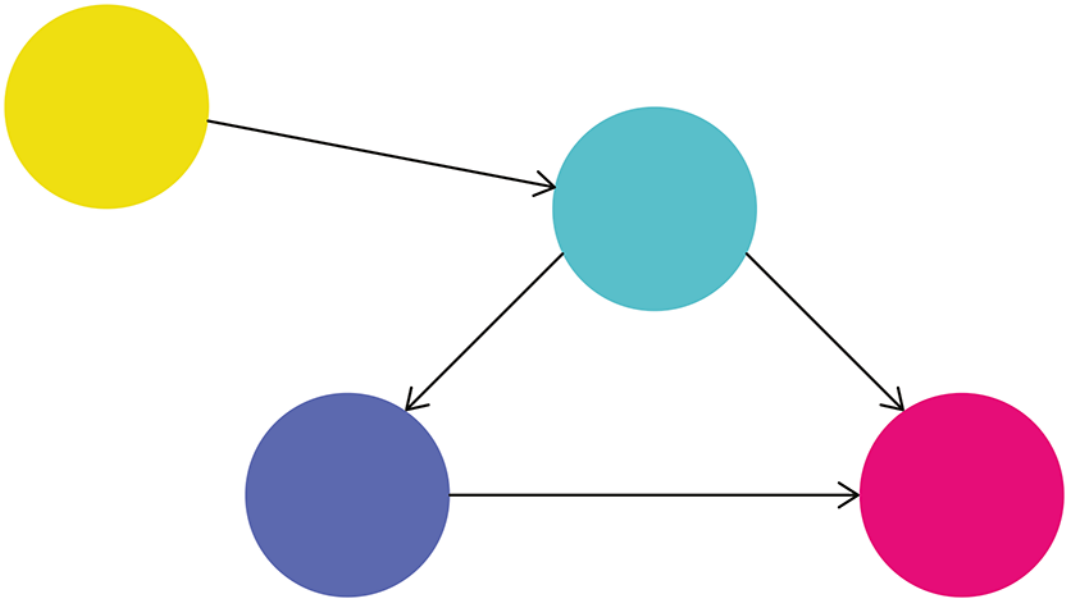
‹packt›

Wnioskowanie i związki przyczynowe w Pythonie

Nowoczesne uczenie maszynowe z wykorzystaniem bibliotek DoWhy, EconML, PyTorch i nie tylko



ALEKSANDER MOLAK

Helion

Tytuł oryginału: Causal Inference and Discovery in Python – Machine Learning and Pearlian Perspective: Unlock the secrets of modern causal machine learning with DoWhy, EconML, PyTorch and more

Tłumaczenie: Radosław Meryk

ISBN: 978-83-289-0832-1

Drogi Czytelniku!
Jeżeli chcesz ocenić tę książkę, zajrzyj pod adres
https://helion.pl/user/opinie/wnizwi
Możesz tam wpisać swoje uwagi, spostrzeżenia, recenzję.

Helion S.A.
ul. Kościuszki 1c, 44-100 Gliwice
tel. 32 230 98 63
e-mail: *helion@helion.pl*
WWW: *https://helion.pl* (księgarnia internetowa, katalog książek)

Printed in Poland.

Spis treści

CZĘŚĆ 3. Odkrywanie związków przyczynowych

Związki przyczynowe? Przecież jest uczenie maszynowe, więc po co zawracać sobie tym głowę?

Rozdział 1

Tutaj zaczyna się nasza podróż.

W tym rozdziale zadam kilka pytań dotyczących analizy związków przyczynowych.

Czym ona jest? Czy wnioskowanie przyczynowe różni się od wnioskowania statystycznego? Jeśli tak, to czym?

Czy analiza przyczynowa w ogóle jest potrzebna? Uczenie maszynowe wydaje się wystarczająco dobre.

Jeśli śledziłeś szybko zmieniający się krajobraz uczenia maszynowego przez ostatnie 5 – 10 lat, prawdopodobnie zauważyłeś wiele przykładów — jak lubimy to nazywać w społeczności uczenia maszynowego — *irracjonalnej skuteczności* nowoczesnych algorytmów uczenia maszynowego w dziedzinach rozpoznawania obrazów, przetwarzania języka naturalnego i innych.

Takie algorytmy jak DALL-E 2 czy GPT-3/4 trafiły do świadomości nie tylko środowiska naukowców, ale także ogółu społeczeństwa.

Mógłbyś zadać sobie pytanie — jeśli te algorytmy tak dobrze się sprawdzają, to po co w ogóle zawracać sobie głowę czymś innym?

