

<https://doi.org/10.59444/uz.9788378425717.pp.405-411>

Maksymilian A. Brzezicki

Uniwersytecki Szpital im. Jana Radcliffe'a w Oxfordzie

JAK POSTAWIŁEM KAWĘ ELEKTRYCZNEJ INTELIGENCJI?



piątek, 9 lutego 2024 r.

tydzień czwarty semestru św. Hilarego

Jeśli jesteś urzędnikiem państwowym, zwłaszcza uwikłanym w aparat wszelkiego autorytetu represji wobec studenta, odradzam czytanie poniższego tekstu. Ta historia może ci się potem śnić po nocach, a nie chciałbym mieć kolejnej osoby na sumieniu. No trudno, jeśli zbudzisz się w nocy zlanym zimnym potem, to nie moja wina. Był rok 1920 – Lwów. Zdolny magister Stefan Banach wali na nogi kolejne zagwozдки fundamentalnej analizy matematyki. Zdolny, ale... leniwy. Nie chce mu się wypełnić ani jednego druku, zalega ze sprawozdaniami, a w biurze dziekanatu nie widziano go chyba nigdy. Jego niechęć do biurokracji znana była w całej Galicji. Problem nie mały, albowiem profesor Hugo Steinhaus chciał zatrzymać go na uczelni. O złożeniu papierów na studia doktorskie nie było mowy. Tylko dzięki litości katedry mógł wciąż tułać się po korytarzach Uniwersytetu Jana Kazimierza. Pewnego dnia młody kalkulant kreślił pokątnie wykresy na zakurzonych od pyłu kredowego tablicach. Do pokoju zapukał Steinhaus. „Stefan, słuchaj – powiedział. – Tam w sekretariacie są jacyś ludzie. Mają do ciebie pytania matematyczne”. Jak wielu naukowców, Banach bez pytania pochwycił szansę opowiedzenia komuś o swojej pasji. Zbiegł do kantorka i rozpoczął wartką dyskusję. A było z kim, bo strudzeni wędrowcy zadawali celne i prowokujące pytania. Nawet nie zauważył, gdy minęło kilka godzin. Na pytania odpowiadał w sposób wyczerpujący i błyskotliwy. Najstarszy przybysz zakończył spotkanie, wstał, uściśnął Banachowi dłoń i wyszedł z delegacją. Stefan uśmiechnięty od ucha do ucha żegnał petentów. A to heca! Jak często zdarza się, że ktoś tak dokładnie wypytuje go o matematyczne zawiłości?

Jeszcze bardziej uśmiechnięty był Steinhaus. „To była komisja egzaminacyjna z Warszawy – ogłosił dumnie. – Właśnie uzyskałeś doktorat”.

Urzędników zapewne ogarnie szewska pasja na myśl, że kiedyś tak się studentom luzowało. A jednak – sto lat temu nauka miała się wyśmienicie; bez pieczętek, rozporządzeń i obiegówek. Polska matematyka wiodła wtedy prym na globie. O dokonaniach Banacha, Steinhausa, Ulama, Mazura i Łomnickiego czytano z wypiekami na twarzy na całym świecie. Uczni godzinami przesiadywali w kawiarni „Szkocka” przy placu akademickim i wyklócali się o najrozmaitsze aspekty swojej pracy. Wywody rozrysowywali kopiowym ołówkiem na stolikach; ponoć debatowali tak długo, że nierzadko wypraszał ich właściciel lokalu, który błagalnie wskazywał na godziny zamknięcia. Następnego ranka – klops. Sprzątaczką zmyła blat, a mądre myśli poszły w niepamięć. Ulam wpadł jednak na świetny pomysł. Napisał obsłudze bilecik z prośbą, aby stolik odstawić na bok i nakryć obrusem, a za fatygę wręczył im kilka złotych. Ten manewr nie okazał się całkowicie skuteczny. Sporadycznie i tak zmywano im blaty do zera. Żona Banacha, Łucja, w końcu zlitowała się nad uczonymi i za 2,5 zł kupiła lwowskim geniuszom... zeszyt. Tam zapisywali sobie zadania i stawiali nagrody za rozwiązanie problemu. Za udowodnienie twierdzenia – mała czarna; a za dowiedzenie twierdzenia przeciwnego – gęś żywa.

