

## Głębokie uczenie i nie tylko

*Jeśli chodzi o idee, byty, abstrakcje i transcendentaly, nigdy nie mogłem wbić im do głowy najmniejszej koncepcji.* - JONATHAN SWIFT, PODRÓŻE GULIWERA

Duża część obecnego entuzjazmu dla sztucznej inteligencji wynika z prostego faktu: inne rzeczy są równe, im więcej masz danych, tym lepiej. Jeśli chcesz przewidzieć wynik następnych wyborów, a możesz ankietować tylko 100 osób, powodzenia; jeśli możesz przeprowadzić wywiad z 10 000, twoje szanse są znacznie większe. W rzeczywistości we wczesnych dniach sztucznej inteligencji nie było zbyt wielu danych, a dane nie stanowiły głównej części obrazu. Większość badań opierała się na podejściu „opartym na wiedzy”, czasami nazywanym GOFAI - Good Old Fashioned AI lub „klasyczną sztuczną inteligencją”. W klasycznej sztucznej inteligencji naukowcy zazwyczaj ręcznie kodowali wiedzę, której sztuczna inteligencja potrzebowałaby do wykonania określonego zadania, a następnie pisali programy komputerowe, które wykorzystywały tę wiedzę, stosując ją do różnych wyzwań poznawczych, takich jak rozumienie historii lub planowanie robotów lub dowodzenie twierdzenia. Big data nie istniały, a te systemy rzadko skupiały się przede wszystkim na wykorzystaniu danych. Chociaż możliwe było (często przy dużym wysiłku) zbudowanie prototypów laboratoryjnych przy użyciu tego podejścia, przejście tego etapu było często zbyt trudne. Całkowita liczba klasycznych systemów AI dowolnych praktycznych znaczenie jest niewielkie. Takie techniki są nadal szeroko stosowane w niektórych obszarach, takich jak planowanie tras dla robotów i nawigacja oparta na GPS. Ogólnie rzecz biorąc, klasyczne podejście skoncentrowane na wiedzy zostało w dużej mierze wyparte przez uczenie maszynowe, które zazwyczaj próbuje nauczyć się wszystkiego na podstawie danych, zamiast polegać na specjalnie zaprojektowanych programach komputerowych, które wykorzystują ręcznie zakodowaną wiedzę. To podejście również sięga lat 50. XX wieku, kiedy Frank Rosenblatt zbudował „sieć neuronową” - jeden z pierwszych systemów uczenia maszynowego - który miał na celu nauczenie się rozpoznawania obiektów wokół niej bez konieczności wcześniejszego przewidywania przez programistów każdej ewentualności. Jego systemy nabrały początkowo ogromnego szumu i zostały ogłoszone w The New York Times z wielkim uznaniem w 1958 roku, ale te zbyt szybko wyblakły, nękane własnymi problemami. Jego sieci (które musiały działać na sprzęcie z lat 50.) były słabe: w dzisiejszej terminologii nie były wystarczająco głębokie (za chwilę wyjaśnimy, co to dokładnie oznacza). Jego aparat również nie miał wystarczającej liczby pikseli; przy 20x20 miał 400, około 1/30 000 rozdzielczości iPhone'a X, co dawało bardzo pikselowe zdjęcia. Z perspektywy czasu Rosenblatt miał dobry pomysł, ale rzeczywiste systemy, które mógł wtedy zbudować, po prostu nie były w stanie wiele zdziałać. Ale sprzęt był tylko częścią problemu. Z perspektywy czasu uczenie maszynowe zależy również w dużej mierze od posiadania dużych ilości danych, takich jak zdjęcia opatrzone etykietami, a Rosenblatt nie miał ich zbyt wiele; nie było internetu, z którego mógłby wyciągnąć miliony przykładów. Mimo to wielu ludzi kontynuowało tradycję Rosenblatta przez dziesięciolecia. A do niedawna także jego następcy walczyli z wielkimi trudnościami. Dopóki big data nie stało się powszechne, w społeczności AI panował ogólny konsensus, że tak zwane podejście sieci neuronowej było beznadziejne. Systemy po prostu nie działały tak dobrze w porównaniu z innymi metodami. Kiedy nadeszła rewolucja big data, na początku 2010 roku, sieci neuronowe wreszcie miały swój dzień. Ludzie tacy jak Geoff Hinton, Yoshua Bengio, Yann LeCun i Jürgen Schmidhuber, którzy utknęli w sieciach neuronowych, nawet w mrocznych czasach lat 90. i 2000, kiedy większość ich kolegów odwróciła się gdzie indziej, zaatakowali. Pod pewnymi względami najważniejszy postęp nie pochodził z jakiegoś przełomu technicznego w matematyce sieci neuronowych - z których większość została wypracowana w latach 80. - ale z gier komputerowych lub, dokładniej, ze specjalnego sprzętu znanego jako GPU (skrót od jednostki przetwarzania grafiki), który społeczność sieci neuronowych przekształciła w sztuczną inteligencję. Procesory graficzne zostały pierwotnie opracowane dla gier wideo, począwszy od lat 70. XX wieku, a od początku XXI wieku były stosowane w sieciach neuronowych. Do 2012 roku

stały się niezwykle wydajne i do pewnych celów były bardziej wydajne niż procesory, tradycyjny rdzeń większości komputerów. Rewolucja nastąpiła w 2012 roku, kiedy wiele osób, w tym zespół badaczy współpracujących z Hintonem, wypracowało sposób na wykorzystanie mocy GPU do ogromnego zwiększenia mocy sieci neuronowych. Nagle po raz pierwszy zespół Hinton'a i inni zaczęli ustanawiać rekordy, przede wszystkim w rozpoznawaniu obrazów w bazie danych ImageNet, o której wspominaliśmy wcześniej. Konkurenci Hinton'a skupili się na podzbiorze bazy danych - 1,4 miliona zdjęć, wylosowanych z tysiąca kategorii. Każdy zespół przeszkolił swój system na około 1,25 miliona z nich, pozostawiając 150 000 do testów. Wcześniej, w przypadku starszych technik uczenia maszynowego, wynik 75% poprawny był dobrym wynikiem; Zespół Hinton'a uzyskał wynik w 84 procentach poprawnie, wykorzystując głęboką sieć neuronową, a inne zespoły wkrótce poradziły sobie jeszcze lepiej; do 2017 r. wyniki etykietowania obrazów, oparte na głębokim uczeniu się, osiągnęły 98 procent. Kluczem do tego nowo odkrytego sukcesu był fakt, że procesory graficzne pozwoliły Hintonowi i innym „uczyć” sieci neuronowe, które były znacznie głębsze - w sensie technicznym posiadania większej liczby warstw (zestawów niejasno podobnych do neuronów elementów, które wyjaśnimy za chwilę) – niż było wcześniej możliwe. Aby wyszkolić głęboką sieć, należy podać sieci kilka przykładów wraz z poprawnymi etykietami dla tych przykładów: to zdjęcie przedstawia psa, to zdjęcie przedstawia kota i tak dalej, paradygmat znany jako uczenie nadzorowane. Wykorzystanie procesorów graficznych oznaczało, że więcej warstw można było trenować szybciej, a wyniki były lepsze. Pomiędzy procesorami graficznymi i rozmiarem biblioteki ImageNet, głębokie uczenie było wyłączone z wyścigu. Niedługo potem Hinton i niektórzy studenci założyli firmę i sprzedali ją na aukcji. Google licytował najwięcej, a dwa lata później kupił startup o nazwie DeepMind za ponad 500 milionów dolarów. Rozpoczęła się rewolucja głębokiego uczenia się.

