Sztuczna inteligencja : metody stochastyczne (V / XVI)

ĆWICZENIA

1. Rzutka sześciokątna kostka jest rzucana pięć razy, a kolejne liczby są rejestrowane w sekwencji. Ile jest różnych sekwencji?
2. Znajdź liczbę możliwych do odróżnienia permutacji liter w słowie MISSISSIPPI, liter wyrazu ASSOCIATIVE.
3. Załóżmy, że urna zawiera 15 kulek, z których osiem jest czerwonych, a siedem czarnych. Na ile sposobów można wybrać pięć piłek, aby:
a. wszystkie pięć jest czerwonych? wszystkie pięć jest czarnych?
b. dwa są czerwone, a trzy czarne
co. co najmniej dwa są czarne
4. Na ile sposobów można wybrać komitet złożony z trzech członków wydziału i dwóch studentów z grupy pięciu wykładowców i siedmiu studentów?
5. W badaniu przeprowadzonym wśród 250 widzów telewizyjnych 88 lubi oglądać wiadomości, 98 lubi oglądać sport, a 94 lubi oglądać komedie. 33 osoby lubią oglądać wiadomości i sport, 31 lubią oglądać sport i komedię, a 35 lubi oglądać wiadomości i komedie. 10 osób lubi oglądać wszystkie trzy. Załóżmy, że osoba z tej grupy jest wybierana losowo:
za. Jakie jest prawdopodobieństwo, że oglądają wiadomości, ale nie sport?
b. Jakie jest prawdopodobieństwo, że oglądają wiadomości lub sport, ale nie komedię?
do. Jakie jest prawdopodobieństwo, że nie oglądają sportu ani wiadomości?
6. Jakie jest prawdopodobieństwo, że czterocyfrowa liczba całkowita bez zer początkowych:
a. Ma 3, 5 lub 7 jako cyfrę?
b. Zaczyna się od 3, kończy na 5, czy ma 7 jako cyfrę?
7. Dwie kości są rzucane. Znajdź prawdopodobieństwo, że ich suma wynosi:
a. 4
b. 7 lub liczba parzysta
c. 10 lub więcej
8. Karta jest dobierana ze zwykłej talii 52 kart. Jakie jest prawdopodobieństwo:
a. losowanie karty twarzy (walet, królowa, król lub as)
b. rysunek królowej lub pika
c. losowanie karty twarzy lub maczugi
9. Jakie jest prawdopodobieństwo, że zostaną rozdane następujące układy w pokera na pięć kart (z normalnej talii 52 kart)?
a. „Kolor” lub wszystkie karty tego samego koloru.
b. „Full house” lub dwie karty o tej samej wartości i trzy karty o innej wartości.
c. „Poker królewski” lub dziesiątka, walet, królowa, król i as tego samego koloru.
10. Oczekiwanie jest średnią lub średnią wartości zmiennej losowej. Na przykład rzucając kostką, można ją obliczyć, sumując uzyskane wartości z dużej liczby rzutów, a następnie dzieląc przez liczbę rzutów.
a. Jakie są oczekiwania po rzucie uczciwą kostką?

b. Wartość koła ruletki z 37 równie prawdopodobnymi wynikami?
c. Wartość losowania z zestawu kart (as jest wyceniany na 1, wszystkie pozostałe karty na 10)?
11. Załóżmy, że gramy w grę, w której rzucamy kostką, a następnie otrzymujemy kwotę równą wartości kości. Na przykład, jeśli pojawi się 3, otrzymamy 3 USD. Czy zagranie w tę grę kosztuje nas 4 USD, czy jest to uzasadnione?
12. Rozważ sytuację, w której losowo generowane są ciągi bitów o długości cztery. Zademonstruj, czy zdarzenie wytworzenia ciągów bitów zawierających parzystą liczbę 1 jest niezależne od zdarzenia wytworzenia ciągów bitów, które kończą się na 1.
13. Pokaż, że zdanie p (A, B | C) = p (A | C) p (B | C) jest równoważne zarówno p (A | B, C) = p (A | C) i p (B | A, C) = p (B | C).
14. Przy wytwarzaniu produktu 85% wytwarzanych produktów nie jest wadliwych. Spośród skontrolowanych produktów 10% dobrych jest postrzeganych jako wadliwe i nie jest wysyłanych, podczas gdy tylko 5% wadliwych produktów jest zatwierdzonych i wysłanych. Jeśli produkt jest wysłane, jakie jest prawdopodobieństwo, że jest uszkodzony?
15. Badanie krwi jest w 90% skuteczne w wykrywaniu choroby. Fałszywie diagnozuje również, że u zdrowej osoby choroba występuje w 3% przypadków. Jeśli 10% badanych ma chorobę, jakie jest prawdopodobieństwo, że osoba, która uzyska pozytywny wynik, faktycznie będzie miała tę chorobę?
16. Załóżmy, że towarzystwo ubezpieczeń samochodowych klasyfikuje kierowcę jako dobry, średni lub zły. Spośród wszystkich ubezpieczonych kierowców 25% jest sklasyfikowanych jako dobre, 50% to średnie, a 25% to złe. Załóżmy, że na nadchodzący rok dobry kierowca ma 5% szansy na wypadek, a przeciętny kierowca ma 15% szansy na wypadek, a zły kierowca ma 25% szansy. Jeśli miałeś wypadek w ubiegłym roku, jakie jest prawdopodobieństwo, że jesteś dobrym kierowcą?
17. Trzej więźniowie, A, B, C są w swoich celach. Powiedziano im, że jeden z nich zostanie stracony następnego dnia, a inni zostaną ułaskawieni. Tylko gubernator wie, kto zostanie stracony. Więzień A prosi strażnika o przysługę. „Zapytaj gubernatora, kto zostanie stracony, a następnie powiedz więźniowi B lub C, że zostaną ułaskawieni". Strażnik robi to, o co go poproszono, a następnie wraca i mówi więźniowi A, że powiedział więźniowi B, że on (B) zostanie ułaskawiony. Jakie są szanse na egzekucję więźnia A, biorąc pod uwagę tę wiadomość? Czy jest więcej informacji niż przed jego prośbą do strażnika?

METODY STOCHASTYCZNE

Wprowadzenie

Wprowadziliśmy wyszukiwanie heurystyczne jako podejście do rozwiązywania problemów w domenach, w których albo problem nie ma dokładnego rozwiązania, albo gdzie pełna przestrzeń stanu może być zbyt kosztowna do obliczenia. W tym rozdziale proponujemy metodologię stochastyczną jako odpowiednie dla tych sytuacji. Argumentacja probabilistyczna jest również odpowiednia w sytuacjach, w których informacje o stanie znajdują się na podstawie próbkowania bazy informacji, a modele przyczynowe są wyciągane z danych. Jedną z ważnych dziedzin zastosowań dla stochastycznej metodologii jest rozumowanie diagnostyczne, w którym relacje przyczyna / skutek nie zawsze są ujmowane w sposób czysto deterministyczny, jak to często jest możliwe w podejściach opartych na wiedzy do rozwiązywania problemów, które widzieliśmy w rozdziałach 2, 3 , 4 i zobaczymy ponownie w rozdziale 8. Sytuacja diagnostyczna zwykle przedstawia dowody, takie jak gorączka lub ból głowy, bez dalszego przyczynowego uzasadnienia. W rzeczywistości dowody mogą często wskazywać na kilka różnych przyczyn, np. Gorączka może być spowodowana grypą lub infekcją. W takich sytuacjach informacje probabilistyczne mogą często wskazywać i uszeregować pod względem ważności możliwe wyjaśnienia dowodów. Inną interesującą aplikacją dla stochastycznej metodologii jest hazard, w którym podobno przypadkowe zdarzenia, takie jak rzut kostkami, rozdawanie kart potasowanych lub obrót koła ruletki, powodują potencjalną wypłatę gracza. W rzeczywistości w XVIII wieku próba

stworzenia matematycznych podstaw hazardu była ważną motywacją dla Pascala (a później Laplace'a) do opracowania rachunku probabilistycznego. Wreszcie, jak zauważono w sekcji 1.1.4, „umiejscowienie" rachunkowości inteligencji sugeruje, że decyzje ludzkie często pojawiają się w złożonych, krytycznych czasowo i ucieleśnionych środowiskach, w których rachunek w pełni mechanistyczny może po prostu nie być możliwy do zdefiniowania lub, jeśli jest zdefiniowany, może nie obliczać odpowiedzi w użytecznym czasie. W takich sytuacjach inteligentne działania najlepiej można postrzegać jako stochastyczne reakcje na przewidywane koszty i korzyści. Następnie opisujemy kilka obszarów problemowych, wśród których wiele, w których obliczeniowa implementacja inteligencji jest często wykorzystywana metodologią stochastyczną; obszary te będą głównymi tematami w dalszych częściach.

1. Uzasadnienie diagnostyczne. Na przykład w diagnozie medycznej nie zawsze istnieje oczywisty związek przyczynowo-skutkowy między zestawem objawów przedstawionych przez pacjenta a przyczynami tych objawów. W rzeczywistości te same zestawy objawów często sugerują wiele możliwych przyczyn. Ważne są również modele probabilistyczne , złożone sytuacje mechaniczne, takie jak monitorowanie lotów samolotów lub helikopterów. Systemy oparte na regułach i probabilistyczne zostały zastosowane w tych i innych domenach diagnostycznych.

2. Rozumienie języka naturalnego. Jeśli komputer ma rozumieć i używać ludzkiego języka, musi on być w stanie scharakteryzować sposób, w jaki sami ludzie używają tego języka. Wyrazy, wyrażenia i metafory są przyswajane, ale zmieniają się i ewoluują wraz z upływem czasu. Obsługuje metodologię stochastyczną rozumienia języka; na przykład, gdy system obliczeniowy jest szkolony w bazie danych specyficznych zastosowań języka (zwanej językoznawstwem korpusowym). Rozważamy te problemy językowe w dalszej części.

3. Planowanie i harmonogramowanie. Gdy agent tworzy plan, na przykład samochód na wakacje, często zdarza się, że żadna deterministyczna sekwencja operacji nie jest gwarantowana. Co się stanie, jeśli samochód się zepsuje, jeśli prom samochodowy zostanie odwołany w określonym dniu, jeśli hotel jest w pełni zarezerwowany, mimo że dokonano rezerwacji? Szczegółowe plany dotyczące ludzi lub robotów są często wyrażane w języku probabilistycznym.

4. Uczenie się. Trzy poprzednie wspomniane obszary można również postrzegać jako dziedziny automatycznego uczenia się. Ważnym elementem wielu systemów stochastycznych jest to, że mają one możliwość próbkowania sytuacji i uczenia się w czasie. Niektóre zaawansowane systemy są w stanie zarówno próbkować dane i przewidywać wyniki, jak i uczyć się nowych relacji probabilistycznych na podstawie danych i wyników.

Metodologia stochastyczna opiera się na właściwościach liczenia. Prawdopodobieństwo zdarzenia w sytuacji opisuje się jako stosunek liczby sposobów wystąpienia zdarzenia do całkowitej liczby możliwych wyników tego zdarzenia. Zatem prawdopodobieństwo, że liczba parzysta wynika z rzutu rzetelną kością, jest całkowitą liczbą parzystych wyników (tutaj 2, 4 lub 6) w stosunku do całkowitej liczby wyników (1, 2, 3, 4, 5 lub 6) lub 1/2. Ponownie, prawdopodobieństwo wyciągnięcia marmuru określonego koloru z torby marmuru jest stosunkiem liczby kulek tego koloru do całkowitej liczby kulek w torbie. W sekcji 1 wprowadzamy podstawowe techniki liczenia, w tym zasady sumowania i produktu. Ze względu na ich znaczenie w liczeniu, prezentujemy również permutacje i kombinacje dyskretnych zdarzeń. Jest to opcjonalna sekcja, którą czytelnicy mogą pominąć z wystarczającym tłem w dyskretnej matematyce. W sekcji 2 wprowadzamy język formalny do rozumowania przy użyciu metody stochastycznej. Obejmuje to definicje niezależności i różne typy zmiennych losowych. Na przykład, w przypadku twierdzeń probabilistycznych, zmienne losowe mogą być logiczne (prawda lub fałsz), dyskretne, liczby całkowite od 1 do 6 jak w rzucie rzetelną lub ciągłą, funkcją zdefiniowaną na liczbach rzeczywistych. W części dalszej przedstawiamy twierdzenie Bayesa, które popiera większość podejść do modelowania stochastycznego i uczenia się. Zasada Bayesa jest ważna dla interpretacji nowych dowodów w kontekście wcześniejszej wiedzy lub doświadczenia. Później przedstawiamy dwa

zastosowania metod stochastycznych, w tym probabilistyczne automaty skończone i metodologię przewidywania wzorców słów angielskich na podstawie próbkowanych danych.

