Motywacja

Wiedza o świecie jaką posiada agent inteligentny jest z konieczności niepełna i niepewna. Nawet w przypadkach kiedy mógłby on zdobyć wiedzę kompletną i pewną, może to być niepraktyczne.

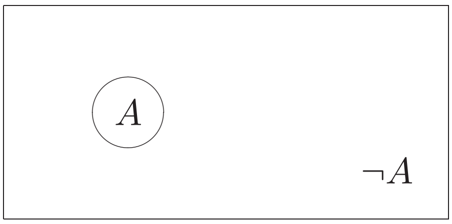
W sztucznej inteligencji od dawna próbowano budować mechanizmy i formalizmy pozwalające wnioskować i działać w takich warunkach, poprzez dodanie oszacowania wiarygodności posiadanych faktów do wnioskowania logicznego. Przykładami mogą być: logiki modalne, logika trójwartościowa, logiki niemonotoniczne, logika rozmyta, logika probabilistyczna, i inne.

Praktyczne zastosowania tych metod okazują się jednak ograniczone. Dopiero stosunkowo niedawno wzrosło zainteresowanie wykorzystaniem prawdopodobieństwa w sposób bezpośredni. To podejście przyniosło duży sukces, i metody oparte na reprezentowaniu wiedzy agenta o świecie w postaci prawdopodobieństw są jednymi z najbardziej dynamicznie rozwijających się technik sztucznej inteligencji. W tym schemacie reprezentacji metodą wnioskowania jest matematyczny rachunek prawdopodobieństwa.

Prawdopodobieństwo bezwarunkowe

Prawdopodobieństwo bezwarunkowe (a priori) określa liczbowo szansę wystąpienia jakiegoś zjawiska, gdy nie są znane żadne okoliczności związane z tym zjawiskiem (np. czy ono w rzeczywistości się wydarzyło).

Graficzna wizualizacja zdarzeń i ich prawdopodobieństw:



$P(A) =$ powierzchnia kółka

$P(\neg A) =$ dopełnienie do prostokąta

powierzchnia prostokąta $= 1$

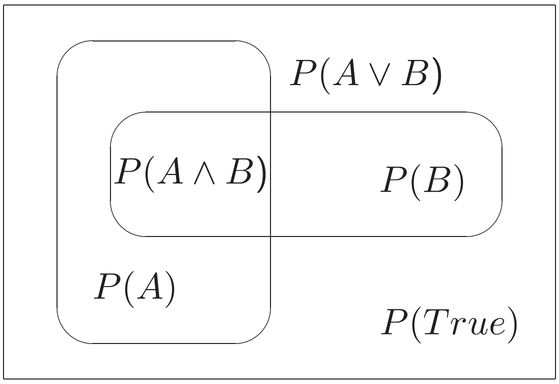
Np.: prawdopodobieństwo, że zgłaszający się do lekarza pacjent jest chory na nietypowe zapalenie płuc SARS (*Severe Acute Respiratory Syndrome*, zwane również nietypowym zapaleniem płuc) może wynosić $P(\text{SARS}) = 0.0001$

Jednak gdyby lekarz wiedział, że pacjent właśnie przyjechał z Hong-Kongu[1] i ma typowe objawy nietypowego zapalenia płuc, to prawdopodobieństwo posiadania przez niego choroby wywołanej tym wirusem należałoby określić zupełnie inaczej.

[1] Wyjaśnienie: ten przykład powstał w roku 2003 kiedy w Chinach szalała epidemia SARS.

Aksjomaty prawdopodobieństwa

- $0 \leq P(A) \leq 1$
- $P(\text{True}) = 1$
- $P(\text{False}) = 0$
- $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$



Więcej o aksjomatach prawdopodobieństwa

Z danych aksjomatów można wyprowadzić wiele użytecznych zależności:

$$P(\neg A) = 1 - P(A) \quad (1)$$
$$P(A) = P(A \wedge B) + P(A \wedge \neg B) \quad (2)$$

(i inne).

Aksjomaty prawdopodobieństwa mają głęboki sens — ścisłe trzymanie się ich gwarantuje niepopełnienie błędu w obstawianiu swoich szans. Inaczej mówiąc, gdyby w jakiejś grze losowej agent zastosował w swoim rozumowaniu prawdopodobieństwa naruszające te aksjomaty, i gotów był przyjmować zakłady zgodne z tymi prawdopodobieństwami, to istnieje strategia obstawiania w tych zakładach, gwarantująca wygraną jego przeciwnikowi.

Zmienne losowe

Zmienna losowa reprezentuje jakieś zjawisko losowe, które może przyjmować wartości z pewnego zbioru (dziedziny zmiennej losowej).

Np.: chcąc określić jaka będzie dziś pogoda i z jakim prawdopodobieństwem, możemy potraktować dzisiejszą pogodę (*Pogoda*$_{\text{DZIŚ}}$) jako zmienną losową, której wartości należą do zbioru: $\{$*Słońce,Chmury,Deszcz,Śnieg*$\}$

Zestaw wartości prawdopodobieństw wszystkich możliwych wartości zmiennej losowej nazywamy **rozkładem prawdopodobieństwa** tej zmiennej losowej. Rozkład prawdopodobieństwa dla zmiennej losowej *Pogoda*$_{\text{DZIŚ}}$ można zapisać: $\mathbf{P}(\mathit{Pogoda}_{\text{DZIŚ}}) = \{0.8, 0.1, 0.09, 0.01\}$

Łączny rozkład prawdopodobieństw

Możemy brać pod uwagę kilka zmiennych losowych opisujących różne zjawiska losowe. Zdarzeniem atomowym nazywamy przypisanie wartości wszystkim zmiennym losowym, czyli kombinacja tych wartości. Na przykład, dla dwóch zmiennych losowych X i Y można skonstruować tabelę zdarzeń atomowych:

	$X = x_1$	$X = x_2$	...	$X = x_n$
$Y = y_1$				
$Y = y_2$				
...				
$Y = y_k$				

Łączny rozkład prawdopodobieństwa (JPD) dla zbioru zmiennych losowych jest tabelą prawdopodobieństw wszystkich zdarzeń atomowych. W polu tabeli w rzędzie j i kolumnie i znajduje się prawdopodobieństwo jednoczesnego przyjęcia przez zmienną X wartości x_i i przez zmienną Y wartości y_j, czyli $P(X = x_i \wedge Y = y_j)$. Sumując w tej tabeli wzdłuż rzędów lub kolumn możemy otrzymać prawdopodobieństwa dla poszczególnych wartości pojedynczych zmiennych. Suma wszystkich prawdopodobieństw całej tabeli daje 1.0.

