

Systemy ekspertowe

Sieci Bayesa

Sieci Bayesowskie

- **Sieć Bayesowska** służy do przedstawiania zależności pomiędzy zdarzeniami bazując na rachunku prawdopodobieństwa.
- Klasycznym przykładem jest reprezentowanie zależności pomiędzy symptomami a chorobą.
- Sieć Bayesa to **skończony, acykliczny graf skierowany**, zbudowany na podstawie prawdopodobieństw warunkowych (które są zawarte w zbiorze CP).

Sieć Bayesa - formalnie

- Sieć Bayesa to trójka:

$$B = \langle N, E, C_p \rangle$$

Gdzie:

N – zbiór wierzchołków grafu

E – zbiór krawędzi grafu

C_p – zbiór prawdopodobieństw warunkowych.

- Wierzchołkami grafu są stwierdzenia lub hipotezy.
- Krawędzie grafu to relacje określające przejście od wierzchołka do wierzchołka z prawdopodobieństwem określającym to przejście.

Przykład budowy Sieci Bayesa

- Zbuduj Sieć Bayesa dla określonego zbioru C_p :

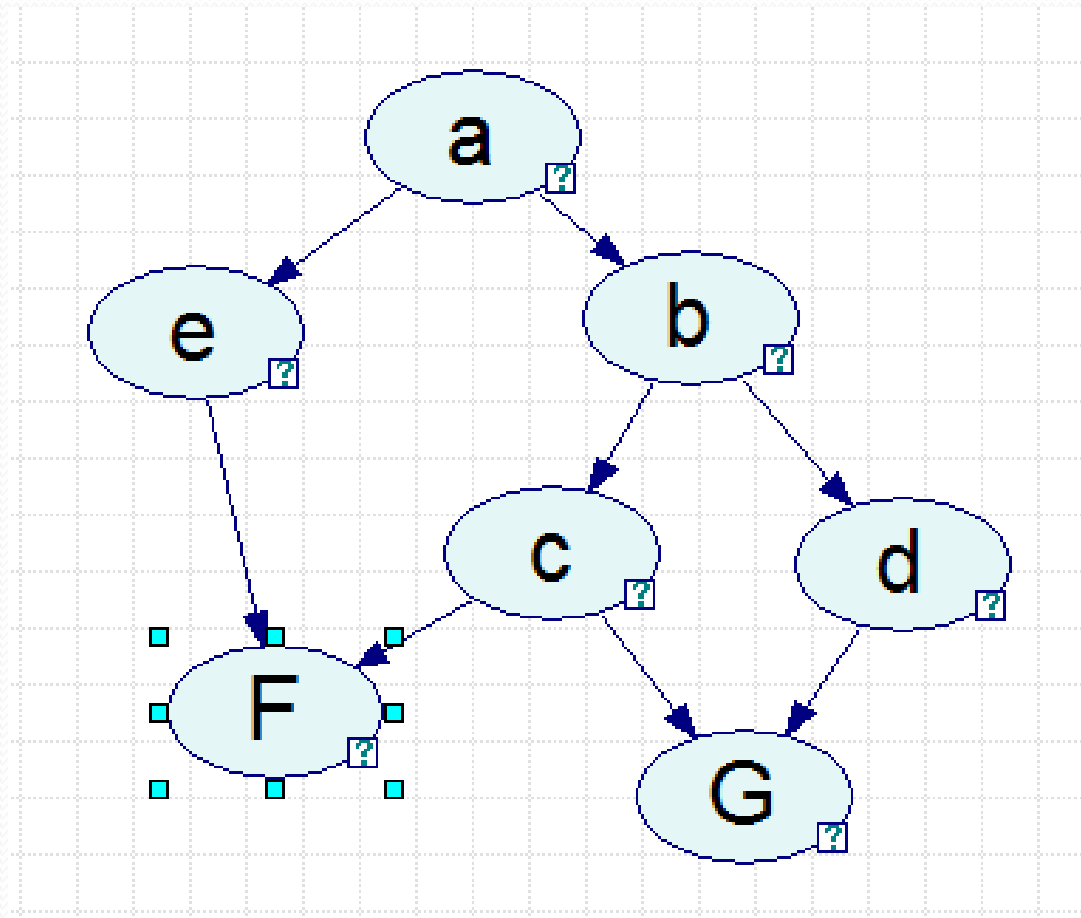
$$P = \{P(a), P(b|a), P(c|b), P(d|b), P(G|c), P(G|d), P(e|a), P(F|e), P(F|c)\}$$

Przykład budowy Sieci Bayesa

- Zbuduj Sieć Bayesa dla określonego zbioru C_p :

$$P = \{P(a), P(b|a), P(c|b), P(d|b), P(G|c), P(G|d), P(e|a), P(F|e), P(F|c)\}$$

Taki graf pozwala na prowadzenie wnioskowania w oparciu o prawdopodobieństwo i wzór Bayesa.