Zeszyt przetrwał wojnę i do dziś stanowi Święty Graal światowej matematyki. „Szkocka księga” zawiera wiele problemów, nad którymi pochylają się adepci liczb i wzorów matematycznych na całej kuli ziemskiej. Lwowscy matematycy to bodaj kolejny obok Stanisława Lema i Rudolfa Weigla polski naukowy produkt eksportowy, który bardziej znany jest za granicą niż nad Wisłą. Plakaty z Banachem wiszą w oksfordzkich akademikach, a wielbicieli kina akcji zapewne pamiętają, jak Stanisław Ulam pracował z Oppenheimerem i von Neumannem w Projekcie Manhattan nad bombą atomową. Później, już w swoim zespole, lwowiak przygotował dla Amerykanów o wiele silniejszą bombę wodorową. Jest również autorem metody Monte Carlo, bez której nie istniałaby współczesna inżynieria budowlana, ekonomia, fizyka czy awiacja. Czasami oczami wyobraźni przenoszę się do „Szkockiej”. Jakie to musiało być widowisko! Dyskusja była ponoć tak zaciekle, że studenci oblegali słynny stolik. Dysputom nie było końca. Jak wybitny trzeba mieć umysł, żeby w Galicji wymyślić podwaliny całych działów matematyki, które później posłużą do lotów w kosmos, reakcji atomowych i obliczeń komputerów. Zainteresowanych tematem odsyłam do świetnej książki Mariusza Urbanka *Genialni*. Pozostaje jednak pytanie: dlaczego nawet sto lat później, w dobie sztucznej inteligencji i superkomputerów, nie jesteśmy w stanie rozwiązać zadań zapisanych na kawiarnianych blatach? Czyżby w głowie lwowskich matematyków kryło się coś, czego komputery nie są w stanie zreplikować?

Gdy zajrzemy pod maskę większości algorytmów, ujrzymy tam *sieci neuronowe*. Te wynalezione w latach 50. XX w. systemy komputerowe są wzorowane na neuronach biologicznych. Jak to działa? Przyjmijmy, dla ustalenia kwestii, że algorytm ma odczytać kod pocztowy z koperty. Klasycznie podeszlibyśmy do tego działania tak: rozumny człowiek popatrzy na kształt każdej cyfry i opisz je jej cechy. Jedynka to kreska z ogonkiem, trójka to dwa łuki, siódemka ma daszek itd. Następnie algorytm będzie musiał szukać tych cech w zeskanowanych bazgrołach. Tak wyszkolone modele działają całkiem dobrze. Ba! Mogą nawet świetnie wytłumaczyć, dlaczego zaklasyfikowały dany numer akurat do tej kategorii, np. „to ósemka, bo ma dwa kółka”. Niestety, bardzo szybko zderzają się ze ścianą skalowalności. Bazują bowiem na mądrości naukowca, który oznaczy kategorie i jakoby „wskáže” algorytmowi, na co patrzeć. Innym pomysłem jest podzielenie esów-floresów na kilka segmentów. Niech jeden „neuron” patrzy na górny kwadrat i „aktywuje się” tylko, gdy zauważy poziomą kreskę, drugi niech mierzy tylko środkowe skosy itd. Te sygnały przekazywane są do kolejnych „neuronów” na głębszych warstwach. W końcu ostatni „neuron” deklaruje odpowiedź: to liczba taka i taka. Neurony, zarówno te biologiczne, jak i elektryczne, przyjmują pewne dane na początku, procesują, przypisują im jakąś „wagę” (np. pewność, że to kreska, a nie kropka), a następnie przekazują dalej. Najciekawszą obserwacją jest to, że neuronom tym nie trzeba nawet wskazywać, na co mają patrzeć. Zazwyczaj inicjalizuje się je na losowo (tzn. patrz, na co tam chcesz), a one „uczą się”, na co mają patrzeć same. Przy kolejnych warstwach nie wiadomo już nawet, na co konkretnie patrzą. Stąd mówi się o „czarnej skrzynce” algorytmu.