Posługiwanie się tabelą JPD

Mając wypełnioną tabelę JPD możemy obliczać prawdopodobieństwa dowolnych zdarzeń. Na przykład:

• Prawdopodobieństwo zdarzenia polegającego na przyjęciu przez zmienną X wartości x_i $P(X = x_i)$ możemy obliczyć przez zsumowanie wszystkich wartości w kolumnie i tabeli JPD.

• Prawdopodobieństwo zdarzenia polegającego na tym, że zmienna X przyjmie wartość x_i **lub** że zmienna Y przyjmie wartość y_j możemy obliczyć przez zsumowanie wszystkich wartości w kolumnie i i rzędzie j tabeli JPD, licząc zawartość pola (i, j) tabeli tylko raz. Jak widać wynik będzie dokładnie ten sam, jak gdyby obliczać z tabeli wartości według wzoru:

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$$

Jednak aby w ten sposób posługiwać się prawdopodobieństwami musimy obliczyć prawdopodobieństwa wszystkich zdarzeń atomowych, i kompletnie wypełnić tabelę JPD, co może być kosztowne.

Obliczanie prawdopodobieństw atomowych

Skąd pochodzą dane o prawdopodobieństwach? Można je zgromadzić statystycznie, można dokonać analizy i obliczyć jako inherentne cechy zjawiska fizycznego, można również związać te prawdopodobieństwa z agentem, charakteryzując jego punkt widzenia na świat.

Na przykład, jakie jest prawdopodobieństwo zdarzenia, że słońce będzie istniało jutro? Można próbować to obliczyć na wiele sposobów, przyjmując różne punkty widzenia:

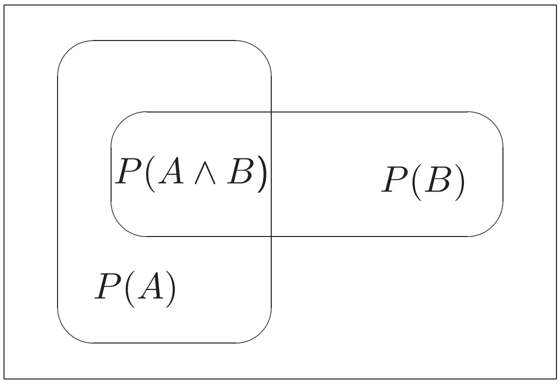
- nie da się określić, bo nie sposób przeprowadzić niezbędnych eksperymentów,
- poprzednie „podobne" eksperymenty dowodzą, że słońce „zawsze" istnieje, więc prawdopodobieństwo wynosi 1,
- prawdopodobieństwo wynosi $1 - \epsilon$ gdzie ϵ jest prawdopodobieństwem wybuchu gwiazdy danego dnia,
- prawdopodobieństwo wynosi $d/(d + 1)$ gdzie d jest liczbą dni dotychczasowego istnienia słońca,
- prawdopodobieństwo można określić budując model istnienia i rozpadu słońca na podstawie zachowania innych, podobnych gwiazd.

Prawdopodobieństwo warunkowe

Prawdopodobieństwo warunkowe (a posteriori) $P(A|B)$ — prawdopodobieństwo zdarzenia A obliczane tylko w sytuacjach, w których B jest spełnione. Jest związane z bezwarunkowym wzorem:

$$P(A|B) = \frac{P(A \wedge B)}{P(B)} \tag{3}$$

Wzór ten można wytłumaczyć następująco: aby obliczyć prawdopodobieństwo $P(A|B)$ musimy wziąć ułamek przypadków zdarzenia $A \wedge B$ we wszystkich przypadkach zdarzenia B.



Inne wytłumaczenie można przedstawić na podstawie wzoru odwróconego:

$$P(A \wedge B) = P(A|B)P(B) \tag{4}$$

Aby obliczyć $P(A \wedge B)$ musimy wiedzieć, że nastąpiło B, i wiedząc to, wtedy obliczyć prawdopodobieństwo A. (Albo na odwrót.)

Ważny, często przydatny wzór wiążący bezwarunkowe prawdopodobieństwo zdarzenia z warunkowym otrzymujemy z połączenia wzorów (2, str.4) i (4):

$$P(A) = P(A|B)P(B) + P(A|\neg B)P(\neg B) \tag{5}$$

Należy podkreślić, że prawdopodobieństwo warunkowe dla ustalonego warunku spełnia wszystkie aksjomaty prawdopodobieństwa, a zatem posiada wszystkie własności prawdopodobieństwa bezwarunkowego, na przykład:

$$P(A|B) + P(\neg A|B) = 1 \tag{6}$$

Dlaczego prawdopodobieństwo warunkowe?

Rozważmy klasyczną zagadkę, znaną jako *Monty Hall problem* (1975):

> Bierzemy udział w grze telewizyjnej. Mamy wybrać jedne z trojga drzwi, gdzie za jednymi z nich stoi samochód do wygrania. Nie posiadamy żadnych dodatkowych informacji, więc wybieramy np. drzwi numer 1. Wtedy prowadzący grę otwiera jedne z pozostałych dwojga drzwi — załóżmy, że są to drzwi numer 3 — za którymi jest pusto, i daje nam możliwość zmiany pierwotnego wyboru, lub pozostania przy swoim. Co powinniśmy zrobić, żeby zmaksymalizować szansę wygrania auta?

