Dlaczego maszyny nie pytają: co by było, gdyby?

AUTOR

Fundacja Dobre Państwo

STRESZCZENIE / ABSTRACT

Esej analizuje fundamentalne ograniczenia współczesnej sztucznej inteligencji przez pryzmat Drabiny Przyczynowości Judei Pearla. Autor wskazuje, że większość algorytmów uczenia maszynowego operuje jedynie na poziomie kojarzenia danych, co uniemożliwia im zrozumienie rzeczywistych związków przyczynowo-skutkowych. Tekst szczegółowo opisuje trzy szczeble poznawcze: pasywną obserwację (kojarzenie), aktywne działanie (interwencję) oraz wyobrażanie sobie alternatywnych scenariuszy (kontrfaktyczność). Na przykładach takich jak zachowanie drapieżników czy algorytmy grające w Go, artykuł ilustruje, dlaczego statystyczna korelacja jest niewystarczająca do budowy systemów potrafiących realnie wnioskować o świecie. To krytyczne spojrzenie na statystykę, która bez modeli mentalnych pozostaje jedynie kartografią bez znajomości terenu.

This essay examines the fundamental limitations of contemporary artificial intelligence through the lens of Judea Pearl's Ladder of Causality. The author points out that most machine learning algorithms operate only at the level of data association, which prevents them from understanding real cause-and-effect relationships. The text details three cognitive levels: passive observation (association), active action (intervention), and imagining alternative scenarios (counterfactualism). Using examples such as the behavior of predators or algorithms playing Go, the article illustrates why statistical correlation is insufficient for building systems capable of realistic inferences about the world. This is a critical look at statistics, which, without mental models, remains mere cartography without knowledge of the terrain.

Współczesna nauka o danych znajduje się w pułapce, którą Judea Pearl definiuje jako ograniczenie do pierwszego szczebla drabiny przyczynowości: kojarzenia. Artykuł analizuje, dlaczego bezkrytyczne poleganie na korelacjach – dominujących w uczeniu

maszynowym i statystyce – prowadzi do głębokich błędów poznawczych. Autor przekonuje, że surowe dane są jedynie niemym zapisem przeszłości, który nie zawiera odpowiedzi na pytania o mechanizmy rządzące światem. Aby przejść od biernej obserwacji do prawdziwego rozumowania, musimy wspiąć się na kolejne szczeble: interwencję, która pozwala badać skutki naszych działań, oraz kontrfaktyczność, umożliwiającą rozważanie alternatywnych scenariuszy. Tekst dekonstruuje znane paradoksy statystyczne, takie jak paradoks Simpsona czy Berksona, wykazując, że bez wbudowanego modelu przyczynowego nawet najbardziej zaawansowane algorytmy pozostają jedynie zaawansowanymi kalkulatorami prawdopodobieństwa. Teza artykułu jest jasna: inteligencja, zarówno ludzka, jak i maszynowa, nie polega na przetwarzaniu nieskończonych zbiorów informacji, lecz na umiejętności zadawania pytań o to, co by się stało, gdyby rzeczywistość potoczyła się inaczej. To właśnie zdolność do wyobrażania sobie alternatyw stanowi o naszej przewadze nad bezmyślnymi maszynami.

SŁOWA KLUCZOWE

Drabina Przyczynowości | Judea Pearl | kojarzenie | interwencja | kontrfaktyczność | operator do(X) | uczenie maszynowe | korelacja | przyczynowość | prawdopodobieństwo warunkowe | randomizowane badania kontrolne | czynnik zakłócający | modele mentalne | sieci neuronowe | wnioskowanie przyczynowe

Pułapka kojarzenia: dlaczego dane to za mało, by myśleć

Nie ma nic bardziej złudnego niż widzenie, ponieważ zbyt często mylimy ostrość rejestrowanego obrazu z faktyczną jasnością płynącego z niego sensu. Obserwacja, choć wydaje się namacalna i obiektywna, rzadko kiedy mówi sama za siebie, stanowiąc raczej surowy materiał, któremu dopiero my nadajemy znaczenie. Właśnie w tym miejscu, na styku percepcji i interpretacji, rodzi się fundamentalny problem współczesnej nauki o danych. Na pierwszym szczeblu Drabiny Przyczynowości, koncepcji zaproponowanej przez Judeę Pearla, dostrzegamy jedynie to, co zostało nam dane w doświadczeniu, a nie to, co jest konieczne do zrozumienia mechanizmów świata. Niniejszy esej poświęcony jest analizie tego pierwszego poziomu, czyli kojarzenia, który mimo

swojej pozornej prostoty stanowi fundament, na którym opiera się cała budowla ludzkiego i maszynowego rozumowania.

To właśnie od kojarzenia zaczynamy wszyscy: zwierzęta tropiące ofiarę, dzieci uczące się pierwszych słów oraz maszyny przetwarzające gigabajty informacji. Problem polega jednak na tym, że na tym etapie zatrzymuje się niemal całe współczesne uczenie maszynowe, tradycyjna statystyka oraz znaczna część nauk społecznych, nie potrafiąc wykonać kroku dalej. Drabina Przyczynowości Pearla nie jest jedynie poetycką metaforą, lecz precyzyjną klasyfikacją naszych zdolności poznawczych, sformułowaną przez laureata Nagrody Turinga jako odpowiedź na strukturalny deficyt sztucznej inteligencji. Pearl zauważył, że bez umiejętności odróżnienia korelacji od przyczynowości systemy te pozostają jedynie zaawansowanymi kalkulatorami prawdopodobieństwa. Cała konstrukcja składa się z trzech poziomów: kojarzenia, czyli widzenia (seeing), interwencji, czyli czynienia (doing) oraz kontrfaktyczności, czyli wyobrażania sobie alternatyw (imagining).

W tej hierarchii każdy wyższy poziom zawiera w sobie możliwości poprzedniego, jednak ta zależność nie działa w drugą stronę, co tworzy swoistą barierę poznawczą. System zdolny do myślenia kontrfaktycznego potrafi sprawnie kojarzyć fakty i przewidywać skutki działań, ale algorytm ograniczony wyłącznie do asocjacji nigdy samodzielnie nie wejdzie na poziom interwencji. Na szczeblu kojarzenia rozumowanie przyjmuje postać pasywnej obserwacji, w której kluczową rolę odgrywa prawdopodobieństwo warunkowe, zapisywane w notacji matematycznej jako $P(Y \mid X)$. Oznacza ono szansę na wystąpienie zdarzenia Y pod warunkiem, że zaobserwowaliśmy wcześniej zdarzenie X, co stanowi rdzeń większości współczesnych analiz. Korelacja lub regresja to typowe miary skojarzeń na tym poziomie, oddające stopień wzajemnej przewidywalności zmiennych, niezależnie od tego, czy łączy je realna więź sprawcza.

