie może być prawdziwym uczeniem się, przynajmniej w takim stopniu, w jakim prawdziwe uczenie się wymaga poznania. Innymi słowy, może mieć miejsce pewne faktyczne uczenie się, ale to człowiek przetwarzający wyniki „uczenia maszynowego” jest tym, który to robi. Maszyna po prostu uwydatnia pewne wzorce w danych, chociaż może to być zadanie nietrywialne obliczeniowo. Stanowi to poważny zarzut co najmniej do tytułu „uczenie maszynowe” dla tych algorytmów, gdyż argument ten kwestionuje stosowanie wszelkich terminów kognitywnych w odniesieniu do operacji maszynowych. Jednakże argument ten nie wydaje się przedstawiać żadnego konkretnego zarzutu wobec samego uczenia maszynowego, ale raczej odnosi się do uczenia maszynowego jako działania maszyny. Oznacza to, że konkretna reakcja danej osoby na ten argument (lub akceptacja) – na przykład odwołanie się do jakiegoś procesu ugruntowania symboli, specjalnych mocy przyczynowych mózgu lub odrzucenie jakiejś przesłanki – będzie wynikać z bardziej ogólnych podstaw filozoficznych, a nie z głębszą refleksją na temat natury uczenia maszynowego w izolacji. Ogólny zarzut jest wyraźnie związany z uczeniem maszynowym, ale wydaje się równie jasne, że jakiegokolwiek rozwiązanie tego problemu musi

uwzględniać wiele kwestii wykraczających poza zakres uczenia maszynowego. Istnieje bardziej specyficzna forma pytania „czy to jest nauka?” zarzut, który bezpośrednio odnosi się do uczenia maszynowego. Wnikliwość i kreatywność są często uważane za centralną, jeśli nie najważniejszą cechę ludzkiego uczenia się. Wydaje się, że czasami nasza nauka zależy od kluczowych intuicyjnych skoków, których nie jesteśmy w stanie wyjaśnić ani przewidzieć. Introspekcyjnie wydaje się, że w twórczym wglądzie jest coś „niealgorytmicznego”. Algorytmy uczenia maszynowego wydają się nie zapewniać takiej możliwości wglądu, ponieważ są „tylko” złożonymi sekwencjami prostych operacji. Praktyka uczenia maszynowego nieuchronnie angażuje element ludzki w celu określenia i kontrolowania algorytmu, testowania różnych założeń i interpretacji wyników algorytmu. Obserwacje te nasuwają wniosek, że uczenie maszynowe (znowu) nie jest wcale prawdziwym uczeniem się, ale raczej szybkim, użytecznym wykrywaniem różnych wzorców w danych. Z tego powodu człowiek kontrolujący i weryfikujący algorytmy uczy się „prawdziwie”. Zarzut ten różni się znacząco od poprzedniego: nie formułuje się tu żadnych twierdzeń o niemożności poznania maszynowego, a jedynie o niepowodzeniu obecnych algorytmów uczenia maszynowego do poziomu prawdziwego uczenia się. Zarzut ten jest całkowicie zgodny z możliwością, że bardziej wyrafinowane i refleksyjne algorytmy, uzupełnione odpowiednią wiedzą podstawową, mogłyby przeprowadzić prawdziwe uczenie się. Zamiast tego argument opiera się na twierdzeniu, że żaden z obecnie dostępnych algorytmów nie spełnia tego standardu prawdziwego uczenia się. W poprzednim akapicie użyto wyrażen „ludzkie uczenie się” i „prawdziwe uczenie się” bez objaśnień; Zakładano po prostu, że czytelnik rozumie, co miał na myśli. Można się jednak zastanawiać, czy nasze zrozumienie natury ludzkiego uczenia się jest wystarczająco jasne, aby zapewnić standard „prawdziwego uczenia się”, którego uczenie maszynowe nie spełnia. Nie ma ugruntowanego modelu tego, jak ludzie faktycznie się uczą, dlatego nie jest jasne, jakie kryteria musiałyby zostać spełnione, aby algorytm maszynowy można było uznać za „uczący się”. Nie ma wątpliwości, że w pewnych sytuacjach