Ten rozdział rozpocznę krótkim omówieniem historii przyczynowości. Następnie podam kilka powodów stosowania w modelowaniu podejścia przyczynowego zamiast czysto statystycznego i wprowadzę pojęcie zakłóceń (ang. *confounding*).

Na koniec zaprezentuję przykłady zastosowania podejścia przyczynowego do rozwiązywania problemów związanych z marketingiem i medycyną. Po zakończeniu lektury tego rozdziału czytelnik powinien mieć jasny obraz obszarów, w których może być przydatne wnioskowanie przyczynowe, oraz powodów tego stanu. Powinien umieć wyjaśnić, czym jest zakłócanie i dlaczego jest ono ważne.

W tym rozdziale omówię następujące zagadnienia:

- Krótka historia przyczynowości.
- Motywacje do stosowania przyczynowego podejścia do modelowania.
- Jak nie stracić pieniędzy… i ludzkich istnień.

Krótka historia przyczynowości

Przyczynowość (ang. *causality*) ma długą historię. Była przedmiotem zainteresowania większości znanych kultur. Jeden z najpłodniejszych filozofów starożytnej Grecji, Arystoteles, twierdził, że zrozumienie przyczynowej struktury procesu jest niezbędnym elementem wiedzy o tym procesie. Ponadto argumentował, że istotą wyjaśniania naukowego jest umiejętność odpowiadania na pytania w rodzaju *dlaczego* (Falcon, 2006, 2022). Arystoteles wyróżnił cztery typy przyczyn (materialne, formalne, sprawcze i ostateczne). O ile taka klasyfikacja pozwala uwzględnić w równym stopniu pewne interesujące aspekty rzeczywistości, o tyle dla współczesnego czytelnika może sprawiać wrażenie sprzecznej z intuicją.

Słynny szkocki filozof z XVIII wieku, David Hume, zaproponował bardziej ujednolicone ramy związków przyczynowo-skutkowych. Hume wyszedł z założenia, że w rzeczywistym świecie nigdy nie obserwujemy związków przyczynowo-skutkowych. Jedyne, czego doświadczamy, to obserwacja, że niektóre zdarzenia są ze sobą powiązane:

„Dostrzegamy jedynie fakt, że jedno zdarzenie wynika z drugiego. Uderzenie jednej kuli bilardowej powoduje ruch drugiej. To wszystko, co ukazuje się zewnętrznym zmysłom. Umysł nie odczuwa ani żadnego uczucia, ani wewnętrznego wrażenia wynikających z tego następstwa zdarzeń. W związku z tym nie ma, w żadnym pojedynczym, konkretnym przypadku przyczyny i skutku, niczego co mogłoby sugerować ideę mocy lub koniecznego związku" (Hume i Millican, 2007; pierwotnie opublikowany w 1739 r.).

Jedna z interpretacji teorii przyczynowości Hume'a (tu uproszczona dla zachowania przejrzystości) brzmi następująco:

- Dostrzegamy jedynie to, że ruch lub pojawienie się obiektu *A* poprzedza ruch lub pojawienie się obiektu *B*.
- Jeśli zaobserwujemy takie następstwo zdarzeń wystarczającą liczbę razy, rozwija się w nas poczucie oczekiwania.
- To poczucie oczekiwania jest sednem pojmowania przyczynowości (nie dotyczy świata, lecz uczucia, które rozwijamy).

Teoria przyczynowości Hume'a

Interpretacja Hume'owskiej teorii przyczynowości, którą podałem powyżej, nie jest jedyna. Warto zauważyć, że nawet Hume w swoim późniejszym dziele *An Inquiry Concerning the Human Understanding* (1758) podał inną definicję przyczynowości. Co więcej, nie wszyscy naukowcy podzielają moją wizję (na przykład Archie, 2005).

Teoria Hume'a jest bardzo interesująca z co najmniej dwóch punktów widzenia.

Po pierwsze elementy jego teorii wykazują duże podobieństwo do bardzo ważnego terminu w psychologii zwanego warunkowaniem (ang. *conditioning*). **Warunkowanie** jest formą uczenia się. Choć istnieje wiele rodzajów warunkowania, wszystkie opierają się na wspólnym fundamencie — mianowicie na skojarzeniach, nazywanych również **asocjacjami** (stąd nazwa tego typu uczenia się — **uczenie asocjacyjne**). W każdym typie warunkowania bierzemy jakieś zdarzenie lub obiekt (zwykle nazywane bodźcem) i kojarzymy je z pewnym zachowaniem lub reakcją. Uczenie asocjacyjne obserwuje się w przyrodzie u różnych gatunków. Jest właściwe ludziom, małpom człekokształtnym, psom i kotom, ale także występuje u znacznie prostszych organizmów, takich jak ślimaki (Alexander, Audesirk i Audesirk, 1985).