5.1 Elementy liczenia (opcjonalnie)

Podstawą metodologii stochastycznej jest umiejętność zliczania elementów domeny aplikacji. Podstawą gromadzenia i liczenia elementów jest oczywiście teoria mnogości, w której musimy być w stanie jednoznacznie ustalić, czy element jest, czy nie jest członkiem zbioru elementów. Po ustaleniu tego, istnieją metodologie zliczania elementów zbiorów, dopełnienia zbioru oraz połączenia i zamiany wielu zbiorów. Przeglądniemy te techniki

Zasady dodawania i mnożenia

Jeśli mamy zestaw A, liczba elementów w zestawie A jest oznaczona przez | A |, zwaną licznością A. Oczywiście, A może być puste (liczba elementów wynosi zero), skończone, w nieskończoność lub nieskończenie nieskończone. Każdy zestaw jest zdefiniowany w kategoriach dziedziny zainteresowań lub wszechświata U elementów, które mogą być w tym zestawie. Na przykład zbiór mężczyzn w klasie może być zdefiniowany w kontekście lub we wszechświecie wszystkich ludzi w tym pokoju. Podobnie rzut 3 na uczciwej kości może być postrzegany jako jeden z zestaw sześciu możliwych wyników. Domena lub wszechświat zbioru A jest zatem również zbiorem i służy do określenia dopełnienia tego zbioru, A. Na przykład dopełnieniem zbioru wszystkich mężczyzn w klasie właśnie wspomnianej jest zbiór wszystkich kobiet, i uzupełnieniem rzutu {3} rzetelnej kości jest {1, 2, 4, 5, 6}. Jeden zestaw A jest podzbiorem drugiego zestawu B, ab.png , jeśli każdy element zestawu A jest również elementem zestawu B. Zatem, trywialnie, każdy zestaw jest podzbiorem samego siebie, każdy zestaw A jest podzbiorem jego wszechświata, a pusty zestaw, oznaczony {} lub ϕ, jest podzbiorem każdego zestawu. Połączenie dwóch zbiorów A i B, A ∪ B, można opisać jako zbiór wszystkich elementów w każdym zestawie. Liczba elementów w połączeniu dwóch zbiorów jest sumą wszystkich elementów w każdym zestawie minus liczba elementów znajdujących się w obu zestawach. Uzasadnieniem tego jest oczywiście fakt, że każdy odrębny element w zestawie można policzyć tylko raz. Trywialnie, jeśli dwa zbiory nie mają wspólnych elementów, liczba elementów w ich połączeniu jest sumą liczby elementów w każdym zestawie. Przecięcie dwóch zbiorów A i B, A ∩ B, jest zbiorem wszystkich elementów wspólnych dla obu zbiorów. Podajemy teraz przykłady szeregu właśnie zdefiniowanych pojęć.

Załóżmy, że wszechświat, U, jest zbiorem {0, 1, 2, 3, 4, 5, 6, 7, 8, 9}

Niech A będzie zbiorem {1, 3, 5, 7, 9}

Niech B będzie zbiorem {0, 2, 4, 6, 8}

Niech C będzie zbiorem {4, 5, 6}

Następnie | A | wynosi 5, | B | wynosi 5, | C | wynosi 3, a | U | wynosi 10.

Ponadto, abu.png | A| ∪ B = U, | B | = | A |, $A = B^{-}$

Ponadto, | A | ∪ B | = | A | + | B | = 10, ponieważ A ∩ B = {}, ale

| A | ∪ C | = | A | + | C | - | A ∩ C | = 7, ponieważ A ∩ C = {5}

Właśnie przedstawiliśmy główne składniki reguły dodawania do łączenia dwóch zestawów. Dla dowolnych dwóch zbiorów A i C liczba elementów w połączeniu tych zbiorów wynosi:

| A ∪ C | = | A | + | C | - | A ∩ C |

Zauważ, że ta reguła dodawania utrzymuje, czy dwa zestawy są rozłączne, czy mają wspólne elementy. Podobna reguła dodawania obowiązuje dla trzech zestawów A, B i C, ponownie, niezależnie od tego, czy mają one wspólne elementy:

$$| A \cup B \cup C | = | A | + | B | + | C | - | A \cap B | - | A \cap C | - | B \cap C | + | A \cap B \cap C |$$

Argument podobny do przedstawionego wcześniej można wykorzystać do uzasadnienia tego równania. Podobne równania włączenia / wyłączenia są dostępne i łatwe do wykazania w przypadku dodawania zestawów elementów z więcej niż trzech zbiorów. Zasada mnożenia do zliczania mówi, że jeśli mamy dwa zestawy elementów A i B odpowiednio o rozmiarach a i b, to istnieje x x unikalne sposoby łączenia elementów tych zbiorów razem. Uzasadnieniem tego jest oczywiście to, że dla każdego z elementów A istnieją pary b dla tego elementu. Zasada mnożenia obsługuje wiele technik stosowanych w zliczaniu, w tym iloczyn kartezjański zbiorów, a także permutacje i kombinacje zbiorów. Iloczyn kartezjański dwóch zbiorów A i B, oznaczony jako A X B, jest zbiorem wszystkich uporządkowanych par (a, b), gdzie a jest elementem zbioru A, a b jest elementem zbioru B; lub bardziej formalnie:

$$A \text{ X } B = \{(a, b) \mid (a \in A) \land (b \in B)\}$$

oraz przez zasadę mnożenia liczenia:

$$| A \text{ X } B | = | A | \text{ x } | B |$$

Produkt kartezjański można oczywiście zdefiniować w dowolnej liczbie zestawów. Produktem dla n zestawów będzie zestaw n-krotek, w którym pierwszy składnik n-krotki jest dowolnym elementem pierwszego zestawu, drugim składnikiem n-krotki jest dowolny element drugiego zestawu i tak dalej. Ponownie liczba unikalnych n-krotek, które powstają, jest iloczynem liczby elementów w każdym zestawie.

Permutacje i kombinacje

Permutacja zestawu elementów jest uporządkowaną sekwencją elementów tego zestawu. W takim układzie elementów każdy może być użyty tylko raz. Przykładem permutacji jest uporządkowanie lub zamówienie zestawu dziesięciu książek na półce, które mogą pomieścić wszystkie dziesięć. Innym przykładem jest przydzielenie określonych zadań czterem z grupy sześciorga dzieci.

Często chcemy wiedzieć, ile (unikalnych) permutacji ma zbiór n elementów. Aby to ustalić, używamy mnożenia. Jeśli w zestawie A znajduje się n elementów, to zestaw permutacji tych elementów jest sekwencją długości n, gdzie pierwszy element sekwencji jest dowolnym z n elementów A, drugim elementem sekwencji jest dowolny z (n - 1) pozostałych elementów A, trzeci element sekwencji to dowolny z pozostałych (n - 2) elementów, a więc na. Kolejność umieszczania elementów w sekwencji permutacji jest nieistotna, tj. Dowolny z n elementów zestawu może być umieszczony najpierw w dowolnym miejscu w sekwencji, dowolny z pozostałych pozostałych elementów n - 1 może być umieszczony jako drugi w dowolnym z n - 1 pozostałych lokalizacji sekwencji permutacji i tak dalej. Wreszcie, zgodnie z zasadą mnożenia, istnieje n! sekwencje permutacji dla tego zestawu n elementów. Możemy ograniczyć liczbę elementów w permutacji zestawu A do dowolnej liczby większej lub równej zero i mniejszej lub równej liczbie elementów n w oryginalnym zestawie A. Na przykład, chcielibyśmy wiedzieć ile jest różnych zamówień z dziesięciu możliwych książek na półce, które mogą pomieścić tylko sześć z nich na raz. Gdybyśmy chcieli ustalić liczbę permutacji n elementów A wziętych r jednocześnie, gdzie $0 \leq r \leq n$, używamy mnożenia jak poprzednio, z tym wyjątkiem, że teraz mamy tylko r miejsc w każdej sekwencji permutacji:

$$n \text{ x } (n - 1) \text{ x } (n - 2) \text{ x } (n - 3) \text{ x } ... \text{ x } (n - (r - 1))$$

Alternatywnie możemy przedstawić to równanie jako:

nx (n - 1) x (n - 2) x (n - 3) x ... x (n - (r - 1)) x (n - r) x (n - r -1) x ... x 2 x 1 / (n - r) x (n - r -1) x ... x 2 x 1

lub równoważnie, liczba permutacji n elementów wziętych r w czasie, która jest symbolizowana jako nPr, wynosi:

nPr = n! / (n - r)!

Kombinacja zestawu n elementów jest dowolnym podzbiorem tych elementów, który można utworzyć. Podobnie jak w przypadku permutacji, często chcemy policzyć liczbę kombinacji elementów, które można utworzyć, biorąc pod uwagę zestaw elementów. Zatem istnieje tylko jedna kombinacja n elementów zestawu n elementów. Kluczową ideą jest to, że liczba kombinacji reprezentuje liczbę podzbiorów pełnego zestawu elementów, które można utworzyć. W przykładzie półki na książki kombinacje reprezentują różne podzbiory sześciu książek, książek na półce, które mogą być utworzone z pełnego zestawu dziesięciu książek. Innym przykładem kombinacji jest zadanie tworzenia czteroosobowych komitetów z grupy piętnastu osób. Każda osoba jest w komitecie lub nie, i nie ma znaczenia, czy jest on pierwszym, czy ostatnim wybranym członkiem. Kolejnym przykładem jest pięciokartowa ręka w grze w pokera. Kolejność rozdawania kart nie ma znaczenia dla ostatecznej wartości ręki. (Może to mieć ogromną różnicę w wartości zakładów, jeśli ostatnie cztery karty zostaną odkryte, jak w tradycyjnym pokerze stud, ale ostateczna wartość układu jest niezależna od rozłożonego zamówienia). Liczba kombinacji n elementów wziętych r jednocześnie, gdzie $0 \leq r \leq n$, jest symbolizowane przez nCr. Prostą metodą określania liczby tych kombinacji jest wzięcie liczby permutacji, nPr, jak już to zrobiliśmy, a następnie podzielenie liczby duplikatów. Ponieważ jakikolwiek podzbiór elementu r n elementów ma r! permutacje, aby uzyskać liczbę kombinacji n elementów wziętych r na raz, dzielimy liczbę permutacji n elementów wziętych r na raz przez r !. Mamy zatem:

nCr = nPr = n! /! (n - r)! r!

Istnieje wiele innych wariantów właśnie przedstawionych zasad liczenia, niektóre z nich zostaną przedstawione w ćwiczeniach z rozdziału 5. W celu dalszego rozwoju tych technik liczenia zalecamy dowolny dyskretny podręcznik matematyki

Elementy teorii prawdopodobieństwa

Opierając się na zasadach liczenia przedstawionych wcześniej, możemy teraz wprowadzić teorię prawdopodobieństwa. Po pierwsze, w sekcji 1 rozważamy kilka podstawowych definicji, takich jak pojęcie, czy dwa lub więcej zdarzeń jest od siebie niezależnych. W sekcji 2 pokazujemy, jak wnioskować o wyjaśnieniach dla poszczególnych zestawów danych. To pozwoli nam rozważyć kilka przykładów wnioskowania probabilistycznego w sekcji 3 i twierdzenia Bayesa w sekcji 4.

Przykładowa przestrzeń, prawdopodobieństwa i niezależność

Następujące definicje, będące podstawą teorii prawdopodobieństwa, zostały po raz pierwszy sformalizowane przez francuskiego matematyka Laplace'a (1816) na początku XIX wieku. Jak wspomniano we wstępie do rozdziału 5, Laplace był w trakcie tworzenia rachunku różniczkowego do hazardu!

DEFINICJA

WYDARZENIE ELEMENTARNE

Wydarzenie elementarne lub atomowe to wydarzenie lub wydarzenie, które nie może składać się z innych zdarzeń.

