Zaawansowane Metody Inteligencji Obliczeniowej

Lab 14: Wnioskowanie probabilistyczne – Sieci Bayesowskie

Michał Kempka Marek Wydmuch

25 maja 2022

Fundusze Europejskie Polska Cyfrowa
Rzeczpospolita Polska
Unia Europejska Europejski Fundusz Rozwoju Regionalnego

"Akademia Innowacyjnych Zastosowań Technologii Cyfrowych (AI Tech)",
projekt finansowany ze środków Programu Operacyjnego Polska Cyfrowa POPC.03.02.00-00-0001/20

1 Wstęp

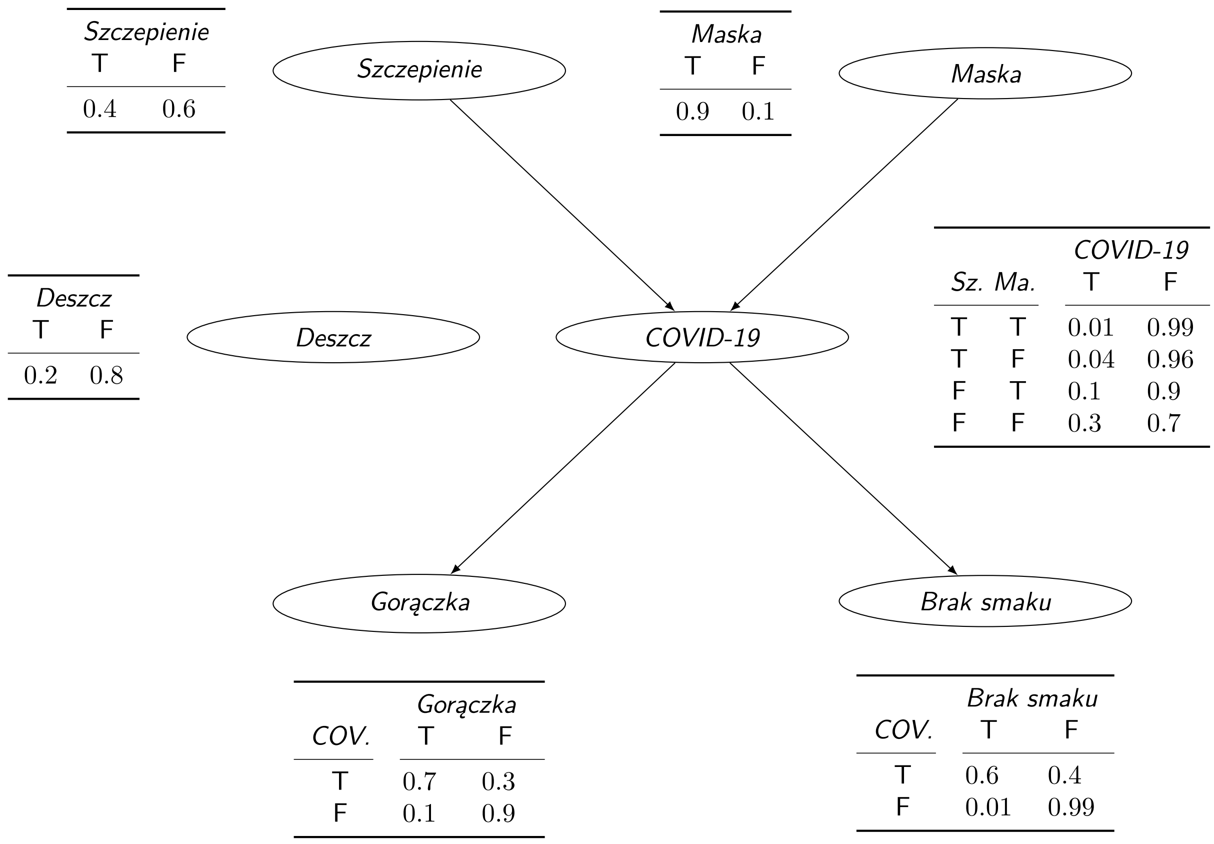
Na poprzednich zajęciach wprowadziliśmy prostą metodę wnioskowania probabilistycznego opartą o pełen łączny rozkład prawdopodobieństw, dzięki której mogliśmy wyliczyć prawdopodobieństwa dla dowolnego zestawu zmiennych losowych pod warunkiem posiadanych obserwacji. Niestety złożoność tej metody rośnie wykładniczo z ilością zmiennych losowanych. Pokazaliśmy też sobie, że dzięki wykorzystaniu niezależności pomiędzy zmiennymi ilość zmiennych losowych potrzebnych by zdefiniować pełen rozkład łączny maleje. Na dzisiejszych zajęciach wprowadzimy strukturę danych nazywaną **siecią Bayesowską (ang. Bayesian network)**. Nazwa ta posiada wiele synonimów: belief network, probabilistic network, casual network, knowledge map. W statystyce modele graficzne (ang. graphical models, są szerszą klasą, do której należą sieci Bayesowskie). Sieci Bayesowskie modelują relację pomiędzy zmiennymi losowymi i są w stanie reprezentować dowolny rozkład łączny.