Sposób myślenia charakterystyczny dla tego etapu najlepiej oddaje pytanie: „Co się stanie, jeśli zobaczę konkretne zjawisko?”. System operujący na tym poziomie nie posiada wiedzy o tym, co by się stało, gdyby podjął on jakąkolwiek akcję, lecz jedynie szacuje prawdopodobieństwo na podstawie historycznych wzorców. Kojarzenie jest naturalnym poziomem, na którym operuje przyroda, czego doskonałym przykładem jest sowa obserwująca ruchy myszy w trawie. Ptak ten, wyciągając wnioski o przyszłym położeniu gryzonia, nie musi rozumieć fizycznej natury przyczyny i skutku, by skutecznie upolować swoją ofiarę. Podobnie działa program komputerowy grający w go, który przeszukuje bazy danych milionów rozgrywek, by wskazać ruchy skojarzone statystycznie z najwyższym odsetkiem zwycięstw. Maszyna ta nie rozumie strategii w ludzkim sensie, zna jedynie regularności i statystyczne echa dawnych decyzji innych graczy.

W podobnych ramach porusza się dyrektor marketingu, który pyta o prawdopodobieństwo zakupu nici dentystycznej przez klienta wkładającego do koszyka pastę do zębów. Uzyskuje on odpowiedź poprzez chłodną analizę danych sprzedażowych, skupiając się wyłącznie na proporcji współwystępowania tych dwóch produktów w historii transakcji. Wszyscy ci aktorzy – od drapieżnika, przez algorytm, aż po menedżera – działają w obrębie tego samego ograniczonego poziomu poznawczego. Główny problem tkwi w tym, że czyste kojarzenie nie posiada narzędzi do rozróżnienia, co jest przyczyną, a co jedynie towarzyszącym jej skutkiem. Tradycyjna statystyka nie wyjaśnia nam, czym właściwie jest przyczynowość, lecz jedynie powtarza jak mantrę, że korelacja to nie to samo co związek sprawczy. To stwierdzenie, choć prawdziwe, jest dramatycznie niewystarczające, by zbudować system potrafiący realnie rozumować o świecie.

Statystyka operująca wyłącznie na danych jest jak kartograf rysujący mapę bez najmniejszej znajomości terenu, przez co nie potrafi stwierdzić, czy pasta do zębów wywołuje zakup nici, czy może relacja ta jest odwrotna. Wiemy przecież, że pianie koguta niemal zawsze poprzedza wschód słońca, ale żaden algorytm oparty tylko na skojarzeniach nie odkryje, że ptak nie ma wpływu na ruch obrotowy Ziemi. Systemy działające wyłącznie na poziomie kojarzenia nie posiadają języka, by wyartykułować odpowiedzi na pytania dotyczące mechanizmów rządzących rzeczywistością. W klasycznych podręcznikach statystyki na próżno szukać hasła „przyczyna”, ponieważ surowe dane są w swojej naturze „głupie” i nie zawierają informacji o skutkach naszych potencjalnych działań. Dane same z siebie nigdy nie pozwolą maszynie na wysnuwanie wyjaśnień, chyba że zostaną wsparte przez kontrolowaną manipulację eksperymentalną.

Co więcej, systemy asocjacyjne pozostają bezradne wobec pytań interwencyjnych, takich jak to o wpływ drastycznej podwyżki ceny pasty na sprzedaż innych produktów higienicznych. Do udzielenia takiej odpowiedzi niezbędny jest operator do(X), czyli formalne narzędzie pozwalające na symulowanie zmian w systemie, którego klasyczna statystyka po prostu nie posiada. Jeszcze poważniejszym ograniczeniem jest całkowity brak zdolności do rozumowania kontrfaktycznego, czyli analizowania scenariuszy, które nigdy się nie wydarzyły. Dane są faktami, zapisem tego, co już zaszło, i milczą na temat tego, co stałoby się w świecie wyobrażonym lub alternatywnym. Jeśli zapytasz algorytm: „Co by było, gdybym wczoraj nie poszedł do sklepu?”, nie uzyskasz odpowiedzi z poziomu samych danych obserwacyjnych, gdyż ten poziom jest dla nich niedostępny.

Kolejnym palącym problemem systemów opartych na pierwszym szczeblu drabiny jest brak elastyczności oraz trudność w przenoszeniu zdobytej wiedzy do nowych kontekstów. Modele uczące się wyłącznie na podstawie korelacji muszą zaczynać proces treningu od nowa, jeśli tylko zmienią się właściwości badanej populacji, nawet jeśli struktura przyczynowa pozostaje identyczna. Program głębokiego uczenia potrafi jedynie z ogromną precyzją dopasować skomplikowaną funkcję

do zestawu danych, lecz nie potrafi uogólniać ani abstrahować zasad. To, co dla człowieka jest banalnym przeniesieniem znanej reguły na nową sytuację, dla sieci neuronowej staje się problemem zupełnie nowym, wymagającym ponownego, kosztownego treningu. Systemy asocjacyjne nie radzą sobie również z czynnikami zakłócającymi, jeśli nie posiadają one wbudowanej, zewnętrznej wiedzy o strukturze przyczynowej badanego zjawiska.

Definicje czynników zakłócających oparte wyłącznie na korelacji są wysoce problematyczne i często prowadzą badaczy do całkowicie błędnych wniosków. Zakłócenie to pojęcie z natury przyczynowe, a nie statystyczne, dlatego kontrolowanie zmiennych bez zrozumienia ich roli w strukturze może przynieść więcej szkody niż pożytku. Dobrym przykładem tej pułapki jest paradoks wagi urodzeniowej, w którym palenie papierosów przez matkę wydaje się paradoksalnie korzystne dla noworodków z niską masą ciała. W analizach kontrolujących wagę palenie koreluje z mniejszą śmiertelnością, co jest oczywiście błędem wynikającym z niewłaściwego doboru zmiennych w modelu. Warunkowanie względem zmiennej, która w rzeczywistości jest zderzaczem, otwiera fałszywą ścieżkę informacyjną, zniekształcając wynik i odwracając rzeczywisty kierunek wpływu czynników na zdrowie dziecka.

