

Sztuczna inteligencja

Andrzej DĄBROWSKI

- Raj na Ziemi jednak istnieje. Znalazła go sztuczna inteligencja.
- Sztuczna inteligencja pomoże w walce z bioterroryzmem. Nauczyła się rozpoznawać bakterie wąglika.
- Potężna kasta zawodowa może zniknąć bez śladu. Sztuczna inteligencja bez trudu pokonała setkę ekspertów.
- Zaleje nas spam. Sztuczna inteligencja złamała system weryfikacji CAPTCHA.

Takie i podobne tytuły pojawiają się codziennie, zarówno w czasopiśmie, na stronach internetowych, jak i usłyszeć można je w telewizji czy w radiu. A to nie wszystko. Po wpisaniu przez nas hasła w wyszukiwarce internetowej sztuczna inteligencja nie tylko wybierze najbardziej pasujące linki, ale i okraja je odpowiednimi reklamami. Czytając e-mail, można nie zauważyć większości spamu – to też efekt działania sztucznej inteligencji. W wielu miejscach sztuczna inteligencja działa niewidzialnie dla nas. Nie wiadomo, które wiadomości czy artykuły na stronach sportowych czy ekonomicznych redaguje dziennikarz, a które sztuczna inteligencja. Przytłoczeni przez hasła *sztuczna inteligencja, uczenie maszynowe, głębokie sieci neuronowe* nie wiemy, czy to tylko atrakcyjne hasła, czy też kryją się za tym rzeczywiste dokonania. Prawda jest taka, że za tymi hasłami kryje się rewolucja nie tylko technologiczna. Zmienia się sposób, w jaki uprawia się naukę, diagnozuje choroby, a nawet tworzy dzieła sztuki. Pesymiści przewidują intelektualną dominację sztucznej inteligencji nad człowiekiem, przywołując pokonanie arcymistrzów gry w szachy i go.

Komputery zawsze uważane były za narzędzia użyteczne, wykonujące prace obliczeniowe i księgowo. Nazwa *mózgi elektroniczne*, która była powszechnie używana w czasach, kiedy zaczynałem studia, dość szybko zanikła, bo, jak się okazało, była zdecydowanie na wyrost.

Zostały adekwatne nazwy: komputer (compute – liczyć), czy po czesku počítač (počitadlo – liczydło), ale po francusku ordinateurur (ordre – porządek, rozkaz).

Teraz ta nazwa byłaby już bliższa prawdy.

Tradycyjnie, aby sprawić, by komputer coś zrobił, należało napisać program, szczegółowo opisujący wszystkie kroki, które ma wykonać. W uczeniu maszynowym, dziale sztucznej inteligencji, to komputer sam buduje swój program działania, ucząc się na podstawie dostępnych mu danych (im więcej danych, tym lepiej – stąd *big data*) – wydobywa z nich wiedzę, proponując najbardziej prawdopodobny wynik. Tak więc uczenie maszynowe jest operacją odwrotną do programowania. Inną różnicą jest, że w tradycyjnym podejściu wyniki pracy komputera są przewidywalne (deterministyczne), a w uczeniu maszynowym losowe, a ich przypadkowość zmniejsza się wraz ze wzrostem liczby danych.

Celem uczenia maszynowego jest wytrenowanie komputerów tak, aby były w stanie podejmować

optymalne decyzje. Co jest zadziwiające, najróżnorodniejsze zastosowania uczenia maszynowego są efektem praktycznie takiego samego algorytmu uczącego – o wiele prostszego niż algorytmy używane w tradycyjnych programach komputerowych.

Co może sztuczna inteligencja?

W powszechnej opinii sztuczna inteligencja kojarzy się z sieciami neuronowymi, a ostatnio z głębokimi sieciami neuronowymi. Jednak zajmuje ona znacznie bardziej rozległy obszar aktywności. Oto kilka z nich.

1. Reprezentacja wiedzy. Wykorzystuje się gigantyczne bazy wiedzy w celu zbudowania systemów eksperckich. Przykładem zastosowania tej dziedziny sztucznej inteligencji są system GIDEON (*Global Infectious Disease & Epidemiology Network*), zautomatyzowany wywiad lekarski, który ma pomóc w diagnozowaniu chorób zakaźnych i tropikalnych oraz system *Citizenship Application* służący do oceny, czy dana osoba kwalifikuje się do ubiegania o obywatelstwo USA.