Samo uczenie głębokie jest ostatecznie tylko jednym z wielu sposobów podejścia do wyzwania polegającego na próbie nakłonienia maszyn do uczenia się rzeczy na podstawie danych, najczęściej za pomocą metod statystycznych. Załóżmy, że prowadzisz księgarnię internetową i chcesz polecać produkty swoim klientom. Jednym z podejść byłoby ręczne określenie, jakie są twoje ulubione książki. Możesz umieścić je na swojej pierwszej stronie, podobnie jak księgarze umieszczają ulubione książki na froncie księgarni. Ale innym podejściem jest poznanie tego, co ludzie lubią z danych, a nie tylko tego, co ludzie ogólnie lubią, ale co może działać dla konkretnego klienta na podstawie tego, co kupili wcześniej. Możesz zauważyć, że ludzie, którzy lubią książki o Harrym Potterze, często kupują Hobbita, a ludzie, którzy lubią Tołstoja, często kupują Dostojewskiego. Wraz ze wzrostem ekwipunku rośnie liczba możliwości i śledzenie wszystkiego z osobna staje się zbyt trudne, więc piszesz program komputerowy, który to wszystko śledzi. Śledzisz statystyki: prawdopodobieństwo, że klient, który kupi książkę 1, kupi również książkę 2, książkę 3 i tak dalej. Kiedy już opanujesz ten temat, zaczynasz śledzić bardziej skomplikowane prawdopodobieństwa, takie jak prawdopodobieństwo, że ktoś, kto kupił zarówno powieść o Harrym Potterze, jak i Hobbita, ale nie książkę o Gwiezdnym wojnach, kupi również powieść science fiction Roberta Heinleina. Ta sztuka robienia świadomych domysłów na podstawie danych to duża i dobrze prosperująca poddziedzina sztucznej inteligencji, znana jako uczenie maszynowe. Przydatnym sposobem myślenia o związku między głębokim uczeniem, uczeniem maszynowym i sztuczną inteligencją jest diagram Venna. Sztuczna inteligencja obejmuje uczenie maszynowe, ale obejmuje również na przykład dowolny niezbędny algorytm lub wiedzę, która jest ręcznie kodowana lub budowana przy użyciu tradycyjnych technik programowania, a nie wyuczona. Uczenie maszynowe obejmuje dowolną technikę, która umożliwia maszynie uczenie się na podstawie danych; deep learning jest najbardziej znaną z tych technik, ale nie jedyną. Skupiamy się tutaj na głębokim uczeniu się, ponieważ jest to z dużym marginesem główny cel większości obecnych inwestycji w sztuczną inteligencję, zarówno w akademii, jak i w przemyśle. Jednak głębokie uczenie nie jest jedynym podejściem, zarówno do uczenia maszynowego, jak i ogólnie do sztucznej inteligencji. Na

przykład jednym z podejść do uczenia maszynowego jest konstruowanie drzew decyzyjnych, które są zasadniczo systemami prostych reguł charakteryzujących dane. Maszyna wektorów nośnych, technika organizująca dane w złożone, abstrakcyjne hipersześciiany, dominowała w uczeniu maszynowym w pierwszej dekadzie XXI wieku, a ludzie używali jej do obliczania wszystkiego, od tematów artykułów prasowych po struktury białek. Modele probabilistyczne próbują obliczyć prawdopodobieństwo różnych możliwych odpowiedzi i wyprowadzić odpowiedź, którą uważają za najbardziej prawdopodobną, podejście, które było kluczowe dla sukcesu Watsona IBM i prawdopodobnie będzie miało ciągły wpływ. Jeszcze inne podejście, czasami nazywane algorytmami genetycznymi, jest wzorowane na procesie ewolucji; różne algorytmy są wypróbowywane i „mutowane” w jakiejś formie. Najlepsi przetrwają i rozmnażają się. Algorytmy genetyczne są wykorzystywane w różnych zastosowaniach, od projektowania anten radiowych po granie w gry wideo, gdzie czasami osiągają ten sam poziom sukcesu, co uczenie głębokie. A lista jest długa; nie będziemy omawiać ich wszystkich, wolimy skupić się na głębokim uczeniu, ponieważ stało się ono tak dominujące w ostatnich latach/ A wiele problemów, takich jak planowanie tras przejazdów i działania robotów, nadal wykorzystuje techniki zaczerpnięte z klasycznej sztucznej inteligencji, które wykorzystują niewiele lub wcale nie wykorzystują uczenia maszynowego. (Często pojedyncza aplikacja, podobnie jak algorytmy routingu ruchu używane przez Waze, będzie obejmować wiele technik, zaczerpniętych zarówno z klasycznej sztucznej inteligencji, jak i uczenia maszynowego). W ostatnich latach uczenie maszynowe stało się wszechobecne, napędzane w dużej mierze przez duże zbiory danych. IBM Watson polegał na ogromnych bazach danych i wykorzystał połączenie klasycznych technik sztucznej inteligencji i probabilistycznego uczenia maszynowego, aby dostroić swój system, aby wygrać w Jeopardy! Miasta wykorzystują uczenie maszynowe do przydzielania usług, usługi współdzielenia samochodów wykorzystują uczenie maszynowe do przewidywania zapotrzebowania kierowców, a wydziały policji wykorzystują uczenie maszynowe do przewidywania przestępstw. Jeśli chodzi o komercyjną stronę, Facebook wykorzystuje uczenie maszynowe do decydowania, jakie wiadomości chcesz zobaczyć w swoim kanale, a także do wnioskowania, jakie reklamy prawdopodobnie klikniesz. Google wykorzystuje systemy uczące się, by polecać filmy, umieszczać reklamy, rozumieć Twoją mowę i próbować odgadnąć, czego możesz szukać w swoich wyszukiwaniach w sieci. Witryna Amazon wykorzystuje uczenie maszynowe do polecenia produktów i interpretowania wyszukiwań. Urządzenie Alexa firmy Amazon wykorzystuje uczenie maszynowe do dekodowania Twoich żądań i tak dalej. Żaden z tych produktów nie jest doskonały; później omówimy kilka przykładów dobrze znanych komercyjnych wyszukiwarek, które są zdezorientowane podstawowymi żądaniami. Ale praktycznie wszystkie z nich są znacznie lepsze niż nic, a więc wyraźnie cenne ekonomicznie. Nikt nie potrafił ręcznie napisać wyszukiwarki w skali całej sieci; Google po prostu nie istniałby bez uczenia maszynowego. Rekomendacje produktów Amazon byłyby znacznie gorsze, gdyby opierały się wyłącznie na ludziach. (Najbliższą rzeczą, o której możemy pomyśleć, jest Pandora, usługa rekomendacji muzyki, która jest w dużej mierze wykonywana ręcznie przez ekspertów, ale w konsekwencji ogranicza się do znacznie mniejszej biblioteki muzycznej niż porównywalne systemy, takie jak Google Play, które w większym stopniu opierają się na maszynach. ) Zautomatyzowane systemy rekomendacji reklam, które mogą personalizować swoje rekomendacje dla poszczególnych użytkowników na podstawie historycznych statystyk tego, co kupili ludzie o podobnej historii, nie muszą być doskonałe; nawet jeśli popełniają okazjonalny błąd, są o wiele bardziej precyzyjnie ukierunkowani niż stara strategia polegająca na umieszczaniu dużych reklam w gazecie. W 2017 r. Google i Facebook razem zarobiły ponad 80 miliardów dolarów na umieszczaniu reklam; uczenie maszynowe, oparte na wnioskowaniu statystycznym, było głównym motorem, który to umożliwił.