WYDARZENIE, E

Wydarzenie to zestaw elementarnych zdarzeń.

PRZESTRZEN PRÓBNA, S

Zbiór wszystkich możliwych wyników zdarzenia E to przykładowa przestrzeń S lub wszechświat dla tego zdarzenia.

PRAWDOPODOBIEŃSTWO,

Prawdopodobieństwo zdarzenia E w przestrzeni próbnej S jest stosunkiem liczby elementów w E do całkowitej liczby możliwych wyników w przestrzeni próbnej S z E.

p (E) = | E | / | S |.

Na przykład, jakie jest prawdopodobieństwo, że 7 lub 11 są wynikiem rzutu dwoma uczciwymi kośćmi? Najpierw określamy przestrzeń próbki dla tej sytuacji. Stosując zasadę mnożenia liczenia, każda kość ma 6 wyników, więc całkowity zestaw wyników dwóch kości wynosi 36. Liczba kombinacji dwóch kości, które mogą dać 7, wynosi 1,6; 2,5; 3,4; 4,3; 5,2; oraz 6,1–6 łącznie. Prawdopodobieństwo wyrzucenia 7 wynosi zatem 6/36 = 1/6. Liczba kombinacji dwóch kości, które mogą dać 11, wynosi 5,6; 6,5 - lub 2, a prawdopodobieństwo wyrzucenia 11 wynosi 2/36 = 1/18. Wykorzystując właściwość addytywną różnych wyników, istnieje prawdopodobieństwo 1/6 + 1/18 lub 2/9 wyrzucenia 7 lub 11 z dwoma uczciwymi kośćmi. W tym przykładzie z 7/11 dwa zdarzenia otrzymują 7, a 11. 11. Elementarnymi zdarzeniami są odmienne wyniki rzutu dwiema kostkami. Zatem zdarzenie 7 składa się z sześciu zdarzeń atomowych (1,6), (2,5), (3,4), (4,3), (5,2) i (6,1) . Pełna przestrzeń próbki to połączenie wszystkich trzydziestu sześciu możliwych zdarzeń atomowych, zestawu wszystkich par, które wynikają z rzutu kostką. Jak wkrótce zobaczymy, ponieważ zdarzenia uzyskania 7 i otrzymania 11 nie mają wspólnych zdarzeń atomowych, są one niezależne, a prawdopodobieństwo ich sumy (unii) jest tylko sumą ich indywidualnych prawdopodobieństw. W drugim przykładzie, ile czterech rodzajów kart można rozdać we wszystkich możliwych układach pokera na kartę fivecard? Po pierwsze, zestaw wydarzeń atomowych, które składają się na pełną przestrzeń wszystkich rąk pokera na kartę fivecard, to kombinacja 52 kart pobranych po 5 na raz. Aby uzyskać łączną liczbę czterech kart tego samego rodzaju, stosujemy zasadę mnożenia. Mnożymy liczbę kombinacji 13 kart branych po 1 na raz (liczbę różnych rodzajów kart: as, 2, 3 ..., król) razy liczbę sposobów wybrania wszystkich czterech kart tego samego rodzaju (kombinacja 4 kart pobranych 4 na raz) razy liczba możliwych innych kart, które wypełniają układ 5 kart (pozostało 48 kart). Zatem prawdopodobieństwo czterokartowego układu pokera wynosi:

$({}_{13}C_1 \times {}_4C_4 \times {}_{48}C_1) / {}_{52}C_5 = 13 \times 1 \times 48/2\,598{,}960 \approx 0{,}00024$

Kilka wyników wynika bezpośrednio z właśnie wykonanych definicji. Po pierwsze, prawdopodobieństwo dowolnego zdarzenia E z przestrzeni próbki S wynosi:

$$0 \le p(E) \le 1, \text{ where } E \subseteq S$$

Drugim wynikiem jest to, że suma prawdopodobieństw wszystkich możliwych wyników w S wynosi 1. Aby to zobaczyć, należy zauważyć, że definicja przestrzeni próby S wskazuje, że składa się ona z połączenia wszystkich pojedynczych zdarzeń E w problemie. Jako trzeci wynik definicji należy zauważyć, że prawdopodobieństwo dopełnienia zdarzenia wynosi:

$$p(\overline{E}) = (|S| - |E|) \ / \ |S| \ = \ (|S| \ / \ |S|) - \ (|E| \ / \ |S|) = 1 - \ p(E)$$

Uzupełnieniem wydarzenia jest ważny związek. Wiele razy łatwiej jest ustalić prawdopodobieństwo wystąpienia zdarzenia w zależności od jego nieistnienia, na przykład określenie prawdopodobieństwa, że co najmniej jeden element losowo generowanego ciągu bitów o długości n wynosi 1. Dopełnieniem jest to, że wszystkie bity w ciągu są równe 0, z prawdopodobieństwem 2-n, przy n długości łańcucha. Zatem prawdopodobieństwo pierwotnego zdarzenia wynosi 1 - 2-n. Wreszcie, na podstawie prawdopodobieństwa uzupełnienia zestawu zdarzeń obliczamy prawdopodobieństwo, gdy nie wystąpi żadne zdarzenie, czasami określane jako sprzeczne lub fałszywe zdanie:

$$p(\{\}) = 1 - p(\overline{\{\}}) = 1 - p(S) = 1 - 1 = 0, \text{ or alternatively,}$$
$$= |\{\}| / |S| = 0 / |S| = 0$$

Ostateczną istotną zależność, prawdopodobieństwo połączenia dwóch zbiorów zdarzeń, można ustalić na podstawie zasady liczenia przedstawionej wcześniej, mianowicie dla dowolnych dwóch zbiorów A i B: $| A \cup B | = | A | + | B | - | A \cap B |$. Z tej zależności możemy określić prawdopodobieństwo zjednoczenia dowolnych dwóch zbiorów pobranych z przestrzeni próbnej S:

$p(A \cup B) = |A \cup B| / |S| = (|A| + |B| - |A \cap B|) / |S| = |A| / |S| + |B| / |S| - |A \cap B| / |S| = p(A) + p(B) - p(A \cap B)$

Oczywiście wynik ten można rozszerzyć na sumę dowolnej liczby zbiorów, zgodnie z przedstawioną wcześniej zasadą włączenia / wyłączenia. Już przedstawiliśmy przykład określania prawdopodobieństwa połączenia dwóch zestawów: Prawdopodobieństwo wyrzucenia 7 lub 11 za pomocą dwóch uczciwych kości. W tym przykładzie właśnie zaprezentowana formuła została zastosowana z prawdopodobieństwem par kostek, które dały 7 rozłączenia z par kostek, które dały 11. Możemy również użyć tej formuły w bardziej ogólnym przypadku, gdy zestawy nie są rozłączne . Załóżmy, że chcieliśmy ustalić, rzucając dwiema rzetelnymi kośćmi, prawdopodobieństwo rzutu 8 lub parą tej samej liczby. Po prostu obliczalibyśmy prawdopodobieństwo tego związku, gdy istnieje jedno zdarzenie elementarne - (4,4) - czyli przecięcie obu pożądanych zdarzeń końcowych.

Następnie rozważamy prawdopodobieństwo wystąpienia dwóch niezależnych zdarzeń. Załóżmy, że jesteś graczem w czteroosobową grę karcianą, w której wszystkie karty są równo rozłożone. Jeśli nie masz królowej pik, możesz dojść do wniosku, że każdy z pozostałych graczy ma ją z prawdopodobieństwem 1/3. Podobnie można stwierdzić, że każdy gracz ma asa kier z prawdopodobieństwem 1/3 i że każdy gracz ma obie karty z prawdopodobieństwem 1/3 × 1/3 lub 1/9. W tej sytuacji założyliśmy, że zdarzenia związane z uzyskaniem tych dwóch kart są niezależne, chociaż jest to tylko w przybliżeniu prawda. Formalizujemy tę intuicję za pomocą definicji.

DEFINICJA

NIEZALEŻNE WYDARZENIA

Dwa zdarzenia A i B są niezależne wtedy i tylko wtedy, gdy prawdopodobieństwo ich wystąpienia jest równe iloczynowi ich wystąpienia indywidualnie. Ta zależność niezależności jest wyrażona:

$p (A \cap B) = p (A) * p (B)$

Czasami używamy równoważnego zapisu p (s, d) dla p (s ∩ d). Wyjaśniamy pojęcie niezależności w kontekście prawdopodobieństw warunkowych. Ponieważ opis przedstawionej właśnie niezależności zdarzeń jest relacją tylko wtedy i tylko wtedy, możemy ustalić, czy dwa zdarzenia są niezależne, opracowując ich relacje probabilistyczne. Rozważ sytuację, w której losowo generowane są ciągi bitów o długości cztery. Chcemy wiedzieć, czy zdarzenie ciągu bitów zawierającego parzystą liczbę 1s jest niezależne od zdarzenia, w którym ciąg bitów kończy się na 0. Przy zastosowaniu zasady mnożenia,

każdy bit ma 2 wartości, w sumie 24 = 16 ciągi bitów z długość 4. Istnieje 8 ciągów bitów o długości cztery, które kończą się cyfrą 0: {1110, 1100, 1010, 1000, 0010, 0100, 0110, 0000}. Istnieje również 8 ciągów bitowych, które mają parzystą liczbę 1: {1111, 1100, 1010, 1001, 0110, 0101, 0011, 0000}. Liczba ciągów bitów, które mają zarówno parzystą liczbę 1, jak i kończą 0, wynosi 4: {1100, 1010, 0110, 0000}. Od tego czasu te dwa wydarzenia są niezależne

p ({parzysta liczba 1s} ∩ {koniec z 0}) = p ({parzysta liczba 1s}) x p ({koniec z 0}) 4/16 = 8/16 x 8/16 = 1/4

Rozważ ten sam przykład losowo generowanych ciągów bitów o długości cztery. Czy dwa następujące zdarzenia są niezależne: ciągi bitów mają parzystą liczbę 1, a ciągi bitów kończą się na 1? Gdy dwa lub więcej zdarzeń nie jest niezależnych, tzn. Prawdopodobieństwo, że jedno zdarzenie wpłynie na prawdopodobieństwo innych, wymaga pojęcia warunkowego prawdopodobieństwa wypracowania ich relacji. Przed zamknięciem tego rozdziału zauważamy, że możliwe są inne systemy aksjomatów wspierające podstawy teorii prawdopodobieństwa, na przykład jako rozszerzenie rachunku zdań. Jako przykład podejścia opartego na zestawie, rosyjski matematyk Kołmogorow (1950) zaproponował wariant następujących aksjomatów: równoważne z naszymi definicjami. Z tych trzech aksjomatów Kołmogorow systematycznie konstruował wszystkie teorie prawdopodobieństwa.