Pierwotne prawdopodobieństwo wygranej wynosiło $1/3$. Po otwarciu drzwi nr 3 musimy uznać, że wzrosło, tylko pytanie o ile?

Możnaby przyjąć, że teraz gra jakby zaczyna się od nowa, mamy tylko dwoje drzwi do wyboru, i prawdopodobieństwo wygranej byłoby równe $1/2$.

Ale można też przyjąć inny punkt widzenia, że prowadzący, wiedząc gdzie stoi samochód, otworzył inne drzwi, w ten sposób przekazując nam część swojej wiedzy. Prawdopodobieństwo, że wygrana jest za drzwiami nr 2 lub 3 wynosiło $2/3$, i teraz nadal tyle wynosi, ponieważ wynika to z losowego jej rozmieszczenia. Tylko my teraz wiemy, których z drzwi 2 lub 3 nie należy wybierać.

Który z powyższych punktów widzenia jest słuszny? Czy to jest tylko kwestia naszego subiektywnego wyboru który punkt widzenia przyjmiemy?

Otóż nie, jest to rzecz najzupełniej obiektywna. Można przeprowadzić serię eksperymentów, i obliczyć prawdopodobieństwo znalezienia samochodu za drzwiami pierwotnie wybranymi, i za „tymi drugimi". Obliczona wartość prawdopodobieństwa może potwierdzić słuszność jednego możliwych wyjaśnień.[2]

Musimy posługiwać się prawdopodobieństwem warunkowym, ilekroć chcemy wyliczyć prawdopodobieństwo jakiegoś zdarzenia w sytuacji, gdy posiadamy jakąś wiedzę o innych, być może zależnych zdarzeniach. $P(A)$ jest poprawnym prawdopodobieństwem zdarzenia A o ile nie posiadamy żadnej wiedzy. Jeśli jednak wiemy, że B, to poprawnym prawdopodobieństwem zdarzenia A jest $P(A|B)$, a gdybyśmy dowiedzieli się, że jeszcze C, to musimy już posługiwać się prawdopodobieństwem $P(A|B \wedge C)$. W ten sposób możemy uważać, że prawdopodobieństwo bezwarunkowe $P(A)$ jest prawdopodobieństwem warunkowym $P(A|)$ w sytuacji, gdy nie posiadamy żadnej wiedzy.

Prawdopodobieństwa warunkowe można obliczać z tablicy łącznego rozkładu prawdopodobieństwa JPD za pomocą wzoru (3).
Jednak nie tak się zwykle robi.

[2] Oczywiście słuszne jest drugie wyjaśnienie, i zmiana wyboru na drzwi nr 2 zwiększa szanse wygranej do $2/3$.

Reguła Bayesa

Z dwukrotnego zastosowania wzoru (3) możemy uzyskać następującą prostą zależność, zwaną regułą Bayesa, będącą podstawą wielu procesów wnioskowania probabilistycznego:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \tag{7}$$

Dlaczego ta reguła ma znaczenie? Wróćmy do przykładu z pacjentem z objawami SARS, niezwykle groźnej choroby. Załóżmy, że u pacjenta przeprowadzono test na obecność wirusa, i wypadł on pozytywnie. Czy pacjenta należy koniecznie hospitalizować i rozpocząć leczenie? Okazuje się, że to zależy!

Przeprowadzony test nigdy nie jest całkowicie niezawodny. Jeśli jest dobry, to zapewnia wysokie prawdopodobieństwo wyniku pozytywnego (potwierdzającego obecność wirusa) w przypadkach, kiedy wirus rzeczywiście jest obecny. Równie ważne okazuje się wymaganie, żeby test z wysokim prawdopodobieństwem dawał wynik negatywny w przypadkach braku wirusa.

Czyli test zapewnia odpowiednio wysoką wartość $P(T^{\oplus}|\mathit{SARS})$ jak również $P(T^{\ominus}|\neg\mathit{SARS})$. Jednak to co interesuje lekarza, a przede wszystkim jego pacjenta, to jest wartość $P(\mathit{SARS}|T^{\oplus})$ albo $P(\neg\mathit{SARS}|T^{\ominus})$.

Reguła Bayesa — przykład

Jak widać, aby na podstawie przeprowadzonego badania próbki krwi wnioskować o prawdopodobieństwie choroby, konieczne jest odwrócenie warunków prawdopodobieństwa warunkowego, czyli właśnie skorzystanie z reguły Bayesa.

Załóżmy, że test na SARS daje wynik pozytywny w 95% przypadków obecności wirusa. W przypadku braku wirusa, test daje wynik negatywny (tzn. prawidłowy) w 90% przypadków. Wiadomo, że wirus występuje u 0.01% ogółu ludności.

$$\begin{aligned} P(\mathit{SARS}) &= 0.0001 \\ P(T^{\oplus}|\mathit{SARS}) &= 0.95 \\ P(T^{\ominus}|\neg\mathit{SARS}) &= 0.90 \end{aligned}$$

Rozważmy pacjenta, dla którego test dał wynik pozytywny.
Jakie jest prawdopodobieństwo, że pacjent ma SARS?
Musimy obliczyć $P(\mathit{SARS}|T^{\oplus})$!

$$P(\mathit{SARS}|T^{\oplus}) = \frac{P(T^{\oplus}|\mathit{SARS})P(\mathit{SARS})}{P(T^{\oplus})}$$

brakuje nam wartości $P(T^{\oplus})$, którą możemy wyliczyć z wzoru (5, str.10):

$$P(T^{\oplus}) = P(T^{\oplus}|\mathit{SARS})P(\mathit{SARS}) + P(T^{\oplus}|\neg\mathit{SARS})P(\neg\mathit{SARS})$$

$$P(T^{\oplus}) = 0.95 \times 0.0001 + 0.10 \times 0.9999$$
$$P(T^{\oplus}) = 0.000095 + 0.09999$$
$$P(T^{\oplus}) = 0.100085$$

i w końcu obliczamy interesującą wartość:

$$P(\mathit{SARS}|T^{\oplus}) = \frac{0.95 \times 0.0001}{0.100085}$$

$$P(\mathit{SARS}|T^{\oplus}) = 0.00094919$$

czyli poniżej jednego promila! Prawie dziesięć razy powyżej przeciętnej, ale czy dosyć aby rozpocząć być może kosztowną i nieobojętną dla zdrowia terapię??

