

Poprawa sztucznej inteligencji dzięki głębokiemu uczeniu się

Gazety, czasopisma biznesowe, sieci społecznościowe i nietechniczne strony internetowe mówią to samo: AI to fajne rzeczy i zrewolucjonizuje świat dzięki głębokiej nauce. Sztuczna inteligencja to znacznie większa dziedzina niż uczenie maszynowe, a głębokie uczenie się to tylko niewielka część uczenia maszynowego. Ważne jest, aby rozróżnić szumy używane do zwabienia inwestorów i pokazać, co ta technologia może zrobić, co jest ogólnym celem tej części. Ta część pomoże zrozumieć głębokie uczenie się z praktycznego i technicznego punktu widzenia oraz zrozumieć, co można osiągnąć w najbliższym czasie, badając jego możliwości i ograniczenia. Rozpoczyna się historią i podstawami sieci neuronowych. Następnie przedstawia aktualny stan wynikający ze splotowych sieci neuronowych, rekurencyjnych sieci neuronowych (zarówno w przypadku nadzorowanego uczenia się), jak i generatywnych sieci przeciwników (rodzaj uczenia się bez nadzoru).

Kształtowanie sieci neuronowych podobnych do ludzkiego mózgu

Poniższe sekcje przedstawiają rodzinę algorytmów uczenia się, które czerpią inspirację z działania mózgu. Są to sieci neuronowe, podstawowy algorytm plemienia łączników, który najlepiej naśladuje neurony w ludzkich mózgowach na mniejszą skalę. Łączność to podejście oparte na uczeniu maszynowym oparte na neuronauce, a także przykład biologicznie połączonych sieci.

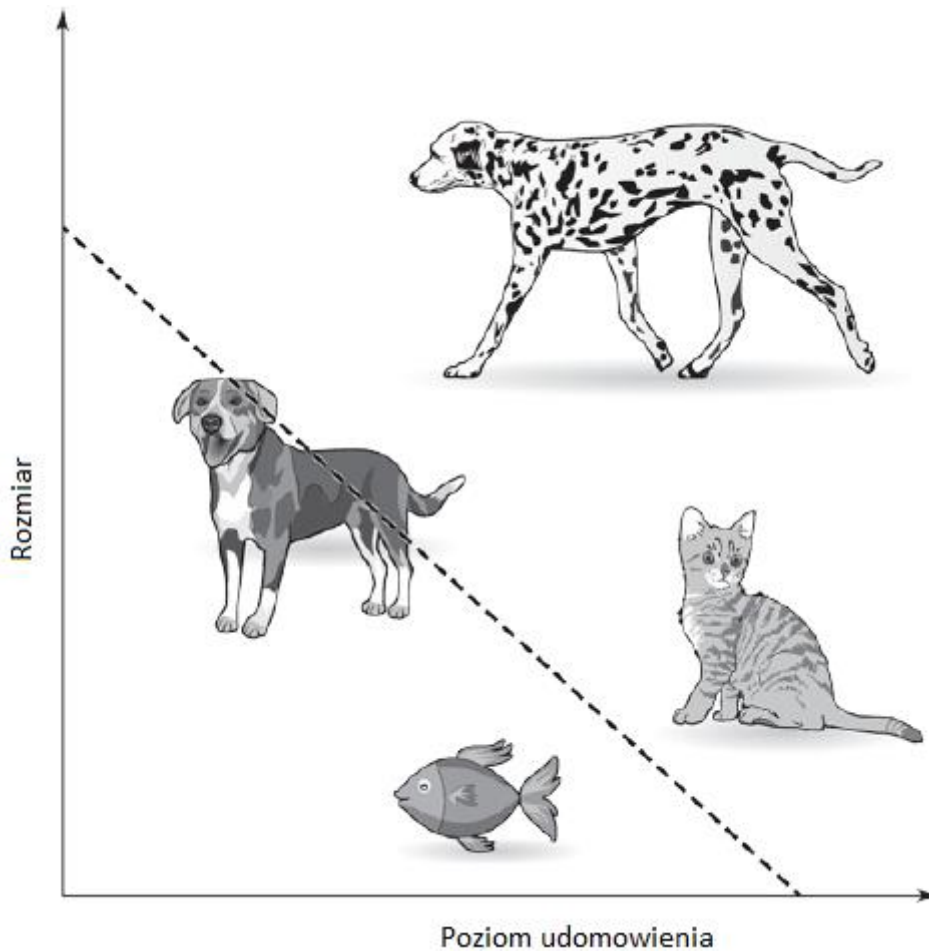
Przedstawiamy neuron

Ludzki mózg ma miliony neuronów, które są komórkami, które odbierają, przetwarzają i przekazują sygnały elektryczne i chemiczne. Każdy neuron posiada jądro z włóknami, które działają jako dane wejściowe, dendryty, które odbierają sygnały z innych neuronów, oraz jedno włókno wyjściowe, akson, który kończy się synapsami przeznaczonymi do komunikacji zewnętrznej. Neurony łączą się z innymi neuronami i przesyłają informacje między nimi za pomocą substancji chemicznych, podczas gdy informacje w samym neuronie są przetwarzane elektrycznie.. Odwrotna inżynieria przetwarzania sygnałów przez mózg pomaga łącznikom w definiowaniu sieci neuronowych opartych na analogiach biologicznych i ich składnikach, przy użyciu terminów mózgowych, takich jak neurony, aktywacja i połączenia jako nazw operacji matematycznych. Sieci neuronowe przypominają serię mnożenia i sumowania, gdy sprawdzasz ich formuły matematyczne. Jednak algorytmy te są niezwykle skuteczne w rozwiązywaniu złożonych problemów, takich jak rozpoznawanie obrazu i dźwięku lub tłumaczenie języka maszynowego. Używając specjalistycznego sprzętu, mogą szybko wykonywać obliczenia predykcyjne.

Począwszy od cudownego perceptronu

Algorytmem rdzenia sieci neuronowej jest neuron (zwany także jednostką). Wiele neuronów ułożonych w połączoną strukturę tworzy sieć neuronową, przy czym każdy neuron łączy się z wejściami i wyjściami innych neuronów. Tak więc neuron może wprowadzać dane z przykładów lub przekazywać wyniki innych neuronów, w zależności od swojej lokalizacji w sieci neuronowej. Frank Rosenblatt z Cornell Aeronautical Laboratory stworzył pierwszy przykład tego rodzaju neuronu, perceptron, kilka dekad temu. Opracował perceptron w 1957 r. Pod patronatem Laboratorium Badań Marynarki Wojennej Stanów Zjednoczonych (NRL). Rosenblatt był zarówno psychologiem, jak i pionierem w dziedzinie sztucznej inteligencji. Biegły w kognitywistyce, jego pomysłem było stworzenie komputera, który mógłby uczyć się metodą prób i błędów, tak jak robi to człowiek. Perceptron był po prostu sprytnym sposobem na prześledzenie linii oddzielającej w prostej przestrzeni utworzonej przez dane wejściowe, jak pokazano na rysunku, w której masz dwie cechy (w tym przypadku rozmiar i poziom udomowienia zwierzęcia), aby rozróżnić dwie klasy (psy i koty w tym przykładzie). Formuła perceptronowa tworzy linię w przestrzeni kartezjańskiej, w której przykłady mniej lub bardziej idealnie

dzielą się na grupy. Podejście jest podobne do Naive Bayesa, opisanego w rozdziale 10, który sumuje prawdopodobieństwa warunkowe pomnożone przez ogólne w celu sklasyfikowania.