Głębokie uczenie się opiera się na dwóch fundamentalnych ideach. Pierwsza, którą można nazwać rozpoznawaniem wzorców hierarchicznych, wywodzi się częściowo z zestawu eksperymentów z lat 50-

tych, które doprowadziły neurologów Davida Hubela i Torstena Wiesela do zdobycia Nagrody Nobla w dziedzinie fizjologii lub medycyny w 1981 roku. Hubel i Wiesel odkryli, że różne neurony w układzie wzrokowym reagowały na bodźce wzrokowe w bardzo różny sposób. Niektóre reagowały najaktywniej na bardzo proste bodźce, takie jak linie w określonych orientacjach, podczas gdy inne reagowały bardziej energicznie na bodźce bardziej złożone. Zaproponowali teorię, zgodnie z którą złożone bodźce mogą być rozpoznawane poprzez hierarchię rosnącej abstrakcji, na przykład od linijek do liter i słów. W latach 80., w ważnym punkcie zwrotnym w historii sztucznej inteligencji, japoński pionier sieci neuronowych, Kunahiko Fukushima, zbudował rzeczywistą realizację obliczeniową pomysłu Hubela i Wiesela, zwaną Neocognitron, i pokazał, że może działać w niektórych aspektach wizji komputerowej. Neocognitron składał się z zestawu warstw (które wyglądają jak prostokąty). Idąc od lewej do prawej na diagramie obok, najpierw znajduje się warstwa wejściowa, która byłaby prezentowana z bodźcem, zasadniczo piksele w obrazie cyfrowym; następnie kolejne warstwy, które analizowałyby obraz, szukając różnic w kontraście, krawędzi itp., ostatecznie kończąc się warstwą wyjściową, która identyfikuje kategorię, do której należy dane wejściowe. Połączenia między warstwami pozwoliły na przeprowadzenie wszystkich istotnych procesów. Wszystkie te pomysły - warstwy wejściowe, warstwy wyjściowe i warstwy wewnętrzne wraz z połączeniami między nimi - są teraz podstawą głębokiego uczenia się. Takie systemy nazywane są sieciami neuronowymi, ponieważ każda warstwa składa się z elementów zwanych węzłami, które można (bardzo luźno) przyrównać do znacznie uproszczonych neuronów. Istnieją połączenia między tymi węzłami, zwane również wagami połączeń lub po prostu wagami; im większa waga na połączeniu z węzła A do węzła B, tym większy wpływ A na B. To, co robi sieć, jest funkcją tych wag.

Drugim kluczowym pomysłem jest uczenie się. Wzmacniając wagi dla określonej konfiguracji danych wejściowych do określonego wyjścia, można na przykład „wytrenować” sieć, aby nauczyła się kojarzyć określone dane wejściowe z odpowiadającymi im danymi wyjściowymi. Załóżmy na przykład, że chcesz, aby sieć uczyła się nazw różnych liter prezentowanych na siatce pikseli. Początkowo system nie miał pojęcia, który wzór pikseli odpowiada której literze. Z biegiem czasu, poprzez proces prób i błędów oraz regulacji, dochodziło do kojarzenia pikseli na górze siatki z literami takimi jak T i E, a piksele na lewej krawędzi z literami takimi jak E, F i H, stopniowo opanowując korelacje między pikselami w różnych lokalizacjach i odpowiednimi etykietami. W latach pięćdziesiątych Rosenblatt miał już podstawową intuicję, że może to być wykonalne, ale sieci, z których korzystał, były proste i ograniczone, zawierały tylko warstwę wejściową i warstwę wyjściową. Jeśli twoje zadanie było wystarczająco proste (jak klasyfikowanie kółek i kwadratów), dość prosta matematyka gwarantowała, że zawsze możesz dostosować swoje wagi, aby „zbliżyć się” (obliczyć) właściwą odpowiedź. Ale w przypadku bardziej skomplikowanych zadań dwie warstwy nie wystarczały, aby wszystko było proste - wymagane byłyby warstwy pośrednie reprezentujące kombinacje rzeczy - a w tamtym czasie nikt nie miał praktycznego rozwiązania do niezawodnego uczenia głębszych sieci, które miały więcej niż dwie warstwy. Prymitywne sieci neuronowe w tamtych czasach miały wejścia (np. obrazy) i wyjścia (etykiety), ale nic pomiędzy. W swojej wpływowej książce Perceptrons z 1969 r. Marvin Minsky i Seymour Papert udowodnili matematycznie, że proste sieci dwuwarstwowe nie są w stanie uchwycić wielu rzeczy, które systemy mogą chcieć sklasyfikować. Zauważyli, że dodanie kolejnych warstw dałoby więcej mocy, ale wiązałoby się również z kosztami: możliwość wytrenowania sieci w celu znalezienia satysfakcjonującego rozwiązania nie była już gwarantowana. Nieco pesymistycznie napisali, że ich intuicyjny osąd był taki, że rozszerzenie na wiele warstw byłoby „sterylne”, chociaż pozostawili otwartą możliwość, że „być może [może] zostać odkryte jakieś interesujące twierdzenie o zbieżności”. Pomędzy tą nutą pesymizmu a brakiem naprawdę przekonujących wyników, dziedzina sieci neuronowych wkrótce uschła. Proste problemy były nudne i miały ograniczone zastosowanie, a złożone problemy wydawały się nie do rozwiązania. Ale nie wszyscy się poddali. Jak przyznali Minsky i

