Najbliższe wykłady

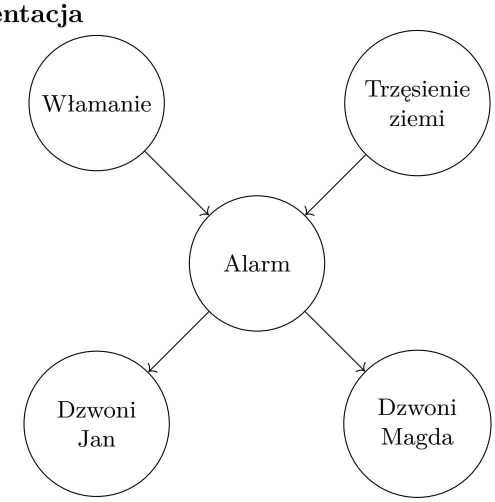
W jaki sposób możemy reprezentować wiedzę i wnioskować, w sytuacji gdy nasze informacje są niepewne (i dyskretne)?

Przykład: alarm

Przykład wzorowany na: J. Pearl, "Probabilistic reasoning in intelligent systems"

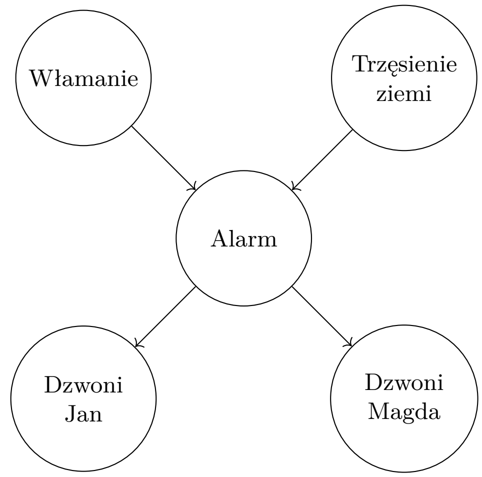
- Mamy w domu zamontowany alaram przeciwwłamaniowy.
- Czasem alarm uruchamia się w przypadku niewielkich trzęsień ziemi.
- Umówiliśmy się z sąsiadami, aby dzwonili gdy usłyszą alarm.
  - Jan dzwoni do nas zawsze gdy usłyszy sygnał alarmu – czasem myli go nawet z innymi dźwiękami.
  - Magda dzwoni rzadziej – w domu lubi słuchać dość głośnej muzyki.
- Przykładowe pytanie: jak ocenimy szanse na to, że było włamanie w zależności od tego kto zadzwonił?

Przykład: alarm — reprezentacja



Przykład: alarm

Oszacowania naszych przekonań: $P(W\text{ł}.) = 0.01,\ P(Trz.) = 0.02$



Wł.	Trz.	$P(Alarm\|W,T)$
1	1	0.95
1	0	0.94
0	1	0.29
0	0	0.001

Alarm	$P(J\|A)$
1	0.90
0	0.05

Alarm	$P(M\|A)$
1	0.70
0	0.01

Probabilistyka z perspektywy wnioskowania bayesowskiego

- Możliwe w danym scenariuszu **zdarzenia** dot. wnioskowania reprezentujemy przez (zazwyczaj) dyskretne zmienne losowe.
- **Wiedzę/Przekonania** odnośnie zdarzeń reprezentujemy w postaci rozkładów prawdopodobieństwa.
- Jeśli dwa zdarzenia są zależne/powiązane ze sobą – opisuje je rozkład dwuwymiarowy; jeśli trzy – trójwymiarowy, itd.
- Jeśli zdarzenia A zależy od zdarzenia B, które **zaobserwujemy** – naszą wiedzę o zdarzeniu A opisuje rozkład warunkowy $P(A|B)$.