Perceptron nie zdawał sobie sprawy z pełnych oczekiwań twórcy lub zwolenników finansów. Wkrótce wykazał ograniczoną pojemność, nawet w specjalizacji rozpoznawania obrazów. Ogólne rozczarowanie zapaliło pierwszą zimę AI i porzucenie łączności do 1980 roku. Jednak niektóre badania kontynuowano pomimo utraty funduszy.

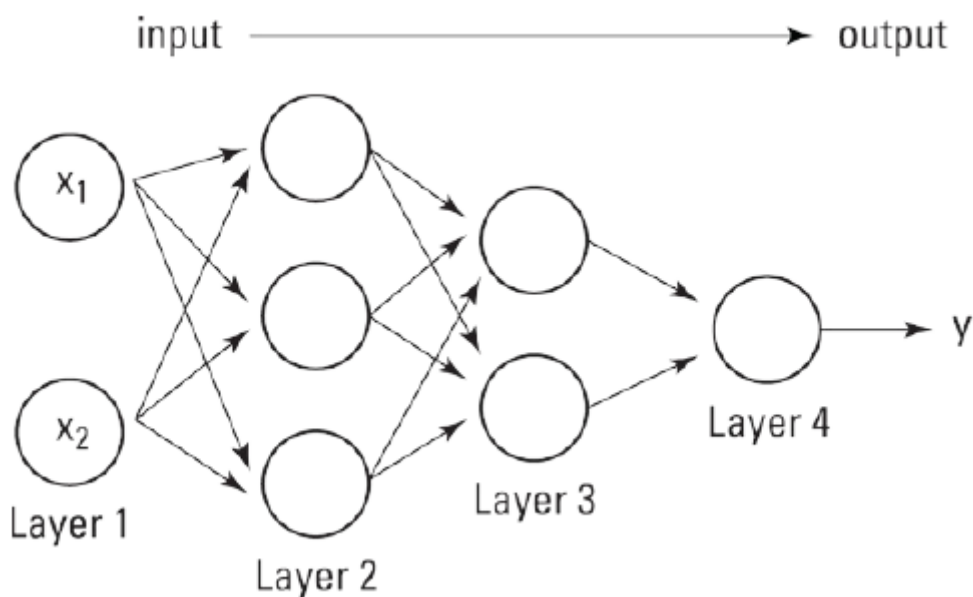
Później eksperci próbowali stworzyć bardziej zaawansowany perceptron i udało im się. Neurony w sieci neuronowej są dalszą ewolucją perceptronu: jest ich wiele, łączą się ze sobą i naśladują nasze neurony, gdy aktywują się pod pewnym bodźcem. Obserwując funkcje ludzkiego mózgu, naukowcy zauważyli, że neurony odbierają sygnały, ale nie zawsze uwalniają swój własny sygnał. Uwolnienie sygnału zależy od ilości odbieranego sygnału. Kiedy neuron otrzymuje wystarczającą liczbę bodźców, wyzwala odpowiedź; w przeciwnym razie milczy. W podobny sposób neurony algorytmiczne po otrzymaniu danych sumują je i wykorzystują funkcję aktywacji do oceny wyniku. Jeśli odbierany sygnał wejściowy osiąga określony próg, neuron przekształca i przekazuje wartość wejściową; w przeciwnym razie po prostu umiera. Sieci neuronowe używają specjalnych funkcji zwanych funkcjami aktywacyjnymi do odpalenia wyniku. Wszystko, co musisz wiedzieć, to to, że są kluczowym składnikiem sieci neuronowej, ponieważ umożliwiają sieci rozwiązywanie złożonych problemów. Są jak drzwi, pozwalające na przejście lub zatrzymanie sygnału. Nie pozwalają jednak po prostu na przejście sygnału; przekształcają to w użyteczny sposób. Na przykład głębokie uczenie się nie jest możliwe bez skutecznych funkcji aktywacyjnych, takich jak Rectified Linear Unit (ReLU), dlatego funkcje aktywacyjne są ważnym aspektem tej historii.

Naśladowanie uczenia się mózgu

W sieci neuronowej należy najpierw wziąć pod uwagę architekturę, czyli rozmieszczenie elementów sieci neuronowej. W poniższych sekcjach omówiono zagadnienia dotyczące architektury sieci neuronowej.

Biorąc pod uwagę proste sieci neuronowe

W przeciwieństwie do innych algorytmów, które mają ustalony potok, który określa sposób, w jaki algorytmy odbierają i przetwarzają dane, sieci neuronowe wymagają, abyś zdecydował o sposobie przepływu informacji poprzez ustalenie liczby jednostek (neuronów) i ich rozmieszczenia w warstwach zwanych architekturą sieci neuronowej, ponieważ pokazane na rysunku.



Rysunek pokazuje prostą architekturę sieci neuronowej. Zauważ, jak warstwy filtrują i przetwarzają informacje w sposób progresywny. Jest to sygnał zwrotny, ponieważ dane przekazują jeden kierunek do sieci.

Połączenia łączą wyłącznie jednostki w jednej warstwie z jednostkami w następnej warstwie (informacje przepływają od lewej do prawej). Nie ma połączeń między jednostkami w tej samej warstwie lub z jednostkami poza kolejną warstwą. Ponadto informacja przesuwana się do przodu (od lewej do prawej). Przetwarzane dane nigdy nie wracają do poprzednich warstw neuronów. Korzystanie z sieci neuronowej przypomina korzystanie z warstwowego systemu filtrowania wody:

Wlewasz wodę z góry, a woda jest filtrowana na dole. Woda nie ma możliwości powrotu; po prostu idzie do przodu i prosto w dół, i nigdy z boku. W ten sam sposób sieci neuronowe wymuszają przepływ danych przez sieć i mieszanie się ze sobą, zgodnie z architekturą sieci. Dzięki wykorzystaniu najlepszej architektury do mieszania elementów sieć neuronowa tworzy nowe złożone elementy na każdej warstwie i pomaga osiągnąć lepsze przewidywania. Niestety nie można ustalić najlepszej architektury bez empirycznego wypróbowania różnych rozwiązań i sprawdzenia, czy dane wyjściowe pomagają przewidzieć wartości docelowe po przepłynięciu przez sieć. Czasami pojęcia można lepiej zrozumieć, jeśli zostaną bezpośrednio przetestowane w rzeczywistości. Google oferuje sieć neuronową, w której

faktycznie możesz przetestować, jak działa sieć neuronowa w intuicyjny sposób, co robisz dodając lub usuwając warstwy i zmieniając rodzaje aktywacji.