Widać, że posiadając wiedzę **przyczynowo-skutkową** o mechanizmach choroby i wynikach testów, możemy obliczać interesujące nas prawdopodobieństwa **diagnostyczne**. Może nasuwać się pytanie, dlaczego trzeba te prawdopodobieństwa każdorazowo obliczać; czemu producent testu podaje wartości $P(T^{\oplus}|\mathit{SARS})$ i $P(T^{\ominus}|\neg\mathit{SARS})$, zamiast od razu wygodnie wyliczyć potrzebną użytkownikowi testu wartość $P(\mathit{SARS}|T^{\oplus})$?

Odpowiedź wynika z łatwiejszej dostępności danych przyczynowych niż diagnostycznych, których określanie może być złożone. Na przykład, gdyby wystąpił nagły wzrost zachorowań na *SARS* (epidemia — Epi), to wartość $P(\mathit{SARS})$ gwałtownie by wzrosła, a za nią również $P(\mathit{SARS}|T^{\oplus})$. Jednak wartość $P(T^{\oplus}|\mathit{SARS})$ powinna pozostać bez zmian, ponieważ odzwierciedla ona jedynie fizjologię choroby i działanie testu. Zatem wcześniejsze obliczenia pozostaną słuszne, po uwzględnieniu zwiększonej wartości $P(\mathit{SARS})$. [3]

[3] Zmianie ulegnie wtedy również wartość $P(T^{\oplus})$ obliczane jako $P(T^{\oplus}|Epi)$, jednak możemy ją obliczyć:

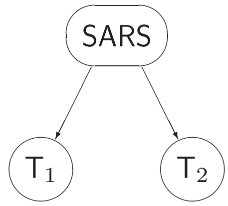
$$P(T^{\oplus}|Epi) = P(T^{\oplus}|\mathit{SARS}, Epi)P(\mathit{SARS}|Epi) + P(T^{\oplus}|\neg\mathit{SARS}, Epi)(1 - P(\mathit{SARS}|Epi))$$

Reguła Bayesa — niezależność warunków

Powróćmy do naszego pacjenta, z pozytywnym wynikiem testu SARS. Być może otrzymana wartość prawdopodobieństwa nie jest wystarczająca do definitywnego stwierdzenia choroby, i zakwalifikowania pacjenta na leczenie. Wyobraźmy sobie, że istnieje drugi test o innych charakterystykach, i oczywiście o innym rozkładzie prawdopodobieństw.

Jeśli potraktujemy ten drugi test jako trzecią zmienną losową, to po uzyskaniu jego wyniku musimy obliczać prawdopodobieństwo SARS jako uwarunkowane wynikami obu testów. W ogólnym przypadku wzór na $P(\mathit{SARS}|\text{Test}_1^{\oplus} \wedge \text{Test}_2^{\oplus})$ będzie uwzględniał zależności pomiędzy wynikami obu testów. To oznacza konieczność obliczania, w przypadku wielu zmiennych losowych, dużej liczby prawdopodobieństw, co teoretycznie niweczy zalety użycia prawdopodobieństwa warunkowego zamiast JPD.

Ważnym elementem jest zauważenie, że wyniki obu testów zależą tylko od występowania wirusa, a nie od siebie nawzajem. Po uwzględnieniu tej obserwacji upraszczają się wzory, i potrzebne jest tylko wyliczenie prawdopodobieństw warunkowych wyników poszczególnych testów.



$$
\begin{aligned}
P(\mathit{SARS}|T_1^{\oplus}, T_2^{\oplus}) &= \frac{P(\mathit{SARS} \cap T_1^{\oplus} \cap T_2^{\oplus})}{P(T_1^{\oplus} \cap T_2^{\oplus})} \\
&= \frac{P(T_1^{\oplus} \cap T_2^{\oplus}|\mathit{SARS})P(\mathit{SARS})}{P(T_1^{\oplus} \cap T_2^{\oplus})} \\
&= \frac{P(T_1^{\oplus}|\mathit{SARS})P(T_2^{\oplus}|\mathit{SARS})P(\mathit{SARS})}{P(T_1^{\oplus})P(T_2^{\oplus})}
\end{aligned}
$$

Gdyby oba testy miały identyczne charakterystyki jak w obliczonym wcześniej przykładzie, to otrzymany pozytywny wynik z obu testów wskazywałby na prawdopodobieństwo choroby równe 0.009025, czyli już prawie 100 razy większe niż przy braku informacji.

Sieci przekonań

Łączny rozkład prawdopodobieństwa pozwala znajdować odpowiedzi na pytania dotyczące dziedziny problemowej, lecz trudno się nim posługiwać przy wielu zmiennych. Ponadto, określanie prawdopodobieństw dla zdarzeń atomowych może wymagać przeprowadzenia kompleksowych badań statystycznych.

Jak wynika z przedstawionego przykładu z wirusem SARS, można zbudować graf przedstawiający rzeczywiste zależności między zmiennymi losowymi, i po wyznaczeniu ich prawdopodobieństw warunkowych efektywnie obliczać prawdopodobieństwa innych zdarzeń. Ściślej, **siecią przekonań** (*belief network, Bayesian network, probabilistic network*) nazywamy następujący graf:

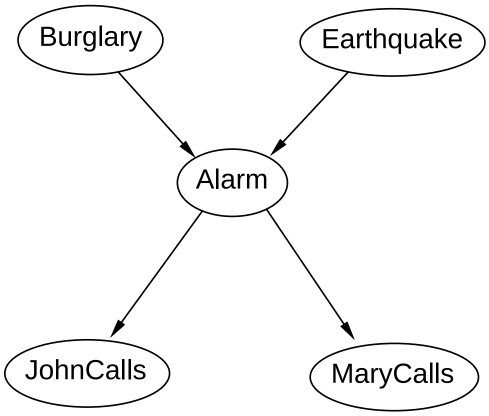
- węzłami sieci są zmienne losowe,
- łuki sieci są skierowane, i łuk $X \longrightarrow Y$ ma intuicyjne znaczenie: „zmienna X ma bezpośredni wpływ na Y",
- każdy węzeł X ma związaną z nim tablicę prawdopodobieństw warunkowych określających wpływ wywierany na X przez jego rodziców (poprzedników w grafie),
- sieć nie może mieć cykli (skierowanych).