uczenie się przez ludzi jest znacznie lepsze od uczenia maszynowego. Nasza zdolność do łączenia odmiennych fragmentów wiedzy i informacji ogólnych, czy to przez analogię, przez przypadek, czy w inny sposób, nie ma sobie równych w uczeniu maszynowym (pomimo wielu prób budowania systemów opartych na zdroworozsądkowym rozumowaniu). Jednak ta obserwacja nie wystarczy, aby stwierdzić, że w procesie uczenia się wykorzystujemy zupełnie inny proces; można jedynie stwierdzić, że w naszej nauce jest coś innego. Prawdopodobnym alternatywnym wyjaśnieniem jest to, że dysponujemy zasobem informacji, uprzedzeń i doświadczeń, któremu po prostu nie ma sobie równych współczesne systemy uczenia maszynowego. W uczeniu maszynowym baza danych zawierająca 10 000 punktów danych jest uważana za dużą; dziecko, które ma tylko jedno doświadczenie na godzinę czuwania (powiedzmy szesnaście dziennie), przekracza tę liczbę w czasie krótszym niż dwa lata. Jeśli ma jedno doświadczenie na minutę przebudzenia, przewyższa bazę danych w około jedenaście dni. Produkty ludzkiego uczenia się są lepsze (w pewnym sensie) od produktów uczenia maszynowego, ale procesy te nie muszą zasadniczo różnić się rodzajem, biorąc pod uwagę, że istnieją ogromne różnice w wiedzy podstawowej, dokładności błędów uprzedzeń, informacjach czasowych i semantycznych oraz Wkrótce. Co być może ważniejsze, istnieją znaczne luki w naszym rozumieniu procesów leżących u podstaw ludzkiego uczenia się. Nie wiemy o tych procesach na tyle, aby na tym etapie określić ich podobieństwo – lub odmienność – do algorytmów proponowanych w uczeniu maszynowym. Można by odpowiedzieć, że introspekcja dotycząca naszego własnego uczenia się dostarcza wszystkich wymaganych dowodów. Taka odpowiedź pomija obszerną literaturę psychologiczną wykazującą zawodność introspekcji w ujawnianiu szczegółów leżących u podstaw podstawowych procesów poznawczych. Może się zdarzyć, że w uczeniu się człowieka jest coś jakościowo innego, tak że algorytmy „uczenia maszynowego” nie zasługują na tę nazwę. Jednakże obecnie takie twierdzenia opierają się w dużej mierze na niewiedzy, a nie na pozytywnych dowodach istnienia różnicy. W poprzednich dyskusjach przyjęto potencjalnie problematyczne założenie: warto martwić się o konkretną etykietę przypisaną algorytmom uczenia maszynowego. Podstawowe

właściwości tych algorytmów – ich niezawodność, zbieżność, złożoność obliczeniowa itd. – są cechami rzeczywistymi niezależnie od używanej nazwy. Co więcej, algorytmy są już określane pod wieloma różnymi nazwami, takimi jak „eksploracja danych”, „statystyka stosowana”, „wyszukiwanie automatyczne” i tak dalej. Można zatem chcieć porzucić etykietę „uczenia maszynowego”, ponieważ w oczywisty sposób nie ma to znaczenia dla leżącej u jego podstaw nauki. Nie należy jednak tak łatwo zrezygnować z etykiety „uczenie maszynowe”, gdyż nazwa ta wskazuje na szereg interesujących zagadnień dotyczących natury poznania i uczenia się oraz znaczenia metod maszynowych w badaniu ludzkiego poznania. Metody uczenia maszynowego są obecnie regularnie stosowane w celu zapewnienia ram i inspiracji dla modeli poznawczych, czasami pod nazwą „kognitywistyka obliczeniowa”. Etykieta jest również ważna, ponieważ utrwała uprzedzenia i oczekiwania u tych, którzy ją słyszą. Czasami te oczekiwania są nieuzasadnione, ale skłaniają poszczególne osoby – zarówno zwolenników, jak i sceptyków – do zadawania ważnych pytań na temat natury i wydajności tych algorytmów. Ostatnia