Odkrywanie sekretu leży w wadze

Sieci neuronowe mają różne warstwy, z których każda ma swoją własną wagę. Wagi reprezentują siłę połączenia między neuronami w sieci. Gdy waga połączenia między dwiema warstwami jest niewielka, oznacza to, że sieć zrzuca wartości przepływające między nimi i sygnalizuje, że wybranie tej trasy raczej nie wpłynie na ostateczną prognozę. Podobnie duża wartość dodatnia lub ujemna wpływa na wartości otrzymywane przez następną warstwę, a tym samym określa pewne prognozy. To podejście jest analogiczne do komórek mózgowych, które nie są samodzielne, ale łączą się z innymi komórkami. W miarę wzrostu doświadczenia połączenia między neuronami mają tendencję do osłabiania lub wzmacniania w celu aktywacji lub dezaktywacji określonej komórki sieci mózgowej regiony, powodując inne przetwarzanie lub działanie (reakcja na niebezpieczeństwo, na przykład, jeśli przetworzone informacje sygnalizują sytuację zagrażającą życiu). Każda kolejna warstwa jednostek sieci neuronowej stopniowo przetwarza wartości zaczerpnięte z cech, jak w przypadku przenośnika taśmowego. Gdy sieć przesyła dane, dociera do każdej jednostki jako zsumowana wartość wytworzona przez wartości obecne w poprzedniej warstwie i ważona przez połączenia w obecnej warstwie. Gdy dane otrzymane z innych neuronów przekroczą pewien próg, funkcja aktywacji zwiększa wartość przechowywaną w jednostce; w przeciwnym razie tłumia sygnał, zmniejszając go. Po przetworzeniu funkcji aktywacji wynik jest gotowy do przejścia do następnej warstwy połączenia. Kroki te powtarzają się dla każdej warstwy, aż wartości osiągną koniec, a otrzymasz wynik. Wagi połączeń zapewniają nowy sposób miksowania i komponowania danych wejściowych, tworząc nowe funkcje poprzez kreatywne miksowanie przetworzonych danych wejściowych ze względu na wagi i funkcje aktywacji. Aktywacja czyni również nieliniową wynikową rekombinację danych wejściowych odbieranych przez połączenia. Oba te składniki sieci neuronowej umożliwiają algorytmowi naukę złożonych funkcji celu, które reprezentują związek między cechami wejściowymi a wynikiem docelowym.

Zrozumienie roli propagacji wstecznej

Uczenie się odbywa się w ludzkim mózgu z powodu tworzenia i modyfikacji synaps między neuronami, w oparciu o bodźce otrzymane w wyniku prób i błędów. Sieci neuronowe zapewniają sposób na odtworzenie tego procesu jako formuły matematycznej zwanej propagacją wsteczną. Oto jak ta architektura połączonych ze sobą jednostek obliczeniowych może rozwiązać problemy: jednostki otrzymują przykład, a jeśli nie zgadną poprawnie, przywracają problem w systemie istniejących wag za pomocą propagacji wstecznej i naprawiają go, zmieniając niektóre wartości. Ten proces trwa wiele iteracji, zanim sieć neuronowa będzie mogła się uczyć. Iteracje w sieci neuronowej nazywane są epokami, nazwą, która idealnie pasuje, ponieważ sieć neuronowa może wymagać dni lub tygodni treningu, aby nauczyć się skomplikowanych zadań. Matematyka propagacji wstecznej jest dość zaawansowana i wymaga znajomości pojęć takich jak pochodne. Propagacja wsteczna jako koncepcja jest wystarczająco intuicyjna, aby uchwycić i przekazać, ponieważ przypomina to, co ludzie robią, wykonując zadanie przy użyciu iterowanych przybliżonych prób i błędów. Od czasu pojawienia się algorytmu propagacji wstecznej w latach 70. deweloperzy naprawiali go wiele razy i obecnie dyskutują, czy go przemyśleć. Propagacja wsteczna jest istotą teraźniejszego renesansu AI. W przeszłości usprawnienia procesu uczenia się każdej sieci neuronowej skutkowały nowymi aplikacjami i ponownym zainteresowaniem tą techniką. Również obecna rewolucja głębokiego uczenia się, która polega na ożywieniu sieci neuronowych (porzuconych na początku lat 90.), wynikała z kluczowych postępów w sposobie uczenia się sieci neuronowych na podstawie ich błędów.

Przedstawiamy Deep Learning

Po propagacji wstecznej kolejna poprawa sieci neuronowych doprowadziła do głębokiego uczenia się. Badania kontynuowano pomimo zimy AI, a sieci neuronowe pokonały problemy techniczne, takie jak zanikający gradient, który ogranicza wymiarowość sieci neuronowych. Programiści potrzebowali większych sieci neuronowych, aby rozwiązać niektóre problemy, tak duże, że były nie do pomyślenia w latach 80. Ponadto naukowcy zaczęli wykorzystywać zmiany w procesorach i procesorach graficznych (procesory graficzne lepiej znane z zastosowania w grach). Znikający gradient występuje, gdy próbujesz przesłać sygnał przez sieć neuronową, a sygnał szybko zanika do wartości bliskich zeru; nie może już przejść przez funkcje aktywacyjne. Dzieje się tak, ponieważ sieci neuronowe są powielonymi łańcuchami. Każde zwielokrotnienie poniżej zera szybko zmniejsza wartości, a funkcje aktywacyjne potrzebują wystarczająco dużych wartości, aby przepuścić sygnał. Im dalej są warstwy neuronu od wyjścia, tym większe prawdopodobieństwo, że zostaną zablokowane w aktualizacjach, ponieważ sygnały są zbyt małe, a funkcje aktywacyjne je zatrzymają. W rezultacie sieć przestaje się uczyć jako całość lub uczy się w niewiarygodnie wolnym tempie. Nowe rozwiązania pomagają uniknąć problemu zanikającego gradientu i wielu innych problemów technicznych, umożliwiając tworzenie większych głębokich sieci w przeciwieństwie do prostszych płytkich sieci z przeszłości. Głębokie sieci są możliwe dzięki badaniom naukowców z University of Toronto w Kanadzie, takich jak Geoffrey Hinton, którzy nalegali na pracę w sieciach neuronowych, nawet gdy wydawały się najbardziej staromodnym podejściem do uczenia maszynowego. Procesory graficzne to potężne macierze i jednostki obliczeniowe wektorowe niezbędne do propagacji wstecznej. Technologie te sprawiają, że szkolenia sieci neuronowych są osiągalne w krótszym czasie i są dostępne dla większej liczby osób. Badania otworzyły także świat nowych aplikacji. Sieci neuronowe mogą uczyć się na ogromnych ilościach danych i korzystać z dużych zbiorów danych (obrazów, tekstu, transakcji i danych w mediach społecznościowych), tworząc modele, które stale działają lepiej, w zależności od przepływu danych, którymi je karmisz. Duży gracz, tacy jak Google, Facebook, Microsoft i IBM, zauważyli nowy trend, a od 2012 r. zaczęli przejmować firmy i zatrudniać ekspertów (Hinton współpracuje teraz z Google; LeCun, twórca Convolutional Neural Networks, kieruje badaniami AI na Facebooku) w nowej dziedzinie głębokiego uczenia się. Projekt Google Brain, prowadzony przez Andrew Ng i Jeffa Deana, zgromadził 16 000 komputerów, aby obliczyć sieć głębokiego uczenia o ponad miliardach wag, umożliwiając w ten sposób bez nadzoru naukę z filmów na YouTube. Sieć komputerowa może nawet określić, czym jest kot sam, bez żadnej interwencji człowieka

Wyjaśnienie różnicy w głębokim uczeniu się

Głębokie uczenie się może wydawać się tylko większą siecią neuronową, która działa na większej liczbie komputerów - innymi słowy, po prostu przełomem w matematyce i technologii obliczeniowej, który udostępnia takie większe sieci. Jednak coś głęboko jakościowego zmieniło się w głębokim uczeniu się w porównaniu z płytkimi sieciami neuronowymi. To coś więcej niż zmiana paradygmatu genialnych technologii w pracy. Dogłębne uczenie się przenosi paradygmat uczenia maszynowego z tworzenia funkcji (funkcji, które ułatwiają uczenie się i które trzeba tworzyć przy użyciu analizy danych) na uczenie się funkcji (złożone funkcje tworzone automatycznie na podstawie rzeczywistych funkcji).