1 Sieci Bayesa

1.1 Definicja i właściwości

Sieć Bayesa

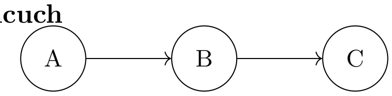
- Graficzny model probabilistyczny reprezentujący relacje między zmiennymi losowymi.
- Zwięzła reprezentacja p-stwa łącznego.
- Sieć Bayesa – acykliczny graf, w którym:
  - wierzchołki odpowiadają dyskretnym zmiennym losowym,
  - krawędzie reprezentują bezpośrednio występujące zależności między tymi zmiennymi,
  - dla każdej zmiennej dany jest jej rozkład prawdopodobieństwa – warunkowy jeżeli do zmiennej prowadzą krawędzie.
- Zazwyczaj – rozkłady przedstawiane są w formie tabelarycznej.

Sieć Bayesa

Założenie Markowa
Typowo, algorytmy i techniki wnioskowania z wykorzystaniem sieci bayesowskich przyjmują, że wszystkie **bezpośrednie zależności** między zmiennymi losowymi w modelowanym systemie są przedstawione za pomocą krawędzi w sieci.

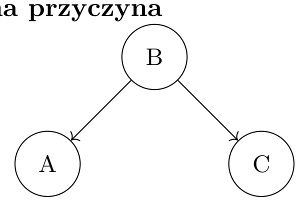
1.2 Warunkowa niezależność

Warunkowa niezależność — łańcuch



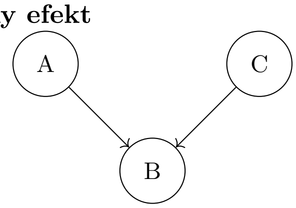
- W powyższym przykładzie mamy dwie bezpośrednie zależności: $B|A$ oraz $C|B$.
- Jeśli zaobserwujemy/ustalimy wartość B, to obserwacja A nie wnosi żadnych nowych informacji o C.
- Przy ustalonej wartości B, zmienna C jest **warunkowo niezależna** od A – oznaczamy $C \perp\!\!\!\perp A|B$.
- Ustalenie wartości B w pewnym sensie blokuje ścieżkę $A \to C$.

Warunkowa niezależność — wspólna przyczyna



- Sytuacja analogiczna do **łańcucha**.
- Jeśli zaobserwujemy/ustalimy wartość B, to obserwacja A nie wnosi żadnych nowych informacji o C.
- Przy ustalonej wartości B, zmienna C jest **warunkowo niezależna** od A.
- Inaczej: $P(C|A \wedge B) = P(C|B) \;\equiv\; C \perp\!\!\!\perp A|B$.

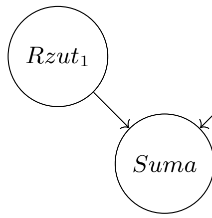
Warunkowa niezależność — wspólny efekt



- Sytuacja w której zdarzenie ma dwie przyczyny.
- Zachowanie **odwrotne** niż w dwóch poprzednich przypadkach, tzn: zdarzenia A i C są niezależne, ale stają się zależne po zaobserwowaniu B.
- $P(A|C \wedge B) \neq P(A|C) \;\equiv\; \neg(A \perp\!\!\!\perp C|B)$

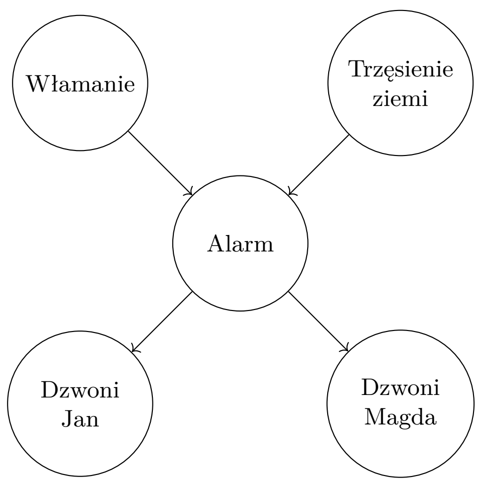
Wspólny efekt — przykład 1

- Rzucamy dwiema monetami i sumujemy „orły”.
- $Rzut_1 \perp\!\!\!\perp Rzut_2$
- Jeśli jednak znamy wartości: $Suma$ i $Rzut_1$, to znamy $Rzut_2$.
- Zdarzenia stają się **warunkowo zależne**: $\neg(Rzut_1 \perp\!\!\!\perp Rzut_2 | Suma)$.