Budowa sieci polega na wyznaczeniu jej topologii, oraz prawdopodobieństw warunkowych dla węzłów, dla których istnieją bezpośrednie zależności.

Idea sieci przekonań zasadza się na względnej łatwości, z jaką możemy wyznaczać prawdopodobieństwa tych bezpośrednich zależności. Prawdopodobieństwa innych zdarzeń będziemy wyznaczać już z gotowej sieci.

Sieci przekonań — przykład

Przykład: system alarmowy w mieszkaniu, reaguje na włamania oraz, niestety, również na drobne trzęsienia (ziemi). Sąsiedzi John i Mary są umówieni, żeby zadzwonić do właściciela gdy usłyszą alarm. John jest nadgorliwy i bierze różne zdarzenia (np. dzwonek telefonu) za sygnał alarmowy (i wtedy zawsze dzwoni). Mary rozpoznaje alarm poprawnie, lecz często słucha głośnej muzyki i może go w ogóle nie dosłyszeć. Będzie nas interesować określenie prawdopodobieństwa tego, że w razie włamania ktoś zadzwoni, żeby nas zawiadomić, jak również tego, że zawiadomienie o włamaniu może być fałszywe.



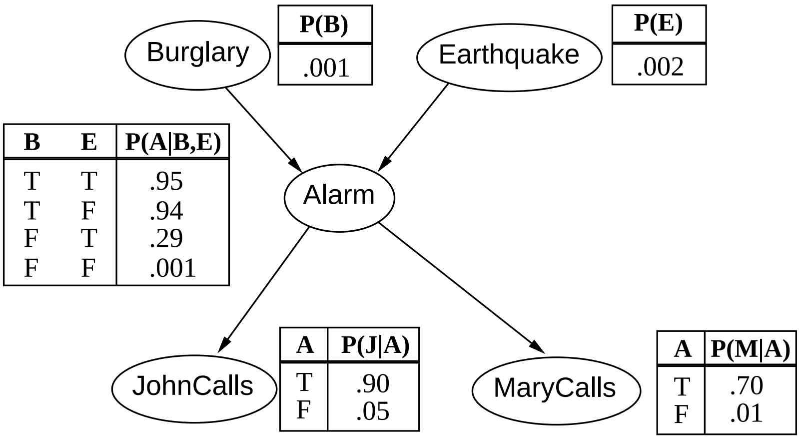
Zauważmy, że ignorujemy tutaj wiele istotnych czynników, np. to czy Mary słucha w danej chwili muzykę czy nie, ponieważ to może być niemożliwe do ustalenia, i reprezentujemy całą niepewność i nieokreśloność sytuacji w prawdopodobieństwach warunkowych danych zmiennych losowych.

Ogólnie, musimy określić prawdopodobieństwa warunkowe dla zmiennych losowych w zależności od innych zmiennych, które są reprezentowane w naszej sieci. Konkretnie, musimy określić prawdopodobieństwa warunkowe dla każdej wartości zmiennej losowej X dla wszystkich kombinacji wartości zmiennych losowych, od których zmienna X zależy.

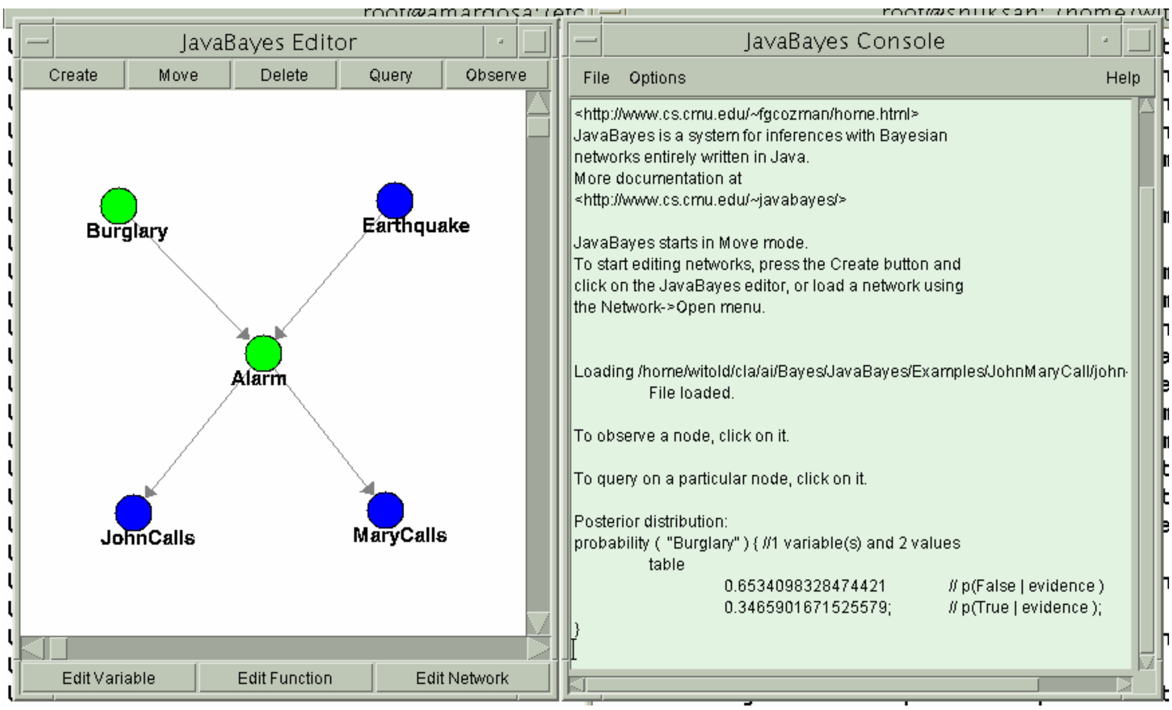
Burglary (włamanie)	*Earthquake* (trz.ziemi)	**P**(*Alarm*|*Burglary,Earthquake*) True	False
True	True	0.950	0.050
True	False	0.940	0.060
False	True	0.290	0.710
False	False	0.001	0.999