Taki aspekt nie mógłby zostać wykryty w inny sposób podczas korzystania z mniejszych sieci, ale staje się widoczny, gdy używasz wielu warstw sieci neuronowej i dużej ilości danych. Kiedy patrzysz w głąb głębokiego uczenia się, możesz być zaskoczony znalezieniem wielu starych technologii, ale, co zadziwiające, wszystko działa tak, jak nigdy wcześniej. Ponieważ badacze w końcu wymyślili, jak sprawić, by niektóre proste rozwiązania Goodola ze sobą współpracowały, duże zbiory danych mogą automatycznie filtrować, przetwarzać i przekształcać dane. Na przykład nowe aktywacje, takie jak ReLU, nie są wcale takie nowe; są znane od perceptronu. Również zdolności rozpoznawania obrazów, które początkowo uczyniły głębokie uczenie się tak popularnymi, nie są nowe. Początkowo głębokie

uczenie się osiągnęło duży rozmach dzięki sieciom neuronowym konwergencji (CNN). Sieci odkryte w latach 80. XX wieku przez francuskiego naukowca Yanna LeCuna, sieci takie przynoszą teraz zadziwiające wyniki, ponieważ używają wielu warstw neuronowych i dużej ilości danych. To samo dotyczy technologii, która pozwala maszynie zrozumieć ludzką mowę lub tłumaczyć z jednego języka na inny; to dziesięciolecia starej technologii, do której badacz powrócił i zaczął pracować nad nowym paradygmatem głębokiego uczenia. Oczywiście część różnicy stanowią również dane (więcej o tym później), zwiększone użycie procesorów graficznych i sieci komputerowe. Wraz z równoległością (więcej komputerów umieszczonych w klastrach i działających równolegle) procesory graficzne pozwalają tworzyć większe sieci i skutecznie trenować je na większej ilości danych. Szacuje się, że procesor graficzny wykonuje pewne operacje 70 razy szybciej niż jakikolwiek procesor, co pozwala na skrócenie czasu szkolenia dla sieci neuronowych z tygodni do dni lub nawet godzin.

Znajdowanie jeszcze inteligentniejszych rozwiązań

Głębokie uczenie się wpływa na skuteczność AI w rozwiązywaniu problemów z rozpoznawaniem obrazu, tłumaczeniem maszynowym i rozpoznawaniem mowy, które początkowo były rozwiązywane przez klasyczną sztuczną inteligencję i uczenie maszynowe. Ponadto przedstawia nowe i korzystne rozwiązania

- * Ciągłe uczenie się przy użyciu uczenia się online
- * Rozwiązania wielokrotnego użytku wykorzystujące uczenie się metodą transferu
- * Więcej demokratyzacji AI przy użyciu platform open source
- * Proste, proste rozwiązania wykorzystujące kompleksowe uczenie się

Korzystanie z nauki online

Sieci neuronowe są bardziej elastyczne niż inne algorytmy uczenia maszynowego i mogą kontynuować trening, pracując nad prognozami i klasyfikacjami. Ta zdolność pochodzi z algorytmów optymalizacyjnych, które umożliwiają uczenie się sieci neuronowych, które mogą pracować wielokrotnie na małych próbkach przykładów (nazywanych uczeniem wsadowym) lub nawet na pojedynczych przykładach (zwanym uczeniem online). Sieci do głębokiego uczenia się mogą budować swoją wiedzę krok po kroku i być otwarci na nowe informacje, które mogą nadejść (np. Umysł dziecka, który jest zawsze otwarty na nowe bodźce i doświadczenia edukacyjne). Na przykład, aplikacja do głębokiego uczenia się na stronie mediów społecznościowych może zostać przeszkolona w zakresie obrazów kotów. Gdy ludzie publikują zdjęcia kotów, aplikacja rozpoznaje je i oznacza je odpowiednią etykietą.

Kiedy ludzie zaczynają publikować zdjęcia psów w sieci społecznościowej, sieć neuronowa nie musi ponownie uruchamiać szkolenia; można kontynuować, ucząc się obrazów psów. Ta funkcja jest szczególnie przydatna do radzenia sobie ze zmiennością danych internetowych. Sieć głębokiego uczenia się może być otwarta na nowości i dostosowywać swoje wagi, aby sobie z nią poradzić.

Korzystanie z nauki transferu

Elastyczność jest przydatna nawet po ukończeniu szkolenia przez sieć, ale należy go ponownie wykorzystać do celów innych niż wstępne uczenie się. Sieci, które rozróżniają obiekty i prawidłowo je klasyfikują, wymagają dużo czasu i dużej mocy obliczeniowej, aby nauczyć się, co robić. Rozszerzenie możliwości sieci na nowe rodzaje obrazów, które nie były częścią poprzedniej nauki, oznacza przeniesienie wiedzy na ten nowy problem (transfer nauki). Na przykład możesz przenieść sieć, która potrafi rozróżniać psy i koty, aby wykonać zadanie polegające na wykrywaniu potraw z makaronem i

serem. Korzystasz z większości warstw sieci takimi, jakie są (zamrażasz je), a następnie pracujesz na końcowych warstwach wyjściowych (dostrajanie). W krótkim czasie i przy mniejszej liczbie przykładów sieć będzie obowiązywać czego nauczył się w rozróżnianiu psów i kotów od makaronu i sera. Będzie działał nawet lepiej niż sieć neuronowa przeszkolona tylko do rozpoznawania makaronu i sera.

Uczenie się przez przeniesienie jest czymś nowym w większości algorytmów uczenia maszynowego i otwiera potencjalny rynek do przenoszenia wiedzy z jednej aplikacji do drugiej, z jednej firmy do drugiej. Google już to robi, udostępniając swoje ogromne repozytorium danych, upubliczniając sieci, w których zbudował. To jest krok w demokratyzacji głębokiego uczenia się poprzez umożliwienie każdemu dostępu do jego potencjału.

Demokratyzacja za pomocą platform open source

Obecnie sieci mogą być dostępne dla wszystkich, w tym dostęp do narzędzi do tworzenia sieci głębokiego uczenia się. Nie chodzi tylko o publiczne publikowanie artykułów naukowych wyjaśniających, jak działa głębokie uczenie się; to kwestia programowania. Na początku głębokiego uczenia musiałeś budować każdą sieć od zera jako aplikację opracowaną w języku takim jak C++, co ograniczało dostęp do kilku dobrze wyszkolonych specjalistów. Dzisiejsze możliwości tworzenia skryptów (na przykład przy użyciu Pythona; przejdź do <http://www.python.org>) są lepsze ze względu na szeroką gamę platform do głębokiego uczenia się typu open source, takich jak Google TensorFlow lub PyTorch przez Facebook. Ramy te umożliwiają replikację najnowszych osiągnięć w głębokim uczeniu się przy użyciu prostych poleceń. Wraz z wieloma światłami pojawiają się cienie. Sieci neuronowe potrzebują ogromnych ilości danych do działania, a dane nie są dostępne dla wszystkich, ponieważ przechowują je większe organizacje. Uczenie się przez transfer może złagodzić brak danych, ale tylko częściowo, ponieważ niektóre aplikacje wymagają rzeczywistych danych. W konsekwencji demokratyzacja AI jest ograniczona. Co więcej, systemy głębokiego uczenia się są tak złożone, że ich wyniki są zarówno trudne do wyjaśnienia (umożliwiając rozkwit stronniczości i dyskryminacji), jak i słabe, ponieważ sztuczki mogą oszukać te systemy (szczegółowe informacje można znaleźć na stronie <https://www.dvhardware.net/article67588.html>). Każda sieć neuronowa może być wrażliwa na ataki nieprzyjacielskie, które są manipulacjami wprowadzanymi w celu oszukania systemu w celu udzielenia błędnej odpowiedzi.