kwestia filozoficzna dotyczy stopnia, w jakim można być realistą co do zawartości lub procesów pośrednich wyuczonego modelu. To znaczy, kiedy można rozumieć, że wewnętrzna struktura lub bogatsza treść semantyczna wyuczonego modelu odpowiada – być może tylko z pewnym prawdopodobieństwem – cechom świata? To pytanie jest szczególnie pilne w przypadku badań nad uczeniem się przyczynowym, które starają się wnioskować o strukturze przyczynowej świata na podstawie zestawów pasywnych obserwacji. Algorytmy wnioskowania przyczynowego przypuszczalnie uczą się zbioru struktur przyczynowych, które mogły wytworzyć określony zbiór danych, lub odkrywają najbardziej prawdopodobną taką strukturę. Istnieje wiele przypadków, w których algorytmy te zostały zastosowane do rzeczywistych zbiorów danych, a wyuczone modele zostały następnie z powodzeniem przypisane światu (np. studia przypadków). Algorytmy te – jak wszystkie metody wnioskowania – są wiarygodne tylko przy określonych założeniach dotyczących natury świata. Co więcej, treść semantyczna wyuczonego modelu – fakt, że możemy go nazwać modelem przyczynowym – wywodzi się z tych założeń, które zapewniają charakterystykę jednego (?) sposobu „rzutowania” struktur przyczynowych na dane obserwowane lub eksperymentalne. Analogia dotyczy takich założeń, jak teoria promieni światła: założenia te wyjaśniają, w jaki sposób trójwymiarowe obiekty są rzutowane na dwuwymiarową płaszczyznę (np. siatkówkę) i są niezbędne, aby każdy system wzrokowy mógł wyciągać wnioski na temat struktury obiektu na podstawie ograniczone, dwuwymiarowe wejście. Tak jak nasz układ wzrokowy doświadcza złudzeń optycznych, gdy różne założenia nie są spełnione (np. prosty kij wygląda na wygięty po włożeniu do wody), tak algorytmy wnioskowania przyczynowego podlegają „złudzeniu wnioskowania przyczynowego”, gdy założenia są naruszane w określony sposób. Te algorytmy wnioskowania przyczynowego, a bardziej ogólnie algorytmy uczenia maszynowego, nie wydają się w tym momencie niczym różnić od standardowych przypadków wnioskowania indukcyjnego: żadne wnioskowanie indukcyjne nie może mieć żadnych gwarancji niezawodności bez różnych założeń na temat świata. Jeśli te założenia są spełnione, algorytmy działają; Jeśli zostaną naruszone, wówczas nie ma szczególnego uzasadnienia, aby wierzyć w wewnętrzną strukturę wyników algorytmu. Czasami jednak twierdzi się, że założenia algorytmów wnioskowania przyczynowego są odmienne, ponieważ wydaje się, że możemy sprawdzić, czy założenia są rzeczywiście spełnione, mając tę samą wiedzę przyczynową, której próbujemy się nauczyć. Argument ten nie koncentruje się na samej możliwości, że założenia mogą być fałszywe, ponieważ wiąże się to z ryzykiem, które musi wiązać się z każdym wnioskowaniem indukcyjnym. Obawa ta nie dotyczy również praktycznej testowalności założeń algorytmów; wnioskowanie indukcyjne – czy to ludzkie, czy maszynowe – nieuchronnie wiąże się z przyjmowaniem założeń, które w danym momencie mogą nie być praktycznie sprawdzalne. Można na przykład potrzebować o rząd wielkości więcej danych niż obecnie. Jeśli jednak założenia są w zasadzie testowalne, można (w pewnym sensie) określić, czy algorytmy są wiarygodne w sytuacji takiej jak ta, i w ten sposób mieć pewne podstawy, aby traktować wyuczony model w sposób realistyczny. Podstawowym zmartwieniem jest to, że założeń w zasadzie nie da się nawet