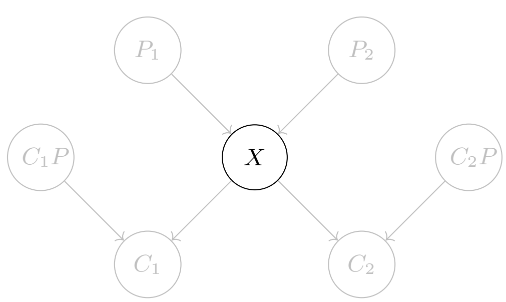
Wspólny efekt — przykład 2

- Włamanie i trzęsienie ziemi to zdarzenia niezależne.
- Jeśli wiemy, że zadzwonił alarm i było trzęsienie ziemi, to włamanie staje się mniej prawdopodobne.



Warunkowa niezależność — otoczka Markowa

- Tzw. **otoczka Markowa** (*Markov blanket*) węzła X, to zestaw tych wezłów, których wartości trzeba ustalić, aby X był warunkowo niezależny od pozostałej części sieci.
- W skład otoczki Markowa węzła X wchodzą: jego rodzice, jego dzieci oraz wszyscy rodzice tychże dzieci.



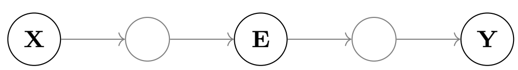
d-separacja

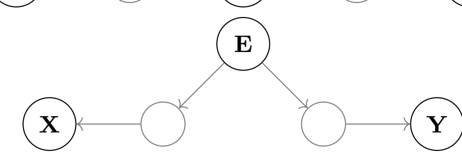
- Pojęcie pozwalające na rozszerzenie warunkowej niezależności na zbiory zmiennych/węzłów sieci.
- O zapisie $A \perp\!\!\!\perp C|B$ mówimy, że znajomość B blokuje przepływ informacji pomiędzy C i A.
- W przypadku wspólnego efektu mówimy, że znajomość B aktywuje przepływ informacji pomiędzy C i A.
- **Ścieżka** pomiędzy zbiorami wierzchołków $\mathbf{X}$ oraz $\mathbf{Y}$ to dowolna (ignorująca kierunki krawędzi i nie zawierająca powtarzającysh się wierzchołków) ścieżka w grafie sieci bayesa, łącząca wierzchołek należący do $\mathbf{X}$, z wierzchołkiem należącym do $\mathbf{Y}$.

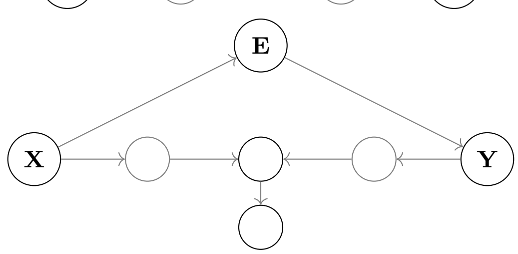
d-separacja

- Ścieżka pomiędzy zbiorami wierzchołków X oraz Y jest **zablokowana** przez zbiór wierzchołków $\mathbf{E}$, gdy istnieje na ścieżce wierzhołek Z taki, że spełniony jest jeden z warunków:
  1. $Z \in \mathbf{E}$ oraz Z jest fragmentem łańcucha w ścieżce $(\ldots \rightarrow Z \rightarrow \ldots)$,
  2. $Z \in \mathbf{E}$ oraz jest wspólną przyczyną dla dwóch sąsiadujących w ścieżce węzłów $(\ldots \leftarrow Z \rightarrow \ldots)$,
  3. Z ani jego potomkowie nie należą do $\mathbf{E}$ oraz Z występuje jako wspólny efekt w ścieżce $(\ldots \rightarrow Z \leftarrow \ldots)$
- O zbiorze węzłów $\mathbf{E}$ mówimy, że **d-separuje** zbiory $\mathbf{X}$ i $\mathbf{Y}$ jeśli każda ścieżka pomiędzy węzłami z tych zbiorów jest blokowana przez $\mathbf{E}$. Zbiory $\mathbf{X}$ i $\mathbf{Y}$ są wtedy warunkowo niezależne przy zaobserwowaniu $\mathbf{E}$.