Korzystanie z kompleksowego uczenia się

Wreszcie, głębokie uczenie się pozwala na uczenie się od końca do końca, co oznacza, że rozwiązuje problemy w łatwiejszy i bardziej bezpośredni sposób niż poprzednie rozwiązanie głębokiego uczenia się, co może mieć większy wpływ na rozwiązywanie problemów. Możesz rozwiązać trudny problem, na przykład rozpoznanie znanych twarzy przez AI lub prowadzenie samochodu. Stosując klasyczne podejście AI, musiałeś podzielić problem na łatwiejsze do opanowania podproblemy, aby osiągnąć akceptowalny wynik w możliwym czasie. Na przykład, jeśli chcesz rozpoznać twarze na zdjęciu, poprzednie systemy AI umieściły problem w tych częściach:

1. Znajdź twarze na zdjęciu.
2. Przytnij twarze ze zdjęcia.
3. Przetwórz przycięte twarze, aby uzyskać pozę podobną do zdjęcia z dowodu osobistego.
4. Nakarm przetworzone przycięte twarze jako przykłady uczenia się do neuronowej sieci do rozpoznawania obrazów.

Dzisiaj możesz nakarmić zdjęcie architekturą dogłębnego uczenia się i poprowadzić ją, aby nauczyć się znajdować twarze na zdjęciach, a następnie je klasyfikować. Możesz zastosować to samo podejście do tłumaczenia języka, rozpoznawania mowy, a nawet samochodów samojezdnych (jak omówiono w rozdziale 14). We wszystkich przypadkach wystarczy przekazać dane wejściowe do systemu głębokiego uczenia się i uzyskać pożądany rezultat.

Wykrywanie krawędzi i kształtów na podstawie obrazów

Konwolucyjne sieci neuronowe (znane również jako ConvNet lub CNN) przeżywają ostatniorenesans głębokiego uczenia się. Następujące sekcje omówią, w jaki sposób CNN pomagają wykrywać krawędzie i kształty obrazu w zadaniach takich jak odszyfrowanie odręcznego tekstu.

Począwszy od rozpoznawania znaków

CNN nie są nowym pomysłem. Pojawiły się pod koniec lat 80 jako praca Yanna LeCuna (obecnie dyrektora AI w Facebooku), kiedy pracował w AT&T Labs-Research wraz z Yoshua Bengio, Leonem Bottou i Patrickiem Haffnerem w sieci LeNet5. W tym czasie posiadanie maszyny zdolnej do odszyfrowywania odręcznych numerów było sporym wyczynem, który pomógł usłudze pocztowej w automatyzacji wykrywania kodów pocztowych oraz sortowania poczty przychodzącej i wychodzącej

Wcześniej programiści osiągnęli pewne wyniki, łącząc obraz liczbowy w celu wykrycia sieci neuronowej. Każdy piksel obrazu podłączony do węzła w sieci. Problemem zastosowania tego podejścia jest to, że sieć nie jest w stanie uzyskać niezmienności translacji, czyli zdolności do rozszyfrowania liczby w różnych warunkach wielkości, zniekształcenia lub położenia na obrazie. Podobna sieć neuronowa może wykryć tylko podobne liczby - te, które widziała wcześniej. Popętnił także wiele błędów. Przekształcanie obrazu przed wprowadzeniem go do sieci neuronowej częściowo rozwiązało problem, zmieniając rozmiar, przenosząc, czyszcząc piksele i tworząc specjalne fragmenty informacji dla lepszego przetwarzania w sieci. Ta technika, zwana tworzeniem funkcji, wymaga zarówno wiedzy specjalistycznej w zakresie niezbędnych transformacji obrazu, jak i wielu obliczeń w zakresie analizy danych. Zadania rozpoznawania obrazów w tym czasie były bardziej dziełem rzemieślnika niż naukowca. Konwolucje łatwo rozwiązały problem niezmienności tłumaczenia, ponieważ oferują inne podejście do przetwarzania obrazu w sieci neuronowej. Konwolucje są podstawą LeNet5 i zapewniają podstawowe bloki konstrukcyjne dla wszystkich rzeczywistych CNN wykonujących następujące czynności:

*Klasyfikacja obrazu: Określenie, który obiekt pojawia się na obrazie

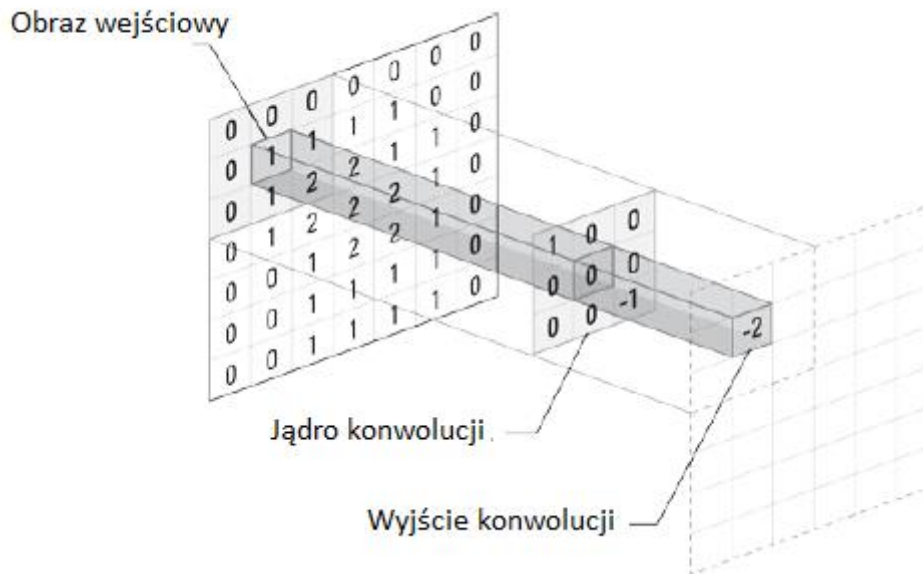
*Wykrywanie obrazu: Wyszukiwanie, gdzie obiekt znajduje się na obrazie

*Segmentacja obrazu: Rozdzielanie obszarów obrazu na podstawie ich treści; na przykład na obrazie drogi, oddzielającej samą drogę od samochodów na niej i pieszych

Wyjaśniając, jak działają konwolucje

Aby zrozumieć, jak działają konwolucje, zaczynasz od wejścia, które jest obrazem złożonym z jednej lub więcej warstw pikseli, nazywane kanałami, przy użyciu wartości od 0 (piksel jest w pełni włączony) do 256 (piksel jest wyłączony). Na przykład obrazy RGB mają osobne kanały dla kolorów czerwonego, zielonego i niebieskiego. Mieszanie tych kanałów generuje paletę kolorów, które widzisz na ekranie. Dane wejściowe podlegają prostym przekształceniom w celu zmiany skali wartości pikseli (na przykład w celu ustawienia zakresu od zera do jednego), a następnie przekazania tych wartości. Przekształcanie danych ułatwia pracę zwojów, ponieważ są to po prostu operacje mnożenia i sumowania, jak pokazano na rysunku 11-4. Splotowa warstwa neuronowa pobiera małe fragmenty obrazu, mnoży wartości

pikseli wewnątrz porcji przez siatkę szczególnie wymyślonych liczb, sumuje wszystko uzyskane z mnożenia i rzutu je na następną warstwę neuronową.