Papert, tak naprawdę nie udowodnili, że nie można zrobić niczego pożytecznego z głębszą siecią, a jedynie, że nie można zagwarantować optymalnych wyników przy użyciu analizowanej przez nich konkretnej matematyki. Z jednej strony to, co napisali Minsky i Papert w 1969 r., nadal pozostaje prawdziwe: do dziś głębokie uczenie się nadal nie oferuje zbyt wiele w zakresie gwarancji formalnych (z wyjątkiem nierealistycznego przypadku, w którym dostępne są zarówno nieskończone dane, jak i nieskończone zasoby obliczeniowe). ); ale z perspektywy czasu widać wyraźnie, że Minsky i Papert nie docenili użyteczności głębszych sieci bez gwarancji. W ciągu następnych dwóch dekad kilka osób, w tym Geoff Hinton i David Rumelhart, niezależnie wymyśliło matematykę, która pozwala głębszym sieciom neuronowym wykonywać zaskakująco dobrą pracę, pomimo braku jakichkolwiek formalnych matematycznych gwarancji doskonałości.

Metaforą, której ludzie często używają, jest wspinanie się na wzgórze. Wyobraź sobie, że podstawa góry jest słabym rozwiązaniem problemu, w którym system ma niską dokładność, a szczyt góry jest najlepszym rozwiązaniem, z dużą dokładnością. (Czasami używa się przeciwnej metafory, zwanej gradientem.) Hinton i inni odkryli, że chociaż nie można zagwarantować doskonałości w „głębszych” sieciach z więcej niż dwiema warstwami, możliwe było zbudowanie systemu, który mógłby dawać wyniki które często były wystarczająco dobre, oportunistycznie wspinając się na górę, wykonując małe kroki odpowiedniego rodzaju, stosując technikę zwaną wsteczną propagacją - teraz koń pociągowy głębokiego uczenia się. Propagacja wsteczna polega na oszacowaniu najlepszego kierunku w górę w dowolnym punkcie. Chociaż znalezienie drogi na najwyższy szczyt góry nie jest gwarantowane - może utknąć na „lokalnym maksimum” (szczyt drugorzędny lub nawet duży głąz na górze, który jest wyższy niż cokolwiek w pobliżu, ale wciąż nie jest wystarczająco dobry) - w praktyce technika jest często wystarczająca. Technika zwana konwolucją, wprowadzona pod koniec lat 80. przez Yanna LeCuna i nadal szeroko stosowana, pozwoliła systemom rozpoznawania obiektów na wydajniejsze dzięki wbudowaniu szeregu połączeń, które umożliwiłyby systemowi rozpoznawanie obiektu bez względu na to, w którym miejscu się pojawia. zdjęcie. Mimo to, chociaż matematyka wydawała się rozsądna, wstępne wyniki dotyczące sieci neuronowych nie były przekonujące. Wiadomo było, że w zasadzie, jeśli można znaleźć odpowiedni zestaw wag (duży, ale często możliwy do opanowania, jeśli), trzy lub więcej warstw może pozwolić na rozwiązanie dowolnego problemu, pod warunkiem, że ma się wystarczającą ilość danych, wystarczającą cierpliwość i ogromną liczbę węzłów. Ale w praktyce to nie wystarczyło; Potrzebna była niemożliwie duża liczba węzłów, a komputery w tamtych czasach nie były w stanie wykonać wszystkich potrzebnych obliczeń w rozsądnym czasie. Ludzie mieli silne przeczucie, że więcej warstw - czyli „głębszych sieci” - pomoże. Ale nikt nie wiedział na pewno. Jeszcze na początku XXI wieku sprzęt po prostu nie nadawał się do pracy. Wytrenowanie typowej głębokiej sieci zajęłoby kilka tygodni, a nawet miesięcy pracy komputera; nie mogłeś (tak jak teraz ludzie) wypróbować stu różnych alternatyw i posortować je, aby znaleźć najlepszą. Wyniki były obiecujące, ale nie konkurencyjne w porównaniu z innymi podejściami. Tu właśnie pojawiły się procesory graficzne. Katalizatorem rewolucji głębokiego uczenia się (poza kilkoma ważnymi poprawkami technicznymi) było wymyślenie, jak efektywnie wykorzystać procesory graficzne do obsługi bardziej złożonych modeli z większą liczbą warstw w rozsądnym czasie. Głębokie uczenie - które obejmuje sieci szkoleniowe składające się z czterech lub więcej warstw, czasem ponad stu - stało się wreszcie praktyczne.

Wyniki głębokiego uczenia się są naprawdę niezwykle. Naukowcy spędzili lata na ręcznym tworzeniu sprytnych reguł, próbując sprawić, by rozpoznawanie obiektów działało. Teraz tę pracę można zastąpić systemem głębokiego uczenia się, który spędza kilka godzin lub dni na komputerze. Głębokie uczenie pozwoliło również ludziom podjąć nowe problemy, nie tylko polecenie reklam, ale także rozpoznawanie mowy i obiektów, których nigdy nie rozwiązano w odpowiedni sposób przy użyciu starszych technik uczenia maszynowego. Głębokie uczenie osiągnęło „najlepszy stan wiedzy” (najlepsze dotychczasowe wyniki) w benchmarku po benchmarku. Na przykład, jak wyjaśniono w