przetestować, ponieważ wydaje się, że jedynym sposobem, aby dowiedzieć się, które testy statystyczne są istotne, jest poznanie leżącej u ich podstaw struktury przyczynowej, ale właśnie to ma znaleźć algorytm wnioskowania przyczynowego. Algorytmy te mogą (kontynuując argument) czasami znajdować przybliżenia prawdziwej struktury przyczynowej, ale tylko w drodze losowego przypadku. Założenia mogą być prawdziwe w każdej konkretnej sytuacji, ale nie można tego wiedzieć, a zatem nie ma podstawy do traktowania wyników algorytmu jako niczego innego niż reprezentacje danych zaobserwowanych lub eksperymentalnych. Algorytmy mogą stworzyć użyteczną „skróconą” wersję danych, którą można wykorzystać na różne sposoby, ale (w tym argumentcie) nie dowiemy się niczego istotnego na temat podstawowej struktury świata. Chociaż dotyczy to głównie algorytmów wnioskowania przyczynowego, ten potencjalny problem nie ogranicza się do nich. Na przykład wiele algorytmów grupowania jest niezawodnych tylko wtedy, gdy można przyjąć pewne założenia dotyczące podstawowych grup. Trzeba zatem móc zbadać grupy, aby określić, czy mają te właściwości, ale wymagałoby to wcześniejszej znajomości grup, co wyeliminowałoby samą potrzebę stosowania algorytmu grupowania. Metody uczenia maszynowego mają wyraźnie mniejszą wartość, jeśli zawsze należy je interpretować instrumentalnie. Teorie instrumentalistyczne – takie, które przewidują zachowanie systemu bez podejmowania jakichkolwiek zobowiązań co do podstawowej struktury lub ontologii teorii – są przydatne na wiele sposobów, ale często pragnie się czegoś więcej niż tylko przewidywania. Przede wszystkim należy posiadać informacje na temat leżących u podstaw mechanizmów, aby móc dokładnie przewidzieć, co się stanie, gdy system ulegnie awarii lub zmieni się na różne sposoby; teorie instrumentalistyczne nie dostarczają takich informacji. Istnieją trzy naturalne reakcje na ten zarzut. Po pierwsze, uważna analiza założeń często ujawnia, że wiedza wymagana do ich sprawdzenia jest słabsza, niż sugeruje to powierzchowne sformułowanie założeń. W szczególnym przypadku algorytmów wnioskowania przyczynowego, aby przetestować założenia, trzeba posiadać pewien rodzaj wiedzy przyczynowej, ale niezbędną wiedza to nie to samo, co wiedza o poszukiwanej strukturze przyczynowej. Na przykład może być potrzebna wiedza, że dana populacja jest „jednorodna przyczynowo” (tj. wszystkie jednostki mają te same relacje przyczynowe, choć niekoniecznie te same wartości zmiennych). Wiedza ta wymaga wiedzy przyczynowej, ale niekoniecznie o przyczynach lub skutkach tej konkretnej zmiennej. W drugiej, pokrewnej odpowiedzi zauważono, że argument ten określa testowalność założeń jako „wszystko albo nic”: albo wiadomo dokładnie, co jest wymagane do przetestowania założenia (choć w rzeczywistości można tego nie testować), albo nie wiadomo wcale. Bardziej realistyczna charakterystyka sytuacji jest taka, że często znane są niektóre, ale nie wszystkie, testy założeń, a także liczba możliwych ścieżek przyszłych testów. Można również mieć podstawy sądzić, że założenie zostało przetestowane w sposób niedoskonały. Jeśli ktoś ma tego rodzaju ograniczoną wiedzę, może mieć ograniczone potwierdzenie założenia, uznając jednocześnie, że wynik algorytmu musi zatem być interpretowany lub akceptowany w ograniczony sposób. Można sprzeciwić się wielu metodom uczenia maszynowego, uzasadniając to tym, że wymagają one jakiejś silnej, nie do końca ustalonej właściwości, ale nie należy odrzucać wyników tych metod tylko dlatego, że na krótką metę nie ma