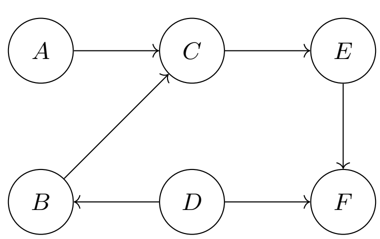
d-separacja: przykłady







Przykładowe zadanie



1. Czy $P(C|B \wedge D) = P(C|B)$?
2. Czy $P(A|B) = P(A|B \wedge E)$?
3. Jaki zbiór węzłów zapewni d-separację zmiennych E oraz B?

1.3 Wnioskowanie

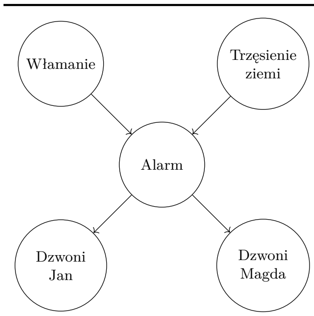
Wnioskowanie w sieciach Bayesa

Wnioskowanie na podstawie tego typu modeli zakłada, że:

- mamy sieć reprezentującą (najlepiej: przyczynowo-skutkową) strukturę zależności między zmiennymi,
- dostępne są dane/obserwacje/dowody ustalające wartości części zmiennych losowych w sieci,
- mamy pytanie – chcemy poznać zaktualizowany rozkład zmiennej/zmiennych spośród tych niezaobserwowanych.

Typy wnioskowania

Typ	Obserwacje/Dane	Pytanie
Diagnostyczne	Magda zadzwoniła	Czy było włamanie?
Predykcyjne	Włączył się alarm	Czy zadzwoni Jan lub Magda?
Wyjaśniające	Włączył się alarm i było trzęsienie z.	Czy nie było również włamania?
Mieszane	Było trzęsienie ziemi i zadzwonił Jan	Czy włączył się też alarm?



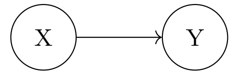
Twierdzenie Bayesa

Dla pary zmiennych losowych X i Y spełniona jest równość:

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

gdzie $P(Y|X)$ to p-stwo zaobserwowania zdarzenia Y pod warunkiem, że zaobserowano zdarzenie X, przy czym $P(Y|X) = \frac{P(Y,X)}{P(X)}$ dla $P(X) > 0$.

Przebieg wnioskowania – sieć dwuelementowa



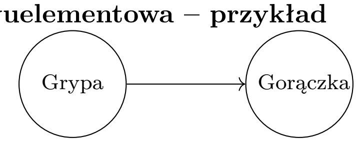
- Jeśli obserwacje dotyczą rodzica (np. $X = x$) – $Bel(Y) = P(Y|X = x)$ odczytujemy bezpośrednio z tablicy prawdopodobieństwa warunkowego.
- $Bel(Y)$ - to „przekonanie” odnośnie stanu Y na podstawie obserwacji.

- Jeśli obserwacje dotyczą dziecka (np. $Y = y$) – przekonania o X aktualizujemy stosując tw. Bayesa:

$$Bel(X = x) = \frac{P(Y = y|X = x)P(X = x)}{P(Y = y)} \tag{1}$$
$$= \alpha P(x)P(Y = y|X = x) \tag{2}$$

- $\alpha = 1/P(Y = y)$ - traktujemy jako współczynnik normalizujący (bo $\sum_x Bel(X = x) = 1$), a $P(x)$ - to nasza wiedza pierwotna, czyli przed aktualizacją przekonań.