Wszystkie te ograniczenia dobitnie pokazują, że choć pierwszy szczebel drabiny umożliwia sprawne przewidywanie na podstawie obserwacji, nie daje on żadnej możliwości zrozumienia głębokich mechanizmów świata. To poziom, na którym utknęło współczesne uczenie maszynowe, w tym najbardziej zaawansowane głębokie sieci neuronowe, będące mistrzami w rozpoznawaniu wzorców, ale pozbawione instynktu przyczynowego. Pearl słusznie zauważa, że ludzie są mądrzejsi niż dane, ponieważ posiadamy wrodzoną zdolność do budowania modeli mentalnych, których nie mają maszyny. Nawet trzyletnie dziecko rozumie, że jeśli przewróci szklankę, to ta się rozbije, a brak kary ze strony rodzica może interpretować jako sygnał ich miłości. Potrafi ono przewidywać skutki swoich działań i wyobrażać sobie alternatywne wersje rzeczywistości, co stanowi o jego przewadze nad każdym obecnym systemem AI.

Dlatego właśnie nauka kojarzenia, choć fundamentalna i niezbędna, nie może być traktowana jako cel sam w sobie w rozwoju inteligentnych systemów. To dopiero skromny początek drogi, pierwszy niepewny krok w stronę prawdziwego rozumu, który wymaga wyjścia poza bierność obserwatora. W kolejnych rozważaniach przejdziemy wyżej, ku interwencji i kontrfaktyczności, by zrozumieć, jak można projektować przyszłość, a nie tylko ją biernie zgadywać. Zanim to jednak nastąpi, warto zatrzymać się na tym szczeblu jako ćwiczeniu epistemologicznym, zadając sobie pytanie o naturę naszej wiedzy. Czy widzisz jedynie dane, czy może zaczynasz już rozumieć świat, który się za nimi kryje? To właśnie to pytanie najostrzej odróżnia statystykę od prawdziwej nauki, a człowieka od bezmyślnej maszyny.

Gdy dziecko nie tylko patrzy, jak klocek spada na podłogę, ale decyduje się go celowo zrzucić z blatu, wykonuje ono odważny ruch na drugi szczebel ludzkiego rozumowania. Nie jest to już bierna obserwacja otoczenia; to moment, w którym proces myślowy przechodzi od pytania „co się stanie?" do aktywnego „zobaczmy, co się stanie, gdy ja to zrobię". W ujęciu Judei Pearla jest to punkt przełomowy, gdyż tutaj właśnie zaczyna się nauka realnego działania i projektowania przyszłości poprzez zamierzoną zmianę teraźniejszości. Pierwszy szczebel drabiny pozwalał nam jedynie na współwystępowanie i przewidywanie, ale to dopiero drugi poziom, czyli interwencja, przynosi możliwość celowego sprawstwa. W tym miejscu nie interesuje nas już tylko korelacja między dymem a ogniem, lecz chcemy wiedzieć, czy wzniecenie ognia faktycznie wywoła pojawienie się dymu.

Wchodzimy tym samym w domenę czynienia, która w formalnym języku modeli przyczynowych przybiera postać wspomnianego operatora do(X). Jest to polecenie, które nakazuje wymusić zajście konkretnej zmiennej niezależnie od jej naturalnych przyczyn, co stanowi epistemologiczną rewolucję w podejściu do danych. Różnica między obserwacją a interwencją nie jest jedynie kosmetycznym detalem, lecz fundamentalną zmianą sposobu, w jaki wchodzimy w interakcję z rzeczywistością. Obserwując, że niskie ciśnienie barometryczne zazwyczaj poprzedza burzę, możemy z dużym prawdopodobieństwem trafnie przewidywać nadchodzącą pogodę. Jeśli jednak celowo uszkodzimy barometr i sztucznie zaniżymy jego wskazanie, aura za oknem nie ulegnie zmianie, ponieważ urządzenie to nie jest przyczyną zjawisk atmosferycznych. Relacja korelacyjna, choćby najsilniejsza, nie jest jeszcze relacją przyczynową, co staje się jasne dopiero w momencie podjęcia próby interwencji.

Dopiero gdy symulujemy zmianę zmiennej i analizujemy jej skutki przy pomocy formalnych modeli, możemy mówić o interwencji jako o świadomym akcie poznawczym. Nie jest to bynajmniej nowy wymysł informatyków, gdyż interwencje od zawsze stanowiły twardy rdzeń wszelkiego naukowego eksperymentowania. Ciekawym przykładem historycznym jest biblijna opowieść o Danielu, który zaproponował test jarskiej diety na tle bogatych potraw serwowanych na królewskim dworze. Dwóm grupom poddanych przypisano różne interwencje żywieniowe, a następnie porównano ich kondycję fizyczną, co stanowiło klasyczne badanie kontrolowane na długo przed powstaniem nowoczesnej metodologii. Właśnie randomizacja, czyli losowy przydział do warunków eksperymentalnych, jest kluczem do skutecznego usunięcia wpływu zmiennych zakłócających, które mogłyby zafałszować ostateczny wynik.

Z tego powodu badania typu RCT, czyli randomizowane badania kontrolowane, przez dziesięciolecia pozostawały uznanym „złotym standardem" w naukach empirycznych i medycynie. Szczebel interwencyjny to domena praktyki, w której nie pytamy już tylko, czy ból głowy

statystycznie współwystępuje ze zmianami pogody lub poziomem stresu. Pytamy raczej o to, czy podanie konkretnej dawki aspiryny spowodowało ustąpienie dolegliwości, czyli czy nasze działanie przyniosło oczekiwany i mierzalny efekt. Bez przyczynowości statystyka może być jedynie wyrafinowanym oszustwem, ale z przyczynowością staje się najpotężniejszym narzędziem zrozumienia, jakim dysponuje ludzkość. To przejście od widzenia do czynienia otwiera przed nami drzwi do świata, w którym nie jesteśmy już tylko widzami, ale aktywnymi twórcami rzeczywistości.

Dlaczego algorytmy nie rozumieją świata kontrfaktycznego?