długim artykule w The New York Times Magazine, głębokie uczenie radykalnie ulepszyło Tłumacza Google. Do 2016 roku Tłumacz Google wykorzystywał klasyczne techniki uczenia maszynowego, wykorzystując ogromne tabele pasujących wzorców w dwóch językach, oznaczonych prawdopodobieństwami. Nowsze podejście oparte na sieciach neuronowych, wykorzystujące głębokie uczenie, przyniosło znacznie lepsze tłumaczenia. Głębokie uczenie doprowadziło również do znacznych ulepszeń w uzyskiwaniu maszyn do transkrybowania mowy i oznaczenia zdjęcia. Poza tym na wiele (choć nie wszystkie) sposoby głębokiego uczenia jest łatwiej w użyciu. Tradycyjne uczenie maszynowe często w dużym stopniu opiera się na wiedzy z zakresu inżynierii funkcji. Na przykład w dziedzinie widzenia wykwalifikowani inżynierowie, którzy mają wiedzę na temat widzenia, próbowali znaleźć wspólne właściwości obrazów wizualnych, które mogą być pomocne dla maszyn próbujących poznać obrazy, takie jak krawędzie, narożniki i plamy. W 2011 roku to, co czyniło kogoś dobrym inżynierem uczenia maszynowego, to często umiejętność znajdowania odpowiednich funkcji wejściowych dla danego problemu. W pewnym stopniu głębokie uczenie się to zmieniło; w wielu problemach (choć, jak zobaczymy, nie we wszystkich) głębokie uczenie może działać dobrze bez rozbudowanej inżynierii funkcji. Systemy, które zaczęły wygrywać w konkursie ImageNet, nauczyły się kategoryzować obiekty - na najnowocześniejszych poziomach - bez znaczącej ilości prac inżynierskich. Zamiast tego systemy nauczyły się wszystkiego, co musiały wiedzieć, po prostu patrząc na piksele na obrazach i etykiety, których miały się nauczyć. Inżynieria funkcji nie wydawała się już potrzebna. Aby zacząć, nie potrzebowałeś doktoratu z nauk o wizji. Co więcej, uczenie głębokie okazało się zadziwiająco ogólne, przydatne nie tylko przy problemach z rozpoznawaniem obiektów i rozpoznawaniem mowy, ale także przy wielu zadaniach, o których ludzie mogliby nawet nie przypuszczać, że są osiągalne. Głębokie uczenie odniosło niezwykły sukces w tworzeniu sztuki syntetycznej w stylu dawnych mistrzów (zamień swój krajobraz w Van Gogha) i kolorowaniu starych obrazów. Może być stosowany, szczególnie w połączeniu z ideą znaną jako generatywne sieci przeciwników, w celu rozwiązania problemu „nienadzorowanego uczenia się”, wyzwania, w których nie ma nauczyciela, który mógłby nazwać przykłady. Głębokie uczenie może być również wykorzystywane jako element systemów grających w gry, czasami na poziomie nadludzkim. Oba początkowe sukcesy DeepMind - najpierw w grach komputerowych Atari, później w Go - polegały częściowo na wykorzystaniu głębokiego uczenia połączonego z uczeniem ze wzmocnieniem, aby uzyskać nową technikę znaną jako głębokie uczenie ze wzmocnieniem, sposób na uczenie metodą prób i błędów z dane na dużą skalę. (AlphaGo zapożyczyła również kilka innych technik, o czym będziemy rozmawiać później.) Czasami sukces może wydawać się odurzający. Kiedy Andrew Ng, wiodący badacz sztucznej inteligencji, który kierował wówczas badaniami nad sztuczną inteligencją w chińskiej firmie wyszukiwawczej Baidu, napisał na łamach Harvard Business Review w 2016 r.: „Jeśli typowa osoba może wykonać zadanie umysłowe z mniej niż jedną sekundą namysłu, prawdopodobnie możemy go zautomatyzować za pomocą sztucznej inteligencji teraz lub w najbliższej przyszłości”, odniósł się w dużej mierze do sukcesu głębokiego uczenia się. Niebo wydawało się granicą.

Mimo to od początku byliśmy sceptyczni. Nawet jeśli od razu było oczywiste, że głębokie uczenie się jest potężniejszym narzędziem niż którykolwiek z jego poprzedników, wydawało nam się, że jest wyprzedane. W 2012 roku napisano to w The New Yorker, kierując się badaniami przeprowadzonymi kilkanaście lat wcześniej na temat poprzedników głębokiego uczenia się:

Realistycznie głębokie uczenie jest tylko częścią większego wyzwania, jakim jest budowanie inteligentnych maszyn. Takim technikom brakuje sposobów przedstawiania związków przyczynowych (takich jak między chorobami a ich objawami) i prawdopodobnie napotkają trudności w przyswajaniu abstrakcyjnych idei, takich jak „rodzeństwo” lub „identyczny z”. Nie mają oczywistych sposobów przeprowadzania wnioskowań logicznych, a od integracji wiedzy abstrakcyjnej są jeszcze daleko. Kilka lat później wszystko to nadal wydaje się być prawdą - pomimo widocznego postępu w niektórych

obszarach, takich jak rozpoznawanie mowy, tłumaczenie języka i rozpoznawanie głosu. Głębokie uczenie wciąż nie jest żadnym uniwersalnym rozpuszczalnikiem. Nie ma też wiele wspólnego z ogólną inteligencją, jakiej potrzebujemy w przypadku systemów otwartych.