Zestaw takich prawdopodobieństw tworzy **tablicę prawdopodobieństw warunkowych CPT** (*conditional probability table*). Dla zmiennych, które nie zależą od niczego musimy określić prawdopodobieństwa a priori. W takim przypadku tabela CPT ma tylko jeden rząd z wartościami prawdopodobieństw dla możliwych wartości zmiennej losowej (sumującymi się do 1.0).

Kompletna sieć przekonań dla przykładu z systemem alarmowym:



Przykładowa sieć w systemie JavaBayes



Konstrukcja sieci przekonań

Można widzieć sieć przekonań jako pewną reprezentację łącznego rozkładu prawdopodobieństw zmiennych losowych. Ten rozkład jest tabelą określającą pojedyncze prawdopodobieństwa zdarzeń typu $P(X_1 = x_1, ..., X_n = x_n)$. W skrócie zapisujemy to prawdopodobieństwo jako: $P(x_1, ..., x_n)$. Korzystając z faktu, że prawdopodobieństwo koniunkcji możemy wyrazić przez iloczyn prawdopodobieństw warunkowych przez prawdopodobieństwa zależności (wzór (3) na stronie 9), mamy:

$$P(x_1, ..., x_n) = \prod_{i=1}^{n} P(x_i|\text{Poprzedniki}(X_i)) \tag{8}$$

Zatem każda pozycja w tablicy prawdopodobieństwa łącznego jest iloczynem odpowiednich elementów w tablicy CPT, czyli CPT jest elementarną reprezentacją łącznego rozkładu prawdopodobieństwa JPD.

Dla poprzedniego przykładu, obliczmy prawdopodobieństwo, że rozległ się alarm, przy czym nie wystąpiło ani trzęsienie ziemi ani włamanie, ale oboje John i Mary zadzwonili.

$$\begin{aligned}
&P(J \land M \land A \land \neg B \land \neg E) \\
&= P(J|A)P(M|A)P(A|\neg B \land \neg E)P(\neg B)P(\neg E) \\
&= 0.90 \times 0.70 \times 0.001 \times 0.999 \times 0.998 \\
&= 0.00062
\end{aligned}$$

W ten sposób można odpowiadać na dowolne zapytania wyliczając pozycje łącznego rozkładu prawdopodobieństwa, np. przez wyliczenie całej tabeli JPD (*joint probability distribution*), z tabeli CPT. Jednak jeśli mamy wiele zmiennych to ta metoda jest bardzo pracochłonna i istnieją bardziej bezpośrednie i efektywne metody.

Algorytm budowy sieci przekonań

Otrzymany wzór na prawdopodobieństwo łączne można w ogólności przedstawić w następujący sposób:

$$
\begin{aligned}
P(x_1, ..., x_n) &= P(x_n | x_{n-1}, ..., x_1) P(x_{n-1}, ..., x_1) \\
&= ... \\
&= P(x_n | x_{n-1}, ..., x_1) \cdots P(x_2 | x_1) P(x_1) \\
&= \prod_{i=1}^{n} P(x_i | x_{i-1}, ..., x_1)
\end{aligned}
$$

Z porównania powyższego równania z równaniem (8) na stronie 25 możemy wyciągnąć wniosek, że:

$$\mathbf{P}(X_i | X_{i-1}, ..., X_1) = P(X_i | \text{Poprzedniki}(X_i)) \tag{9}$$

o ile tylko Poprzedniki$(X_i) \subseteq \{x_{i-1}, ..., x_1\}$

Ostatnią zależność łatwo jest osiągnąć numerując zmienne losowe zgodnie z częściowym porządkiem określonym przez łuki na sieci.

Te wyniki można zinterpretować w ten sposób, że sieć przekonań jest poprawną reprezentacją dziedziny pod warunkiem, że każdy węzeł jest **warunkowo** niezależny od swoich (dalszych) przodków, prócz bezpośrednich rodziców. (Inaczej: cała zależność jednej zmiennej od drugiej wyrażona jest w jawnej zależności od rodziców, inne zależności są wtórne.)

Wskazuje nam to w jaki więc sposób musimy konstruować sieci przekonań. Intuicyjnie, bezpośrednimi rodzicami węzła X_i powinny być wszystkie te węzły $X_1, ..., X_{i-1}$, które bezpośrednio wpływają na X_i, i żadne inne.

Dla zmiennych z przedstawionego wcześniej przykładu, można przypuszczać, że B wpływa na M, ale nie wpływa bezpośrednio. Można to podsumować następująco:

$$\mathbf{P}(M | J, A, B, E) = \mathbf{P}(M | A)$$

Ogólny algorytm konstrukcji sieci:

1. Wybierz zbiór zmiennych losowych X_i opisujących dziedzinę.
2. Wybierz porządek na tych zmiennych.
3. Dopóty, dopóki pozostały jeszcze zmienne:
   (a) Wybierz zmienną X_i, która zależy bezpośrednio tylko do zmiennych już wybranych, i dodaj do sieci węzeł dla niej
   (b) Ustal Poprzedniki(X_i) jako minimalny zbiór węzłów już umieszczonych w sieci, tak by była spełniona własność niezależności (9) na stronie 27
   (c) Określ prawdopodobieństwa warunkowe dla X_i.