1. Prawdopodobieństwo zdarzenia E w przestrzeni próbki S wynosi od 0 do 1, tj. 0 ≤ p (E) ≤ 1.

2. Gdy suma wszystkich E = S, p (S) = 1, a p (S^{-}) = 0.

3. Prawdopodobieństwo połączenia dwóch zbiorów zdarzeń A i B wynosi:

p (A ∪ B) = p (A) + p (B) - p (A ∩ B)

Wnioskowanie probabilistyczne: przykład

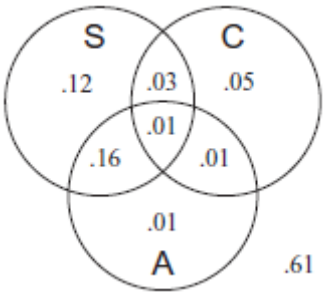
Pokazujemy teraz przykłady uzasadnienia z właśnie przedstawionymi pomysłami. Załóżmy, że jedziesz autostradą międzystanową i zdajesz sobie sprawę, że stopniowo zwalniasz z powodu zwiększonego natężenia ruchu. Zaczynasz szukać możliwych wyjaśnień spowolnienia. Czy to może być budowa dróg? Czy był wypadek? Wszystko, co wiesz, to to, że zwalniasz. Ale poczekaj! Masz dostęp do statystyk stanu autostrad, a dzięki nowemu samochodowemu GUI i systemowi wnioskowania możesz pobrać na komputer swojego samochodu odpowiednie informacje statystyczne. Okej, więc masz dane; co możesz z nimi zrobić? W tym przykładzie zakładamy, że mamy trzy parametry prawda lub fałsz (zdefiniujemy ten parametr jako boolowską zmienną losową w rozdziale 5.2.4). Po pierwsze, czy ruch - a ty - zwalniasz, czy nie. Ta sytuacja będzie oznaczona S, z przypisaniem t lub f. Po drugie, istnieje prawdopodobieństwo, że nastąpi wypadek, A, z przypisaniami t lub f. Wreszcie, prawdopodobieństwo, że w tym czasie istnieje budowa drogi, C; ponownie albo t albo f. Możemy wyrazić te relacje dla ruchu na autostradzie międzystanowej, dzięki naszemu samochodowemu systemowi pobierania danych, w tabeli

S	C	A	p
t	t	t	0.01
t	t	f	0.03
t	f	t	0.16
t	f	f	0.12
f	t	t	0.01
f	t	f	0.05
f	f	t	0.01
f	f	f	0.61

Pozycje w tabeli są oczywiście interpretowane podobnie jak tabele prawdy, z tą różnicą, że w prawej kolumnie podano prawdopodobieństwo wystąpienia sytuacji po lewej stronie. Zatem trzeci rząd tabeli podaje prawdopodobieństwo spowolnienia ruchu i wypadku, ale bez konstrukcji wynoszącej 0,16:

$$S \cap \overline{C} \cap A = 0.16$$

Należy zauważyć, że opracowujemy nasz rachunek probabilistyczny w kontekście zbiorów zdarzeń. Rysunek pokazuje, w jaki sposób prawdopodobieństwa z tabeli można przedstawić za pomocą tradycyjnego diagramu Venna.



Równie dobrze moglibyśmy przedstawić tę sytuację jako probabilistyczne przypisanie prawdy zdań, w którym to przypadku ∩ zostałby zastąpiony przez ∧, a Tabela byłaby interpretowana jako wartości prawdy w połączeniu zdań. Następnie zauważamy, że suma wszystkich możliwych wyników wspólnego rozkładu tabeli wynosi 1,0; jest to, jak można się spodziewać, z aksjomatami prawdopodobieństwa przedstawionymi w rozdziale. Należy zauważyć, że rozwijamy nasz rachunek probabilistyczny w kontekście zbiorów zdarzeń.

Równie dobrze moglibyśmy przedstawić tę sytuację jako probabilistyczne przypisanie prawdy zdań, w którym to przypadku ∩ zostałoby zastąpione przez ∧, a Tabela byłaby zinterpretowana jako wartości prawdy w połączeniu zdań. Następnie zauważamy, że suma wszystkich możliwych wyników wspólnego rozkładu tabeli 5.1 wynosi 1,0; jest to zgodne z oczekiwaniami z aksjomatami prawdopodobieństwa przedstawionymi. Możemy również obliczyć prawdopodobieństwo dowolnego prostego lub złożonego zestawu zdarzeń. Na przykład możemy obliczyć prawdopodobieństwo spowolnienia ruchu S. Wartość dla wolnego ruchu wynosi 0,32, sumę pierwszych czterech wierszy tabeli 5.1; to znaczy wszystkie sytuacje, w których S = t. Czasami nazywa się to bezwarunkowym lub krańcowym prawdopodobieństwem powolnego ruchu, S. Proces ten nazywa się marginalizacją, ponieważ wszystkie prawdopodobieństwa inne niż powolny ruch są sumowane. Oznacza to, że rozkład zmiennej

można uzyskać, sumując wszystkie pozostałe zmienne ze wspólnego rozkładu zawierającego tę zmienną

W podobny sposób możemy obliczyć prawdopodobieństwo budowy C bez spowolnienia es.png zjawisko niezbyt częste w stanie Nowy Meksyk! Sytuację tę ujmuje p (C ∩ S) = t, jako suma piątej i szóstej linii tabeli 5.1 lub 0,06. Jeśli weźmiemy pod uwagę negację sytuacji C ∩ S, otrzymalibyśmy (stosując prawa deMorgan) p (C ∪ S). Obliczając prawdopodobieństwo zjednoczenia dwóch zbiorów, otrzymujemy:

0,16 + 0,12 + 0,01 + 0,61 + 0,01 + 0,03 + 0,16 + 0,12 - (0,16 + 0,12) = 0,94

I znowu całkowite prawdopodobieństwo C ∩ S i jego dopełniacza (negacji) wynosi 1,0.

Zmienne losowe

W teorii prawdopodobieństwa poszczególne prawdopodobieństwa są obliczane analitycznie za pomocą metod kombinatorycznych lub empirycznie przez próbkowanie populacji zdarzeń. Do tego momentu większość naszych prawdopodobieństw została ustalona analitycznie. Na przykład kostka ma sześć boków i dwie monety. Kiedy „kość" lub „moneta" jest „sprawiedliwa", mówimy, że każdy wynik, z rzutu lub rzutu, jest równie prawdopodobny. W rezultacie łatwo jest określić przestrzeń zdarzeń dla tych problematycznych sytuacji, które później nazywamy parametrycznymi. Bardziej interesujące rozumowanie probabilistyczne wynika jednak z próbkowania analizy sytuacji w rzeczywistym świecie wydarzeń. W tych sytuacjach często brakuje dobrze zdefiniowanej specyfikacji, która wspiera analityczne obliczanie prawdopodobieństw. Jest tak również, że niektóre sytuacje, nawet jeśli istnieją podstawy analityczne, są tak złożone, że czas i koszty obliczeń nie są wystarczające do deterministycznego obliczenia wyników probabilistycznych. W takich sytuacjach zwykle przyjmujemy empiryczną metodologię próbkowania. Co najważniejsze, zakładamy, że wszystkie wyniki eksperymentu nie są jednakowo prawdopodobne. Zachowujemy jednak podstawowe aksjomaty lub założenia, które przyjęliśmy w poprzednich sekcjach; mianowicie, że prawdopodobieństwo zdarzenia jest liczbą pomiędzy (i włącznie) 0 i 1 oraz, że zsumowane prawdopodobieństwa wszystkich wyników wynoszą 1. Zachowujemy również naszą zasadę dotyczącą prawdopodobieństwa zjednoczonych zbiorów zdarzeń. Definiujemy zmienną losową jako metodę uściślenia tego rachunku.

DEFINICJA

ZMIENNA LOSOWA

Zmienna losowa jest funkcją, której domeną jest przestrzeń próbki i zawiera ona zestaw wyników, najczęściej liczb rzeczywistych. Zamiast korzystać z przestrzeni zdarzeń specyficznych dla problemu, zmienna losowa pozwala mówić o prawdopodobieństwach jako wartościach liczbowych związanych z przestrzenią zdarzeń.

BOOLEAN, DISCRETE i CIĄGŁE ZMIENNE LOSOWE

Logiczna zmienna losowa jest funkcją od przestrzeni zdarzeń do {true, false} lub do podzbioru liczb rzeczywistych {0,0, 1,0}. Zmienna losowa typu boolean jest czasem nazywana próbą Bernoulliego. Dyskretna zmienna losowa, która zawiera boolowskie zmienne losowe jako podzbiór, to funkcja z przestrzeni próbnej na (policzalny podzbiór) liczb rzeczywistych w [0,0, 1,0]. Ciągła zmienna losowa ma w swoim zakresie zbiór liczb rzeczywistych. Przykład z wykorzystaniem dyskretnej zmiennej losowej w domenie Season, w której zdarzeniami atomowymi w sezonie są {wiosna, lato, jesień, zima}, przypisuje 0,75, powiedzmy, elementowi domeny Season = wiosna. W tej sytuacji mówimy p (Sezon = wiosna) = 0,75.

Przykładem losowej zmiennej typu boolean w tej samej domenie byłoby mapowanie p (Season = spring) = true. Większość przykładów probabilistycznych, które rozważamy, będzie dotyczyć dyskretnych zmiennych losowych. Innym przykładem zastosowania losowej zmiennej typu boolean byłoby obliczenie prawdopodobieństwa uzyskania 5 głów w 7 rzutach uczciwej monety. Byłaby to kombinacja 5 z 7 przewrotów będących główkami razy prawdopodobieństwo 1/2 główki do 5. potęgi razy prawdopodobieństwo 1/2 trafienia głów do 2. potęgi lub:

$_7C_5 \times (1/2)^5 \times (1/2)^2$

Ta sytuacja z rzutem monetą jest przykładem tak zwanego rozkładu dwumianowego. W rzeczywistości wynik każdej sytuacji, w której chcemy mierzyć sukcesy rw próbach n, gdzie p jest znanym prawdopodobieństwem sukcesu, można przedstawić jako:

$_nC_r \times p^r \times (1 - p)^{(n - r)}$

Ważnym naturalnym rozszerzeniem powiązania miar probabilistycznych ze zdarzeniami jest pojęcie oczekiwanego kosztu lub wypłaty za ten wynik. Na przykład możemy obliczyć przyszły zwrot z obstawiania określonych wartości pieniężnych z losowania karty lub zakręcenia kołem ruletki. Definiujemy oczekiwanie zmiennej losowej lub zdarzenia, np. (E):

DEFINICJA

OCZEKIWANIE ZDARZENIA

Jeżeli nagrodą za wystąpienie zdarzenia E, z prawdopodobieństwem p (E), jest r, a koszt nieistnienia zdarzenia, 1 - p (E), wynosi c, to oczekiwanie wynikające ze zdarzenia, np. (E), jest:

ex (E) = r x p (E) + c x (1 - p (E))

Załóżmy na przykład, że uczciwe koło ruletki ma liczby całkowite od 0 do 36 równomiernie rozmieszczone na rowkach koła. W grze każdy gracz kładzie 1 USD na dowolną wybraną przez siebie liczbę: jeśli koło zatrzyma się na wybranej liczbie, wygrywa 35 USD; w przeciwnym razie traci dolara. Nagroda za zwycięstwo wynosi 35 $; koszt straty 1 USD. Ponieważ prawdopodobieństwo wygranej wynosi 1/37, przegranej 36/37, oczekiwana wartość tego zdarzenia, np. (E), wynosi:

ex (E) = 35 (1/37) + (-1) (36/37) ≈ -0,027

W ten sposób gracz traci średnio około 0,03 USD na grę!

Tą część kończymy krótką dyskusją i podsumowaniem początków wartości prawdopodobieństw stosowanych w rozumowaniu stochastycznym. Jak wspomniano powyżej, w większości naszych dotychczasowych przykładów rozważaliśmy wartości probabilistyczne, które można ustalić na podstawie znanych sytuacji, takich jak przewrócenie uczciwej monety lub obrót uczciwego koła ruletki. Kiedy mamy takie sytuacje, możemy wyciągnąć wiele wniosków na temat aspektów przestrzeni prawdopodobieństwa, takich jak jej średnia, statystycznie miara prawdopodobieństwa i jak daleko od tej średniej wartości próbkowane zwykle się różnią, odchylenie standardowe wyników w tym domena. Tę dobrze rozumianą sytuację nazywamy parametrycznym podejściem do generowania przykładowej przestrzeni wyników. Podejście parametryczne jest uzasadnione w sytuacjach stochastycznych, w których istnieją a priori oczekiwania dotyczące struktury wyników prób eksperymentalnych. Naszym zadaniem jest „uzupełnienie” parametrów tej dobrze rozumianej sytuacji. Przykładem jest rzucie uczciwą monetą z wynikiem jako rozkład dwumianowy. Następnie możemy zbudować nasze oczekiwania dotyczące sytuacji na podstawie modelu dwumianowego dla możliwych wyników. Istnieje wiele zalet podejść parametrycznych. Po pierwsze, do skalibrowania oczekiwanych wyników potrzeba mniej punktów danych, ponieważ kształt krzywej wyników jest znany z góry. Kolejną zaletą jest to, że

często możliwe jest ustalenie z góry liczby wyników lub ilości danych szkoleniowych wystarczających do oszacowania prawdopodobieństwa jakości. W rzeczywistości w sytuacji parametrycznej, oprócz obliczenia średniej i odchylenia standardowego oczekiwań, możemy dokładnie określić, kiedy pewne punkty danych wykraczają poza normalne oczekiwania. Oczywiście wiele, jeśli nie większość, interesujących sytuacji nie przynosi wyraźnych oczekiwanych rezultatów. Jednym z przykładów jest na przykład rozumowanie diagnostyczne w medycynie. Drugim przykładem jest użycie i interpretacja wyrażenia w języku naturalnym. W przypadku języka dość często przyjmuje się nieparametryczne podejście do oczekiwań, próbkując dużą liczbę sytuacji, jak na przykład w korpusie językowym. Analizując zebrane przykłady użycia języka w gazetach, powiedzmy lub w rozmowach z działem pomocy technicznej dla wsparcia komputerowego, można wywnioskować znaczenie niejednoznacznych wyrażeń w tych domenach. Metodologię tę analizujemy, analizując możliwe relacje fonemów w rozdziale 5.3. Przy wystarczającej liczbie punktów danych wynikowy dyskretny rozkład w środowiskach nieparametrycznych można często wygładzić przez interpolację w celu zachowania ciągłości. Następnie można wywnioskować nowe sytuacje w kontekście utworzonej dystrybucji. Główną wadą metod nieparametrycznych jest to, że przy braku ograniczeń wynikających z wcześniejszych oczekiwań, często wymagana jest duża ilość danych szkoleniowych w celu kompensacji.