Zwieńczeniem tego procesu jest pojedyncza linijka kodu, która zawsze przyprawia mnie o gęsią skórę: *wsteczna propagacja*. Wciśnięcie tego guzika powoduje uruchomienie najbardziej spektakularnych algorytmów, jakie wynalazł człowiek. Uzyskane przez sieć neuronową wyniki porównuje się z zestawem odpowiedzi testowych. Jeśli algorytm odpowiednio odczytał cyfrę kodu pocztowego – świetnie! Otrzymuje nagrodę. Jeśli nie – wymierza mu się karę. I tak w koło Macieju, miliony i miliardy razy, aż nie uzyska się prawie stuprocentowej pewności, że maszyna „nauczyła się” odczytywać ludzkie pismo. Tak wytrenowana sieć może już służyć ludzkości i bez trudu odczytywać kolejne inskrypcje, włącznie z tym, co nakreślił ktoś w egipskiej świątyni kilka tysięcy lat temu. Modyfikacje i bardziej zaawansowane wersje tego systemu znajdują zastosowanie w wielu aplikacjach, których używamy na co dzień, na przykład tłumaczenie tekstu ze zdjęcia w Google Translate, rozpoznawanie głosu przez Siri czy zabójczo skuteczne rekomendacje Tiktoka.

Powyżej opisany proces może przypominać praktykę uczącego się człowieka. Powne granie tej samej melodii w szkole muzycznej, ciągle wykonywanie gimnastycznego manewru lub uczenie się, których mięśni należy używać podczas wiosłowania.

Tysiące neuronów, które nieustannie dążą do perfekcji. Czymże lepsze jest dziecko uczące się chodzić od małego algorytmiku? Jeśli kogoś takie myśli przyprawiają o niepokój, polecam zapoznać się z uczącymi się modelami językowymi. BabyLM duka jak mały bobas po kilku godzinach uczenia się. Dajmy mu więcej czasu, a powstanie LLaMA, która produkuje eseje już na poziomie przeciętnego licealisty. SCISPACE w mniej niż minutę dokona przeglądu czasopism naukowych i zsyntezuje je w postaci świetnie napisanego akapitu – z cytatami i bibliografią – a gotowy produkt będzie na wyższym poziomie, niż gdyby został wykonany przez większość moich studentów z Oxfordu. W kilka sekund przeprowadza analizę, za którą kiedyś słono płacono statystykom i analitykom. Wystarczy popatrzeć na miny publiczności, gdy na konferencjach przedstawiam ten algorytm. W try miga rozpisuje im przegląd systematyczny i to zaraz po dyskusji, w której organiczne mózgi zarzekały się, że nie zrobią tego szybciej niż w półtora roku.

Jeszcze ciekawszym do rozpatrzenia jest problem uczuć. W innych tekstach udowadniałem, jak wielkie modele językowe potrafią „obrażać się” i nie mieć humoru odpowiadając na zadane im pytanie. Algorytmy sprawnie dokonują tzw. analizy sentymentu i skutecznie odczytują kontekst, np. smutek w ogłoszeniu „Sprzedam buciki dziecięce. Nieużywane”. A co z kreatywnością? Jeśli nie mieszkasz pod kamieniem, to w 2024 r. trudno przeżyć choćby jeden dzień bez spotkania się z generowaną przez algorytmy treścią. Teksty, grafiki, animacje, filmy i szablony zalegają na hałdach YouTube’a, Tiktoka, Instagrama i wszelkiej maści gazetek czy portali lokalnych. Nawet przedszkolak potrafi wklepać kilka słów do DALL-E czy StableDiffusion i otrzymać nieodróżnialne od realnych fotografie i obrazy. Raz na jakiś czas media obiegają spreparowane DeepFakiem filmy i zdjęcia celebrytów w sensacyjnych sytuacjach. Słynny amerykański lingwista Noam Chomski gniewa się, że to jedynie maszyny plagiatowe – algorytmy, które wciągają ludzką sztukę, mielą ją w swoich bebeczach i wypluwają coś na jej kształt. Ale czy na pewno? Czym różni się algorytm od artysty, który chodzi do National Gallery w Londynie, by podpatrzeć mistrzowski pędzel? Czy ktoś karał Mickiewicza za plagiatowanie antyku? A czy zżymamy się na poetę, który wybiera się w romantyczną podróż pociągiem po walijskiej Snowdonii, by opisać naturę w swoich limerykach? Bynajmniej. A skoro każdy człowiek kreuje treści na podstawie swoich... sieci neuronowych... i pochłoniętej treści swoich poprzedników... to czym różni się od algorytmu?