Przebieg wnioskowania – sieć dwuelementowa – przykład



$$P(Grypa = Prawda) = 0.05$$
$$P(Gorączka = P|Grypa = P) = 0.9$$
$$P(Gorączka = P|Grypa = F) = 0.2$$

- Zaobserowaliśmy, że pacjent ma gorączkę – czy ma grypę?
- (przykład z Korb i Nicholson)

Przebieg wnioskowania – sieć dwuelementowa – przykład c.d.

$$\begin{aligned}
P(Grypa = Prawda) &= 0.05\\
P(Gorączka = P|Grypa = P) &= 0.9\\
P(Gorączka = P|Grypa = F) &= 0.2\\
Bel(Grypa = T) &= \alpha P(Grypa = P)P(Goraczka = P|Grypa = P)\\
&= \alpha \times 0.05 \times 0.9\\
&= \alpha 0.045\\
Bel(Grypa = F) &= \alpha P(Grypa = F)P(Goraczka = P|Grypa = F)\\
&= \alpha \times 0.95 \times 0.2\\
&= \alpha 0.19
\end{aligned}$$

Przebieg wnioskowania – sieć dwuelementowa – przykład c.d.

$$\begin{aligned}
Bel(Grypa = T) + Bel(Grypa = F) = \alpha 0.045 + \alpha 0.19 = 1 &\implies\\
\alpha &= \frac{1}{0.045 + 0.19} \approx 4.25\\
Bel(Grypa = T) &= \alpha 0.045 \approx 0.19\\
Bel(Grypa = F) &= \alpha 0.19 \approx 0.81
\end{aligned}$$

Po uwzględnieniu pomiaru temperatury, przekonanie o tym, że pacjent ma grypę wzrosło z poziomu 0.05 do 0.19.

Przebieg wnioskowania – łańcuch

$$X_1 \longrightarrow X_2 \longrightarrow \cdots \longrightarrow X_n$$

Analogiczna sytuacja:

- jeżeli wnioskujemy od zgodnie ze kierunkiem zależności – możemy bezpośredni korzystać z tabel prawdopodobieństwa,
- w kierunku odwrotnym – stosujemy tw. Bayesa.

Przebieg wnioskowania – przypadek ogólny

- W przypadku ogólnym dokładne wnioskowanie okazuje się problemem NP-trudnym.
- Są dedykowane algorytmy zdefiniowane dla sieci, w których wierzchołki są połączone co najwyżej jedną ścieżką (drzewa lub lasy).
- Dokładne algorytmy bazują na tzw. propagacji przekonań: wnioskowanie przebiega na podstawie wymiany komunikatów pomiędzy węzłami sieci. Klasyczne podejście - algorytm Kima i Pearla.
- W przypadku większych sieci, w praktyce stosuje się algorytmy przybliżone.

Przebieg wnioskowania – algorytmy przybliżone

- Podejścia bazujące na wykonywaniu stochastycznych symulacji.
- Podczas symulacji przy pomocy sieci generowane są próbki losowe, na podstawie których szacowane są prawdopodobieństwa przypisane do poszczególnych węzłów.
- Jeżeli chcielibyśmy przybliżać wynik z dowolną dokładnością – nadal będzie to problem NP-trudny. W praktyce, zazwyczaj obserwacje/dowody na których odbywa się warunkowanie nie są bardzo mało prawdopodobne (zgodnie z intuicją) – podejścia przybliżone działają wtedy dość szybko.

Literatura

- J. Pearl, „Probabilistic reasoning in intelligent systems”.
- Kevin B. Korb, Ann E. Nicholson, „Bayesian Artificial Intelligence”, Chapman&Hall
- A. Pfeffer, „Practical probabilistic programming”. Manning Publications Co.