Taka operacja jest elastyczna, ponieważ propagacja wsteczna stanowi podstawę mnożenia liczbowego wewnątrz splotu oraz wartości, które filtry splotu są cechami obrazu, które są ważne dla sieci neuronowej, aby mogła wykonać swoje zadanie klasyfikacji. Niektóre zwoje przechwytyją tylko linie, niektóre tylko krzywe lub specjalne wzory, bez względu na to, gdzie pojawiają się na obrazie (i jest to właściwość niezmienności translacji zwojów). Gdy dane obrazu przechodzą przez różne konwolucje, są przekształcane, składane i renderowane w coraz bardziej złożone wzory, aż do splotu, który tworzy obrazy referencyjne (na przykład obraz przeciętnego kota lub psa), których wyszkolony CNN używa później do wykrywania nowych obrazów. Co ciekawe, ustawienie podstawowych architektur ConvNet nie jest trudne. Wyobraź sobie, że im więcej masz warstw, tym lepiej. Ustawiasz liczbę warstw splotu i niektóre cechy zachowania splotu, takie jak sposób tworzenia siatki (filtr, jądro lub wartości detektora cech), jak siatka przesuwa się na obrazie (krok) i jak zachowuje się wokół granic obrazu (wyściółka). Spojrzenie na działanie zwojów wskazuje, że głęboka głęboka nauka oznacza, że dane przechodzą głębsze transformacje niż w przypadku nowego algorytmu uczenia maszynowego lub płytkiej sieci neuronowej. Im więcej warstw, tym więcej transformacji przechodzi obraz i tym głębiej się on staje.

Postęp przy użyciu wyzwań związanych z obrazem

CNN to sprytny pomysł. AT&T faktycznie zaimplementował LeNet5 w bankomatach aby sprawdzać czytelników. Jednak kolejna zima AI zaczęła się w połowie lat 90, kiedy wielu badaczy i inwestorów straciło wiarę, że sieci neuronowe mogą zrewolucjonizować AI. Ponadto dane w tym czasie nie były skomplikowane. Badaczom udało się osiągnąć wyniki porównywalne z LeNet5, stosując nowe algorytmy uczenia maszynowego o nazwie Maszyny wektorów wsparcia (z grupy Analogerów) i Lasy losowe, wyrafinowanie drzew decyzyjnych z plemienia symbolistów). Tylko garstka badaczy, takich jak Geoffrey Hinton, Yann LeCun i Yoshua Bengio, rozwijała technologie sieci neuronowych, dopóki nowy zestaw danych nie zapewnił przełomu i zakończył zimę AI. Tymczasem w 2006 r. Fei-Fei Li, profesor informatyki na University of Illinois Urbana-Champaign (a teraz główny naukowiec w Google Cloud oraz profesor w Stanford), zapewnił więcej zestawów danych w świecie rzeczywistym w celu ulepszenia algorytmów testowych. Zaczęła gromadzić niesamowitą liczbę obrazów reprezentujących dużą liczbę klas obiektów. Ona i jej zespół wykonali tak ogromne zadanie, korzystając z Amazonka Mechanical Turk, usługi, której używasz, aby prosić ludzi o wykonanie dla ciebie mikro-zadań (takich jak klasyfikacja

obrazu) za niewielką opłatą. Wynikowy zestaw danych, ukończony w 2009 r., nazywał się ImageNet i zawierał 3,2 miliona oznaczonych obrazów, ułożonych w 5 247 hierarchicznie zorganizowanych kategorii. ImageNet wkrótce pojawił się na konkursie w 2010 roku, w którym sieci neuronowe udowodniły swoją zdolność do prawidłowej klasyfikacji obrazów ułożonych w 1000 klas. W ciągu siedmiu lat rywalizacji (wyzwanie definitywnie zakończone w 2017 roku) zwycięskie algorytmy podniosły dokładność przewidywania obrazów z 71,8 procent do 97,3 procent, co przewyższa ludzkie możliwości (tak, ludzie popełniają błędy w klasyfikacji obiektów). Na początku naukowcy zauważyli, że ich algorytmy zaczęły działać lepiej z większą ilością danych (wtedy nie było nic takiego jak ImageNet), a następnie zaczęli testować nowe pomysły i ulepszone architektury sieci neuronowych. Nawet jeśli konkursy ImageNet już się nie odbywają, badacze opracowują więcej architektur CNN, zwiększając dokładność lub możliwości wykrywania, a także niezawodność. W rzeczywistości wiele rozwiązań w zakresie głębokiego uczenia się jest nadal w fazie eksperymentalnej i nie jest jeszcze stosowanych w aplikacjach o znaczeniu krytycznym, takich jak bankowość lub bezpieczeństwo, nie tylko z powodu trudności w ich interpretacji, ale również z powodu możliwych luk.

Luki w zabezpieczeniach występują we wszystkich postaciach. Naukowcy odkryli, że dodając specjalnie opracowany szum lub zmieniając pojedynczy piksel na obrazie, CNN może radykalnie zmienić swoje odpowiedzi, nieukierunkowane (wystarczy oszukać CNN) lub ukierunkowane (chcesz, aby CNN podał konkretną odpowiedź) ataki. OpenAI jest niedochodową firmą badawczą zajmującą się sztuczną inteligencją. Chodzi o to, że CNN nie są jeszcze bezpieczną technologią. Nie możesz ich po prostu użyć zamiast oczu; musisz zachować przy tym szczególną ostrożność.

Nauka naśladowania sztuki i życia

CNN nie wpływał tylko na zadania widzenia komputerowego, ale są również ważne w wielu innych aplikacjach (na przykład są niezbędne do widzenia w samochodach z własnym napędem). CNN przekonało wielu badaczy do zainwestowania czasu i wysiłku w rewolucję głębokiego uczenia się. Konsekwentne badania i rozwój zrodziły nowe pomysły. Późniejsze testy w końcu wprowadziły innowacje do sztucznej inteligencji, pomagając komputerom nauczyć się rozumieć język mówiony, tłumaczyć pisane języki obce oraz tworzyć zarówno tekst, jak i zmodyfikowane obrazy, pokazując w ten sposób, jak złożone obliczenia dotyczące rozkładów statystycznych można przełożyć na rodzaj sztuki, kreatywności i wyobraźni. Jeśli mówisz o głębokim uczeniu się i jego możliwych zastosowaniach, musisz również wspomnieć o nawracających sieciach neuronowych (RNN) i generatywnych sieciach przeciwnych (GAN), w przeciwnym razie nie będziesz miał jasnego obrazu tego, co głębokie uczenie może zrobić dla sztucznej inteligencji.

Zapamiętywanie sekwencji, które mają znaczenie

Jedną ze słabości CNN jest brak pamięci. Dobrze sobie radzi ze zrozumieniem pojedynczego obrazu, ale próba zrozumienia obrazu w kontekście, np. klatki w filmie, przekłada się na niemożność uzyskania właściwej odpowiedzi na trudne wyzwania związane z AI. Wiele ważnych problemów to sekwencje. Jeśli chcesz zrozumieć książkę, czytasz ją strona po stronie. Sekwencje są zagnieżdżone. Na stronie znajduje się ciąg słów, a na słowo - ciąg liter. Aby zrozumieć książkę, musisz zrozumieć sekwencję liter, słów i stron. RNN jest odpowiedzią, ponieważ przetwarza rzeczywiste dane wejściowe podczas śledzenia poprzednich danych. Dane wejściowe w sieci nie tylko postępują naprzód, jak zwykle w sieci neuronowej, ale również wewnątrz niej zapętłają się. To tak, jakby sieć słyszy swoje echo. Jeśli podasz RNN sekwencję słów, sieć nauczy się, że kiedy zobaczy słowo poprzedzone pewnymi innymi słowami, może określić sposób uzupełnienia frazy. RNN nie są po prostu technologią, która może zautomatyzować kompilację danych wejściowych (tak jak w przypadku, gdy przeglądarka automatycznie uzupełnia wyszukiwane hasła podczas wpisywania słów). Ponadto RNN mogą

przekazywać sekwencje i zapewniać tłumaczenie jako dane wyjściowe, takie jak ogólne znaczenie frazy (tzn. AI może rozróżniać frazy, w których ważne jest sformułowanie) lub przetłumacz tekst na inny język (ponownie tłumaczenie działa w kontekście). Działa to nawet z dźwiękami, ponieważ możliwe jest interpretowanie niektórych modulacji dźwięku jako słów. RNN pozwalają komputerom i telefonom komórkowym z wielką precyzją zrozumieć nie tylko to, co powiedziałeś (to ta sama technologia, która automatycznie podpisuje napisy), ale także to, co chciałeś powiedzieć, otwierając drzwi do programów komputerowych, które z tobą czatują, oraz asystentów cyfrowych, takich jak Siri, Cortana i Alexa.