Warunkowanie

Więcej informacji na temat różnych rodzajów warunkowania znajdziesz na stronie *https://bit.ly/MoreOnConditioning*. Możesz również poszukać na przykład frazy *warunkowanie klasyczne* kontra *warunkowanie instrumentalne* i nazwisk, np. Ivan Pavlov i Burrhus Skinner.

Po drugie asocjacje są również podstawą większości klasycznych algorytmów uczenia maszynowego. Kiedy szkolimy sieć neuronową z wykorzystaniem uczenia nadzorowanego, staramy się znaleźć funkcję, która przekształca określone wejście na określone wyjście. Aby zrobić to skutecznie, trzeba ustalić, które elementy danych wejściowych są przydatne do prognozowania wyników. W większości przypadków asocjacje są wystarczające do osiągnięcia tego celu.

Dlaczego przyczynowość? Zapytaj dzieci!

Czy Hume'owskiej teorii przyczynowości czegoś brakuje? Chociaż na to pytanie próbowało odpowiedzieć wielu innych filozofów, w tej książce skupię się na jednej szczególnie interesującej odpowiedzi, która pochodzi od… dzieci.

Interakcje ze światem

Alison Gopnik to amerykańska psycholog dziecięca zajmująca się badaniami nad rozwijaniem modeli świata przez dzieci. Alison Gopnik pomaga również informatykom w zrozumieniu sposobu budowania zdroworozsądkowego rozumienia przez dzieci świata zewnętrznego. Dzieci korzystają z uczenia asocjacyjnego w jeszcze większym stopniu niż dorośli, ale są także nienasyconymi eksperymentatorami.

Czy widziałeś kiedyś rodzica próbującego przekonać swoje dziecko, aby przestało rzucać zabawką? Niektórzy rodzice interpretują tego typu zachowanie jako *niegrzeczne*, *destrukcyjne* lub *agresywne*, ale dzieci często kierują się innymi motywacjami. Prowadzą systematyczne eksperymenty, które pozwalają im poznać prawa fizyki i zasady społecznych interakcji (Gopnik, 2009). Już 11-miesięczne niemowlęta wolą przeprowadzać eksperymenty z przedmiotami, które wykazują nieprzewidywalne właściwości (na przykład sprawiają wrażenie przechodzenia przez ścianę), niż z obiektami, które zachowują się przewidywalnie (Stahl i Feigenson, 2015). Ta preferencja pozwala im budować skuteczne modele świata.

Od dzieci możemy się nauczyć, że nie powinniśmy ograniczać się, jak sugerował Hume, do obserwacji świata. Możemy także wchodzić z nim w interakcje. W kontekście wnioskowania przyczynowego te interakcje są nazywane **interwencjami**. Więcej informacji na ich temat można znaleźć w rozdziale 2. Interwencje stanowią sedno tego, co wielu uważa za Świętego Graala metody naukowej: **randomizowanych badań kontrolowanych** (ang. *randomized controlled trial* — RCT).

Zakłócenia — związki, które nie są prawdziwe

Zdolność przeprowadzania eksperymentów poszerza paletę możliwości w stosunku do tego, o czym myślał Hume. Eksperymenty stwarzają wielkie możliwości! Chociaż nie pozwalają rozwiązać wszystkich problemów filozoficznych związanych ze zdobywaniem nowej wiedzy, mogą rozwiązać część z nich. Bardzo ważną cechą odpowiednio zaprojektowanego, randomizowanego eksperymentu jest możliwość unikania **zakłóceń** (ang. *confounding*). Dlaczego to jest ważne?

Zmienna zakłócająca wpływa na dwie lub więcej innych zmiennych i generuje pomiędzy nimi *fałszywe* powiązanie. Z czysto statystycznego punktu widzenia takie powiązania są nie do odróżnienia od tych wynikających z mechanizmu przyczynowego. Dlaczego to stwarza problemy? Przyjrzyjmy się przykładowi.