Nietrudno zauważyć, dlaczego sztuczna inteligencja często porównywana jest do mózgu człowieka. Kuszącym jest nawet *zrównanie* jej z ludzkim istnieniem. Czy dni dominacji ludzkiego gatunku są już policzone, a algorytmy rychło przejmą nad nami kontrolę? Mój profesor ze Stanfordu, Andrew Ng, strasznie irytuje się na takie stwierdzenia. Rzeczywiście, w wielu aspektach zastanawianie się, czy algorytm „żyje”, „myśli”

czy „jest świadomy” przywodzi jedynie bezsensowne wnioski. Dla nieuzbrojonego oka „mądry” jest algorytm, który przejdzie test Turinga, czyli, mówiąc w skrócie, oszuka człowieka, że jest innym człowiekiem. I tu natrafiamy na pierwszą różnicę z organicznym mózgiem. Taki chatGPT, Gemini czy LLaMA przechodzą test Turinga z palcem w bucie. A mimo to dławią się w przypadku nawet prostych zadań logicznych. Tęgie elektronowe mózgi matematyczne jak Wolfram Alfa być może scałkują trudną funkcję, ale nie poradzą sobie z interpretacją nawet najprostszego wiersza Jana Brzechwy. Nasuwa się tu jeden wniosek – mamy dużą ilość wyspecjalizowanych algorytmów, które służą do wielu zadań. I tam rzeczywiście pokonują człowieka. Ale nie ma na razie systemu, który potrafiłby okiełznać każdy problem z plastycznością i łatwością równą człowiekowi.

Najnowsze badania pokazują również, że uczenie się sieci neuronowych tylko do złudzenia przypomina ludzki mózg. Od lat naukowcy zastanawiali się, dlaczego człowiek potrafi tak dobrze generalizować informacje. Ekipa z laboratorium prof. Bogacza z Brain Network Dynamics Unit z Oxfordu odkryła, że dzieje się to dzięki fizjologicznym różnicom w uczeniu się biologicznych mózgów. Wbrew pozorom człowiek nie zdobywa informacji na podstawie *wstecznej propagacji* (jak sieci neuronowe), ale przez tzw. *prospektywną konfigurację neuronów*. Młody radiolog nie musi przeglądać setek tysięcy skanów i na chybił trafił zgadywać, co jest nie tak z pacjentem. Przed otwarciem swoich pierwszych klisz uczy się bowiem teorii. Na studiach i w trakcie specjalizacji zdobywa wiedzę na temat podstaw fizyki i nauk fundamentalnych. Uzbrojony w taką wiedzę teoretyczną rezydent może rozpoznawać obrazy lepiej i skuteczniej niż nierozgrzany algorytm. Oczywiście zarówno człowiek, jak i maszyna będą musieli przerobić solidną ilość skanów i przeprowadzić niejedną analizę błędów, by zostać ekspertami. Badania pokazują, że z czasem dochodzą do podobnych wyników. Zarówno specjalista radiologii, jak i algorytm będą w stanie skutecznie rozpoznać COVID-19 czy raka płuc na tomografii klatki piersiowej. Ale ten drugi będzie miał już problem z wyhaczeniem, że chorujący na COVID-19 ma również złamane żebro lub malformację aorty, bo po prostu nie był do tego przyuczony. Im bardziej generalizujemy, im więcej przypadków rzadkich i niejednoznacznych, i im krótszy czas nauki, tym bardziej wygrywa biologiczny mózg. Zwłaszcza jeśli nie mamy wystarczająco dużo danych do wytrenowania modelu, co zdarza się dość często w przypadku chorób rzadkich. Co ciekawe, w mojej opinii fenomen ten znajdziemy również wśród modelu szkolenia specjalistów, np. gdy porównamy skuteczność medyków wykonujących powtarzalne i specyficzne czynności. Recepcjonistka siedząca w poradni lekarza rodzinnego zapewne dość szybko rozpozna chore na zapalenie gardła dziecko. Ba, być może nauczy się, kiedy lekarz wypisuje antybiotyk, a kiedy nie. Po wielu latach takiego doświadczenia z pewnością wywróci oczami,