Odkrywanie magii rozmów AI

Chatbot to oprogramowanie, które może rozmawiać z tobą na dwa sposoby: słuchowy (mówisz z nim i słuchasz odpowiedzi) lub tekstowy (wpisujesz, co chcesz powiedzieć i czytasz odpowiedzi). Być może słyszałeś o tym pod innymi nazwami (agent konwersacyjny, chatterbot, talkbot i inni), ale chodzi o to, że możesz już użyć jednego na naszym smartfonie, komputerze lub specjalnym urządzeniu. Siri, Cortana i Alexa są dobrze znanymi przykładami. Możesz również wymieniać słowa z chatbotem, kontaktując się z obsługą klienta firmy przez Internet lub telefon, lub za pośrednictwem aplikacji na telefon komórkowy, gdy używasz Twittera, Slacka, Skype'a lub innych aplikacji do rozmowy. Chatboty to duży biznes, ponieważ pomagają firmom zaoszczędzić pieniądze na operatorach obsługi klienta - utrzymując stały kontakt z klientem i obsługując tych klientów - ale pomysł nie jest nowy. Nawet jeśli nazwa jest nowa (opracowana w 1994 roku przez Michaela Mauldina, wynalazcę wyszukiwarki Lycos), chatboty są uważane za szczyt sztucznej inteligencji. Według wizji Alana Turinga wykrycie silnej sztucznej inteligencji przez rozmowę z nią nie powinno być możliwe. Turing opracował słynny test oparty na rozmowach, aby ustalić, czy sztuczna inteligencja uzyskała inteligencję równoważną istocie ludzkiej. Masz słabą AI, gdy AI wykazuje inteligentne zachowanie, ale nie jest świadomy jak człowiek. Silna AI występuje, gdy AI naprawdę może myśleć jak człowiek. Test Turinga wymaga od człowieka oceny interakcji z dwoma podmiotami za pośrednictwem terminala komputerowego: jednego człowieka i jednej maszyny. Na podstawie rozmowy sędzia ocenia, która z nich jest sztuczną inteligencją. Turing zapewnił, że jeśli sztuczna inteligencja może skłonić człowieka do myślenia, że rozmowa jest prowadzona z innym człowiekiem, można uwierzyć, że sztuczna inteligencja jest na ludzkim poziomie sztucznej inteligencji. Problem jest trudny, ponieważ nie chodzi tylko o prawidłowe i poprawne gramatycznie udzielenie odpowiedzi, ale także o włączenie kontekstu (miejsce, czas i cechy osoby, z którą rozmawia AI) i wykazanie spójnej osobowości (AI powinna być jak prawdziwa osoba, zarówno w tle, jak i postawie). Od lat sześćdziesiątych wyzwanie testu Turinga stanowi motywację do opracowania chatbotów opartych na idei modeli opartych na wyszukiwaniu. Oznacza to, że użycie przetwarzania języka naturalnego (NLP) przetwarza język wprowadzony przez człowieka przesłuchującego. Niektóre słowa lub zestawy słów przywołują wstępnie ustawione odpowiedzi i informacje zwrotne z pamięci chatbota. NLP to analiza danych skoncentrowana na tekście. Algorytm dzieli tekst na tokeny (elementy frazy, takie jak rzeczowniki, czasowniki i przymiotniki) i usuwa wszelkie mniej przydatne lub mylące informacje. Tokenizowany tekst jest przetwarzany za pomocą operacji statystycznych lub uczenia maszynowego. Na przykład NLP może pomóc w oznaczaniu części mowy i identyfikowaniu słów i ich znaczenie lub określić, czy jeden tekst jest podobny do innego. Joseph Weizenbaum zbudował pierwszego tego typu chatbota, ELIZA, w 1966 roku jako formę psychoterapeuty komputerowego. ELIZA została wykonana z prostych heurystyk, które są podstawowymi frazami do dostosowania do kontekstu i słowami kluczowymi, które skłoniły ELIZA do przywołania odpowiedniej odpowiedzi z ustalonego zestawu odpowiedzi. Modele oparte na wyszukiwaniu działają dobrze, gdy są przesłuchiwane przy użyciu predefiniowanych tematów, ponieważ zawierają wiedzę ludzką, podobnie jak system ekspercki, dzięki czemu mogą odpowiadać odpowiednimi, poprawnymi gramatycznie zwrotami. Problemy pojawiają się w przypadku pytań nietypowych. Chatbot może próbować odeprzeć te pytania, odbijając je w innej formie (tak jak zrobiła