W szczególności boryka się z trzema podstawowymi problemami, z których każdy wpływa zarówno na samo głębokie uczenie się, jak i na inne popularne techniki, takie jak głębokie uczenie przez wzmacnianie, które w dużej mierze na nim polega: głębokie uczenie się jest chciwe. Aby poprawnie ustawić wszystkie połączenia w sieci neuronowej, głębokie uczenie często wymaga ogromnej ilości danych. AlphaGo wymagało 30 milionów gier, aby osiągnąć nadludzką wydajność, o wiele więcej gier niż jakikolwiek człowiek rozegra w życiu. Przy mniejszych ilościach danych uczenie głębokie często działa słabo. Jego mocną stroną jest praca z milionami lub miliardami punktów danych, stopniowo łądząc na zestawie wag sieci neuronowych, które uchwycą relacje między tymi przykładami. Jeśli system głębokiego uczenia podaje tylko kilka przykładów, wyniki rzadko są solidne. I oczywiście wiele z tego, co robimy jako ludzie, uczymy się w ciągu zaledwie kilku chwil; Gdy po raz pierwszy otrzymasz parę okularów 3-D, prawdopodobnie możesz je założyć i z grubsza wywnioskować, co się dzieje, bez konieczności przymierzania ich sto tysięcy razy. Głębokie uczenie po prostu nie jest przystosowane do tego rodzaju szybkiego uczenia się. Tam, gdzie Andrew Ng obiecał, że maszyny mogą wkrótce zautomatyzować wszystko, co dana osoba może zrobić w ciągu sekundy, bardziej realistyczny pogląd mógłby brzmieć: „Jeśli typowa osoba może wykonać zadanie umysłowe w czasie krótszym niż jedna sekunda, a my możemy zebrać ogromną ilość bezpośrednio istotne dane, mamy szansę na walkę – o ile problemy, które faktycznie napotykamy, nie różnią się zbyt od danych treningowych, a domena nie zmienia się zbyt w czasie.” To dobrze w przypadku gier planszowych, takich jak Go lub szachy z regułami, które były statyczne przez tysiąclecia, ale jak zauważyliśmy we wstępie, w wielu rzeczywistych problemach z uzyskaniem odpowiedniej ilości danych są nierealne, a nawet niemożliwe. Na przykład duża część tego, dlaczego głębokie uczenie się zmagają z językiem, polega na tym, że w języku nowe zdania o świeżym znaczeniu są nieskończone, a każde z nich subtelnie różni się od poprzedniego. Im bardziej różnią się Twoje rzeczywiste problemy od danych, których użyłeś do trenowania systemu, tym mniej prawdopodobne jest, że system będzie niezawodny. Uczenie głębokie jest nieprzejrzyste. Podczas gdy klasyczne systemy eksperckie składają się z reguł, które można stosunkowo łatwo zrozumieć (np. „Jeśli dana osoba ma podwyższoną liczbę białych krwinek, prawdopodobnie ma infekcję”), sieci neuronowe składają się z ogromnych macierzy liczb, z których praktycznie żadna nie ma intuicyjnego sensu dla zwykłych ludzi. Nawet przy użyciu zaawansowanych narzędzi eksperci mają trudności ze zrozumieniem, dlaczego poszczególne sieci neuronowe podejmują decyzje, które podejmują. Istnieje nierozwiązana zagadka dotycząca tego, dlaczego sieci neuronowe działają tak dobrze, jak działają, oraz brak jasności co do dokładnych okoliczności, w których tak nie jest. Sieć neuronowa, która uczy się konkretnego zadania, może uzyskać, powiedzmy, 95 procent poprawności w jakimś teście. Ale co wtedy? Często bardzo trudno jest zrozumieć, dlaczego sieć popełnia błędy w pozostałych 5 procentach, nawet jeśli niektóre z tych błędów są dramatycznymi błędami, których nigdy nie popełniłby żaden człowiek, jak na przykład pomylenie lodówki i znaku parkingowego. A jeśli te błędy mają znaczenie i nie możemy zrozumieć, dlaczego system je popełnia, mamy problem. Problem jest szczególnie dotkliwy, ponieważ sieci neuronowe nie są w stanie podać wyjaśnień w ludzkim stylu swoich odpowiedzi, poprawnych lub innych. Zamiast tego sieci neuronowe są „czarnymi skrzynkami”; robią to, co robią i trudno jest zrozumieć, co jest w środku. W takim stopniu, w jakim moglibyśmy chcieć na nich polegać w sprawach takich jak prowadzenie samochodu lub wykonywanie prac domowych, jest to poważny problem. Jest to również problem, jeśli chcemy korzystać z głębokiego uczenia w kontekście większych systemów, ponieważ tak naprawdę nie potrafimy określić parametrów pracy - żargon inżyniera, kiedy coś działa, a kiedy nie. Leki na receptę zawierają wszelkiego rodzaju informacje o tym, które potencjalne skutki uboczne są niebezpieczne, a

które po prostu nieprzyjemne, ale ktoś, kto sprzedaje policji system rozpoznawania twarzy oparty na uczeniu głębokim, może nie być w stanie powiedzieć klientowi z dużym wyprzedzeniem, kiedy to zadziała, a kiedy nie. Może okazać się, że w słoneczne dni całkiem dobrze sprawdza się w przypadku rasy białej, ale zawodzi w przypadku Afroamerykanów w przyćmionym świetle; trudno to poznać, chyba że poprzez eksperymentowanie. Kolejną konsekwencją nieprzejrzystości głębokiego uczenia się jest to, że nie pasuje ono naturalnie do wiedzy o tym, jak działa świat. Nie ma łatwego sposobu, aby powiedzieć głębokiej sieci, że „jabłka rosną na drzewach” lub „jabłka mają tendencję do spadania, a nie w górę, kiedy spadają z drzew”. Jest to w porządku, jeśli wszystko, co musisz zrobić, to rozpoznać jabłko; to druzgocące, jeśli chcesz, aby system głębokiego uczenia się interpretował, co dzieje się w urządzeniu Rube Goldberga, gdy piłka spada z rampy, zjeżdża po zsywie i na schodach ruchomych. Głębokie uczenie się jest kruche. Jak widzieliśmy w części początkowej, głębokie uczenie się może być idealne w jednej sytuacji, a całkowicie błędne w innej. Halucynacje, takie jak nieistniejąca lodówka, którą widzieliśmy wcześniej, nie są czymś jednorazowym, którego się czepiamy; pozostają uporczywym problemem, lata po pierwszym zauważeniu problemu. Według jednego z badań nowsze modele nadal podlegają temu samemu, popełniając takie błędy w 7 do 17 procent przypadków. W typowym przykładzie zdjęcie pokazuje dwie uśmiechnięte kobiety rozmawiające przez telefony komórkowe, z kilkoma nieostrymi drzewami za nimi, z jedną kobietą twarzą do aparatu, a drugą odwróconą tak, że widoczny był tylko jej policzek. Powstały podpis - Kobieta rozmawiająca przez telefon komórkowy siedząc na ławce - jest mieszanką częściowo dokładnych szczegółów i całkowitej konfabulacji, być może napędzanej przez dziwactwa statystyczne w zestawie treningowym; jedna z kobiet zniknęła, a ławka pojawiła się znikąd. Nietrudno sobie wyobrazić, jak podobne mechanizmy mogą sprawić, że zautomatyzowany robot bezpieczeństwa błędnie zinterpretuje telefon komórkowy jako broń. Istnieją dosłownie dziesiątki sposobów na oszukanie głębokich sieci. Niektórzy badacze MIT na przykład zaprojektowali trójwymiarowego żółwia, którego system głębokiego uczenia pomylił z karabinem. A wbicie żółwia pod wodę (gdzie zwykle nie można znaleźć karabinu) nie pomogło. Podobnie zespół nałożył trochę pianki na piłkę baseballową umieszczoną w brązowej rękawiczce, a sieć uznała, że piłka to espresso – pod każdym kątem – nawet wtedy, gdy piłka znajdowała się tuż przed rękawicą baseballową. Inny zespół naukowców dodał dyskretnie małe plamki losowego szumu w małych zakamarkach obrazów i oszukał sieci neuronowe, by sądziły, że świnka-skarbonka to „tygrysi kot”. Inny zespół dodał małe naklejki z psychodelicznymi obrazkami tosterów do rzeczywistych obiektów, takich jak banan, i oszukał system głębokiego uczenia się, by sądził, że ta scena to toster, a nie, powiedzmy, banan obok małego, zniekształconego obrazu tosteru. Jeśli twój syn lub córka przegapiły banana, spieszyłybyś z nimi do neurologa. Do tego dochodzi celowo zmieniony znak stopu, który system głębokiego uczenia błędnie odczytuje jako znak ograniczenia prędkości. Jeszcze inny zespół badawczy porównał systemy głębokiego uczenia się z ludźmi w dwunastu różnych zadaniach, w których obrazy były zniekształcane na tuzin różnych sposobów, takich jak zamiana obrazów kolorowych na czarno-białe, zastępowanie kolorów, obracanie obrazu i tak dalej. Ludzie prawie zawsze radzili sobie lepiej. Nasze systemy wizualne są zbudowane tak, aby były solidne; głębokie uczenie, nie tak bardzo. Inne badanie wykazało, że głębokie uczenie ma problemy z rozpoznawaniem zwykłych obiektów w nietypowych pozach, takich jak odwrócony autobus szkolny, który pomyłono z pługiem śnieżnym. Sprawy stają się jeszcze dziwniejsze, gdy przechodzimy do języka. Informatycy ze Stanford, Robin Jia i Percy Liang, przeprowadzili badanie systemów, które pracują nad zadaniem SQuAD, w którym systemy głębokiego uczenia próbowały podkreślić odpowiedzi na pytania w tekście. Systemy głębokiego uczenia otrzymały ten tekst-