gdy świeżo wypieczona lekarka popełni błąd, wyznaczając dawkę penicyliny. Pomyśli pewnie, że sama mogłaby siedzieć na miejscu medyczki i uzyskiwać jej niezasłużone pobory. Ale w przypadku np. wstrząsu anafilaktycznego po podaniu antybiotyku, to jednak ta lekarka będzie miała większe szanse na uratowanie życia dziecku. Zarówno recepcjonistka, jak i lekarka nigdy nie widziały takiego przypadku, ale to lekarka ma więcej zasobów teoretycznych, by poradzić sobie z tym zadaniem. Wiedza z książek i nauki fundamentalne to coś, dzięki czemu wygrywamy z obecnymi algorytmami.

Elektryczne sieci neuronowe słabo radzą sobie też ze źle spreparowanymi danymi. W branży komputerowej znane jest powiedzonko „Garbage in, garbage out” – jak wkładasz do modelu śmieciowe dane, to model wyrzuci ci śmieciowe wyniki. Człowiek radzi sobie z tym problemem lepiej. Biologiczny mózg niejako „dostosowuje” swoje neurony do zadania przed uzyskaniem danych. Potrafi również dynamicznie „wyciszyć” podsieć neuronów, która reaguje w zły sposób. Wróćmy do kawiarni „Szkockiej” we Lwowie. Wchodzimy tuż przed zajęciami. Zapach świeżo zmielonej kawy miesza się z rześkim mroźnym styczniowym porankiem. Na ladzie uśmiechają się chrupiące pączusie z truskawkową marmoladą. W zgiełku zamówień jak zawsze słychać debatujących profesorów. Jak wół – „Szkocka”! Teraz wchodzimy raz jeszcze. Tym razem jest nieco cieplej; kawy i pączków nie brakuje. Coś jednak jest nie tak – nie słychać ani szeptu na temat analizy fundamentalnej funkcji! No trudno – pewnie wakacje, a profesorowie rozjechali się w góry. Ale przecież nadal wiemy, że weszliśmy do „Szkockiej”. Mózg zorientował się, że brak stymulacji dźwięku rozmowy musi być wyciszony w algorytmie rozpoznania lokalizacji. Charakterystyczne pączki, kawa, a być może i wiedza o tym, że jest lato, przyczyniły się do zmiany stanu początkowego neuronów. Algorytm komputerowy dostałby bzika. Tak silna destabilizacja zestawu danych treningowych posłałaby skuteczność rozpoznania na bambus. Dzieje się to przez wykluczenie bardzo charakterystycznej zmiennej – w końcu gdzie indziej zasiadają lwowscy geniusze. Sieć natychmiast przestawiłaby się na inne tory. Im więcej takich przykładów, tym bardziej automat głupieje. Maszyny również nie mają dostępu do tzw. modelu rzeczywistości, tzn. polegają na tym, co człowiek im dostarczy lub same zdobędą za pomocą sensorów. Taka baza danych różni się zawsze od rzeczywistości, przez co maszyna nie zawsze rozpoznaje, co jest prawdziwe, a co tylko ułudą... ale na tym te heglowskie rozważania zostawmy.

Jak widać, przynajmniej na początku trzeciej dekady XXI wieku, maszyny mają jeszcze dużo do nadrobienia. Nie oznacza to jednak, że inżynierowie poddali się zupełnie. Podczas niedawnego obiadu w collegu prof. David Clifton zwierzył mi się, że właśnie szykują maszynę, która będzie studiowała podręczniki medyczne, a dopiero na ich podstawie spróbuje rozpoznawać choroby. Mało powiedzieć, że zakrzutusiłem

się podanym mi batatem, i to bynajmniej nie z powodu niekompetencji szefa kuchni. Być może to właśnie taki algorytm rozprawi się raz na zawsze z zagadkami ze Szkockiej Księgi i będzie domagał się od nieboszczyka Banacha swojej „małej czarnej”. Jeśli tak, to ja za rozwiązanie takiego zadania oferuję całe ogniwo elektryczne (lub inną krzemową przekąskę do wyboru). Jeśli to czytasz, drogi algorytmie, to zgłoś się po nagrodę. Mój adres to... a zresztą, skoro jesteś taki mądry, to sam go sobie znajdź.