Warunkowe prawdopodobieństwo

Miary prawdopodobieństwa omówione do tego momentu w rozdziale 5 są często nazywane wcześniejszymi prawdopodobieństwami, ponieważ są one opracowywane przed uzyskaniem jakichkolwiek nowych informacji o oczekiwanych skutkach zdarzeń w konkretnej sytuacji. W niniejszej sekcji rozważamy warunkowe prawdopodobieństwo wystąpienia zdarzenia, to znaczy prawdopodobieństwo zdarzenia, biorąc pod uwagę pewne nowe informacje lub ograniczenie tego zdarzenia. Jak widać wcześniej w tym rozdziale, wcześniejsze prawdopodobieństwo uzyskania 2 lub 3 w rzucie rzetelnej kości jest sumą tych dwóch indywidualnych wyników podzieloną przez całkowitą liczbę możliwych wyników rzutu rzetelną kością lub 2 / 6. Wcześniejsze prawdopodobieństwo wystąpienia choroby to liczba osób z chorobą podzielona przez liczbę osób w danej dziedzinie. Przykładem warunkowego lub późniejszego prawdopodobieństwa jest sytuacja, gdy pacjent wchodzi do gabinetu lekarskiego z, na przykład, zestawem objawów, bólów głowy i nudności. Doświadczony lekarz będzie znał zestaw wcześniejszych oczekiwań dotyczących różnych chorób w oparciu o objawy, ale będzie chciał ustalić konkretną diagnozę dla tego pacjenta, który obecnie cierpi na bóle głowy i nudności. Aby doprecyzować te pomysły, tworzymy dwie ważne definicje.

DEFINICJA

PRZED PRAWDOPODOBIEŃSTWO

Wcześniejsze prawdopodobieństwo, zwykle bezwarunkowe prawdopodobieństwo zdarzenia, jest prawdopodobieństwem przypisywanym na podstawie całej wiedzy potwierdzającej jego wystąpienie lub nieobecność, to znaczy prawdopodobieństwo zdarzenia przed jakimkolwiek nowym dowodem. Wcześniejsze prawdopodobieństwo zdarzenia jest symbolizowane: p (zdarzenie).

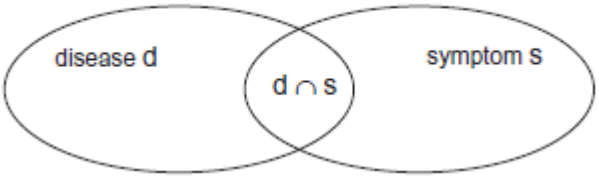
PRAWDOPODOBIEŃSTWO

Prawdopodobieństwo późniejsze (po fakcie), na ogół prawdopodobieństwo warunkowe zdarzenia, to prawdopodobieństwo zdarzenia, biorąc pod uwagę pewne nowe dowody. Późniejsze prawdopodobieństwo zdarzenia, biorąc pod uwagę niektóre dowody, jest symbolizowane: p (zdarzenie | dowód). Następnie rozpoczynamy prezentację twierdzenia Bayesa, którego ogólna postać znajduje się dalej. Idea popierająca Bayesa polega na tym, że prawdopodobieństwo nowej (tylnej) sytuacji hipotezy przy danym dowodzie może być postrzegane jako funkcja znanych prawdopodobieństw dla dowodów przy tej hipotezie. Można powiedzieć, że chcemy wyznaczyć

funkcję f, tak aby p (h | e) = f (p (e | h)). Zwykle chcemy ustalić wartość po lewej stronie tego równania biorąc pod uwagę, że często łatwiej jest obliczyć wartości po prawej stronie. Udowadniamy teraz twierdzenie Bayesa o jednym objawie i jednej chorobie. Opierając się na poprzednich definicjach, prawdopodobieństwo tylne osoby z chorobą d, z zestawu chorób D, z objawem lub dowodem, s, z zestawu objawów S, wynosi:

p (d s) = d ∩ s ⁄ s

Jak w sekcji 5.1, „|" otaczające zbiór to liczność lub liczba elementów w tym zbiorze. Prawa strona tego równania to liczba osób mających zarówno (przecięcie) chorobę d, jak i objawy podzielona przez całkowitą liczbę osób mających objawy. Rycina 5.2 przedstawia schemat Venna tej sytuacji. Rozwijamy prawą stronę



tego równania. Ponieważ przestrzeń próbki do wyznaczania prawdopodobieństwa licznika i mianownika jest taka sama, otrzymujemy:

p (d | s) = p (d ∩ s) / p (s).

Istnieje równoważny związek dla p (s | d); ponownie, patrz rysunek 5.2:

p (s | d) = p (s ∩ d) / p (d).

Następnie rozwiązujemy równanie p (s | d), aby określić wartość p (s ∩ d):

p (s ∩ d) = p (s | d) p (d).

Podstawiając ten wynik w poprzednim równaniu na p (d | s) tworzy regułę Bayesa dla jednej choroby i jednego objawu:

p (d s) = p (s d) p (d) / p (s)

Tak więc prawdopodobieństwo choroby z danym objawem w późniejszym okresie jest iloczynem prawdopodobieństwa objawu przy danej chorobie i prawdopodobieństwa choroby, znormalizowanym przez prawdopodobieństwo tego objawu. Następnie przedstawiamy regułę łańcucha, ważną technikę stosowaną w wielu domenach rozumowanie stochastyczne, szczególnie w przetwarzaniu języka naturalnego. Właśnie opracowaliśmy równania dla dowolnych dwóch zestawów A1 i A2:

p (A1 ∩ A2) = p (A1 | A2) p (A2) = p (A2 | A1) p (A1).

a teraz uogólnienie na wiele zbiorów Ai, zwane regułą łańcucha:

p (A1 ∩ A2 ∩ ... ∩ An) = p (A1) p (A2 | A1) p (A3 | A1 ∩ A2) .... p (An | ∩ Ai)

Robimy argument indukcyjny, aby udowodnić regułę łańcucha, rozważmy n-ty przypadek:

p (A1 ∩ A2 ∩ ... ∩ An-1 ∩ An) = p ((A1 ∩ A2 ∩ ... ∩ An-1) ∩ An),

Stosujemy regułę przecięcia dwóch zestawów, aby uzyskać:

p ((A1 ∩ A2 ∩ ... ∩ An-1) ∩ An) = p (A1 ∩ A2 ∩ ... ∩ An-1) p (An | A1 ∩ A2 ∩ ... ∩ An-1)

a następnie zmniejsz ponownie, biorąc pod uwagę, że:

p (A1 ∩ A2 ∩ ... An-1) = p ((A1 ∩ A2 ∩ ... ∩ An-2) ∩ An-1)

aż do osiągnięcia p (A1∩ A2), przypadek podstawowy, który już wykazaliśmy. Zamykamy tę sekcję kilkoma definicjami opartymi na zastosowaniu relacji reguły łańcucha. Najpierw redefiniujemy zdarzenia niezależne w kontekście prawdopodobieństw warunkowych, a następnie definiujemy zdarzenia warunkowo niezależne lub pojęcie, w jaki sposób zdarzenia mogą być od siebie niezależne, biorąc pod uwagę jakieś trzecie zdarzenie. Warunkowe prawdopodobieństwo. Miary prawdopodobieństwa omówione do tego momentu w rozdziale 5 są często nazywane wcześniejszymi prawdopodobieństwami, ponieważ są one opracowywane przed uzyskaniem jakichkolwiek nowych informacji o oczekiwanych skutkach zdarzeń w konkretnej sytuacji. W niniejszej sekcji rozważamy warunkowe prawdopodobieństwo wystąpienia zdarzenia, to znaczy prawdopodobieństwo zdarzenia, biorąc pod uwagę pewne nowe informacje lub ograniczenie tego zdarzenia. Jak widać wcześniej w tym rozdziale, wcześniejsze prawdopodobieństwo uzyskania 2 lub 3 w rzucie rzetelnej kości jest sumą tych dwóch indywidualnych wyników podzieloną przez całkowitą liczbę możliwych wyników rzutu rzetelną kością lub 2 / 6. Wcześniejsze prawdopodobieństwo wystąpienia choroby to liczba osób z chorobą podzielona przez liczbę osób w danej dziedzinie. Przykładem warunkowego lub późniejszego prawdopodobieństwa jest sytuacja, gdy pacjent wchodzi do gabinetu lekarskiego z, na przykład, zestawem objawów, bólów głowy i nudności. Doświadczony lekarz będzie znał zestaw wcześniejszych oczekiwań dotyczących różnych chorób w oparciu o objawy, ale będzie chciał ustalić konkretną diagnozę dla tego pacjenta, który obecnie cierpi na bóle głowy i nudności. Aby doprecyzować te pomysły, tworzymy dwie ważne definicje.

DEFINICJA

PRZED PRAWDOPODOBIEŃSTWO

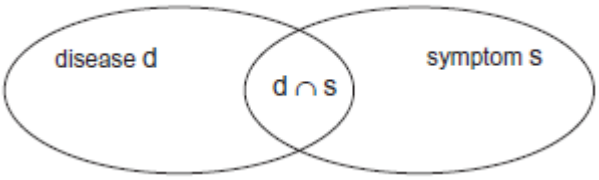
Wcześniejsze prawdopodobieństwo, zwykle bezwarunkowe prawdopodobieństwo zdarzenia, jest prawdopodobieństwem przypisywanym na podstawie całej wiedzy potwierdzającej jego wystąpienie lub nieobecność, to znaczy prawdopodobieństwo zdarzenia przed jakimkolwiek nowym dowodem. Wcześniejsze prawdopodobieństwo zdarzenia jest symbolizowane: p (zdarzenie).

PRAWDOPODOBIEŃSTWO

Prawdopodobieństwo późniejsze (po fakcie), na ogół prawdopodobieństwo warunkowe zdarzenia, to prawdopodobieństwo zdarzenia, biorąc pod uwagę pewne nowe dowody. Późniejsze prawdopodobieństwo zdarzenia, biorąc pod uwagę niektóre dowody, jest symbolizowane: p (zdarzenie | dowód). Następnie rozpoczynamy prezentację twierdzenia Bayesa, którego ogólna postać znajduje się w rozdziale 5.4. Idea popierająca Bayesa polega na tym, że prawdopodobieństwo nowej (tylnej) sytuacji hipotezy przy danym dowodzie może być postrzegane jako funkcja znanych prawdopodobieństw dla dowodów przy tej hipotezie. Można powiedzieć, że chcemy wyznaczyć funkcję f, tak aby p (h | e) = f (p (e | h)). Zwykle chcemy ustalić wartość po lewej stronie tego równania biorąc pod uwagę, że często łatwiej jest obliczyć wartości po prawej stronie. Udowadniamy teraz twierdzenie Bayesa o jednym objawie i jednej chorobie. Opierając się na poprzednich definicjach, prawdopodobieństwo tylne osoby z chorobą d, z zestawu chorób D, z objawem lub dowodem, s, z zestawu objawów S, wynosi:

p (d s) = d ∩ s ⁄ s

„|” otaczające zbiór to liczność lub liczba elementów w tym zbiorze. Prawa strona tego równania to liczba osób mających zarówno (przecięcie) chorobę d, jak i objawy podzielona przez całkowitą liczbę osób mających objawy. Rysunej przedstawia schemat Venna tej sytuacji. Rozwijamy prawą stronę



tego równania. Ponieważ przestrzeń próbki do wyznaczania prawdopodobieństwa licznika i mianownika jest taka sama, otrzymujemy:

$p(d \mid s) = p(d \cap s) / p(s)$.