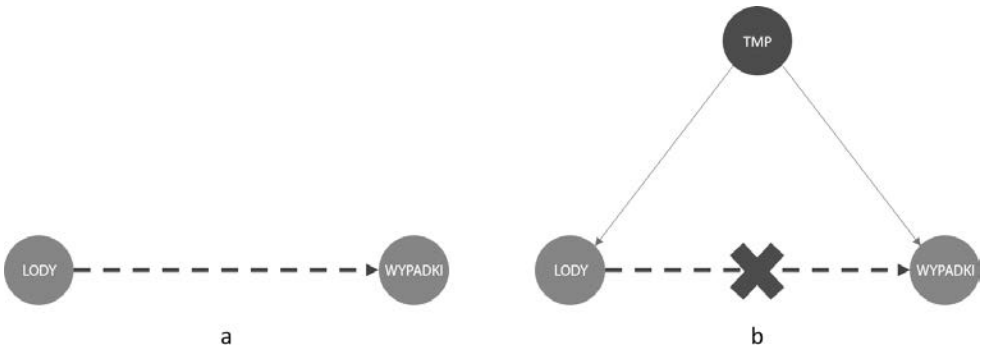
Wyobraź sobie, że pracujesz w instytucie badawczym i próbujesz zrozumieć przyczyny tonięcia ludzi. Twoja organizacja udostępniła Ci ogromną bazę danych zmiennych społeczno-ekonomicznych. Aby przewidzieć liczbę dziennych utonięć w obszarze zainteresowania, zdecydowałeś się na skorzystanie z modelu regresji na obszernym zbiorze tych zmiennych. Po sprawdzeniu wyników okazało się, że największy uzyskany współczynnik dotyczy dziennej sprzedaży lodów. To bardzo interesujące! Lody zwykle zawierają duże ilości cukru, więc być może cukier wpływa na koncentrację lub kondycję fizyczną ludzi przebywających w wodzie.

Ta hipoteza może mieć sens, ale zanim przejdziemy dalej, zadajmy kilka pytań. Co z innymi zmiennymi, których nie uwzględniliśmy w modelu? Czy w celu opisania wszystkich istotnych aspektów problemu dodaliśmy do modelu wystarczającą liczbę predyktorów? A co, jeśli dodaliśmy ich za dużo? Czy dodanie do modelu tylko jednej zmiennej może całkowicie zmienić wynik?

> **Dodanie zbyt wielu predyktorów**
>
> Dodanie do modelu *zbyt wielu* predyktorów może być szkodliwe zarówno z punktu widzenia statystycznego, jak i przyczynowego. Więcej na ten temat dowiemy się w rozdziale 3.

Okazuje się, że jest to możliwe.

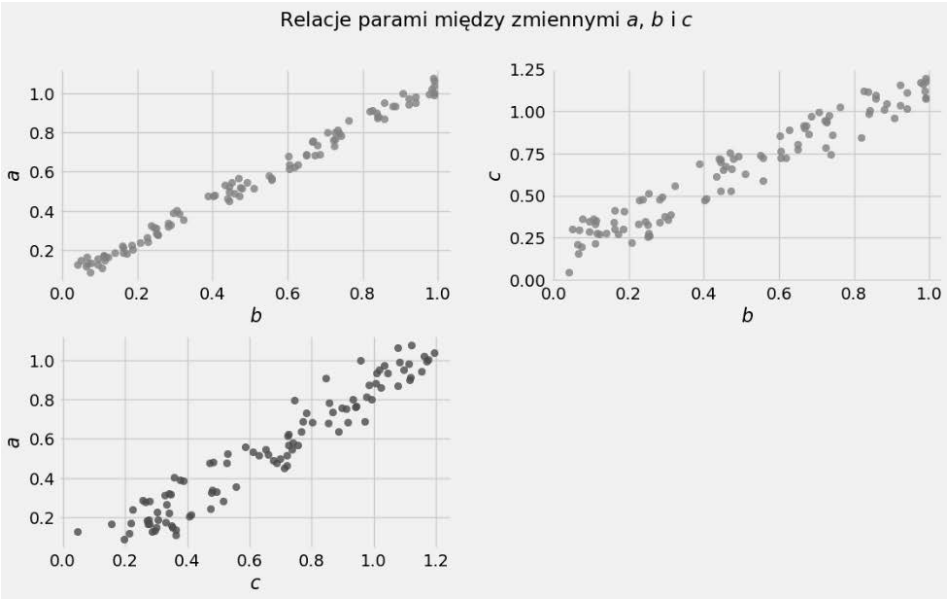
Pozwólcie, że przedstawię *czynnik zakłócający* — średnią dobową temperaturę. Wyższa dzienna temperatura sprawia, że ludzie chętniej kupują lody i chętniej pływają. Im więcej osób pływa, tym częściej dochodzi do wypadków. Spróbujmy zobrazować tę zależność (rysunek 1.1):



Rysunek 1.1. Graficzna reprezentacja modeli z dwiema (a) i trzema zmiennymi (b). Linie przerywane przedstawiają powiązanie, linie ciągłe przedstawiają związek przyczynowy. LODY = sprzedaż lodów, WYPADKI = liczba wypadków, TMP = temperatura

Na rysunku 1.1 widać, że dodanie do modelu średniej dziennej temperatury usuwa związek między sprzedażą lodów a dzienną liczbą utonięć. Dla niektórych czytelników może to być zaskakujące, dla innych nie. Więcej o mechanizmie tego efektu dowiemy się w rozdziale 3.