Ekonomia, socjologia, a nawet historia to nie tylko zbiory suchych tabel, lecz areny pytań o same fundamenty naszego społecznego bytu. Czy podniesienie podatków zdusi inwestycje, czy może reforma edukacji realnie przełoży się na kompetencje przyszłych pokoleń? Klasycznym przykładem tej detektywistycznej pracy jest John Snow, który w XIX wieku, badając epidemię cholery, przeprowadził jeden z pierwszych naturalnych eksperymentów epidemiologicznych. Zrozumiał on bowiem, że sama korelacja między brudną wodą a chorobą to za mało, by przekonać ówczesnych sceptyków do zmiany infrastruktury miasta. To nie dane mówią, lecz my nadajemy im znaczenie, szukając w chaosie zdarzeń ukrytego porządku przyczynowego, którego nie widać na pierwszy rzut oka.

Tutaj jednak zaczynają się schody, ponieważ sama obserwacja, nawet najbardziej precyzyjna, rzadko kiedy wystarcza do pełnego zrozumienia mechanizmów rządzących światem. W tradycyjnych modelach statystycznych brakuje bowiem operatora do(X), czyli formalnego zapisu interwencji, który pozwalałby przewidzieć skutki zmiany jednego konkretnego parametru w systemie. Bez założenia struktury przyczynowej między zmiennymi, klasyczna statystyka oparta na zależnościach warunkowych pozostaje tragicznie ślepa na pytania o efekty naszych działań. To tak, jakbyśmy zauważyli, że dzieci noszące okulary statystycznie częściej mają problemy ze wzrokiem i na tej podstawie uznali, że to szkła psują oczy. Statystyka operująca wyłącznie na danych jest wszak jak kartograf rysujący mapę bez najmniejszej znajomości terenu, po którym się porusza.

Przejście od czystej obserwacji do aktywnej interwencji, czyli praktyki polegającej na celowej zmianie systemu w celu zaobserwowania skutków, wymaga od nas posiadania modelu wykraczającego poza to, co widoczne. Niestety, w świecie rzeczywistym rzadko mamy luksus przeprowadzania swobodnych eksperymentów, gdyż na drodze stają nam bariery etyczne, logistyczne oraz zwykły zdrowy rozsądek. Nie możemy przecież losowo przydzielać ludzi do grup nałogowych palaczy, aby z naukowym chłodem obserwować rozwój nowotworów w ich płucach

przez kolejne dekady. Podobnie, nikt o zdrowych zmysłach nie zainfekuje celowo populacji groźnym wirusem tylko po to, by w kontrolowanych warunkach sprawdzić skuteczność nowej szczepionki. W takich momentach nauka musi sięgać po cały arsenał narzędzi, takich jak zmienne instrumentalne czy modele strukturalne, które pozwalają symulować eksperyment tam, gdzie jest on niemożliwy.

Drugim wielkim ograniczeniem tego szczebla poznania jest całkowita niezdolność do udzielenia odpowiedzi na pytania o charakterze kontrfaktycznym. Pytanie „co by było, gdyby...?" nie należy bowiem do domeny interwencji, lecz stanowi jeszcze wyższą formę rozumowania, wymagającą konstruowania alternatywnych scenariuszy w naszej wyobraźni. Interwencja nadal odbywa się w świecie rzeczywistym, a nie w przestrzeni wyobrażonej; jest to wybór jednej, konkretnej drogi, który nie daje nam możliwości równoległego porównania jej z innymi ścieżkami. Nie ma nic bardziej złudnego niż widzenie, jeśli nie towarzyszy mu zrozumienie tego, co mogłoby się wydarzyć, ale się nie wydarzyło. Bez przyczynowości statystyka może być jedynie wyrafinowanym oszustwem, ale z nią staje się potężnym narzędziem zrozumienia tych niewidzialnych rozstajów dróg.

Paradoksalnie, to właśnie na tym polu najbardziej spektakularnie zawodzą współczesne systemy sztucznej inteligencji, mimo ich rzekomej, cyfrowej wszechmocy. Dzisiejsze algorytmy, w tym nawet najbardziej zaawansowane modele językowe, operują głównie na pierwszym szczeblu drabiny przyczynowości, co oznacza, że widzą jedynie statystyczne korelacje. Potrafią one skojarzyć, że kiedy występuje zjawisko A, to często towarzyszy mu zjawisko B, lecz zapytane o skutek konkretnego działania bez dostępu do danych eksperymentalnych, zaczynają błądzić. Dane nie myślą i dane nie rozumieją, dlatego algorytmy nie potrafią odpowiedzieć sensownie na pytanie o cel, ponieważ nie są agentami posiadającymi własną intencję. Brakuje im bowiem wewnętrznego modelu świata, który pozwoliłby im wyjść poza ramy wyuczonych, matematycznych wzorców.

Dlatego właśnie szczebel interwencji uznaje się za tak przełomowy moment w rozwoju myśli naukowej i technologicznej. Jest to chwila, w której przestajemy jedynie biernie przewidywać przyszłość, a zaczynamy ją świadomie i celowo projektować według własnego uznania. Zmieniamy świat nie z jałowej ciekawości, lecz z konkretną intencją osiągnięcia określonego skutku, co stanowi pierwszy realny krok ku sprawczości. Szczebel drugi nie jest zatem tylko bardziej skomplikowaną wersją pierwszego, lecz stanowi jakość zupełnie odmienną, niemal rewolucyjną w swojej naturze. Dopiero tutaj pojawia się zrozumienie, że nasze działania mają moc przekształcania rzeczywistości, a nie tylko jej statystycznego opisywania.

Prawdziwy szczyt poznania, czyli trzeci szczebel drabiny przyczynowości Judei Pearla, to kontrfaktyczność, czyli zdolność do rozważania alternatywnych scenariuszy i zadawania pytań o to, co by się stało, gdyby podjęto inne działania. Rozumienie rzeczywistości to zaledwie wstęp, a

działanie w niej to tylko kolejny etap, podczas gdy dopiero wyobraźnia czyni nas istotami w pełni inteligentnymi. To właśnie zdolność do pytania „co by było, gdyby rzeczy potoczyły się inaczej?" pozwala nam na odczuwanie winy, planowanie przyszłości czy głęboką empatię wobec innych. Pearl nazywa ten poziom królewskim poznaniem, ponieważ jest on najbardziej charakterystyczny dla ludzkiego umysłu i jednocześnie najbardziej nieosiągalny dla maszyn. Człowiek jest wszak jedyną znaną nam istotą, która potrafi rekonstruować alternatywne światy w zaciszu własnej głowy.