Peyton Manning został pierwszym rozgrywającym, który poprowadził dwie różne drużyny do wielu Super Bowl. Jest także najstarszym rozgrywającym, który kiedykolwiek grał w Super Bowl w wieku 39 lat. Poprzedni rekord został uzyskany przez Johna Elwaya, który poprowadził Broncos do zwycięstwa w



Super Bowl XXXIII w wieku 38 lat i jest obecnie wiceprezesem wykonawczym ds. operacji piłkarskich w Denver i Główny menadżer.

i to pytanie -

Jak nazywa się rozgrywający, który miał 38 lat w Super Bowl XXXIII?

Jeden z systemów głębokiego uczenia poprawnie podkreślił „John Elway”. Jak na razie dobrze. Jednakże, kiedy Jia i Liang podali systemowi ten sam akapit, ale dodali na końcu jedno dodatkowe zdanie, zawierające całkowicie nieistotne informacje – „Quarterback Jeff Dean miał koszulkę numer 17 w Champ Bowl XXXIV” - a następnie zadali to samo pytanie dotyczące Super Bowl. , system był całkowicie zdezorientowany, przypisując zwycięstwo Jeffowi Deanowi, a nie Johnowi Elwayowi, łącząc dwa zdania o dwóch różnych mistrzostwach i nie wykazując prawdziwego zrozumienia żadnego z nich. Inne badanie wykazało, jak łatwo jest sfalszować systemy odpowiedzi na pytania, zadając pytania częściowe. Poleganie przez systemy głębokiego uczenia się na korelacji, a nie na zrozumieniu, regularnie powoduje, że uderzają brzęczyk z losowym odgadnięciem, nawet jeśli pytanie nie zostało jeszcze zakończone. Na przykład, jeśli zapytasz system „Ile”, otrzymasz odpowiedź „2”; jeśli zapytasz „Jaki sport?” otrzymasz odpowiedź „tenis”. Pobaw się tymi systemami przez kilka minut i poczujesz, że wchodzisz w interakcję z skomplikowaną sztuczką salonową, a nie z prawdziwą inteligencją. Jeszcze dziwniejsza wersja tego samego problemu może wystąpić w tłumaczeniu maszynowym. Kiedy Tłumacz Google otrzymał tekst wejściowy pies pies pies pies pies pies pies pies pies pies pies pies pies i poproszono o przetłumaczenie tego z języka joruba (i kilku innych języków) na angielski, zwrócono to tłumaczenie:

Zegar końca świata wskazuje trzy minuty do dwunastej. Doświadczamy charakterów i dramatycznych wydarzeń na świecie, które wskazują, że coraz bardziej zbliżamy się do czasów ostatecznych i powrotu Jezusa.

W końcu głębokie uczenie się nie jest tak głębokie. Ważne jest, aby pamiętać, że w określeniu głębokiego uczenia słowo „głębokie” odnosi się do liczby warstw w sieci neuronowej i nic więcej. „Głęboko” w tym kontekście nie oznacza, że system nauczył się czegoś szczególnie bogatego koncepcyjnie o danych, które widział. Algorytm uczenia głębokiego wzmocnienia, który napędza system gier DeepMind Atari, może na przykład grać w miliony gier w Breakout i nigdy tak naprawdę nie nauczyć się, czym jest wiosło, co elegancko zademonstrował niedawno start-up AI Vicarious. W Breakout gracz porusza wiosłem tam i z powrotem wzdłuż poziomej linii. Jeśli zmienisz grę tak, aby wiosło było teraz kilka pikseli bliżej cegieł (co w ogóle nie przeszkadzałoby człowiekowi), cały system DeepMind rozpadnie się. (Zespół z Berkeley wykazał coś podobnego w przypadku Space Invaders: małe fragmenty hałasu zakłóciły wydajność, pokazując, jak powierzchowny był system do nauki gier.) Wydaje się, że niektórzy ludzie w tej dziedzinie w końcu dostrzegają te punkty. Profesor Uniwersytetu w Montrealu Yoshua Bengio, jeden z pionierów głębokiego uczenia się, przyznał niedawno, że „głębokie [sieci neuronowe] mają tendencję do uczenia się powierzchniowych regularności statystycznych w zbiorze danych, a nie abstrakcyjnych pojęć wyższego poziomu”. W wywiadzie pod koniec 2018 r. Geoff Hinton i Demis Hassabis, założyciel DeepMind, podobnie zasugerowali, że ogólna sztuczna inteligencja nie była nawet bliższa rzeczywistości. Jeden kończy się rodzajem problemu „długiego ogona”, w którym jest kilka pospolitych przypadków (gruba głowa zbioru danych), które są łatwe, ale także duża liczba rzadkich przedmiotów (długi ogon), które są bardzo trudno się nauczyć. Łatwo jest uzyskać głębokie uczenie, które powie Ci, że grupa małych dzieci gra we Frisbee, ponieważ jest tak wiele oznaczonych przykładów. O wiele trudniej jest zmusić go do powiedzenia ci, co jest niezwykłego na tym zdjęciu:



Psy, koty, konie-zabawki i powozy są zupełnie powszechne, ale ta konkretna konfiguracja elementów nie jest, a głębokie uczenie się nie ma pojęcia, co robić.