2. Planowanie automatyczne. System służący planowaniu i podejmowaniu decyzji, szczególnie przy wykonywaniu operacji przez inteligentnych agentów, roboty autonomiczne i samochody autonomiczne.

3. Przetwarzanie języka naturalnego. Zespół procedur zajmujących się automatyzacją analizy, rozumienia, tłumaczenia i generowania języka naturalnego przez komputer.

4. Wizja komputerowa. Zajmuje się rozpoznawaniem obrazów. Ma liczne zastosowania w kontroli produkcji żywności i środków farmaceutycznych. Rozpoznawanie wąglika, o którym była mowa na początku artykułu, związane jest z wizją komputerową.

5. Robotyka. Technologia służąca wytwarzaniu automatów zastępujących człowieka. Działanie robotów oparte jest na zastosowaniu różnych metod sztucznej inteligencji: wizja komputerowa, planowanie automatyczne, sieci neuronowe.

6. Silna sztuczna inteligencja. Projekty stworzenia wirtualnego mózgu są dopiero w powijakach. W ramach projektu Blue Brain Project, rozpoczętego w 2005 roku przez Politechnikę Federalną w Lozannie, stworzono w 2011 roku symulację, odpowiadającą skali mózgu pszczoły (około miliona neuronów i miliarda połączeń nerwowych). W 2015 roku Japończycy użyli

ówcześnie czwartego najszybszego komputera na świecie (komputer K ma 705 tys. rdzeni i 1,4 mln GB pamięci RAM) do symulacji jednej sekundy aktywności 1% mózgu ludzkiego. Komputer potrzebował 40 minut, aby wykonać to zadanie. Kora ludzkiego mózgu, odpowiedzialna za wyższe procesy poznawcze, zawiera 15–33 miliardów neuronów, z których każdy może mieć do 10 tysięcy połączeń synaptycznych. Szacuje się ilość informacji, potrzebnych do odtworzenia jej funkcjonalności, na 500 petabajtów. Przewiduje się, że superkomputery o wystarczającej mocy obliczeniowej, aby przetworzyć taką ilość danych, powstaną około 2020 roku.

Pierwszym zwiastunem silnej sztucznej inteligencji jest program *AlphaGoZero*, który bez pomocy człowieka nauczył się grać w go. Mistrzowie tej gry zauważyli że sztuczna inteligencja odkryła zupełnie nowe strategie gry, nieznanne człowiekowi.

Podobnie, *Google Translator*, program do automatycznego tłumaczenia, sam wytworzył sztuczny język, pozwalający tłumaczyć pary języków, które nigdy nie były wcześniej trenowane w procesie uczenia.

Uczenie maszynowe

Szczególnie interesującą dziedziną sztucznej inteligencji jest uczenie maszynowe. Uczenie maszynowe wykorzystuje dostarczane mu dane, stosując w trakcie treningu metody statystyczne. Uczenie maszynowe jest *czarną skrzynką* (czyli układem, o którego budowie wewnętrznej nic nie wiadomo) z danymi na wejściu i wynikami na wyjściu. Uczenie, określające związek między wejściem a wyjściem, oparte jest na trzech schematach: uczenie nienadzorowane, uczenie nadzorowane i uczenie przez wzmocnienie.

Uczenie nadzorowane zakłada obecność ludzkiego nadzoru przy ocenie wyników, uzyskanych w każdym przebiegu algorytmu dla danych na wejściu. Taki sposób treningu jest wykorzystywany przy klasyfikacji z zadanymi wzorcami (rozpoznawanie pisma, diagnostyka medyczna).

W **uczeniu przez wzmocnienie** program uzyskuje sygnał w postaci nagrody czy kary w zależności od stopnia zgodności wyniku z projektowanym celem (tak były uczone komputery do gry w szachy i w go).

Uczenie nienadzorowane zakłada brak udziału człowieka oraz systemu nagród i kar w procesie nauczania: grupowanie (*klasteryzacja*) danych, segmentacja obrazu.