Zanim przejdę dalej, zwrócę szczególną uwagę na ważną rzecz: zakłócanie jest *pojęciem ściśle przyczynowym*. Co to znaczy? Chodzi o to, że nie da się wiele powiedzieć o zakłóceniach, posługując się językiem czysto statystycznym (zauważmy, że to oznacza, iż definicja Hume'a w takiej formie, w jakiej ją przedstawiłem, *nie jest w stanie tego uchwycić*). Aby to wyraźnie zobaczyć, spójrzmy na rysunek 1.2:

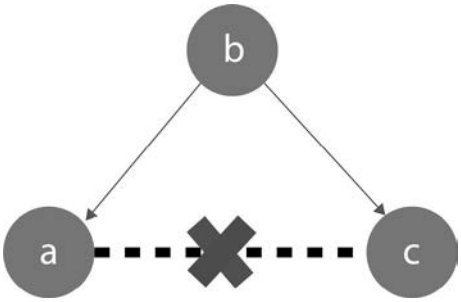


Rysunek 1.2. Wykresy punktowe relacji parami pomiędzy zmiennymi a, b i c. Kod umożliwiający odtworzenie poprzedniego wykresu można znaleźć w notatniku Chapter_01.ipynb (*https://github.com/PacktPublishing/Causal-Inferenceand-Discovery-in-Python/blob/main/Chapter_01.ipynb*)

Niebieskie punkty na rysunku 1.2 oznaczają związek przyczynowy, czerwone oznaczają związek pozorny, a zmienne *a*, *b* i *c* są powiązane w następujący sposób:

- *b* powoduje *a* i *c*,
- *a* i *c* są przyczynowo niezależne.

Graficzną reprezentację tych zależności przedstawia rysunek 1.3:



Rysunek 1.3. Zależności między zmiennymi a, b i c

Czarna linia przerywana z czerwonym krzyżykiem oznacza, że pomiędzy zmiennymi *a* i *c* nie ma związku przyczynowego w żadnym kierunku

Ale przecież na rysunku 1.2 widać pewną zależność! Spróbujmy się jej przyjrzeć!

Przedstawione na rysunku 1.2 zależności niefałszywe (niebieskie) i fałszywe (czerwone) wyglądają dość podobnie, a ich współczynniki korelacji są podobnie duże. W praktyce w większości przypadków po prostu nie można ich rozróżnić na podstawie wyłącznie kryteriów statystycznych. Do takiego rozróżnienia potrzebna jest wiedza przyczynowa.

Asymetrie i odkrywanie przyczyn

W gruncie rzeczy, aby dowiedzieć się, który kierunek jest przyczynowy, w niektórych przypadkach można wykorzystać rozkład szumów lub funkcjonalne asymetrie. Informacje te można wykorzystać do odtworzenia struktury przyczynowej na podstawie danych obserwacyjnych, ale wymagają one również pewnych założeń dotyczących procesu generowania tych danych. Więcej informacji na ten temat podam w części III „Odkrywanie związków przyczynowych" (rozdział 13.).

Napisałem, że w danych naszego eksperymentu niektóre relacje były fałszywe. Następnie dodałem do modelu kolejną zmienną, co zmieniło generowany przez niego wynik. Mimo tego byłem w stanie wykonać użyteczne prognozy bez tej zmiennej. Jeśli to prawda, dlaczego miałoby mnie obchodzić, czy związek jest fałszywy, czy nie? Dlaczego miałoby mnie interesować, czy związek jest przyczynowy, czy nie?

Jak nie stracić pieniędzy… i ludzkich istnień

Dowiedziałeś się, że problemów wynikających z zakłóceń można uniknąć dzięki eksperymentom randomizowanym. Niestety, nie zawsze są one dostępne. Czasami wykonanie takich eksperymentów może być zbyt kosztowne, nieetyczne lub praktycznie niemożliwe (na przykład przeprowadzenie eksperymentu polegającego na migracji dużej grupy jakiejś populacji). W tym podrozdziale omówię kilka scenariuszy, w których pożądane jest wyciąganie wniosków przyczynowych mimo dysponowania wyłącznie danymi obserwacyjnymi. Te przykłady stworzą solidną podstawę do kolejnych rozdziałów.