2 Sieci Bayesowskie

- Sieć Bayesowska to acykliczny graf skierowany (DAG).
- Każdy wierzchołek odpowiada jednej zmiennej losowej, która może być dyskretna lub ciągła.
- Krawędź z wierzchołka X do Y oznacza, że X jest rodzicem Y.
- Każdy wierzchołek reprezentuje warunkowy rozkład prawdopodobieństwa $P(X|Rodzice(X))$.
- Topologia sieci definiuje warunkową niezależność pomiędzy pomiędzy zmiennymi. Krawędź z X do Y oznacza, że X ma bezpośredni wpływ na Y.

Schemat sieci Bayesowskiej zaprezentowany jest na Rysunku 1.

Zauważ, że na Rysunku 1 mamy 6 zmiennych losowych binarnych, model ten więc ma $2^6 = 64$ zdarzeń elementarnych i odpowiadającym im wartości prawdopodobieństwa. W tej sieci Bayesowskiej reprezentujemy ten rozkład przy pomocy jedynie 11 wartości (dla zmiennych binarnych wystarczy przechowywać jedną wartość). Możemy więc postrzegać sieci Bayesowską jako skompresowaną reprezentacje rozkładu łącznego, co pozwala na używanie ich do poznanego wnioskowania przez enumerację. Dodatkowo koduje ona informacje na temat zależności zmiennych, co pozwala na zastosowanie bardziej wydajnego wnioskowania poprzez eliminację zmiennych oraz na bardziej wydajne implementację algorytmów przybliżających rozkład poprzez próbkowanie. Więcej na ten temat możesz przeczytać w Russell and Norvig [1] w Rozdziale 14.4 i 14.5.



Szczepienie T	Szczepienie F
0.4	0.6

Maska T	Maska F
0.9	0.1

Deszcz T	Deszcz F
0.2	0.8

Sz.	Ma.	COVID-19 T	COVID-19 F
T	T	0.01	0.99
T	F	0.04	0.96
F	T	0.1	0.9
F	F	0.3	0.7

COV.	Gorączka T	Gorączka F
T	0.7	0.3
F	0.1	0.9

COV.	Brak smaku T	Brak smaku F
T	0.6	0.4
F	0.01	0.99

Rysunek 1: Schemat typowej sieci Baysowskiej

2.1 Całkowity rozkład łączny w sieciach Bayesowskich

Sieć Bayesowska definiuje rozkład po wszystkich zmiennych losowych jakie zawiera, których liczbę oznaczmy jako n:

$$P(x_1, \dots, x_n) = \prod_{i=1}^{n} P(x_i | rodzice(X_i))\,, \tag{1}$$

gdzie $rodzice(X_i)$ oznacza wartości zmiennych losowych $Rodzice(X_i)$, które pojawiają się w $(x_1, \dots, x_n)$.

Równanie 1 definiuje sieć Bayesowską i posłużmy się nim teraz by przedstawić proces konstruowania sieci Bayesowskiej. Zauważmy że:

$$P(x_1, \dots, x_n) = P(x_n | x_{n-1}, \dots, x_1) P(x_{n-1} | x_{n-2}, \dots, x_1) \dots P(x_2 | x_1) P(x_1) \tag{2}$$

$$= \prod_{i=1}^{n} P(x_i | x_{n-1}, \dots, x_1)\,. \tag{3}$$

Następnie zauważ, że:

$$P(X_i | X_{n-1}, \dots, X_1) = P(X_i | Rodzice(X_i))\,, \tag{4}$$

jeśli $Rodzice(X_i) \subseteq \{X_{n-1}, \dots, X_1\}$. Ten warunek jest spełniany jeśli kolejność zmiennych losowych będzie zgodna z hierarchią w grafie (sieci Bayesowskiej). Równanie 4 mówi nam, że sieć Bayesowska poprawnie reprezentuje domenę jeśli każdy wierzchołek jest warunkowo niezależny od swoich poprzedników w hierarchii sieci pod warunkiem swoich rodziców. Aby zapewnić ten warunek przy konstrukcji sieci Bayesowskiej korzystamy z następującej metodologii:

1. Wierzchołki: Zdecyduj o kolejności zmiennych losowych, dla każdej kolejności da się zbudować sieć Bayesowską, ale wynikowa sieć będzie prostsza jeśli zmienne uporządkujemy w kolejności takiej, że "przyczyna" będzie poprzedzać "efekt".

2. Krawędzie: dla $i = 1, \ldots, n$ wybierz z $X_1, \ldots, X_{i-1}$ minimalny zbiór rodziców dla X_i taki, żeby Równanie 4 było spełnione, dodaj krawędź od każdego rodzica do X_i, wylicz rozkład warunkowy $P(X_i|Rodzice(X_i)$. Intuicyjnie rodzice X_i powinni być zmiennymi, które bezpośrednio wpływają na X_i.

2.2 Interpretowanie sieci Bayswoskiej

Rozważ sieć Baysowskom z Rysunku 1.

1. Ile pojedynczych wartość zawiera rozkład łączny reprezentowany przez tą sieć?
2. Wylicz $P(Szczepienie = T, Maska = T, COVID = F, Gorczka = T, Braksmaku = F)$.
3. Wylicz $P(Gorczka = T|COVID = T, Szczepienie = F)$.
4. Wylicz $P(Gorczka = F, BrakSmaku = T|COVID = F, Maska = T, Deszcz = T)$.
5. Wylicz $P(COVID = T|Szczepienie = T, BrakSmaku = T)$.

2.3 Warunkowa niezależność zmiennych w sieciach Bayesowskich

Na podstawie Równania 4 wiemy, że zmienne są warunkowo niezależne od swoich poprzedników w hierarchii sieci pod warunkiem swoich rodziców. Z hierarchii wynika również inna ważna własność, zmienna losowa jest warunkowo niezależna od innych zmiennych w sieci, pod warunkiem jej rodziców, dzieci oraz wszystkich rodziców jej dzieci, własność tę nazywamy kocykiem Markowa (ang. Markov blanket).

2.4 Prosta sieć dla rzutów monetą

Mamy worek z trzema niewyważonymi monetami a, b i c z następującymi prawdopodobieństwami wypadnięcia reszki: 20%, 60% i 80%. Jedna moneta jest losowana z worka (z równym prawdopodobieństwem) a następnie wykonywane są trzy rzuty wylosowaną monetą oznaczone zmiennymi losowymi X1, X2, X3.

1. Zdefiniuj sieć Bayesowską odpowiadającą temu problemowi, oblicz rozkłady warunkowe.
2. Która moneta najprawdopodobniej została wyciągnięta z worka jeśli zaobserwowaliśmy, że dwa razy wypadła reszka a raz orzeł.

2.5 Niezależność w sieci Bayswoskiej

Rozważ sieć Baysowskom z Rysunku 1.

1. Jeśli nie mamy obserwacji, czy $Szczepienie$ i $Maska$ są niezależne? Przedstaw dowód.
2. Jeśli zaobserwowaliśmy *COVID-19*$= T$ czy $Szczepienie$ i $Maska$ są niezależne. Przedstaw dowód.

Literatura

[1] Russell, S. and Norvig, P. (2010). *Artificial Intelligence: A Modern Approach*. Prentice Hall, third edition.

[2] Sutton, R. S. and Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. The MIT Press, second edition.

Fundusze Europejskie
Polska Cyfrowa
Rzeczpospolita Polska
Unia Europejska
Europejski Fundusz Rozwoju Regionalnego

"Akademia Innowacyjnych Zastosowań Technologii Cyfrowych (AI Tech)",
projekt finansowany ze środków Programu Operacyjnego Polska Cyfrowa POPC.03.02.00-00-0001/20