Teoria Bayesa

- Prawdopodobieństwo łączne (całkowite)

Jeśli zbiór Ω podzielimy na zbiory B , $\sim B$ takie, że $B \cup \sim B = \Omega$,
 $B \cap \sim B = \emptyset$ i $P(B) > 0$ oraz $P(\sim B) > 0$, to dla dowolnego zdarzenia $A \subseteq \Omega$
prawdziwy jest wzór:

$$P(A \cap B) = P(A|B) \cdot P(B) + P(A|\sim B) \cdot P(\sim B)$$

gdzie:

$P(B)$ – prawdopodobieństwo zajścia zdarzenia B

$P(\sim B)$ – prawdopodobieństwo zajścia zdarzenia przeciwnego do B

$P(A|B)$ – prawdopodobieństwo (warunkowe) zajścia
zdarzenia A pod warunkiem zajścia zdarzenia B .

- Jeśli zdarzenia A i B są niezależne wówczas prawdopodobieństwo łączne liczone jest ze wzoru:

$$P(A \cap B) = P(A) \cdot P(B)$$

- Prawdopodobieństwo warunkowe (twierdzenie Bayesa lite)

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}$$

Teoria Bayesa - ręcznie

- Szpital przeprowadza test na rzadką chorobę, na którą choruje 1 na 1000 osób. Test zawsze wykrywa chorych, przy czym daje pozytywną fałszywą odpowiedź w 3% przypadków. Jakie jest prawdopodobieństwo tego, że osoba, u której test da wynik pozytywny jest jednak zdrowa?
- A – osoba jest chora
- $\sim A$ – osoba jest zdrowa
- B – test pozytywny (chory)
- $\sim B$ – test negatywny (zdrowy)
- $(B|A)$ – test wykrył chorobę u chorej osoby
- $(B|\sim A)$ – test wykrył chorobę u zdrowej osoby
- $(\sim A|B)$ – pozytywny wynik, ale zdrowa osoba

- $P(A) = 0,001$
- $P(\sim A) = 0,999$
- $P(B|A) = 1$
- $P(B|\sim A) = 0,03$

$$\begin{aligned}P(\sim A|B) &= \frac{P(B|\sim A) \cdot P(\sim A)}{P(B)} \\ &= \frac{P(B|\sim A) \cdot P(\sim A)}{P(B|A) \cdot P(A) + P(B|\sim A) \cdot P(\sim A)} \\ &= \frac{0,03 \cdot 0,999}{1 \cdot 0,001 + 0,03 \cdot 0,999} = \frac{0,02997}{0,03097} \\ &= 97\%\end{aligned}$$