Dylemat marketera

Wyobraź sobie, że jesteś marketerem znającym się na technologii i chcesz efektywnie alokować swój budżet na marketing bezpośredni. Jak podszedłbyś do tego zadania? Gdy alokujesz budżet na kampanię marketingu bezpośredniego, chciałbyś wiedzieć, jakiego zwrotu możesz oczekiwać, jeśli wydasz na daną osobę określoną kwotę. Mówiąc inaczej, chcesz oszacować wpływ swoich działań na wyniki niektórych klientów (Gutierrez, Gérardy, 2017). Być może, w celu rozwiązania problemu można by skorzystać z technik uczenia nadzorowanego? Aby odpowiedzieć na to pytanie, przyjrzyjmy się bliżej temu, co chcemy prognozować.

Interesuje nas reakcja konkretnej osoby na prezentowane jej treści. Spróbujmy zakodować to w następującym wzorze:

$$\tau_i = Y_i(1) - Y_i(0)$$

Oto opis poszczególnych elementów powyższego wzoru:

- τ_i to efekt eksperymentu na osobę i;
- $Y_i(1)$ to wynik dla osoby i, kiedy została poddana eksperymentowi T (w tym przykładzie otrzymała od Ciebie treści marketingowe);
- $Y_i(0)$ to wynik dla tej samej osoby przy założeniu, że nie została poddana eksperymentowi T.

Zgodnie z powyższym wzorem od wyniku Y_i osoby i, gdy ta osoba nie została poddana eksperymentowi T, chcesz odjąć wynik uzyskany w przypadku, gdyby osobę tę poddano eksperymentowi T.

Interesujące jest to, że aby rozwiązać to równanie, trzeba wiedzieć, która odpowiedź osoby i została uzyskana po przeprowadzeniu eksperymentu, a która bez jego przeprowadzenia. W rzeczywistości nigdy nie można obserwować tej samej osoby w dwóch wzajemnie wykluczających się warunkach jednocześnie. Aby rozwiązać równanie z poprzedniego wzoru, potrzebne są kontrfakty (ang. *counterfactuals*).

Kontrfakty to szacunki wyników po zmianie wartości jednej lub większej liczby zmiennych, gdyby wszystkie inne pozostały niezmienione. Ponieważ kontrfaktów nie można zaobserwować, prawdziwy skutek przyczynowy τ jest nieznany. Jest to jeden z powodów, dla których wyżej sformułowanego problemu nie da się rozwiązać z wykorzystaniem

klasycznych technik uczenia maszynowego. Rodzina technik przyczynowych zwykle stosowanych do takich problemów nazywa się **modelowaniem różnicowym** (ang. *uplift modelling*), które omówię dokładniej w rozdziałach 9. i 10.

Pobawmy się w doktora!

Spróbuję posłużyć się innym przykładem. Wyobraź sobie, że jesteś lekarzem. Jedna z Twoich pacjentek, Joanna, cierpi na rzadką chorobę *D*. Dodatkowo zdiagnozowano u niej wysokie ryzyko wystąpienia zakrzepów krwi. Przestudiowałeś informacje dotyczące dwóch najpopularniejszych leków na chorobę *D*. Obydwa leki mają praktycznie identyczną skuteczność na *D*, ale na podstawie diagnozy postawionej Joannie nie masz pewności, który lek będzie dla niej bezpieczniejszy. Przeglądasz dane badawcze przedstawione w tabeli 1.1:

Tabela 1.1. Dane dla leków A i B

Lek	A		B	
Zakrzepy	Tak	Nie	Tak	Nie
Razem	27	95	23	99
Procent	22%	78%	19%	81%

Liczby w tabeli 1.1 przedstawiają liczbę pacjentów, u których zdiagnozowano chorobę *D*, poddanych leczeniu za pomocą specyfików *A* lub *B*. Wiersz 2 (**zakrzepy**) zawiera informacje o tym, czy u pacjentów wykryto zakrzepy krwi, czy nie. Warto zwrócić uwagę, że wyniki procentowe są zaokrąglone. Który lek byś wybrał na podstawie tych danych? Odpowiedź wydaje się dość oczywista. U 81% pacjentów, którzy otrzymali lek *B*, nie wystąpiły zakrzepy krwi. To samo dotyczyło jedynie 78% pacjentów, którzy otrzymali lek *A*. Ryzyko powstania zakrzepów krwi jest o około 3% niższe u pacjentów otrzymujących lek *B* w porównaniu z pacjentami otrzymującymi lek *A*.