Algorytm ten gwarantuje, że sieć nie będzie miała cykli, jak również, że nie będą określane żadne nadmiarowe wartości prawdopodobieństw, które mogłyby naruszyć aksjomaty prawdopodobieństwa (z wyjątkiem jednej dopełniającej liczby w każdym rzędzie).

Zwartość sieci i nieoptymalne porządki węzłów

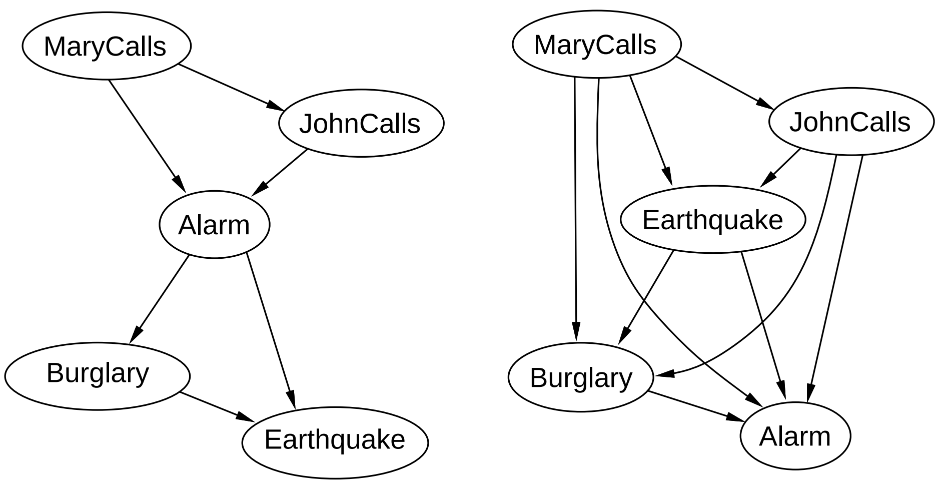
Sieci przekonań są zwykle w naturalny sposób zwarte, ponieważ zwykle tylko niewielka liczba zmiennych losowych, spośród być może wielkiej ich liczby, wpływa na każdą pojedynczą zmienną.

> Na przykład, dla sieci o $n = 20$ węzłach, w której maksymalna liczba zależności dla węzłów wynosi $k = 5$, dla zmiennych binarnych tablice CPT dla węzłów będą miały maksymalnie $2^k = 32$ wartości prawdopodobieństwa do określenia, co daje dla całej sieci $n \times 2^k = 640$ wartości. Kompletna tablica JPD ma $2^n \approx 1,000,000$ wartości.

Ta oszczędność jest możliwa tylko wtedy, gdy zmienne mają bezpośrednią zależność tylko od pewnej (małej) liczby innych zmiennych, czyli warunkową niezależność od większości zmiennych. Gdyby zmienne w sieci miały zależności od wszystkich innych zmiennych to reprezentacja tych zależności w postaci sieci przekonań miałaby niewielki sens. Jednak w większości zagadnień praktycznych istnieje silna struktura problemu, którą można w efektywny sposób wykorzystać w budowie sieci.

Czasami można to osiągnąć, ignorując pewne zależności o niewielkim prawdopodobieństwie (np. bezpośredni wpływ trzęsień ziemi na fakt czy sąsiedzi zadzwonią czy nie, który może być znaczący lub nie). Wprowadza to pewną niedokładność w sieci, ale może znacznie uprościć jej konstrukcję.

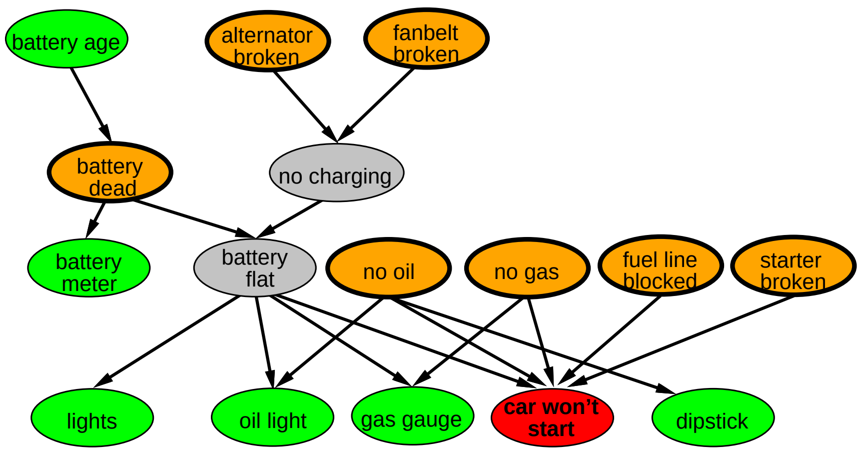
Jednak znacznie bardziej na zwartość wpływa poprawne określenie porządku zmiennych.



Powyższe przykłady ilustrują wyniki otrzymane z konstrukcji sieci przy niewłaściwej kolejności rozpatrywania węzłów (np. M,J,A,B,E).