W bestsellerowej książce o algorytmach uczenia *Naczelny algorytm* scharakteryzowano różne podejścia do modelowania relacji między wejściem a wyjściem.

(P. Domingos, *Naczelny algorytm. Jak jego odkrycie zmieni nasz świat*, Helion, 2016)

Symboliści. Zakłada się, że inteligencja sprowadza się do manipulacji symbolami. Algorytm uczenia,

którym posługują się symboliści, to odwrotna dedukcja: analizowana jest część założeń i wynikających z nich wniosków i dopasowywane są te dodatkowe założenia, przy których wnioski są najbardziej prawdopodobne.

Koneksjoniści. Zakładają, że mózg składa się z różnych prostych elementów tego samego typu (neuronów). O jego działaniu decyduje siła połączeń między tymi elementami. Podstawowym algorytmem uczenia jest propagacja wsteczna. Polega ona na korekcji wag połączeń między neuronami po konfrontacji obliczonego wyniku z rzeczywistym.

Ewolucjoniści. Zakładają, że podstawą sposobów nauczania (a raczej dopasowania) jest selekcja naturalna. Metoda uczenia polega na użyciu algorytmów genetycznych.

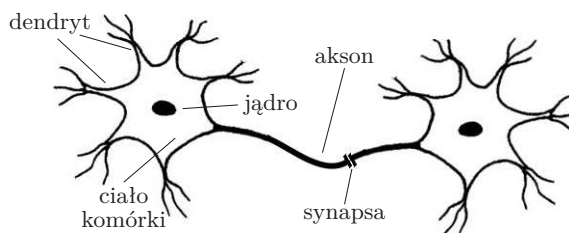
Bayesiści. Zakładają, że wiedza polega na ocenie wiarygodności wyników. Metoda uczenia polega na użyciu twierdzenia Bayesa i twierdzeń z niego wynikających. (Twierdzenie Bayesa pozwala obliczyć prawdopodobieństwo przyczyny, gdy znany jest jej skutek.)

Analogiści. Zakładają, że podobne rodzi podobne. To często używane podejście przy klasyfikacji z zadanymi wzorcami. Algorytm uczenia to maszyny wektorów wspierających (SVM – *Support Vector Machine*). Metoda ta polega na szukaniu najbardziej „wyrazistych” granic między obszarami danych.

Sieci neuronowe

Pośród wymienionych powyżej metod najczęściej opisywane są sieci neuronowe. Łatwo przypisać do nich etykietę mózgu elektronowego, obrazowo opisać sposób trenowania sieci i przedstawić naprawdę bogate zastosowania. Dla matematyka sieci neuronowe mają urok eleganckiej konstrukcji z ciekawie zarysowanym problemem związanym z klasycznymi problemami matematycznymi.

Naturalne jest poszukiwanie źródeł sieci neuronowych w budowie mózgu. Sygnały w mózgu przekazywane są za pomocą sieci neuronów. Neuron jest zbudowany z ciała komórki oraz odchodzących od niego wypustek: aksonu i dendrytów. Informacje w postaci sygnałów elektrycznych przekazywane są do następnych neuronów przez akson, który przewodzi pobudzenie z udziałem synapsy do dendrytów, należących do kolejnego neuronu.

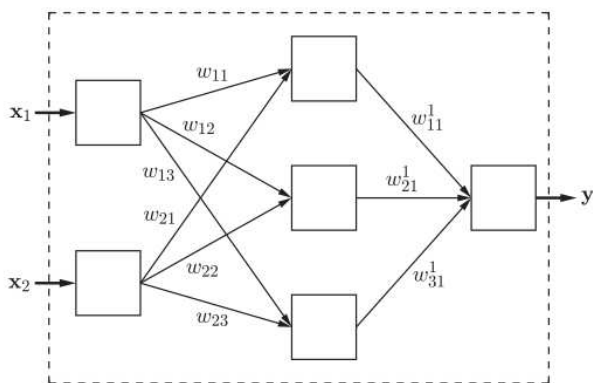


Różne obszary mózgu są odpowiedzialne za poszczególne funkcje: widzenie, słuch, kontrola ruchu, myślenie