To wygląda na dobry wniosek, ale jesteś sceptyczny. Wiesz, że zakrzepy krwi mogą być bardzo ryzykowne i chcesz dowiedzieć się więcej. Znalazłeś bardziej szczegółowe dane, które uwzględniają płeć pacjenta. Spójrzmy na tabelę 1.2:

Tabela 1.2. Dane dla leku A i leku B z dodanymi wynikami dla płci. K = kobieta, M = mężczyzna. Dla ułatwienia interpretacji dodano kodowanie kolorami, lepsze wyniki zaznaczono na zielono, a gorsze na pomarańczowo

Lek	A		B	
Zakrzepy	Tak	Nie	Tak	Nie
Kobiety	24	56	17	25
Mężczyźni	3	39	6	74
Razem	27	95	23	99
Procent	22%	78%	18%	82%
Procent (K)	30%	70%	40%	60%
Procent (M)	7%	93%	7,5%	92,5%

Stało się tu coś dziwnego. Uzyskane liczby są takie same jak poprzednio: lek *B* nadal jest preferowany dla wszystkich pacjentów. Wydaje się jednak, że lek *A* działa lepiej w przypadku i kobiet, i mężczyzn! Czy właśnie znaleźliśmy medycznego kota Schrödingera (*https://en.wikipedia.org/wiki/Schr%C3%B6dinger%27s_cat*), który odwraca działanie leku, gdy zaobserwuje płeć pacjenta?

Jeśli sądzisz, że mogłem pomylić się w obliczeniach, nie wierz mi na słowo, po prostu samodzielnie sprawdź dane. Można je znaleźć w pliku *data/ch_01_drug_data.csv* (*https://github.com/PacktPublishing/Causal-Inference-and-Discovery-in-Python/blob/main/data/ch_01_drug_data.csv*).

To, czego właśnie doświadczyliśmy, nazywa się **paradoksem Simpsona** (znanym również jako **efekt Yule-Simpsona**). Paradoks Simpsona występuje wtedy, gdy podział danych uwzględniający dodatkowe zmienne w ustawieniach regresji znacząco zmienia wynik analizy. W rzeczywistym świecie zwykle istnieje wiele sposobów partycjonowania danych. Mógłbyś zapytać: skąd więc mam wiedzieć, który podział jest *właściwy*?

Można by spróbować odpowiedzieć na to pytanie ściśle z punktu widzenia uczenia maszynowego: przeprowadzić selekcję cech z walidacją krzyżową i wybrać zmienne, które w znaczący sposób wpływają na wynik. To rozwiązanie w niektórych sytuacjach jest wystarczająco dobre. Sprawdzi się np. wtedy, gdy zależy nam jedynie na prognozowaniu (a nie na podejmowaniu decyzji) i gdy wiadomo, że dane produkcyjne będą niezależne i równomiernie rozłożone. Innymi słowy, dane produkcyjne muszą mieć rozkład praktycznie identyczny (lub przynajmniej wystarczająco podobny) do rozkładu danych szkoleniowych i walidacyjnych. Jeśli chcesz czegoś więcej, będziesz potrzebować jakiegoś (przyczynowego) modelu świata.

Asocjacje w realnym świecie

Niektórzy uważają, że relacje czysto asocjacyjne (skojarzeniowe) rzadko występują w prawdziwym świecie lub że zazwyczaj są słabe, więc nie mogą zbytnio wpływać na uzyskiwane wyniki. Aby przekonać się, jak zaskakująco silne i spójne mogą być fałszywe relacje w prawdziwym świecie, odwiedź stronę Tylera Vigena: *https://www.tylervigen.com/spurious-correlations*. Warto zauważyć, że zależności pomiędzy wieloma zmiennymi są czasami bardzo silne i utrzymują się przez długi czas! Osobiście podoba mi się przykład ze startami kosmicznymi i doktoratami z socjologii. Często wykorzystuję go podczas moich wykładów i prezentacji. Który jest Twoim ulubionym? Udostępnij go innym i oznacz mnie na LinkedIn, Twitterze (aby nawiązać kontakt, zapoznaj się z punktem „Pozostańmy w kontakcie" w rozdziale 15.). Możemy o tym porozmawiać!

Podsumowanie

„Niech przemówią dane" to chwytliwe i mocne hasło, ale jak można się było przekonać podczas lektury tego rozdziału, same dane nie zawsze wystarczą. Warto pamiętać, że w wielu przypadkach „dane nie mówią same za siebie" (Hernán, Robins, 2020) i że do odpowiedzi na niektóre pytania mogą być potrzebne dodatkowe informacje poza obserwacjami.

W tym rozdziale dowiedziałeś się, że gdy myślisz o przyczynowości, nie powinieneś, wbrew temu, co uważał David Hume, ograniczać się do obserwacji. Możesz też eksperymentować, zupełnie jak dzieci.