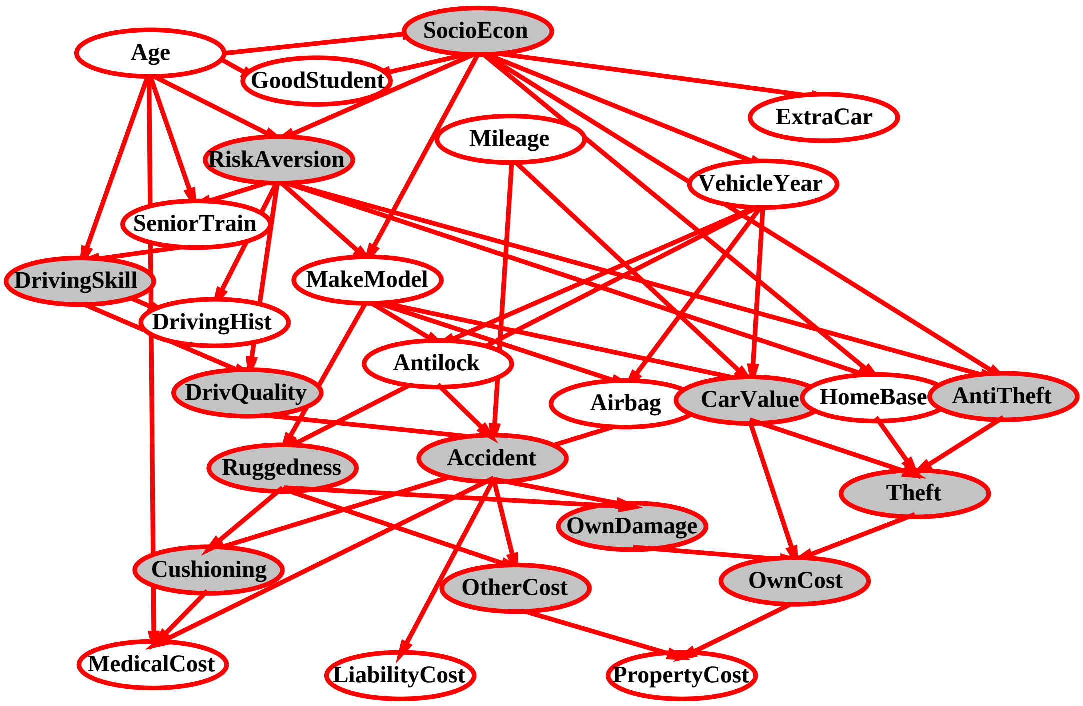
Wnioskowanie w sieciach przekonań — przykład

Przykład wnioskowania diagnostycznego: mamy sieć opisującą podstawowe zjawiska towarzyszące uruchamianiu samochodu. Stan początkowy: auto nie chce odpalić. Mamy zdarzenia obserwowalne (węzły zielone), i zdarzenia identyfikujące konkluzje wnioskowania diagnostycznego (przyczyny awarii — węzły pomarańczowe). Węzły szare są węzłami wewnętrznymi, które, opisując pewne zjawiska wewnętrzne i zależności, pozwalają zmniejszyć wielkość sieci.



Wnioskowanie w sieciach przekonań — przykład (2)

Bardziej rozbudowany przykład, służący do przewidywania kosztów odszkodowania (*medical, liability, property*), na podstawie danych z formularza ubezpieczeniowego (pozostałe niewyszarzone węzły).



Procesy wnioskowania w sieciach przekonań

Mając skonstruowaną sieć przekonań możemy prowadzić różne procesy wnioskowania ogólnie podpadające pod następujący schemat. Część zmiennych losowych uznajemy za **zmienne faktowe**, i mamy dla nich dokładne (pewne) wartości. Inny zbiór zmiennych uznajemy za **zmienne zapytaniowe** i chcemy dla nich obliczyć prawdopodobieństwo warunkowe względem zmiennych faktowych $P(\text{Zapytaniowe}|\text{Faktowe})$.

Jest naturalne, że zmiennymi faktowymi będą zmienne związane z obserwacjami agenta, a zmiennymi zapytaniowymi zmienne istotne dla podejmowania przez agenta decyzji o jego akcjach. Jest to przykład **wnioskowania diagnostycznego**.

Takie wnioskowanie nie zawsze zgodne jest z intuicjami ludzi odnośnie prawdopodobieństwa. Na przykład, wiedząc, że J, chcemy obliczyć $P(B|J)$. Mylny tok rozumowania: jeśli alarm dzwoni to John prawie na pewno do nas zadzwoni, a system alarmowy jest prawie 100%-owo dokładny, zatem $P(B|J)$ będzie duże, prawie 90%. Jednak to wnioskowanie nie bierze pod uwagę faktu, że trzęsienia ziemi też powodują włączenie się systemu alarmowego (i telefon Johna), a są

o wiele bardziej ($50\times$) prawdopodobne. W rzeczywistości, gdy policzymy dokładnie $P(B|J)$ to otrzymamy wartość 0.016.

Załóżmy dalej, że zaraz po telefonie Johna zadzwoniła do nas Mary. Chcemy teraz obliczyć $P(B|J \wedge M)$, która to wartość wzrasta tylko do 0.29. Podobnie $P(E|J \wedge M) = 0.18$, gdy $P(A|J \wedge M) = 0.76$.

Wnioskowanie diagnostyczne nie jest jedynym rodzajem wnioskowania. Innym rodzajem jest wnioskowanie **przyczynowo-skutkowe** polegające na określaniu prawdopodobieństwa skutków, gdy znamy przyczyny. Na przykład $P(J|B) = 0.86$, $P(M|B) = 0.67$.

Jeszcze innym rodzajem wnioskowania jest wnioskowanie **międzyprzyczynowe**, np. wiemy A, określamy najpierw $P(B|A) = 0.376$. Jednak gdybyśmy wiedzieli równocześnie, że E, wtedy $P(B|A \wedge E)$ idzie w dół i wynosi tylko 0.003. Pomimo, iż włamania i trzęsienia ziemi są niezależne, wiedza o wystąpieniu jednego zmniejsza szanse wystąpienia drugiego.

Jak również możliwe są inne schematy wnioskowania, np. $P(A|J \wedge \neg E) = 0.03$ albo $P(B|J \wedge \neg E) = 0.017$.

Zastosowania sieci przekonań

Poza wyliczaniem wartości przekonań o wystąpieniu pewnych faktów, sieci przekonań mogą służyć do innych procesów:

- Podejmowanie decyzji łącznie na podstawie prawdopodobieństw na sieci i innych możliwości agenta.
- Określanie jakie inne fakty należy poznać aby uzyskać użyteczną informację.
- Przeprowadzenie analizy czułości w celu określenia, które elementy modelu mają największy (krytyczny) wpływ na wyniki.
- Wyjaśnianie i prezentacja wyników wnioskowania probabilistycznego użytkownikowi.