to ELIZA) i zostać zauważonym jako sztuczny mówca. Rozwiązaniem jest tworzenie nowych fraz, na przykład w oparciu o modele statystyczne, uczenie maszynowe, a nawet wstępnie przeszkolony RNN, które mogłyby być oparte na neutralnej mowie lub nawet odzwierciedlać osobowość konkretnej osoby. Takie podejście nazywa się modelami generatywnymi i stanowi dziś granicę botów, ponieważ generowanie języka w locie nie jest łatwe. Modele generatywne nie zawsze odpowiadają trafnie i poprawnie na frazy, ale wielu badaczy poczyniło ostatnio postępy, szczególnie w RNN. Jak zauważono w poprzednich znakach, sekret znajduje się w sekwencji: Podajesz sekwencję wejściową w jednym języku i sekwencję wyjściową w innym języku, jak w przypadku problemu tłumaczenia maszynowego. W takim przypadku podajesz sekwencję wejściową i sekwencję wyjściową w tym samym języku. Dane wejściowe są częścią konwersacji, a dane wyjściowe są następującą reakcją. Biorąc pod uwagę aktualny stan wiedzy w budowaniu chatbota, RNN świetnie sprawdzają się w przypadku krótkich wymian, chociaż uzyskanie doskonałych wyników dla dłuższych lub bardziej wyrażonych fraz jest trudniejsze. Podobnie jak w przypadku modeli opartych na wyszukiwaniu, RNN przywołują informacje, które zdobywają, ale nie w sposób zorganizowany. Jeśli zakres dyskursu jest ograniczony, systemy te mogą dawać dobre odpowiedzi, ale ulegają degradacji, gdy kontekst jest otwarty i ogólny, ponieważ potrzebują wiedzy porównywalnej do tego, co człowiek zdobywa w ciągu życia. (Ludzie są dobrymi rozmówcami opartymi na doświadczeniu i wiedzy.) Dane do szkolenia RNN są naprawdę kluczowe. Na przykład Google Smart Reply, chatbot firmy Google, oferuje szybkie odpowiedzi na e-maile. W prawdziwym świecie zwykle odpowiadał na większość rozmów z „Kocham cię”, ponieważ był szkolony na przykładach stronicznych. Coś podobnego stało się z chatbotem Taysem Microsoftu, którego umiejętność uczenia się na podstawie interakcji z użytkownikami popełniła błąd, ponieważ rozmowy były stroniczne i złośliwe. Jeśli chcesz poznać aktualny stan wiedzy w świecie chatbotów, możesz na bieżąco informować o corocznych konkursach chatbotów, w których testy Turinga są stosowane do obecnej technologii. Na przykład nagroda Lobnera jest najbardziej znana i właściwe miejsce na start. Choć wciąż niezdolny do zaliczenia testu Turinga, ostatnim zdobywcą nagrody Lobnera jak pamiętam było Mitsuku, oprogramowanie, które może uzasadniać określone obiekty zaproponowane podczas dyskursu; może także grać w gry, a nawet wykonywać magiczne sztuczki. Zmuszenie AI do konkurowania z innymi RNN AI może sprawić, że komputer z tobą porozmawia, a jeśli nie masz pojęcia, że sieć neuronowa reaktywuje sekwencje słów, których się wcześniej nauczyła, masz wrażenie, że dzieje się coś z inteligencją sceny. W rzeczywistości nie kryje się za tym żadna myśl ani rozumowanie, chociaż technologia nie tylko przypomina zaprogramowane frazy, ale jest dość wyrazista. Generator Adversarial Networks (GAN) to kolejny rodzaj technologii głębokiego uczenia się, która może zapewnić jeszcze silniejszą iluzję, że sztuczna inteligencja może wykazywać kreatywność. Ponownie, technologia ta polega na przywołaniu poprzednich przykładów i zrozumieniu przez maszynę, że przykłady zawierają reguły - zasady, którymi maszyna może się bawić, gdy dziecko bawi się klockami zabawek (technicznie rzecz biorąc, reguły są rozkładami statystycznymi leżącymi u podstaw przykładów). Niemniej jednak sieci GAN są niesamowitą technologią, która pokazała dość dużą liczbę przyszłych aplikacji. GAN powstały w wyniku pracy kilku badaczy z Departament d'informatique et de recherche operationnelle na Uniwersytecie w Montrealu w 2014 roku, a najbardziej znaczącym z nich jest Ian Goodfellow. Zaproponowane nowe podejście do głębokiego uczenia się natychmiast wzbudziło zainteresowanie i obecnie jest jedną z najczęściej badanych technologii, z ciągłym rozwojem i ulepszeniami. Yann LeCun Generative Adversarial Networks to „najciekawszy pomysł w ciągu ostatnich dziesięciu lat w uczeniu maszynowym”. W wywiadzie dla MIT Technology Review Ian Goodfellow wyjaśnia ten poziom entuzjazmu za pomocą tego intrygującego stwierdzenia: „Możesz myśleć o modelach generatywnych jako nadających sztucznej inteligencji formę wyobraźni”. Aby zobaczyć podstawowy GAN w akcji (jest teraz wiele wyrafinowanych wariantów, a więcej jest opracowywanych), potrzebujesz referencyjnego zestawu danych, zwykle składającego się z rzeczywistych danych, których przykładów chciałbyś użyć do nauki sieci GAN. Na przykład, jeśli masz zestaw danych obrazu psa, oczekujesz, że GAN dowie się,

jak wygląda pies z zestawu danych. Po zapoznaniu się z psami GAN może zaproponować wiarygodne, realistyczne obrazy psów, które różnią się od tych z początkowego zestawu danych. (Będą to nowe obrazy; zwykła replikacja istniejących obrazów jest uważana za błąd z GAN). Zbiór danych jest punktem początkowym. Potrzebujesz także dwóch sieci neuronowych, z których każda specjalizuje się w innym zadaniu i obie konkurują ze sobą. Jedna sieć nazywa się generatorem i pobiera dowolne dane wejściowe (na przykład sekwencję liczb losowych) i generuje dane wyjściowe (na przykład obraz psa), który jest artefaktem, ponieważ jest sztucznie tworzony za pomocą sieci generatora. Druga sieć to dyskryminator, który musi poprawnie odróżniać produkty generatora, artefakty, od przykładów w zbiorze danych szkoleniowych. Kiedy GAN rozpoczyna szkolenie, obie sieci próbują poprawić, stosując propagację wsteczną, w oparciu o wyniki dyskryminatora. Błędy popełniane przez dyskryminator przy odróżnianiu rzeczywistego obrazu od artefaktu przenoszą się na dyskryminator (jak w przypadku klasyfikacji sieci neuronowej). Prawidłowe odpowiedzi dyskryminatora propagują się jako błędy w generatorze (ponieważ nie był w stanie uczynić artefaktów podobnymi do obrazów w zestawie danych, a dyskryminator je zauważył). Oryginalne obrazy wybrane przez Goodfellow w celu wyjaśnienia, jak działa GAN, pochodzą od twórcy sztuki i badacza. Badacz ma umiejętności wykrywania wykutej sztuki, ale oszust poprawia się również, aby uniknąć wykrycia przez badacza. Możesz się zastanawiać, jak generator uczy się tworzyć właściwe artefakty, jeśli nigdy nie widzi oryginału. Tylko dyskryminator widzi oryginalny zestaw danych, gdy próbuje odróżnić prawdziwą sztukę od artefaktów generatora. Nawet jeśli generator nigdy nie analizuje niczego z oryginalnego zestawu danych, otrzymuje wskazówki dzięki pracy dyskryminatora. Są to niewielkie wskazówki, kierowane przez wiele nieudanych prób na początku od generatora. To tak, jakby nauczyć się malować Mona Lisę bez jej obejrzenia i tylko z pomocą znajomego, który mówi ci, jak dobrze zgadłeś. Sytuacja przypomina twierdzenie o nieskończonej armii małp, z pewnymi różnicami. W tym twierdzeniu oczekujesz, że małpy piszą wiersze Szekspira przez zwykłe szczęście. W tym przypadku generator używa losowości tylko na początku, a następnie powoli kieruje się informacjami zwrotnymi od dyskryminatora. Po pewnych modyfikacjach tego podstawowego pomysłu, sieci GAN stały się zdolne do:

- * Tworzenie fotorealistycznych obrazów obiektów, takich jak przedmioty modowe, a także projektowanie wnętrz lub wzornictwo przemysłowe na podstawie opisu słów (prosisz o żółty i biały kwiat i otrzymujesz go).

- * Modyfikowanie istniejących obrazów poprzez zastosowanie wyższej rozdzielczości, dodanie specjalnych wzorów i wypełnianie brakujących części (na przykład, chcesz usunąć osobę ze zdjęcia i GAN zastępuje lukę pewnym wiarygodnym tłem, i ukończenie architektury neuronowej:

- * Wiele aplikacji granicznych, takich jak generowanie ruchu ze zdjęć statycznych, tworzenie złożonych obiektów, takich jak pełne teksty (co nazywa się prognozowaniem strukturalnym, ponieważ dane wyjściowe nie są po prostu odpowiedzią, ale raczej zbiorem odpowiedzi powiązanych ze sobą), tworzenie danych dla nadzorowanego uczenia maszynowego, a nawet generowanie potężnej kryptografii

Sieci GAN są technologią pogranicza głębokiego uczenia się i istnieje wiele otwartych i nowych obszarów badań nad jej zastosowaniem w AI. Jeśli sztuczna inteligencja będzie miała kreatywną i kreatywną moc, prawdopodobnie będzie czerpać z technologii takich jak GAN. Możesz dowiedzieć się, co się dzieje z tą technologią, czytając strony GAN z OpenAI, niekomercyjnej firmy badawczej zajmującej się sztuczną inteligencją, założonej przez Grega Brockmana, Ilyę Sutskever, Elona Muska (założyciela PayPal, SpaceX i Tesli) oraz Sama Altmana.