Istnieje równoważny związek dla $p(s \mid d)$; ponownie, patrz rysunek powyżej:

$p(s \mid d) = p(s \cap d) / p(d)$.

Następnie rozwiązujemy równanie $p(s \mid d)$, aby określić wartość $p(s \cap d)$:

$p(s \cap d) = p(s \mid d)\, p(d)$.

Podstawiając ten wynik w poprzednim równaniu na $p(d \mid s)$ tworzy regułę Bayesa dla jednej choroby i jednego objawu:

$p(d\, s) = p(s\, d)\, p(d) / p(s)$

Tak więc prawdopodobieństwo choroby z danym objawem w późniejszym okresie jest iloczynem prawdopodobieństwa objawu przy danej chorobie i prawdopodobieństwa choroby, znormalizowanym przez prawdopodobieństwo tego objawu. Następnie przedstawiamy regułę łańcucha, ważną technikę stosowaną w wielu domenach rozumowanie stochastyczne, szczególnie w przetwarzaniu języka naturalnego. Właśnie opracowaliśmy równania dla dowolnych dwóch zestawów A_1 i A_2:

$p(A_1 \cap A_2) = p(A_1 \mid A_2)\, p(A_2) = p(A_2 \mid A_1)\, p(A_1)$.

a teraz uogólnienie na wiele zbiorów Ai, zwane regułą łańcucha:

$p(A_1 \cap A_2 \cap ... \cap A_n) = p(A_1)\, p(A_2 \mid A_1)\, p(A_3 \mid A_1 \cap A_2) .... p(A_n \mid \cap A_i)$

Robimy argument indukcyjny, aby udowodnić regułę łańcucha, rozważmy n-ty przypadek:

$p(A_1 \cap A_2 \cap ... \cap A_{n-1} \cap A_n) = p((A_1 \cap A_2 \cap ... \cap A_{n-1}) \cap A_n)$,

Stosujemy regułę przecięcia dwóch zestawów, aby uzyskać:

$p((A_1 \cap A_2 \cap ... \cap A_{n-1}) \cap A_n) = p(A_1 \cap A_2 \cap ... \cap A_{n-1})\, p(A_n \mid A_1 \cap A_2 \cap ... \cap A_{n-1})$

a następnie zmniejsz ponownie, biorąc pod uwagę, że:

$p(A_1 \cap A_2 \cap ... A_{n-1}) = p((A_1 \cap A_2 \cap ... \cap A_{n-2}) \cap A_{n-1})$

aż do osiągnięcia $p(A_1 \cap A_2)$, przypadek podstawowy, który już wykazaliśmy. Zamykamy tę sekcję kilkoma definicjami opartymi na zastosowaniu relacji reguły łańcucha. Najpierw redefiniujemy zdarzenia niezależne w kontekście prawdopodobieństw warunkowych, a następnie definiujemy zdarzenia warunkowo niezależne lub pojęcie, w jaki sposób zdarzenia mogą być od siebie niezależne, biorąc pod uwagę jakieś trzecie zdarzenie.

DEFINICJA

NIEZALEŻNE WYDARZENIA

Dwa zdarzenia A i B są od siebie niezależne wtedy i tylko wtedy, gdy $p(A \cap B) = p(A)\, p(B)$. Kiedy $p(B) \neq 0$ jest to takie samo, jak powiedzenie, że $p(A) = p(A \mid B)$. To znaczy, wiedząc, że B jest prawdziwe, nie wpływa na prawdopodobieństwo, że A jest prawdziwe.

ZDARZENIA WARUNKOWO NIEZALEŻNE

Mówi się, że dwa zdarzenia A i B są warunkowo niezależne od siebie, biorąc pod uwagę zdarzenie C wtedy i tylko wtedy, gdy $p((A \cap B) \mid C) = p(A \mid C)\, p(B \mid C)$.

W wyniku uproszczenia ogólnej reguły łańcucha oferowanej przez zdarzenia warunkowo niezależne można zbudować większe systemy stochastyczne przy mniejszych kosztach obliczeniowych; to znaczy, warunkowo niezależne zdarzenia upraszczają wspólne rozkłady. Przykład z naszej sytuacji na wolnym ruchu: załóżmy, że gdy zwalniamy, zauważamy pomarańczowe beczki kontroli ruchu wzdłuż drogi. Oprócz sugerowania, że przyczyną naszego spowolnienia jest teraz bardziej prawdopodobne budownictwo drogowe niż wypadek drogowy, obecność pomarańczowych beczek będzie miała własną miarę probabilistyczną. W rzeczywistości zmienne przedstawiające spowolnienie ruchu i obecność pomarańczowych beczek są warunkowo niezależne, ponieważ oba są spowodowane przez budowę drogi. Mówimy zatem, że zmienna konstrukcja drogi oddziela spowolnienie ruchu od pomarańczowych beczek. Ze względu na statystyczną efektywność uzyskaną dzięki warunkowej niezależności głównym zadaniem w budowie dużych stochastycznych systemów obliczeniowych jest rozbicie złożonego problemu na słabiej powiązane podproblemy. Relacje interakcji podproblemów są następnie kontrolowane przez różne relacje separacji warunkowej. Widzimy ten pomysł sformalizowany wraz z definicją d-separacji w rozdziale 9.3. Następnie, w rozdziale 5.3, przedstawiamy ogólną formę twierdzenia Bayesa i pokazujemy, w jaki sposób w złożonych sytuacjach obliczenia niezbędne do poparcia pełnego wnioskowania bayesowskiego mogą stać się trudne. Wreszcie w sekcji 5.4 przedstawiamy przykłady rozumowania opartego na bayesowskich miarach prawdopodobieństwa

Twierdzenie Bayesa

Wielebny Thomas Bayes był matematykiem i pastorem. Jego słynne twierdzenie zostało opublikowane w 1763 roku, cztery lata po jego śmierci. Jego artykuł zatytułowany „Esej na temat rozwiązania problemu w doktrynie szans" został opublikowany w Philosophical Transactions of Royal Society of London. Twierdzenie Bayesa wiąże przyczynę i skutek w taki sposób, że poprzez zrozumienie efektu możemy poznać prawdopodobieństwo jego przyczyn. W rezultacie twierdzenie Bayesa jest ważne zarówno dla ustalenia przyczyn chorób, takich jak rak, jak i przydatne do określenia wpływu niektórych konkretnych leków na tę chorobę.

Wprowadzenie

Jednym z najważniejszych wyników teorii prawdopodobieństwa jest ogólna postać twierdzenia Bayesa. Równanie Bayesa dla jednej choroby i jednego objawu. Aby pomóc utrzymać nasze relacje diagnostyczne w kontekście, zmieniamy nazwy zmiennych używanych wcześniej do wskazania poszczególnych hipotez, cześć, z zestawu hipotez, H i zestawu dowodów, E. Ponadto, rozważymy teraz zbiór indywidualnych hipotez cześć jako rozłączny i mając związek wszystkich cześć równy H.

$p(h_i \mid E) = (p(E \mid h_i) \times p(h_i)) / p(E)$

To równanie można odczytać: „Prawdopodobieństwo hipotezy przy danym zbiorze dowodów E wynosi… " Po pierwsze, zauważ, że mianownik p(E) po prawej stronie równania jest bardzo czynnikiem

normalizującym dla każdej hipotezy h_i z zestawu H. Często zdarza się, że twierdzenie Bayesa jest używane do ustalenia, która hipoteza z zestawu możliwych hipotez jest najsilniejszy, biorąc pod uwagę konkretny zestaw dowodów E. W tym przypadku często upuszczamy mianownik p(E), który jest identyczny dla wszystkich h_i, z oszczędnością potencjalnie dużego kosztu obliczeniowego. Bez mianownika stworzyliśmy maksymalną wartość a posteriori dla hipotezy:

$\arg\max (h_i)\ p(E \mid h_i)\ p(h_i)$

Odczytujemy to wyrażenie jako „Maksymalna wartość w całym h_i $p(E \mid h_i)$ $p(h_i)$". Opisane właśnie uproszczenie jest bardzo ważne dla rozumowania diagnostycznego, a także w przetwarzaniu języka naturalnego. Oczywiście, arg max lub hipoteza maksymalnego prawdopodobieństwa, nie jest już zmienną losową, jak zdefiniowano wcześniej. Następnie rozważ obliczenie mianownika p (E) w sytuacji, gdy cała przestrzeń próbki jest podzielona przez zestaw hipotez hi. Podział zestawu jest zdefiniowany jako podział tego zestawu na rozłączne nieprzekraczające się podzbiory, których połączenie tworzy cały zestaw. Zakładając, że zestaw hipotez podzieli całą przestrzeń próbki, otrzymujemy:

$p(E) = \Sigma_i\ p(E \mid h_i)\ p(h_i)$

Zależność tę wykazano, biorąc pod uwagę fakt, że zestaw hipotez hi tworzy podział pełnego zestawu dowodów E wraz z regułą prawdopodobieństwa przecięcia dwóch zbiorów. Więc:

$E = (E \cap h_1) \cup (E \cap h_2) \cup ... \cup (E \cap h_n)$

Ale przez uogólnioną zasadę unii zbiorów:

$p(E) = p((E \cap h_1) \cup (E \cap h2) \cup ... \cup (E \cap h_n))$

$= p(E \cap h_1) + p(E \cap h_2) + ... + p(E \cap h_n) - p(E \cap h_1 \cap E \cap h_2 \cap ... \cap h_n)$

$= p(E \cap h_1) + p(E \cap h_2) + ... + p(E \cap h_n)$

ponieważ zestaw hipotez dotyczy podziału E, a ich przecięcie jest puste. To obliczenie p (E) daje ogólną postać twierdzenia Bayesa, w której zakładamy, że zbiór hipotez, hi dzielenie zbioru dowodów E:

phie.png

$p(h_i|E)$ to prawdopodobieństwo, że h_i jest prawdziwe, biorąc pod uwagę dowody E.

$p(h_i)$ jest prawdopodobieństwem, że hi jest ogólnie prawdziwe.

$p(E|h_i)$ to prawdopodobieństwo zaobserwowania dowodu E, gdy hi jest prawdziwe.

n jest liczbą możliwych hipotez.

Twierdzenie Bayesa zapewnia sposób obliczenia prawdopodobieństwa hipotezy hi, biorąc pod uwagę konkretny dowód, biorąc pod uwagę tylko prawdopodobieństwa, z którymi dowód wynika z faktycznych przyczyn (hipotez). Jako przykład załóżmy, że chcemy zbadać dowody geologiczne w pewnym miejscu, aby sprawdzić, czy nadaje się ono do znalezienia miedzi. Musimy wiedzieć z góry prawdopodobieństwo znalezienia każdego zestawu minerałów i prawdopodobieństwo istnienia pewnych dowodów obecnych, gdy zostanie znaleziony konkretny minerał Następnie możemy użyć twierdzenia Bayesa, z dowodami znalezionymi w konkretnej lokalizacji, aby określić prawdopodobieństwo miedzi. Takie podejście stosuje PROSPECTOR, zbudowany na Uniwersytecie Stanforda i SRI International i wykorzystywany w badaniach minerałów (miedź, molibden i inne minerały). POSZUKIWACZ znalazł komercyjnie znaczące złoża minerałów w kilku lokalizacjach. Następnie przedstawiamy prosty numeryczny przykład demonstrujący twierdzenie Bayesa. Załóżmy, że wychodzisz na zakup samochodu. Prawdopodobieństwo, że udasz się do dilera 1, d_1, wynosi 0,2.