pewności, czy dana właściwość zostanie zachowana. Trzecią i najbardziej ogólną odpowiedzią na tę rodzinę zarzutów jest stwierdzenie, że schemat argumentacji w rzeczywistości przemawia przeciwko większości metod indukcyjnych, a nie tylko wnioskowaniu przyczynowemu; sprzeciwia się uczeniu maszynowemu, a nie uczeniu maszynowemu. Rozważmy konkretny wniosek indukcyjny: „Wszystkie elektrony mają ładunek ujemny”. Każda metoda, która wyciąga taki wniosek, musi przyjąć pewne założenia dotyczące świata, na przykład, że elektrony tworzą spójny, stabilny zbiór pod względem posiadania pewnego ładunku elektrycznego. Takie założenie można sprawdzić jedynie poprzez faktyczne określenie ładunków elektrycznych wszystkich elektronów, ale takie testy wyeliminowałyby potrzebę wyciągania jakichkolwiek wniosków z wniosków indukcyjnych. Należy zauważyć, że w tym przykładzie nie było wzmianki o metodach maszynowych ani nie było żadnych specyfikacji poza „metodą indukcyjną”. Problem znajomości warunków potwierdzenia założenia

metody jest problemem ogólnym i przemawia przeciwko prawie wszystkim metodom wnioskowania indukcyjnego, a nie konkretnie metodom uczenia maszynowego. Rzadko można poznać a priori wszystkie warunki potwierdzenia lub testowania założeń, które są wymagane, aby dana metoda wnioskowania dostarczyła wiarygodnych informacji o strukturze wewnętrznej lub uniwersalnej.

Wniosek

Metody uczenia maszynowego są często traktowane z pewną dozą podejrzliwości. Często przedstawia się je jako „czarne skrzynki”, które pobierają dane i bez żadnych wskazówek w jakiś sposób uczą się części prawdziwej struktury świata. Algorytmy te są w praktyce znacznie mniej tajemnicze: etykieta „automatycznych statystyk” jest często trafnym deskryptorem. Metody uczenia maszynowego odkrywają i wykorzystują relacje strukturalne między danymi, a to wnioskowanie strukturalne leży u podstaw zarówno mocnych, jak i słabych stron algorytmów uczenia maszynowego. Metody te można stosować w sposób stosunkowo ogólny dziedzinowy, gdyż konkretne znaczenie zmiennych nie ma znaczenia dla funkcjonowania algorytmu. Jednak ze względu na tę ogólność nie mogą dostarczać informacji specyficznych dla domeny, takich jak mechanizmy leżące u podstaw relacji informacyjnych. Uczenie maszynowe to jedna z najszybciej rozwijających się dziedzin informatyki, a wiele z najważniejszych wyzwań wiąże się z rozszerzeniem algorytmów na nowe typy danych, nowe modele lub słabsze założenia. Istnieją filozoficzne obawy dotyczące uczenia maszynowego, ale większość z nich koncentruje się albo na części „maszynowej”, albo „uczącej się”. Z jednej strony uczenie maszynowe jest przykładem złożonych obliczeń maszynowych, w związku z czym pojawiają się naturalne pytania, czy jakiegokolwiek operacje maszynowe można poprawnie opisać za pomocą terminów kognitywnych. Z drugiej strony algorytmy uczenia maszynowego wykonują złożone, ale jasno określone sekwencje obliczeń, w związku z czym pojawiają się pytania, czy metody te kwalifikują się jako „uczenie się” i czy założenia niezbędne do wnioskowania indukcyjnego można odpowiednio przetestować. Podsumowując, metody uczenia maszynowego otworzyły nowe możliwości poznawania struktury i zachowania naszego świata. Algorytmów tych należy oczywiście używać z odpowiednią świadomością i testowaniem podstawowych założeń. Jednak właściwie stosowane uczenie maszynowe może wykorzystać strukturę danych w celu uzyskania cennej wiedzy o strukturze i relacjach na świecie.