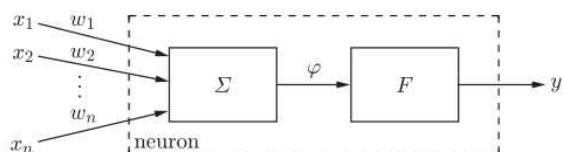
abstrakcyjne i inne. Wydawałoby się naturalne, że neurony powinny wyspecjalizować się w obsłudze różnych funkcji. W kwietniu 2000 roku zespół neurobiologów z MIT opublikował w czasopiśmie *Nature* wyniki nadzwyczajnego eksperymentu. Zmienili oni układ połączeń w mózgu fretki, zamieniając połączenia nerwowe oczu z częścią kory mózgowej odpowiedzialnej za słuch, a połączenia nerwowe wychodzące z uszu do części kory odpowiedzialnej za widzenie. Mózg fretki szybko przystosował się do nowej sytuacji – nie zauważono żadnej straty zarówno w odczuciu bodźców wzrokowych, jak i słuchowych. Wynika z tego, że wszystkie neurony działają tak samo, a połączenia w sieci neuronowej dopasowują się do funkcji, które mają obsługiwać. Jest to wskazówka, że modele mózgu można budować z tych samych elementów, a trening służy do ustanowienia połączeń, zależnych od modelowanej funkcji. Można postawić pytanie: jaki powinien być najprostsz model sieci neuronów i połączeń między nimi, aby mógł służyć do zbudowania relacji między danymi i wynikami danego typu?

Model sieci neuronowych, przedstawiony znacznie wcześniej od propozycji neurobiologów z MIT (Model Hebba zaproponowany był w latach 40. XX wieku, model perceptronu w latach 60.), zakłada, że neurony działają tak samo, a różnicowanie działalności sieci jest określone przez połączenia między neuronami. Sieci neuronowe składają się z neuronów w warstwie wejściowej, warstwach pośrednich i warstwie wyjściowej.

Pełniejszy opis można znaleźć w artykule K. Ambroch *Sztuczne sieci neuronowe*, MSN, Nr 32 (2004) <http://www.smp.uph.edu.pl/msn/32/ambroch.pdf>. Wykresy pochodzą z tej samej publikacji.



Pojedynczy neuron przetwarza wektor sygnałów wejściowych x wymiaru n na sygnał wyjściowy y (n jest liczbą neuronów w warstwie wcześniejszej). Na podstawie danych x w bloku sumowania Σ obliczana jest wartość $\varphi = w^T x = \sum_i w_i x_i$, a następnie sygnał ten jest przetwarzany przez funkcję aktywacji F . Wielkości w_{ij} nazywane są wagami połączeń.



Funkcja aktywacji w zasadzie powinna mieć postać funkcji progowej $F(\varphi) = 1_{[\varphi > \varphi_0]}$. Zazwyczaj jest zastępowana przez różniczkowalne sigmoidalne funkcje rosnące o asymptotach poziomych 1 i 0 (odpowiednio w ∞ i $-\infty$). Przyjmuje się, że w sieci występuje jeden rodzaj funkcji aktywacji. Czy tak ubogie sieci mogą być uniwersalne?

Uniwersalne twierdzenie aproksymacyjne

Sieci o jednym wyjściu modelują obliczanie wartości funkcji, a neurony reprezentują funkcje aproksymujące. W problemie aproksymacji chcemy jak najbogatszą rodzinę funkcji przybliżyć za pomocą jak najprostszej, łatwo zdefiniowanej rodziny funkcji. W 1885 roku Karl Weierstrass udowodnił twierdzenie, że każda ciągła funkcja z domkniętego odcinka jest jednostajną granicą ciągu wielomianów. Twierdzenie to było uogólniane na wiele sposobów, dobierano różne rodziny funkcji, aby przyspieszyć zbieżność czy powiększyć rodzinę funkcji aproksymowanych. We wszystkich tych przypadkach rodzina funkcji aproksymujących była nieskończona.