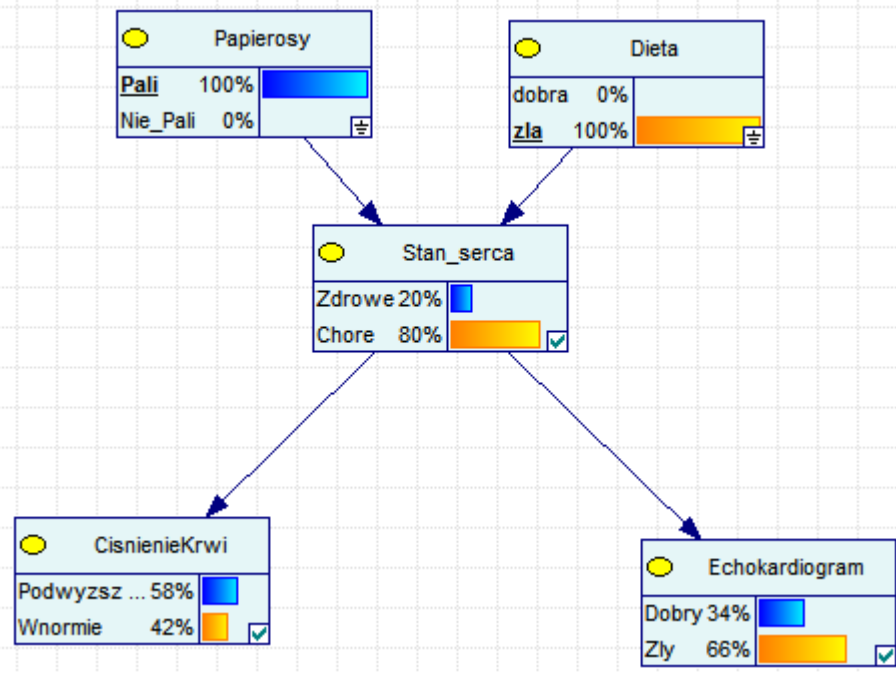
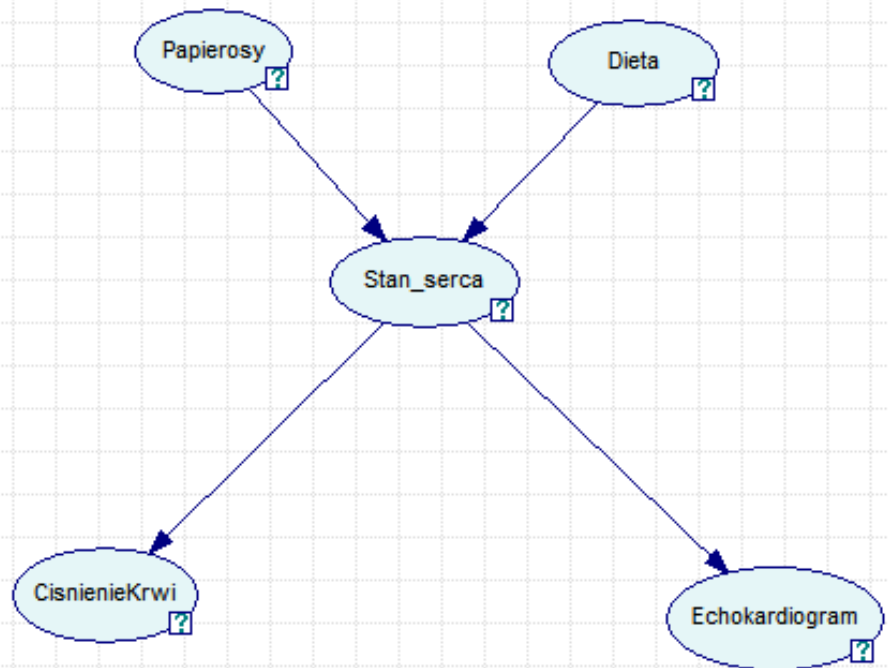
Zadanie 1

- Sprawdź, czy zgodnie z podanymi założeniami możliwe jest utworzenie sieci Bayesa:
 1. $Z = \{A, B, C, D, E, F, G, H\}$
 $C_p = \{P(A), P(B), P(C|A), P(C|B), P(D|C), P(E|D), P(F|C), P(F|C), P(G|C), P(G|B), P(G|F), P(E|D), P(H|F)\}$
 2. $Z = \{A, B, C, D, E, F, G, H\}$
 $C_p = \{P(A), P(D), P(E), P(B|A), P(B|D), P(C|B), P(F|C), P(G|C), P(H|G), P(A|F)\}$
 3. $Z = \{A, B, C, D, E, F, G, H\}$
 $C_p = \{P(A), P(D), P(E), P(B|A), P(B|D), P(C|B), P(F|C), P(G|C), P(H|G)\}$
 4. $O = \{A, B, C, D, E, G\}; H = \{X, Y, Z, F\}$
 $CP = \{P(A), P(B), P(D), P(E), P(F), P(G), P(D|B), P(X|A, B), P(F|G), P(C|X), P(Y|C, D, E), P(Z|C, E), P(B|Z)\}$