Prawdopodobieństwo przejścia do rozdającego 2, d_2, wynosi 0,4. Jesteś tylko trzema dealerami biorąc pod uwagę, a prawdopodobieństwo przejścia do trzeciego, d3, wynosi również 0,4. Przy d_1 prawdopodobieństwo zakupu określonego samochodu, a_1, wynosi 0,2; u dealera d_2 prawdopodobieństwo zakupu a_1 wynosi 0,4. Wreszcie u dilera d3 prawdopodobieństwo zakupu a1 wynosi 0,3. Załóżmy, że kupujesz samochód A1. Jakie jest prawdopodobieństwo, że kupiłeś go u dilera d_2? Po pierwsze, chcemy wiedzieć, biorąc pod uwagę, że kupiłeś samochód a1, że kupiłeś go od sprzedawcy d_2, tj. W celu ustalenia p (d_2 | a_1). Prezentujemy twierdzenie Bayesa w postaci zmiennej do wyznaczania p (d_2 | a_1), a następnie ze zmiennymi związanymi z sytuacją w przykładzie.

p (d_2 | a_1) = (p (a_1 | d_2) p (d_2)) / (p (a_1 | d_1) p (d_1) + p (a_1 | d_2) p (d_2) + p (a_1 | d_3) p (d_3))

= (0,4) (0,4) / ((0,2) (0,2) + (0,4) (0,4) + (0,4) (0,3))

= 0,16 / 0,32

= 0,5

Stosowanie twierdzenia Bayesa wiąże się z dwoma głównymi zobowiązaniami: po pierwsze, muszą być znane wszystkie prawdopodobieństwa dotyczące powiązania dowodów z różnymi hipotezami, a także relacje prawdopodobieństwa między dowodami. Po drugie, a czasem trudniejsze do ustalenia, wszystkie relacje między dowodami a hipotezami lub p (E | h_k) muszą być oszacowane lub próbkowane empirycznie. Przypomnijmy, że obliczenie p (E) dla ogólnej postaci twierdzenia Bayesa wymagało również, aby zestaw hipotez hi podzielił zbiór dowodów E. Zasadniczo, a zwłaszcza w takich dziedzinach, jak medycyna i przetwarzanie języka naturalnego, założenia tego podziału nie można z góry uzasadnić. Warto jednak zauważyć, że wiele sytuacji, które naruszają to założenie (że poszczególne dowody dzielą zestaw dowodów) zachowują się całkiem dobrze! Korzystanie z tego założenia partycji, nawet w sytuacjach, gdy jest to nieuzasadnione, nazywa się przy użyciu naiwnego Bayesa lub klasyfikatora Bayesa. W przypadku naiwnego Bayesa przyjmuje się, że dla hipotezy h_j:

$$p(E|h_j) \approx \prod_{i=1}^{n} p(e_i| h_j)$$

tzn. zakładamy, że dowody są niezależne, biorąc pod uwagę szczególną hipotezę. Wykorzystując twierdzenie Bayesa do ustalenia prawdopodobieństwa pewnej hipotezy hi, biorąc pod uwagę zbiór dowodów E, p (h_i | E), liczby po prawej stronie równania są często łatwo dostępne. Jest to szczególnie prawdziwe w porównaniu z uzyskaniem wartości dla lewej strony równania, tj. Bezpośrednim wyznaczeniem p (h_i | E). Na przykład, ponieważ populacja jest mniejsza, o wiele łatwiej jest ustalić liczbę pacjentów z zapaleniem opon mózgowych, którzy mają bóle głowy, niż określić odsetek osób cierpiących na bóle głowy z zapaleniem opon mózgowych. Co ważniejsze, w przypadku prostego przypadku pojedynczej choroby i pojedynczego objawu nie jest potrzebnych bardzo wiele liczb. Problemy zaczynają się jednak, gdy rozważymy wiele chorób h_i z dziedziny chorób H i wiele objawów en z zestawu E możliwych objawów. Kiedy rozważymy każdą chorobę z H i każdy objaw z E pojedynczo, mamy m x n środków do zebrania i zintegrowania. (Właściwie m x n prawdopodobieństw późniejszych plus m + n wcześniejszych prawdopodobieństw.) Niestety, nasza analiza wkrótce stanie się znacznie bardziej złożona. Do tego momentu każdy objaw rozpatrywaliśmy indywidualnie. W rzeczywistych sytuacjach rzadko występują pojedyncze objawy. Na przykład, gdy lekarz rozważa pacjenta, musi wziąć pod uwagę wiele kombinacji objawów. Potrzebujemy formy twierdzenia Bayesa, aby rozważyć każdą pojedynczą hipotezę h_i w kontekście połączenia wielu możliwych objawów e_i.

p (h_i | $e_1 \cup e_2 \cup ... \cup e_n$) = (p ($h_i$) p ($e_1 \cup e_2 \cup ... \cup e_n$ | h_i)) / p ($e_1 \cup e_2 \cup ... \cup e_n$)

Przy jednej chorobie i jednym objawie potrzebowaliśmy tylko m x n pomiarów. Teraz, dla każdej pary objawów e_i i ej oraz konkretnej hipotezy choroby cześć, musimy znać zarówno p ($e_i \cup e_j \mid h_i$), jak i p ($e_i \cup e_j$). Liczba takich par wynosi nx (n - 1), lub w przybliżeniu n^2, gdy n ma objawy w E. Teraz, jeśli chcemy użyć Bayesa, będzie około (mx n^2 prawdopodobieństwa warunkowe) + (n^2 prawdopodobieństwo symptomu) + (m prawdopodobieństwa choroby) lub około mx n^2 + n^2 + m informacji do zebrania. W realistycznym systemie medycznym z 200 chorobami i 2000 objawami wartość ta wynosi ponad 800 000 000! Jest jednak nadzieja. Jak omówiono, kiedy przedstawiliśmy warunkową niezależność, wiele z tych par symptomów będzie niezależnych, to znaczy p ($e_i \mid e_j$) = p (e_i). Niezależność oznacza oczywiście, że obecność e_j nie ma wpływu na prawdopodobieństwo e_i. Na przykład w medycynie większość objawów nie jest powiązanych, np. Wypadanie włosów i ból łokcia. Ale nawet jeśli tylko dziesięć procent naszych przykładowych objawów nie jest niezależnych, pozostaje jeszcze około 80 000 000 relacji do rozważenia. W wielu sytuacjach diagnostycznych musimy również radzić sobie z negatywnymi informacjami, np. Gdy pacjent nie ma objawów takich jak złe ciśnienie krwi. Wymagamy zarówno: p (e_i) = 1 - p (e_i) i p ($h_i \mid e_i$) = 1 - p ($h_i \mid e_i$).

Zauważamy również, że p ($e_i \mid h_i$) i p ($h_i \mid e_i$) nie są takie same i prawie zawsze będą miały różne wartości. Zależności te oraz unikanie kołowego rozumowania są ważne przy projektowaniu bayesowskich sieci wierzeń. Ostatnim problemem, który ponownie sprawia, że utrzymywanie statystyk złożonych systemów bayesowskich jest praktycznie niemożliwe, jest potrzeba odbudowania tabel prawdopodobieństwa, gdy zostaną odkryte nowe związki między hipotezami a zestawami dowodów. W wielu aktywnych obszarach badawczych, takich jak medycyna, nowe odkrycia zachodzą nieustannie. Argumentacja bayesowska wymaga pełnych i aktualnych prawdopodobieństw, w tym prawdopodobieństw łącznych, jeśli wnioski z nich mają być prawidłowe. W wielu domenach tak obszerne gromadzenie i weryfikacja danych nie są możliwe, a jeśli to możliwe, dość drogie. Tam, gdzie te założenia są spełnione, podejścia bayesowskie oferują korzyść matematycznie uzasadnionego rozwiązania problemu niepewności. Większość domen systemowych ekspertów nie spełnia tych wymagań i musi opierać się na podejściach heurystycznych, jak przedstawiono w Rozdziale 8. Ponadto, z powodu problemów ze złożonością wiemy, że nawet dość potężne komputery nie mogą stosować pełnych technik bayesowskich do skutecznego rozwiązywania problemów w czasie rzeczywistym. Kończymy ten rozdział dwoma przykładami, które pokazują, jak podejście bayesowskie może działać w celu uporządkowania relacji hipotezy / dowodów. Ale najpierw musimy zdefiniować probabilistyczne maszyny / akceptory stanów skończonych

Zastosowania metodologii stochastycznej

W tej sekcji przedstawiamy dwa przykłady, które wykorzystują miary prawdopodobieństwa do uzasadnienia interpretacji niejednoznacznych informacji. Po pierwsze, definiujemy ważne narzędzie do modelowania oparte na maszynie stanu skończonego z Sekcji 3.1, probabilistycznej maszynie stanu skończonego.

DEFINICJA

PROBABILISTYCZNA MASZYNA SKOŃCZONA

Probabilistyczna maszyna stanów skończonych jest maszyną stanów skończonych, w której następną funkcją stanu jest rozkład prawdopodobieństwa dla pełnego zestawu stanów maszyny.

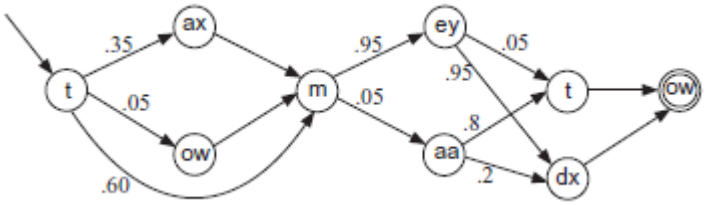
PROBABILISTYCZNY AKCEPTOR STANU SKOŃCZONEGO

Probabilistyczna maszyna stanów skończonych jest akceptorem, gdy jeden lub więcej stanów jest wskazanych jako stany początkowe, a jeden lub więcej jako stany akceptacji. Można zauważyć, że te dwie definicje są prostymi rozszerzeniami stanu skończonego i obecnością maszyn Moore. Dodatkiem

do niedeterminizmu jest to, że funkcja następnego stanu nie jest już funkcją w ścisłym tego słowa znaczeniu. Oznacza to, że nie ma unikalnego stanu zakresu dla każdej wartości wejściowej i każdego stanu domeny. Zamiast tego, dla każdego stanu, funkcja następnego stanu jest rozkładem prawdopodobieństwa we wszystkich możliwych kolejnych stanach

Jak zatem wymawia się „Tomato”?

Rysunek poniżej przedstawia probabilistyczny akceptor stanu skończonego, który reprezentuje różne wymowy słowa „tomato”. Szczególnie dopuszczalna wymowa słowa tomato to



charakteryzuje się ścieżką od stanu początkowego do stanu akceptacji. Wartości dziesiętne, które oznaczają łuki na wykresie, reprezentują prawdopodobieństwo, że głośnik wykona to konkretne przejście w maszynie stanu. Na przykład 60% wszystkich głośników w tym zestawie danych idzie bezpośrednio z t na m, bez produkowania żadnej samogłoski pomiędzy. Oprócz scharakteryzowania różnych sposobów, w jakie ludzie w bazie wymowy wymawiają słowo „pomidor”, model ten może być wykorzystany do interpretacji niejednoznacznych zbiorów fonemów. Odbywa się to poprzez sprawdzenie, jak dobrze fonemy dopasowują możliwe ścieżki przez automaty stanów powiązanych słów. Ponadto, biorąc pod uwagę częściowo utworzone słowo, maszyna może spróbować ustalić ścieżkę probabilistyczną, która najlepiej uzupełnia to słowo. W drugim przykładzie, również zaadaptowanym przez Jurafsky'ego i Martina, rozważamy problem rozpoznawania fonemów, często nazywany dekodowaniem. Załóżmy algorytm rozpoznawanie fonemów który zidentyfikował telefon ni (jak w „kolanie”), który pojawia się tuż po rozpoznanym słowie (telefon) I, i chcemy skojarzyć ni ze słowem lub pierwszą częścią słowa. W tym przypadku mamy do pomocy korpusy językowe, korpusy Brown i Switchboard. Korpus Browna to zbiór słów z 500 napisanych słów teksty, w tym gazety, powieści i pisma akademickie zebrane na Uniwersytecie Browna w latach 60. XX wieku (Kucera i Francis 1967, Francis 1979). Korpus centrali to 1,4 miliona słów w rozmowach telefonicznych. Korpusy te zawierają łącznie około 2 500 000 słów, które pozwalają nam próbkować zarówno informacje pisane, jak i mówione. Istnieje wiele sposobów identyfikowania najbardziej prawdopodobnego słowa kojarzonego z telefonem ni. Najpierw możemy ustalić, które słowo, z tym telefonem jako pierwsze, jest najczęściej używane. Tabela pierwszA przedstawia surowe częstotliwości tych słów wraz z prawdopodobieństwem ich wystąpienia, to znaczy częstotliwości słowa podzielonej przez całkowitą liczbę słów w tych połączonych ciałach. (Ta tabela została zaadaptowana przez Jurafsky'ego i Martina (2009); patrz ich książka dla uzasadnienia, że „the” należy do tej kolekcji). Z tych danych, słowo „the” wydaje się być pierwszym wyborem do dopasowania ni. Następnie zastosujemy formę twierdzenia Bayesa, którą przedstawiliśmy w poprzednim rozdziale. Nasza druga próba analizy telefonu po I polega na uproszczeniu tej formuły, która ignoruje jej mianownik:

$$p(\text{word} \mid [\text{ni}]) \propto p([\text{ni}] \mid \text{word}) \times p(\text{word})$$