Dlaczego więc głębokie uczenie się jest tak wyprzedane, biorąc pod uwagę wszystkie te problemy? Jest oczywiście skuteczny w aproksymacji statystycznej z dużymi zestawami danych i ma pewien rodzaj elegancji - jedno proste równanie, które wydaje się tak wiele rozwiązywać. Ma też znaczną wartość handlową. Jednak z perspektywy czasu jest tak samo oczywiste, że czegoś brakuje. Widzimy cały entuzjazm dla głębokiego uczenia się jako przykład iluzorycznej luki w postępie, o której mówiliśmy w rozdziale otwierającym; w przypadku niektórych zadań głębokie uczenie się może odnieść sukces, ale to nie znaczy, że za zachowaniem kryje się prawdziwa inteligencja. Głębokie uczenie to zupełnie inna bestia niż ludzki umysł. W najlepszym przypadku głębokie uczenie się jest rodzajem idiotycznego uczonego, z cudownymi zdolnościami percepcyjnymi, ale bardzo małym ogólnym zrozumieniem. Łatwo jest znaleźć skuteczne systemy uczenia głębokiego do oznaczania zdjęć — między innymi Google, Microsoft, Amazon i IBM oferują komercyjne systemy, które to robią, a biblioteka oprogramowania sieci neuronowych Google TensorFlow daje szansę każdemu studentowi informatyki wolny. Łatwo jest również znaleźć skuteczne systemy głębokiego uczenia do rozpoznawania mowy, które w tym momencie znów są mniej więcej towarem. Ale rozpoznawanie mowy i rozpoznawanie obiektów to nie inteligencja, to tylko wycinki inteligencji. Aby uzyskać prawdziwą inteligencję, potrzebujesz również rozumowania, języka i analogii, z których żadna nie jest tak dobrze obsługiwana przez obecną technologię. Na przykład nie mamy jeszcze systemów sztucznej inteligencji, które mogłyby rzetelnie rozumieć umowy prawne, ponieważ sama klasyfikacja wzorców nie wystarczy. Aby zrozumieć umowę prawną, musisz być w stanie uzasadnić, co jest powiedziane, a czego nie, jak różne klauzule odnoszą się do wcześniej ustalonego prawa i tak dalej; głębokie uczenie się tego nie robi. Nawet rzetelne podsumowywanie fabuł starych filmów dla Netflix byłoby zbyt wiele, aby zapytać. Rzeczywiście, nawet w tym kawałku poznania, jakim jest percepcja, która najbardziej przypomina mocną stronę głębokiego uczenia się, postęp jest tylko częściowy; głębokie uczenie może identyfikować obiekty, ale nie może zrozumieć relacji między nimi i często może zostać oszukane. W innych obszarach, takich jak rozumienie języka i codzienne rozumowanie, uczenie głębokie pozostaje daleko w tyle za ludźmi. Zdecydowana większość tego, co napisano w popularnych mediach o głębokim uczeniu się, sprawia wrażenie, jakby postęp w jednym z tych obszarów był równoznaczny z postępem we wszystkich. Na przykład MIT Technology Review umieścił Deep Learning na swojej corocznej liście

przełomowych technologii w 2013 roku i podsumował to w następujący sposób: Dzięki ogromnej mocy obliczeniowej maszyny mogą teraz rozpoznawać obiekty i tłumaczyć mowę w czasie rzeczywistym. Sztuczna inteligencja w końcu staje się inteligentna. Ale logika jest błędna; tylko dlatego, że rozpoznasz sylabę lub border collie nie oznacza, że jesteś mądry. Nie wszystkie problemy poznawcze są sobie równe. Ekstrapolacja od sukcesu w jednym aspekcie poznania do sukcesu we wszystkich aspektach poznania jest poddawaniem się iluzji postępu. W ostatecznym rozrachunku głębokie uczenie jest zarówno piękne, jak i tragiczne. Jest piękny, bo w dobry dzień wymaga bardzo mało pracy; często nie trzeba spędzać dużo czasu na nieznośnej pracy związanej z inżynierią funkcji, a w najlepszym przypadku maszyna zajmuje się dużą częścią tego, co należy zrobić. To tragiczne, ponieważ nic nie gwarantuje, że jakikolwiek system w prawdziwym świecie udzieli właściwej odpowiedzi, gdy jej potrzebujesz, ani nie zapewni, że będziesz mógł go debugować, jeśli nie zadziała. Zamiast tego jest pod pewnymi względami bardziej jak sztuka niż nauka; wypróbujesz różne rzeczy i jeśli masz wystarczającą ilość danych, zwykle działają. Ale nie możesz tego udowodnić z góry, tak jak możesz udowodnić pewne twierdzenie o geometrii. Żadna teoria nie przewiduje dokładnie, które zadania głębokie uczenie się może, a których nie może rozwiązać w solidny sposób. Często trzeba to wypróbować, empirycznie. Widzisz, co działa, a co nie, często majstrując przy początkowym systemie i zestawie danych, aż uzyskasz pożądane wyniki. Czasami to proste. Czasami jest ciężko. Głębokie uczenie jest niezwykle cennym narzędziem dla sztucznej inteligencji – spodziewamy się, że będzie nadal odgrywać ważną i integralną rolę w przyszłości, i że ludzie wymyślą dla niego wiele innych kreatywnych zastosowań, o których nawet nie pomyśleliśmy. Ale o wiele bardziej prawdopodobne jest, że będzie to tylko jeden komponent w szeregu narzędzi niż samodzielne rozwiązanie. Ostatecznie stało się tak, że ludzie byli niezwykle podekscytowani konkretnym zestawem algorytmów, które są niezwykle przydatne, ale pozostają bardzo daleko od prawdziwej inteligencji - tak jakby odkrycie śrubokrętu nagle umożliwiło podróże międzygwiazdne. Nic nie może być dalej od prawdy. Potrzebujemy śrubokręta, ale będziemy też potrzebować znacznie więcej.

Nie mamy przez to na myśli, że systemy głębokiego uczenia się nie mogą robić rzeczy, które wydają się inteligentne, ale mamy na myśli, że samo uczenie głębokie nie ma elastyczności i adaptacyjnego charakteru prawdziwej inteligencji. W nieśmiertelnych słowach Prawa 31 Prawa Akina Projektowania Statków Kosmicznych: „Nie możesz dostać się na Księżyc, wspinając się po kolejnych wyższych drzewach”. W dalszej części książki opiszemy, co trzeba zrobić, aby dostać się na Księżyc – lub przynajmniej dostać się do maszyn, które potrafią myśleć, rozumować, rozmawiać i czytać z wszechstronnością przeciętnego człowieka. To, czego potrzebujemy, to nie tylko „głębsze” uczenie się, w sensie posiadania większej liczby warstw w sieci neuronowej, ale głębsze zrozumienie. Potrzebujemy systemów, które naprawdę potrafią wnioskować o złożonej korelacji między bytami, które są ze sobą powiązane przyczynowo w ciągle zmieniającym się świecie. Aby zobaczyć, co przez to rozumiemy, nadszedł czas, aby zagłębić się w dwie z najtrudniejszych dziedzin w sztucznej inteligencji: czytanie i roboty.