Niestety, eksperymenty nie zawsze są dostępne. W takiej sytuacji możesz spróbować wykorzystać dane obserwacyjne do wyciągnięcia wniosku przyczynowego, jednak same dane zwykle nie wystarczą do osiągnięcia tego celu. Potrzebujesz także modelu przyczynowego. W następnym rozdziale przedstawię *drabinę przyczynowości* — zaproponowaną przez Judeę Pearla zgrabną metaforę pozwalającą zrozumieć trzy poziomy przyczynowości.

Bibliografia

Alexander J.E., Audesirk T.E., Audesirk G.J. *Classical Conditioning in the Pond Snail Lymnaea stagnalis*. „The American Biology Teacher", 47(5), 1985, s. 295 – 298. *https://doi.org/10.2307/4448054*.

Archie L., *Hume's Considered View on Causality*. [Preprint]: *http://philsci-archive.pitt.edu/id/eprint/2247*, 2005 (dostęp 23.04.2022).

Falcon A. „Aristotle on Causality", *The Stanford Encyclopedia of Philosophy* (wydanie wiosna 2022), pod red. Edwarda N. Zalta, *https://plato.stanford.edu/archives/spr2022/entries/aristotle-causality/*. Dostęp 23.04.2022.

Gopnik A., *The philosophical baby: What children's minds tell us about truth, love, and the meaning of life*, Nowy Jork, Farrar, Straus and Giroux, 2009.

Gutierrez P., Gérardy, J., *Causal Inference and Uplift Modelling: A Review of the Literature*, materiały z 3rd International Conference on Predictive Applications and APIs, opublikowane w „Proceedings of Machine Learning Research", 67, 2017, s. 1 – 13.

Hernán M.A., Robins J.M., *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC, 2020.

Hume D., Millican P.F., *An enquiry concerning human understanding*. Oxford: Oxford University Press, 2007.

Kahneman D., *Thinking, Fast and Slow*, Farrar, Straus and Giroux, 2011.

Lorkowski C.M., *https://iep.utm.edu/hume-causation/*. Dostęp 23.04.2022.

Stahl A.E., Feigenson L., *Cognitive development. Observing the unexpected enhances infants' learning and exploration*. „Science", 348(6230), 2015, s. 91 – 94. *https://doi.org/10.1126/science.aaa3799*.

Skorowidz

E

F

G

N

O

P

R

S

T

Przyczyna i skutek, nic więcej. Pomyłki jako takie nie istnieją...

José Antonio Cotrina, hiszpański pisarz science fiction

W uczeniu maszynowym odkrywanie związków przyczynowych daje możliwości, jakich nie można uzyskać tradycyjnymi technikami statystycznymi. Najnowsze trendy w programowaniu pokazują, że przyczynowość staje się kluczowym zagadnieniem dla generatywnej sztucznej inteligencji. Niezbędna okazuje się więc znajomość grafów przyczynowych i zapytań konfrontacyjnych.

Dzięki tej książce łatwo przyswoisz teoretyczne podstawy i zaczniesz je płynnie wdrażać w rzeczywistych scenariuszach. Dowiesz się, w jaki sposób myślenie przyczynowe ułatwia rozwiązywanie problemów, i poznasz pojęcia Pearla, takie jak strukturalny model przyczynowy, interwencje, kontrfakty itp. Każde zagadnienie zostało dokładnie wyjaśnione i opatrzone zbiorem praktycznych ćwiczeń z kodem w Pythonie. Nauczysz się także implementować poszczególne modele i zrozumiesz, czym się kierować przy wyborze technik i algorytmów do rozwiązywania konkretnych scenariuszy przyczynowych. To przewodnik, który docenią szczególnie inżynierowie uczenia maszynowego i analitycy danych.

W książce:

- wnioskowanie związków przyczynowych
- budowa i działanie strukturalnych modeli przyczynowych
- czteroetapowy proces wnioskowania związków przyczynowych w Pythonie
- techniki modelowania efektu interwencji
- nowoczesne metody odkrywania związków przyczynowych za pomocą Pythona
- korzystanie z wnioskowania związków przyczynowych

ALEKSANDER MOLAK jest niezależnym badaczem i konsultantem w dziedzinie uczenia maszynowego. Współpracował z licznymi firmami w Europie, USA i Izraelu, gdzie uczestniczył w tworzeniu wielkoskalowych systemów uczenia maszynowego. Jest też współzałożycielem firmy Lespire.io, dostawcy szkoleń z zakresu sztucznej inteligencji dla zespołów korporacyjnych.

Helion
helion.pl
HELION SA
ul. Kościuszki 1c
44-100 Gliwice
tel.: 32 230 98 63
helion@helion.pl
KOD KORZYŚCI
Sięgnij po więcej!
ISBN 978-83-289-0832-1
9 788328 908321
Cena: 109,00 zł