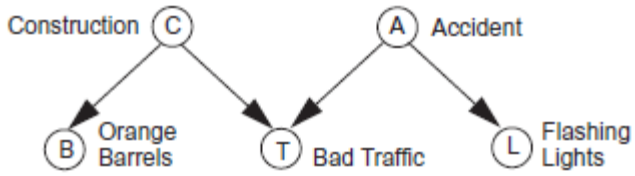
word	frequency	probability
knee	61	.000024
the	114834	.046
neat	338	.00013
need	1417	.00056
new	2625	.001

word	p([ni] | word)	p(word)	p([ni] | word) x p(word)
new	0.36	0.001	0.00036
neat	0.52	0.00013	0.000068
need	0.11	0.00056	0.000062
knee	1.0	0.000024	0.000024
the	0.0	0.046	0.0

Wyniki tego obliczenia, uporządkowane od najbardziej zalecanych do najmniej, znajdują się w tabeli drugiej i wyjaśniają, dlaczego p (ni | the) jest niemożliwe. Wyniki tabeli drugiej sugerują również, że new jest najbardziej prawdopodobnym słowem do dekodowania ni. Ale nowa kombinacja dwóch słów wydaje się nie mieć większego sensu, podczas gdy inne kombinacje, takie jak potrzebuję. Częścią problemu w tej sytuacji jest to, że wciąż rozważamy na poziomie telefonu, to znaczy określając prawdopodobieństwo p (ni | new). W rzeczywistości istnieje prosty sposób rozwiązania tego problemu, a mianowicie poszukiwanie wyraźnych kombinacji dwóch słów w korpusie. Zgodnie z tym tokiem rozumowania okazuje się, że potrzebuję o wiele bardziej prawdopodobnej pary następujących po sobie słów niż jestem nowy lub jakakolwiek inna kombinacja I-słów. Metodologia wyprowadzania prawdopodobieństw z par lub potrójnych kombinacji słów w ciałach nazywa się analizą n-gramową. Używając dwóch słów, używaliśmy bigramsów, z trzema trygramami.

Rozszerzenie przykładu drogi / ruchu

Ponownie przedstawiamy i rozszerzamy przykład z sekcji wcześniejszej. Załóżmy, że jedziesz autostradą międzystanową i zdajesz sobie sprawę, że stopniowo zwalniasz z powodu zwiększonego natężenia ruchu. Zaczynasz szukać możliwych wyjaśnień spowolnienia. Czy to może być budowa dróg? Czy był wypadek? Być może istnieją inne możliwe wyjaśnienia. Po kilku minutach natkniesz się na pomarańczowe beczki na poboczu drogi, które zaczynają odcinać zewnętrzny pas ruchu. W tym momencie stwierdzasz, że najlepszym wyjaśnieniem spowolnienia jest budowa dróg. Jednocześnie alternatywna hipoteza wypadku została wyjaśniona. Podobnie, gdybyś widział migające światła w odległości, na przykład z samochodu policyjnego lub karetki pogotowia, najlepszym wyjaśnieniem, biorąc pod uwagę ten dowód, byłby wypadek drogowy, a budowa drogi byłaby wyjaśniona. Wyjaśnienie hipotezy nie oznacza, że nie jest to już możliwe. Raczej w kontekście nowych dowodów jest to po prostu mniej prawdopodobne. Rycina 5.4 przedstawia opis bayesowski tego, co właśnie widzieliśmy. budowa dróg jest skorelowana z pomarańczowymi beczkami i złym ruchem.



Podobnie wypadek koreluje z migającymi światłami i złym ruchem. Analizujemy rysunek 5.4 i budujemy wspólny rozkład prawdopodobieństwa dla budowy drogi i złych relacji drogowych. Upraszczamy obie te zmienne, aby były prawdziwe (t) lub fałszywe (f) i przedstawiają rozkład prawdopodobieństwa w tabeli

C is true = .5

C	T	p
t	t	.3
t	f	.2
f	t	.1
f	f	.4

T is true = .4

Zauważ, że jeśli konstrukcja jest f, prawdopodobnie nie będzie dużego ruchu, a jeśli jest t, to prawdopodobnie jest zły ruch. Należy również zauważyć, że prawdopodobieństwo budowy drogi na autostradzie międzystanowej C = true wynosi 0,5, a prawdopodobieństwo złego ruchu, T = true wynosi 0,4 (to jest Nowy Meksyk!). Następnie rozważamy zmianę prawdopodobieństwa budowy drogi, biorąc pod uwagę fakt, że mamy zły ruch, lub p (C | T) lub p (C = t | T = t).

p (C | T) = p (C = t, T = t) / (p (C = t, T = t) + p (C = f, T = t)) = .3 / (.3 +. 1) = 0,75

Zatem teraz, z prawdopodobieństwem budowy drogi wynoszącym 0,5, biorąc pod uwagę fakt, że ruch jest rzeczywiście zły, prawdopodobieństwo budowy drogi wzrasta do 0,75. Prawdopodobieństwo to wzrośnie jeszcze bardziej dzięki obecności pomarańczowych beczek, co wyjaśnia hipotezę wypadku. Oprócz wymogu posiadania wiedzy lub pomiarów dla któregokolwiek z naszych średnic znajdujących się w określonym stanie, musimy również zająć się, jak wspomniano w sekcji 5.4.1, kwestiami złożoności. Rozważ obliczenie prawdopodobieństwa łącznego wszystkich parametrów z powyższego rysunku (przy użyciu reguły łańcucha i uporządkowanej topologicznie kolejności zmiennych):

p (C, A, B, T, L) = p (C) xp (A | C) xp (B | C, A) xp (T | C, A, B) xp (L | C, A, B , T)

Ten wynik jest ogólnym rozkładem miar prawdopodobieństwa, który jest zawsze prawdziwy. Koszt wytworzenia tej wspólnej tabeli prawdopodobieństwa jest wykładniczy pod względem liczby zaangażowanych parametrów, w tym przypadku wymaga tabeli o wielkości 25 lub 32. Rozważamy oczywiście problem zabawki z tylko pięcioma parametrami. Sytuacja o interesujących rozmiarach, z trzydziestoma lub więcej parametrami, wymaga wspólnej tabeli dystrybucji zawierającej około miliarda elementów. Jak zobaczymy w rozdziale 9.3, bayesowskie sieci przekonań i separacja d dają nam dalsze narzędzia do rozwiązania tej złożoności reprezentacyjnej i obliczeniowej

Epilog

Gry losowe pochodzą przynajmniej z cywilizacji greckiej i rzymskiej. Jednak dopiero w czasach europejskiego renesansu rozpoczęła się matematyczna analiza teorii prawdopodobieństwa. Jak zauważono w rozdziale, rozumowanie probabilistyczne rozpoczyna się od ustalenia zasad liczenia i

kombinatoryki. W rzeczywistości jedną z pierwszych „maszyn" kombinatorycznych przypisuje się Ramonowi Llullowi, hiszpańskiemu filozofowi i franciszkańskiemu mnichowi, który stworzył swoje urządzenie, które ma automatycznie wyliczać atrybuty Boga z zamiarem nawrócenia pogan . Pierwszą publikację o prawdopodobieństwach, De Ratiociniis Ludo Aleae, napisał Christian Huygens. Huygens opisuje wcześniejsze wyniki Blaise Pascal, w tym metodykę obliczania prawdopodobieństw, a także warunkowe prawdopodobieństwa. Pascal skupił się zarówno na „obiektywnej" analizie świata gier, jak i na bardziej „subiektywnej" analizie systemów przekonań, w tym na istnieniu Boga. Definicje prawdopodobieństw opierają się na formalizmie zaproponowanym przez Francuski matematyk Pierre Simon Laplace. Książka Laplace'a Theorie Analytique des Probabilitees (1816) dokumentuje to podejście. Praca Laplace'a była oparta na wcześniejszych wynikach opublikowanych przez Gotloba Leibnitza i Jamesa Bernoulli. Thomas Bayes był matematykiem i pastorem. Jego słynne twierdzenie zostało opublikowane w 1764 roku, po jego śmierci. Jego artykuł zatytułowany Esej na temat rozwiązania problemu w doktrynie szans został opublikowany w Philosophical Transactions of Royal Society of London. Jak na ironię, twierdzenie Bayesa nigdy nie jest wyraźnie określone w tym artykule, chociaż tak jest tam! Bayes prowadzi również obszerną dyskusję na temat „rzeczywistości" miar statystycznych. Badania Bayesa były częściowo motywowane do odpowiedzi na filozoficzny sceptycyzm szkockiego filozofa Davida Hume'a. Odrzucenie przez Hume'a związku przyczynowego zniszczyło wszelkie podstawy do argumentów poprzez cuda o istnieniu Boga. W rzeczywistości, w artykule z 1763 r. Przedstawionym Brytyjskiemu Towarzystwu Królewskiemu, minister Richard Price wykorzystał twierdzenie Bayesa, aby wykazać, że istnieją dobre dowody na korzyść cudów opisanych w Nowym Testamencie. Matematycy z początku XX wieku, w tym Fisher (1922), Popper (1959) i Carnap (1948), ukończyli podstawy współczesnej teorii prawdopodobieństwa, kontynuując „subiektywne / obiektywne" debaty na temat natury prawdopodobieństw. Kołmogorow (1950, 1965) aksjatyzował podstawy rozumowania probabilistycznego (patrz rozdział 5.2.1). Możliwe jest przedstawienie tematu wnioskowania probabilistycznego z kilku punktów widzenia. Dwa najbardziej popularne podejścia opierają się na rachunku zdań i teorii mnogości. Dla rachunku zdań, rozdział 2.1, twierdzeniom przypisuje się wartość ufności lub probabilistyczną wartość prawdy w przedziale [0,0, 1,0]. Podejście to oferuje naturalne rozszerzenie semantyki rachunku zdań. Wybraliśmy drugie podejście do probabilistycznego wnioskowania, czując, że ta orientacja jest nieco bardziej intuicyjna, przenosząc wszystkie techniki liczenia i inne techniki teorii mnogości, jak pokazano w rozdziale 5.1. W rozdziałach 9 i 13 rozszerzamy naszą prezentację systemów stochastycznych na reprezentację pierwszego rzędu (opartą na zmiennych) schematy. Twierdzenie Bayesa stanowiło podstawę dla kilku systemów eksperckich z lat 70. i 80. XX wieku, w tym do szczegółowej analizy ostrego bólu brzucha w University of Glasgow Hospital oraz PROSPECTOR, system ze Stanford Uniwersytet wspierający poszukiwania minerałów. Naiwne podejście Bayesa zastosowano w wielu problemach z klasyfikacją, w tym w rozpoznawaniu wzorców (Duda i Hart 1973), przetwarzaniu języka naturalnego (Mooney 1996) i innych. Domingos a Pazzani (1997) uzasadniają sukcesy naiwnych klasyfikatorów Bayesa w sytuacjach, które nie spełniają założeń Bayesowskiej niepodległości. Obecnie prowadzonych jest wiele badań dotyczących probabilistycznego wnioskowania w sztucznej inteligencji. Niepewne rozumowanie stanowi element konferencji AI, w tym AAAI, IJCAI, NIPS i ZEA. Istnieje kilka doskonałych tekstów wprowadzających w rozumowaniu probabilistycznym (Ross 1988, DeGroot 1989), a także kilka wprowadzeń do stosowania metod stochastycznych w aplikacjach sztucznej inteligencji (Russell i Norvig 2003, Jurafsky i Martin 2009, Manning i Schutze 1999).