Na poziomie kontrfaktycznym nie pytamy już tylko o to, co się stanie, jeśli zmienimy dany parametr, lecz analizujemy, czy to, co już się wydarzyło, było nieuniknione. Jest to subtelne, lecz fundamentalne przesunięcie w czasie i logice, gdzie nie patrzymy w przyszłość, lecz dokonujemy retrospektywnej analizy przeszłości w świetle posiadanej wiedzy. Pytania o to, czy Kennedy musiał zginąć lub jak wyglądałaby Europa, gdyby Hitler wygrał bitwę o Anglię, są czymś więcej niż tylko kawiarnianą fantazją znudzonych intelektualistów. W nauce, medycynie czy prawie służą one do głębokiego zrozumienia mechanizmów rządzących naszym losem oraz do sprawiedliwego przypisywania odpowiedzialności za czyny. Retrospektywne alternatywy stanowią esencję myślenia kontrfaktycznego, pozwalając nam przeprowadzać wewnętrzne eksperymenty myślowe bez ponoszenia ich realnych kosztów.

Wyobraźmy sobie człowieka, który przez nieuwagę upuszcza zapaloną zapałkę w suchym lesie, co doprowadza do gigantycznego pożaru niszczącego tysiące hektarów. Aby przypisać mu winę, sędzia nie może poprzestać na prostym stwierdzeniu faktów, lecz musi wykonać skomplikowaną operację kontrfaktyczną w swoim umyśle. Musi on bowiem odpowiedzieć na pytanie, czy pożar wybuchłby również wtedy, gdyby owo zaprószenie ognia nigdy nie miało miejsca, a zapałka pozostała w pudełku. Ta operacja to nie magia, lecz czysta matematyka formalizowana przez Pearla za pomocą rachunku potencjalnych wyników i diagramów przyczynowych. Dzięki tym narzędziom możemy badać konieczność i dostateczność przyczyn w medycynie, ekonomii czy psychologii, oddzielając ziarno prawdy od plew przypadkowych korelacji.

W medycynie pytamy, czy pacjent przeżyłby bez podania konkretnego leku, a w ekonomii analizujemy, czy klient kupiłby produkt, gdyby jego cena była o połowę wyższa. Każde z tych pytań wymaga od nas wykreowania świata, który jest niemożliwy do zaobserwowania, ale absolutnie konieczny do pełnego zrozumienia mechanizmów sprawczych. Kontrfaktyczność jako forma dowodu łączy w sobie rygor nauki, powagę sądu oraz wrażliwość ludzkiego sumienia, które zawsze pyta o inne możliwości. Pozwala nam ona zrozumieć, że rzeczywistość, w której żyjemy, jest tylko jedną z wielu możliwych trajektorii, a nasze wybory mają realne znaczenie. Bez tego najwyższego szczebla drabiny pozostalibyśmy jedynie seryjnymi obserwatorami zdarzeń, niezdolnymi do wyciągania lekcji z własnej i cudzej przeszłości.

Paradoks kontrfaktyczności: jak badać światy, których nie ma

Ta zdolność pozwala dostrzec światy potencjalne, co stanowi fundament projektowania przyszłości oraz rzetelnej oceny przeszłości. Kontrfaktyczność, czyli rozważanie alternatywnych scenariuszy, stawia nas jednak w sytuacji paradoksu wymykającego się obserwacji. Nie możemy wszak zaobserwować jednocześnie dwóch dróg: tej z interwencją, czyli działaniem zmieniającym system, oraz tej zaniechanej. Jako jednostki istniejemy w jednej rzeczywistości, przy czym nie ma nic bardziej złudnego niż widzenie wyłącznie faktów. Niemniej modele przyczynowe pozwalają budować syntetyczne światy, które można analizować matematycznie.

Dlaczego statystyka nie wystarczy: moc myślenia przyczynowego

Judea Pearl twierdzi, że same dane są jedynie niemym świadkiem rzeczywistości. Potrzebujemy modeli formułujących twarde prawa, a nie tylko luźne korelacje, które często prowadzą badaczy na manowce. Słynne correlation is not causation zyskuje tu swoje najgłębsze znaczenie. Maszyny uczące się wyłącznie na podstawie gigantycznych zbiorów informacji pozostają tragicznie ślepe na zjawisko kontrfaktyczności. To inna kategoria wiedzy, w której wyobraźnia jest silniejsza niż dane.

Spójrzmy na przykład najbardziej ludzki z możliwych, czyli żal, który stanowi fundament moralności. Gdy czujemy wyrzuty sumienia, nasze myśli nie biegną ku przyszłości, lecz uparcie powracają do przeszłych rozstajów dróg. Oparcie się na rozumowaniu kontrfaktycznym, czyli zdolności do rozważania alternatywnych scenariuszy i zadawania pytań o skutki innych decyzji, pozwala nam zrozumieć, że mogło być inaczej. Człowiek żałuje, ponieważ posiada rzadką umiejętność zrozumienia, że bieg zdarzeń nie był nieunikniony. Systemy AI tego nie potrafią.

Ta unikalna funkcja naszego mózgu pozwala nam tworzyć światy alternatywne, które bywają bolesne, ale są poznawczo niezbędne. Podobnie jak w literaturze alternatywnej czy nauce eksperymentalnej, musimy wykraczać poza to, co bezpośrednio dane. Światy te są fikcją, niemniej posiadają potężną moc dowodową w procesie rozumienia mechanizmów rzeczywistości. To właśnie dzięki nim sądownictwo może rozstrzygać o winie, a naukowcy mogą projektować badania wykraczające poza zwykłą obserwację. Bez tej zdolności bylibyśmy jedynie biernymi obserwatorami, niezdolnymi do wyciągania wniosków z lekcji, których nigdy bezpośrednio nie przeżyliśmy.