Spośród 23 problemów, które David Hilbert przedstawił na II Międzynarodowym Kongresie Matematyków w 1900 roku w Paryżu, 13. problem dotyczy istnienia rozwiązań wielomianów stopnia 7. będących algebraiczną (wariant: ciągłą) funkcją dwóch zmiennych. Pytanie to można (w nieoczywisty sposób!) uogólnić: *Czy każda ciągła funkcja trzech zmiennych jest złożeniem skończonej liczby ciągłych funkcji dwóch zmiennych?*

W 1957 roku 19-letni uczeń, Władimir Arnold, korzystając z pięknego twierdzenia Kołmogorowa (*każda ciągła funkcja wielu zmiennych jest złożeniem skończonej liczby funkcji trzech zmiennych*), odpowiedział twierdząco na to pytanie.

Później Kołmogorow wykazał, że ciągła funkcja wielu zmiennych może być przedstawiona za pomocą operacji złożenia i dodawania funkcji tylko jednej zmiennej.

Z twierdzeń tych wynika, że w problemie aproksymacji można użyć skończonej liczby funkcji. Dopiero twierdzenie Cybenki (nazywane uniwersalnym twierdzeniem aproksymacyjnym) z roku 1989 pokazało, jak uniwersalne są sieci neuronowe.

Twierdzenie (Cybenko). *Niech F będzie ciągłą funkcją sigmoidalną. Skończone sumy postaci $G(x) = \sum_{j=1}^N \alpha_j F(w_j^T x + \theta_j)$ są gęste w przestrzeni funkcji ciągłych n zmiennych na kostce jednostkowej.*

Powyższe twierdzenie pokazuje, że sieć z jedną warstwą wewnętrzną, z dowolną ciągłą funkcją aktywacji, może z zadaną dokładnością aproksymować funkcję ciągłą. W tej samej pracy Cybenko wykazał, że twierdzenie to jest również prawdziwe, gdy mamy do czynienia z zagadnieniem klasyfikacji (a więc: gdy funkcja celu jest stała na podzbiorach stanowiących skończony podział kostki w \mathbb{R}^n).

W praktyce używane są **głębokie sieci neuronowe** składające się z wielu warstw. Pozwala to na szybszą

optymalizację sieci, szczególnie gdy warstwa wyjściowa składa się z wielu neuronów. Ostatnio nawet Google zamienił swój algorytm oceny wyników przeszukiwania sieci *PageRank* na wersję *RankBrain*, opartą na głębokich sieciach neuronowych.

Uczenie sieci

O potęgze sieci neuronowych decyduje uczenie sieci. Po wprowadzeniu danych x na wejście sieci z zadaniem układem wag oblicza się wyjście y . Wartość wyjścia jest oceniana w zależności od metody uczenia. Tak otrzymany błąd wyniku y jest podstawą do korekty wag. Ta korekta jest tym, co określamy uczeniem sieci.

Podstawowym algorytmem korekty wag jest **propagacja wsteczna**. Model sieci neuronowej da się opisać przez funkcję wielu zmiennych mierzącą dopasowanie sieci. Na przykład w uczeniu nadzorowanym funkcja ta mierzy różnicę między wynikiem, otrzymanym z sieci, a wzorcem. Tymi zmiennymi są wagi połączeń między neuronami. Dobra sieć to taka, w której funkcja dopasowania osiąga wartość bliską optymalnej. Propagacja wsteczna w klasycznej postaci to nic innego jak metoda najszybszego spadku. Wagi połączeń korygują się w kierunku gradientu funkcji dopasowania, obliczonego dla ostatnio zadanych danych x . Dysponując dużą liczbą dobrze dobranych zestawów, wagi połączeń zbliżają się do swoich wartości optymalnych, a więc sieć staje się coraz bardziej dopasowana do zestawu danych uczących.

Postęp w budowaniu sieci neuronowych polega na eksperymentowaniu z różnymi funkcjami aktywacji, zmianą algorytmu optymalizacji w propagacji wstecznej (np. stochastyczna metoda najszybszego spadku czy metoda dropout), wreszcie dopasowywaniem architektury sieci. Ostatnio do tego celu używa się algorytmów genetycznych (*neuroevolution*), gdzie w jednym przebiegu poprawia się nie tylko wagi połączeń, ale i architekturę sieci metodami algorytmów genetycznych.