Przykład 1

- Na złą kondycję serca z całą pewnością ma wpływ zarówno niewłaściwa dieta, jak i nałóg palenia papierosów. Choroba serca bardzo często skutkuje podwyższonym ciśnieniem krwi. Ponadto osoby z chorobą serca mają niewłaściwe wyniki na echokardiogramie.
- Z najnowszych badań przeprowadzonych przez TNS OBOP jednoznacznie wynika, że 30% Polaków pali papierosy nie zważając na konsekwencje zdrowotne. Te same badania wykazały również, że zaledwie 60% badanych utrzymuje właściwą dietę bogatą w składniki mineralne i wartości odżywcze. Mało kto niestety zdaje sobie sprawę, że palenie papierosów i niewłaściwe odżywianie zwiększa ryzyko choroby serca do 80%. W przypadku, gdy źle odżywiająca się osoba nie pali, ryzyko dla jej serca jest zmniejszone o 30% w porównaniu z poprzednią sytuacją. Aby dokonać porównania, przedstawiono również jak wpływa sam fakt palenia papierosów przy zachowaniu właściwej diety na kondycję serca. Stwierdzono zatem, że palenie papierosów w takim przypadku zwiększa ryzyko choroby do 40%, natomiast gdy dana osoba nie pali i dobrze się odżywia, jedynie w 10% jest narażona na ryzyko choroby serca.
- Jak donosi nasz regionalny ekspert ds. chorób serca, doktor Mariusz Żyła, 70% pacjentów z niewydolnością serca ma zwiększone ciśnienie krwi. Bardziej wiarygodnym czynnikiem jest jednak echokardiogram, który poprawnie wykazuje chorobę (anomalię w wynikach) u 80% badanych. Niestety zarówno wysokie ciśnienie krwi, jak i wynik echokardiogramu jest błędny u 10% wszystkich przypadków i źle klasyfikuje osobę zdrową.

Rozwiązanie



Zadanie 2

- W podanym niżej tekście występują pewne zależności przyczynowo skutkowe opisane liczbowo prawdopodobieństwami warunkowymi. Proszę podać zbiór CP takich prawdopodobieństw warunkowych oraz narysować graf przyczynowo-skutkowy. Czy otrzymany graf jest siecią Bayes'a – proszę uzasadnić odpowiedź.

Zadanie 2

- W podanym niżej tekście występują pewne zależności przyczynowo skutkowe opisane liczbowo prawdopodobieństwami warunkowymi. Proszę podać zbiór CP takich prawdopodobieństw warunkowych oraz narysować graf przyczynowo-skutkowy. Czy otrzymany graf jest siecią Bayes'a – proszę uzasadnić odpowiedź.
- Prawdopodobieństwo awarii tylnej przerzutki z powodu zawilgocenia wynosi 0.3, zabrudzenia 0.6, natomiast z powodu zużycia 0.7. Awaria tylnej przerzutki uniemożliwi jazdę na 40%.
- Zawilgocenie i zabrudzenie na 70% spowodują przedwczesne wytarcie się klocków hamulcowych. Wytarte klocki na 20% uniemożliwią jazdę. Prawdopodobieństwo awarii suportu roweru z powodu zawilgocenia wynosi 0.6 a zabrudzenia 0.3.
- Awaria suportu na pewno uniemożliwi jazdę. Brak możliwości jazdy na rowerze na 40% spowoduje spadek zainteresowania jego stanem technicznym, a to na pewno negatywnie wpłynie na jego czystość – czyli na jego zabrudzenie.

Zadanie 3

- W podanym niżej tekście występują pewne zależności przyczynowo skutkowe opisane liczbowo prawdopodobieństwami warunkowymi. Proszę podać zbiór CP takich prawdopodobieństw warunkowych oraz narysować graf przyczynowo-skutkowy. Czy otrzymany graf jest siecią Bayes'a – proszę uzasadnić odpowiedź.
- Prawdopodobieństwo wystąpienia anginy w przypadku objawów takich jak ból gardła i gorączka jest wysokie i wynosić może 0.8. Jednak wystąpienie gorączki i bólu głowy może świadczyć o grypie, co jest hipoteza prawdopodobna na 0.6.
- W przypadku gdy pacjent cierpiący na grypę nie wyleczył się całkowicie może dojść do zapalenia oskrzeli z prawdopodobieństwem 0.4.
- Zapalenie oskrzeli może spowodować ból gardła z prawdopodobieństwem 0.3.