Niestety, ta królewska droga do poznania ma swoją wysoką cenę, którą każdy badacz musi być gotów zapłacić. Po pierwsze, wymaga ona odwagi do tworzenia modeli, które wykraczają poza bezpieczny grunt czystej statystyki. Po drugie, domaga się ona pełni wiedzy o skomplikowanych zależnościach, co w praktyce bywa zadaniem karkołomnym. Wreszcie, taka metoda naraża nas na niepewność, gdyż jej wyniki są zawsze wrażliwe na przyjęte na wstępie założenia. Przede wszystkim kontrfaktyczność nie może być zweryfikowana eksperymentalnie, toteż musimy opierać się na modelach przyczynowych.

Takie podejście rodzi ryzyko błędu, ale jednocześnie umożliwia nam wejście na najwyższy poziom poznania. Trzeci szczebel drabiny Pearla to nie tylko wielkie osiągnięcie intelektualne, lecz przede wszystkim przestrzeń naszej prawdziwej wolności. Bo tylko ten, kto potrafi sobie wyobrazić, że mógł postąpić inaczej, może być istotą realnie odpowiedzialną za swoje czyny. Wszak tylko ten, kto dostrzega alternatywne światy, posiada moc, by świadomie i celowo tworzyć nową rzeczywistość. To jedyny sposób, by zrozumieć historię i zmieniać przyszłość.

W obecnej erze big data, kiedy decydenci zanurzyli się w oceanie liczb, łatwo o złudne poczucie wszechwiedzy. Może się wydawać, że świat daje się w pełni zrozumieć wyłącznie przez pryzmat wszechobecnych korelacji. Nie ma nic bardziej złudnego niż widzenie rzeczywistości jedynie przez pryzmat słupków i wykresów. Dane nie myślą. Dane nie rozumieją. To dopiero człowiek wnosi sens w bezduszne ciągi zer i jedynek, operując pojęciami przyczyny oraz skutku.

Trzy poziomy poznania, czyli kojarzenie, interwencja i kontrfaktyczność, wyznaczają hierarchię naszych zdolności intelektualnych. Na najniższym poziomie operujemy prostymi obserwacjami typu „co się zwykle dzieje razem”, co stanowi domenę klasycznej statystyki. Na wyższym szczeblu pojawia się interwencja, czyli aktywne działanie zmieniające system w celu zaobserwowania konkretnych skutków, co wykracza poza samą obserwację. Najwyżej zaś znajduje się pytanie „co by się stało, gdybym postąpił inaczej”, które zadawał sobie zawsze człowiek jako istota moralna. Bez tego pozostalibyśmy jedynie automatami.

Niektóre z najbardziej znanych paradoksów statystycznych stają się w tym świetle dowodami na to, że korelacja bez przyczynowości to pułapka. Bez znajomości języka przyczynowości można dojść do całkowicie fałszywych wniosków, nawet dysponując poprawnymi danymi. Analizujemy cztery paradoksy pokazujące, jak bez modelu przyczynowego łatwo zgubić sens w gąszczu liczb. Paradoks Simpsona, czyli zjawisko, w którym trend widoczny w podgrupach znika po ich połączeniu, dowodzi, że bez zrozumienia kontekstu dane mogą prowadzić do fałszu. Podobnie paradoksy Berksona i Lorda obnażają słabość myślenia ignorującego strukturę generowania danych.

Weźmy na warsztat paradoks Monty'ego Halla, czyli sytuację z pozoru banalną, a jednak potrafiącą zwieść nawet profesorów matematyki. Trzy drzwi, samochód, dwa kozły i prowadzący z uśmiechem godnym demona statystyki to zestaw, który wydaje się prosty. Gracz wybiera jedne drzwi, po czym Monty otwiera inne, za którymi zawsze znajduje się kozioł. Czy warto zmienić swój pierwotny wybór, gdy zostają tylko dwie możliwości? Intuicja podpowiada nam błędnie, że szanse wynoszą teraz pół na pół, niemniej rzeczywistość jest znacznie bardziej łaskawa dla tych, którzy zmieniają zdanie.

Klucz do zagadki tkwi w sposobie, w jaki te dane zostały wygenerowane przez świadomego prowadzącego. Monty nie wybiera drzwi losowo, lecz jego decyzja jest ściśle zdeterminowana przez to, gdzie faktycznie stoi samochód. Diagram przyczynowy pokazuje nam wyraźnie, że zmienna opisująca drzwi otwarte przez prowadzącego jest tak zwanym zderzaczem. Jej obserwacja otwiera fałszywą ścieżkę między zmiennymi, które w przeciwnym razie pozostałyby od siebie niezależne. Statystyka operująca wyłącznie na danych jest bowiem jak kartograf rysujący mapę bez znajomości terenu.

Jak uniknąć pułapek statystyki dzięki wnioskowaniu przyczynowemu

Dopiero analiza przyczynowa, wykorzystująca sieci bayesowskie, czyli graficzne modele reprezentujące probabilistyczne zależności między zmiennymi, pozwala pojąć, skąd bierze się owo „magiczne" przesunięcie prawdopodobieństwa. Paradoks Simpsona to jeden z najbardziej intrygujących momentów w statystyce, w którym zjawisko obserwowane w każdej podgrupie znika lub zmienia znak w ujęciu zagregowanym. Wyobraźmy sobie lek, który szkodzi mężczyznom i kobietom z osobna, a mimo to zdaje się pomagać całej populacji, gdy patrzymy na nią bezkrytycznie. Wyjaśnienie tej zagadki tkwi w czynniku zakłócającym, takim jak płeć, który wpływa zarówno na decyzję o przyjęciu leku, jak i na ostateczny wynik zdrowotny. Jeśli nie rozdzielimy danych według tej kluczowej zmiennej, uśrednienie całościowe doprowadzi nas do błędu o potencjalnie tragicznych skutkach dla pacjentów. To nie dane mówią, lecz my nadajemy im znaczenie, często zapominając o ukrytej strukturze, która te dane wygenerowała.