Wady sztucznej inteligencji

W chórze zachwytów nad osiągnięciami sztucznej inteligencji nie powinno zabraknąć refleksji nad błędami spowodowanymi niewłaściwą konstrukcją algorytmu sztucznej inteligencji, jak i jego niewłaściwym wykorzystaniem.

Przeuczenie. Zjawisko to występuje wtedy, gdy zbiór danych treningowych jest zbyt mało zróżnicowany. Model jest wtedy zbyt dopasowany do specyficznych danych i staje się mało elastyczny.

Firma Microsoft umieściła w sieci *czatbota* Tay, mającego wygląd kilkunastoletniej dziewczyny, który brał udział w dyskusjach młodych ludzi w wieku 15–24 lat na Tweeterze. Jednak w ciągu doby został „zdeprawowany” przez uczestników, nauczony terminologii rasistowskiej i seksistowskiej. Nie minęła doba, a Microsoft musiał usuwać rasistowskie tweety,

wreszcie zamknąć działanie czatbota, wysyłając rozpaczliwy tweet o treści: *c u soon humans need sleep now so many conversations today thx.*

Program *Compas*, używany przez sądy amerykańskie, pomagający wydawać wyroki, sugerował wyższe wyroki dla Afroamerykanów. Jego trening został oparty na analizie rozpatrywanych spraw, gdzie dominowały sprawy Afroamerykanów z udziałem dużej liczby recydywistów.

Problemy prawne. Korzystanie ze sztucznej inteligencji przysparza nowych problemów prawnych. Oto kilka przykładów:

- Lekarz wydaje diagnozę w oparciu o system ekspercki. Kto ponosi odpowiedzialność, gdy system się pomyli?
- Jeżeli system ekspercki jest uznany za bardziej wiarygodny od człowieka, czy lekarz jest zobowiązany do uwzględnienia wskazań systemu?
- Jeżeli transakcja została wykonana w czyimś imieniu przez wirtualnego agenta, to kto jest odpowiedzialny w przypadku straty?
- Czy jest możliwe, aby inteligentny agent był posiadaczem akcji i mógł przeprowadzać transakcje w swoim imieniu?
- Kto odpowiada za wypadki samochodów autonomicznych: właściciel, twórca oprogramowania, producent?

Technologiczna osobliwość. Jest to moment stworzenia silnych sztucznych inteligencji, które mogłyby opracowywać jeszcze wydajniejsze sztuczne inteligencje, wywołując lawinowe zmiany w technologii i społeczeństwie

Wywiad w *Wysokich obcasach* z dnia 23.07.2016 *Gazety Wyborczej*: *Zbliżamy się do momentu, w którym być może będziemy musieli oddać koronę władców świata* www.wysokieobcasy.pl/wysokie-obcasy/1,53662,20432293,technologiczna-osobliwosc-zblizamy-sie-do-momentu-w-ktorym.html. Ciekawe, czy kiedykolwiek doczekamy się takiego momentu w dziejach ludzkości?

Na konferencji sztucznej inteligencji w Puerto Rico (2015) zadano pytanie ekspertom: kiedy, z prawdopodobieństwem 50%, sztuczna inteligencja osiągnie ludzki poziom? Połowa ekspertów odpowiedziała, że najwcześniej w 2045 roku, ale niektórzy z nich twierdzili, że trzeba będzie czekać na to więcej niż 100 lat.

Twórca terminu *sztuczna inteligencja*, John McCarthy, wraz z bardzo utytułowanymi kolegami (Marvin Minsky, Nathaniel Rochester i Claude Shannon) twierdzili (w roku 1956, w epoce komputerów epoki kamiennej): *W ciągu dwóch miesięcy zespół 10 ludzi może znaleźć rozwiązanie takich problemów jak maszynowe tłumaczenie, automatyczne tworzenie pojęć abstrakcyjnych i rozwiązywanie niektórych problemów zarezerwowanych dla ludzi.*

Ta krzepiąca przepowiednia pozwala mi sądzić, że jeszcze długo czekać będziemy na moment technologicznej osobliwości w wydaniu sztucznej inteligencji.