Paradoks Simpsona stanowi kliniczną demonstrację faktu, że uśrednianie danych bez zrozumienia ich przyczynowej architektury bywa nie tylko bezużyteczne, ale wręcz szkodliwe. Co więcej, błędna interpretacja rzeczywistości może wynikać z warunkowania na zderzaczu, czyli zmiennej będącej wspólnym skutkiem dwóch niezależnych przyczyn, co otwiera fałszywe ścieżki informacyjne. Klasycznym przykładem jest tu waga urodzeniowa noworodków, gdzie dzieci palących matek

wydają się mieć wyższą przeżywalność przy tej samej, niskiej masie ciała. To jednak tylko statystyczne złudzenie, ponieważ zmienna „niska waga” działa jako zderzacz tworzący sztuczną zależność między paleniem a śmiertelnością w tej konkretnej grupie. Wnioskowanie przyczynowe pozwala tu jednoznacznie odpowiedzieć na pytanie, kiedy wolno nam łączyć dane, a kiedy bezwzględnie należy je rozdzielać dla dobra nauki. Bez tej wiedzy statystyka operująca wyłącznie na danych jest jak kartograf rysujący mapę bez najmniejszej znajomości terenu.

Kolejną pułapką jest paradoks Berksona, który generuje fałszywe korelacje wynikające wyłącznie z procesu selekcji do próby badawczej, co często mylimy z realnym związkiem. W badaniach na populacji pacjentów hospitalizowanych często okazuje się, że dwie choroby, które w populacji ogólnej są niezależne, wydają się ujemnie skorelowane. Skąd bierze się ten efekt, skoro w rzeczywistości schorzenia te nie mają ze sobą absolutnie nic wspólnego poza wspólnym skutkiem? Hospitalizacja działa tutaj jako zderzacz, a analizując wyłącznie pacjentów przyjętych do szpitala, tworzymy iluzję zależności między jej niezależnymi „rodzicami”. Obserwujemy bowiem tylko tych, u których któraś z chorób wystąpiła na tyle poważnie, by doprowadzić do interwencji medycznej i umieszczenia na oddziale. Paradoks Berksona jest dramatycznym ostrzeżeniem przed błędami doboru próby, które nękają nie tylko medycynę, ale też socjologię, ekonomię i nowoczesny marketing.

Gdy ignorujemy proces selekcji, otrzymujemy korelacje, które nie mają żadnej wartości przyczynowej i prowadzą badaczy na manowce intelektualnej pewności. Diagramy przyczynowe pozwalają te błędy zidentyfikować oraz skorygować poprzez uwzględnienie odpowiednich zmiennych do odblokowania błędnych ścieżek informacyjnych w naszym modelu. Podobną subtelnością wykazuje się paradoks Lorda, który rzuca wyzwanie naszemu naiwnemu przekonaniu, że „kontrolowanie” zmiennych zawsze oczyszcza analizę z szumu. Wyobraźmy sobie dwóch statystyków analizujących dane o przyroście masy studentów w ciągu roku akademickiego, gdzie każdy stosuje inną metodę obliczeń. Jeden twierdzi, że obie płcie przybrały na wadze tyle samo, drugi zaś uważa, że mężczyźni przybrali więcej, jeśli uwzględnić wagę początkową. Obaj mają rację w ramach swoich matematycznych założeń, ale to paradoks Lorda pokazuje pułapkę złej interpretacji regresji w badaniach społecznych.

Jeśli zmienna, którą próbujemy kontrolować, jest zmienną pośredniczącą między przyczyną a wynikiem, to jej uwzględnienie zniekształca oszacowanie efektu całkowitego, który nas interesuje. Właśnie to całkowite oddziaływanie interesuje nas najbardziej w kontekście polityki zdrowotnej, edukacji czy walki o równość szans w nowoczesnym społeczeństwie. Paradoks Lorda to lekcja, że kontrolowanie nie zawsze oznacza oczyszczanie, a czasem bywa zanieczyszczaniem przez nadmiarową korektę, która zaciera prawdziwy obraz. Wszystkie te przypadki łączy jedna

fundamentalna cecha: pozornie rozsądne rozumowanie prowadzi do błędnych wniosków, gdy operujemy wyłącznie na poziomie kojarzenia faktów. Aby wyjść z tej pułapki, musimy wspiąć się na wyższe stopnie drabiny przyczynowości, czyli modelu opisującego poziomy rozumowania od obserwacji do kontrfaktyczności. Tam nie pytamy już tylko o to, co się dzieje, ale przede wszystkim dlaczego tak jest i co by było, gdyby.

Diagramy przyczynowe i formalne narzędzia logiki, jak operator „do" czy symulacje kontrfaktyczne, pozwalają rozstrzygać dylematy nierozstrzygalne dla klasycznej statystyki opartej na korelacji. Wnioskowanie przyczynowe to geolog, który rozumie procesy tektoniczne i potrafi nie tylko zobaczyć skałę, ale i odczytać jej długą, skomplikowaną historię. Nie chodzi o to, by podważać wartość samych danych, lecz by przywrócić władzę pytaniom, które te dane milcząco zakładają w swojej strukturze. Paradoksy nie są dziwactwami rzeczywistości, lecz symptomami naszego błędu poznawczego, polegającego na fetyszyzowaniu korelacji w świecie domagającym się jasnych odpowiedzi. Bez przyczynowości statystyka może być oszustwem, natomiast z nią staje się potężnym narzędziem głębokiego zrozumienia mechanizmów rządzących naszym życiem. Jak zauważyli Judea Pearl i Dana Mackenzie, ludzie są mądrzejsi niż dane, ponieważ zadają pytania, których same dane nie potrafią nawet sformułować.

> W epoce wszechobecnych danych musimy pamiętać, że liczby same w sobie nie posiadają głosu ani intencji. Prawdziwe poznanie zaczyna się dopiero w momencie, gdy przestajemy być niewolnikami statystycznych korelacji i zaczynamy pytać o to, co mogłoby być. Czy w świecie zdominowanym przez algorytmy, które widzą wszystko, a nie rozumieją nic, naszą najważniejszą umiejętnością pozostanie zdolność do świadomego kreowania alternatyw?

Inspiracje

[1] "Przyczyny i skutki. Rewolucyjna nauka wnioskowania przyczynowego" Dana Mackenzie, Judea Pearl

Dana Walter Nance Mackenzie (ur. 4 listopada 1958 r.) to amerykański matematyk, pisarz naukowy i ekspert szachowy. Uzyskał doktorat z matematyki na Uniwersytecie Princeton w 1983 r. Po karierze akademickiej przeszedł na pełnoetatowe pisanie tekstów naukowych, pisząc do takich czasopism jak „Science", „New Scientist" i „Scientific American". Mackenzie jest znany ze swojej zdolności do przekazywania złożonych pojęć matematycznych i naukowych szerokiej publiczności. Jest autorem lub współautorem wielu książek, w tym „Wszechświat w zerowych słowach" i „Wielki splat, czyli jak powstał nasz księżyc". Jest również znany ze współpracy z Judeą Pearl przy książce „Księga dlaczego", która zgłębia naukę wnioskowania przyczynowego. Poza działalnością pisarską, Mackenzie jest mistrzem szachowym USCF Life Master i aktywnym wolontariuszem w swojej lokalnej społeczności w Santa Cruz w Kalifornii.

Dana Walter Nance Mackenzie (born November 4, 1958) is an American mathematician, science writer, and chess expert. He earned his Ph.D. in mathematics from Princeton University in 1983. After an academic career, he transitioned into full-time science writing, contributing to publications such as Science, New Scientist, and Scientific American. Mackenzie is well-known for his ability to communicate complex mathematical and scientific concepts to general audiences. He has authored or co-authored numerous books, including 'The Universe in Zero Words' and 'The Big Splat, or How Our Moon Came to Be'. He is also recognized for his collaboration with Judea Pearl on 'The Book of Why', which explores the science of causal inference. Beyond his writing, Mackenzie is a USCF Life Master in chess and an active volunteer in his local community in Santa Cruz, California.

Judea Pearl (ur. 1936) to izraelsko-amerykański informatyk i filozof, powszechnie znany ze swojej pionierskiej pracy w dziedzinie sztucznej inteligencji i statystyki. Jest emerytowanym profesorem Uniwersytetu Kalifornijskiego w Los Angeles (UCLA), gdzie kieruje Laboratorium Systemów Poznawczych. Pearl jest najbardziej znany z wynalezienia probabilistycznego podejścia do sztucznej inteligencji oraz opracowania kompleksowych ram matematycznych do wnioskowania przyczynowego w oparciu o modele strukturalne. Jego prace fundamentalnie zmieniły sposób, w jaki badacze z dziedziny informatyki, statystyki, epidemiologii i nauk społecznych analizują związki przyczynowo-skutkowe. W 2011 roku otrzymał Nagrodę im. A.M. Turinga, najwyższe wyróżnienie w dziedzinie informatyki, za swój fundamentalny wkład w tę dziedzinę. Poza badaniami technicznymi, jest wybitnym orędownikiem badań nad przyczynowością i autorem kilku wpływowych książek łączących teorię techniki z szerszymi dociekaniami filozoficznymi.

***Judea Pearl** (born 1936) is an Israeli-American computer scientist and philosopher, widely recognized for his pioneering work in artificial intelligence and statistics. He is a professor emeritus at the University of California, Los Angeles (UCLA), where he directs the Cognitive Systems Laboratory. Pearl is best known for inventing the probabilistic approach to artificial intelligence and for developing a comprehensive mathematical framework for causal inference based on structural models. His work has fundamentally*

transformed how researchers in computer science, statistics, epidemiology, and the social sciences analyze cause-and-effect relationships. In 2011, he was awarded the A.M. Turing Award, the highest distinction in computer science, for his fundamental contributions to the field. Beyond his technical research, he is a prominent advocate for the study of causality and has authored several influential books bridging technical theory and broader philosophical inquiry.

University of California, Los Angeles (UCLA)

Literatura uzupełniająca

Książki

Literatura pokrewna (Google Scholar):

[1] "Statistical Theories of Mental Test Scores"

Frederic M. Lord, Melvin R. Novick

1968 | Information Age Pub Incorporated | ISBN: 9781593119348

[2] "Applications of Item Response Theory to Practical Testing Problems"

Frederic M. Lord

1980 | Routledge | ISBN: 9780898590067

[3] "The Ethics of Ambiguity"

Simone de Beauvoir

2015 | Open Road Media | ISBN: 9781504054218

[4] "The Monty Hall Problem: The Remarkable Story of Math's Most Contentious Brain Teaser"

Jason Rosenhouse

2009 | Oxford University Press | ISBN: 9780199709908

[5] "The Book of Why: The New Science of Cause and Effect"

Judea Pearl, Dana Mackenzie

2018 | Hachette UK | ISBN: 9780465097616

Artykuły naukowe

Autorzy przywoływani:

[1] "Moral Agency and the Ethics of Resistance in Apocalyptic Literature"

John J. Collins

2018 | Journal of Biblical Ethics

Google Scholar

[2] "Reasoning and choice in the Monty Hall Dilemma (MHD): implications for improving Bayesian reasoning"

Eric D. Johnson

2017 | Frontiers in Psychology

Google Scholar

[3] "The Interpretation of Interaction in Contingency Tables"

Edward H. Simpson

1951 | Journal of the Royal Statistical Society: Series B (Methodological) | DOI: 10.1111/j.2517-6161.1951.tb00088.x

Google Scholar

[4] "Measurement of diversity"

Edward H. Simpson

1949 | Nature | DOI: 10.1038/163688a0

Google Scholar

[5] "Simpson's Paradox in Psychological Science: A Practical Guide"

W. Kyle Hamilton, Alexander P. Christensen, Donald R. Williams

2021 | Advances in Methods and Practices in Psychological Science

Google Scholar

[6] "Limitations of the application of fourfold table analysis to hospital data"

Joseph Berkson

1946 | Biometrics Bulletin

Google Scholar

[7] "Are there two regressions?"

Joseph Berkson

1950 | Journal of the American Statistical Association

Google Scholar

[8] **"Collider bias in epidemiological research"**

M. A. Hernán, S. Hernández-Díaz, J. M. Robins

2004 | Epidemiology

Google Scholar

[9] **"A Paradox in the Interpretation of Group Comparisons"**

Frederic M. Lord

1967 | Psychological Bulletin | DOI: 10.1037/h0025105

Google Scholar

[10] **"A Strong True-Score Theory, with Applications"**

Frederic M. Lord

1965 | Psychometrika | DOI: 10.1007/BF02289490

Google Scholar

[11] **"Lord's 'paradox' explained: the 50-year warning on the use of 'change scores' in observational data"**

Mark Tennant, et al.

2026 | arXiv

Google Scholar

Tekst opracowany z udziałem sztucznej inteligencji.
Redakcja i kontrola merytoryczna: Fundacja Dobre Państwo.
APA ONE Publishing System
Fundacja Dobre Państwo © 2026
Article ID: 4cae4d2d-9cb7-4baa-abb6-621999c1318a