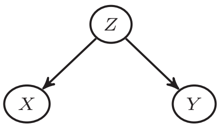
# W sieci Bayesa

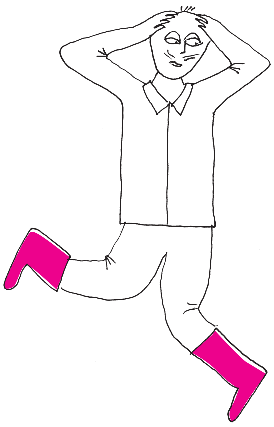
*Łukasz RAJKOWSKI*

Rozpoczniemy od żartobliwej, acz pouczającej historyjki: podczas rozmowy dwóch stałych bywalców lokalnego baru jeden z nich mówi do drugiego „Noszenie kaloszy jest bardzo niezdrowe; ilekroć budzę się rano i mam je na nogach, boli mnie głowa”. Łatwo wskazać lukę w przedstawionym rozumowaniu (psując odrobinę dowcip, ale czegóż nie robi się dla dobra nauki): choć niewątpliwie spędzenie nocy w kaloszach i poranny ból głowy często występowały wspólnie, nie można na tej podstawie powiedzieć, że to pierwsze jest przyczyną drugiego. Oba mają po prostu wspólną praprzyczynę, jaką jest niechlubne zamiłowanie autora uwagi do wysokoprocentowych trunków. Zależność tę można przedstawić graficznie, tak jak na rysunku 1, na którym $X$, $Y$ i $Z$ związane są odpowiednio z kaloszami na nogach, porannym bólem głowy i przesadną konsumpcją poprzedniego wieczora. Zwróćmy uwagę, że (1) nie każdy wieczór spędzony jest w barze (na szczęście), (2) nie każdy wieczór spędzony w barze skutkuje bolącą głową (kaloszami na nogach) następnego ranka i (3) nie zawsze boląca głowa (kalosze) są dowodem na wieczór spędzony w barze. Sugeruje to, aby na $X, Y, Z$ patrzeć jak na *zmienne losowe*, które określone wartości (powiedzmy 0 i 1) przyjmują z określonym prawdopodobieństwem. Jest zatem sensowne pytanie na przykład o prawdopodobieństwo obudzenia się z kaloszami na nogach bez bólu głowy oraz gdy nie byliśmy poprzedniego dnia w barze, czyli




Rys. 1

$$p_{X,Y,Z}(1,0,0) := \mathbb{P}(X=1, Y=0, Z=0).$$



Funkcja $p_{X,Y,Z}$ ma kilka funkcji „pochodnych”, można z niej, na przykład, odczytać prawdopodobieństwo obudzenia się z kaloszami

$$p_X(1) := p_{X,Y,Z}(1,0,0) + p_{X,Y,Z}(1,0,1) + p_{X,Y,Z}(1,1,0) + p_{X,Y,Z}(1,1,1).$$

Podobnie możemy zdefiniować $p_Y, p_Z$, a także $p_{X,Y}, p_{Y,Z}$ i $p_{X,Z}$ – za każdym razem sumujemy $p_{X,Y,Z}$ dla wszystkich możliwych wartości zmiennych, które nie występują w dolnym indeksie (otrzymujemy w ten sposób *rozkłady brzegowe*). Jesteśmy też w stanie obliczyć prawdopodobieństwo obudzenia się bez kaloszy pod warunkiem spędzenia wieczoru w barze, czyli

$$p_{X|Z}(1|1) = p_{X,Z}(1,1)/p_Z(1).$$

Analogicznie możemy określić funkcje $p_{A|B}$, $p_{A,B|C}$ czy $p_{A|B,C}$, gdzie pod $A, B, C$ można dowolnie wstawić $X, Y, Z$; otrzymamy w ten sposób *rozkłady warunkowe*. Zwróćmy uwagę na oczywistą algebraicznie równość

$$(*) \qquad p_{X,Y,Z} = p_Z \cdot p_{Y|Z} \cdot p_{X|Y,Z}.$$

Wspomniany na początku bywalec baru zaobserwował zależność między kaloszami i bólem głowy, możemy jednak podejrzewać, że zjawiska te są *niezależne pod warunkiem* pobytu w barze. Oznacza to, że jeśli dysponujemy informacją o sposobie spędzenia poprzedniego wieczoru, to wiedza o kaloszach na nogach o poranku nie wpłynie na naszą ocenę szansy na ból głowy. Wykorzystując wprowadzone przez nas oznaczenia, możemy to zapisać na trzy algebraicznie równoważne sposoby

$$p_{Y|X,Z} = p_{Y|Z}, \quad p_{X|Y,Z} = p_{X|Z}, \quad p_{X,Y|Z} = p_{X|Z}p_{Y|Z}$$

(wszystkie te równości można szybko sprowadzić do $p_{X,Y,Z}p_Z = p_{X,Z}p_{Y,Z}$). Równość $(*)$ możemy wówczas zapisać jako

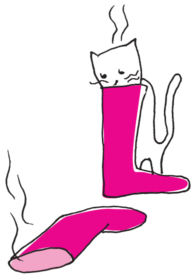
$$(**) \qquad p_{X,Y,Z} = p_Z \cdot p_{Y|Z} \cdot p_{X|Z}.$$

W takiej sytuacji mówimy, że funkcja $p_{X,Y,Z}$ *faktoryzuje się* względem grafu przedstawionego na rysunku 1 – przedstawia się bowiem jako iloczyn, w którym każdy czynnik jest postaci $p_{\text{zmienna}|\text{jej rodzice w grafie}}$. Jeśli w każdym z czynników ze zmiennych, po których warunkujemy, nie da się niczego uszczknąć (tak, jak przechodząc z zawsze prawdziwego $(*)$ do szczególnego $(**)$), to mówimy, że dany graf jest *siecią bayesowską* dla funkcji $p_{X,Y,Z}$. Jeśli graf z rysunku 1 jest

**Rozwiązanie zadania F 954.** Długofalowa granica zjawiska fotoelektrycznego określa pracę wyjścia $W$ elektronu, charakterystyczną dla metalu, z którego zrobiona jest fotokatoda: $W = h\nu_{pr}$, gdzie $h$ to stała Plancka. W rozpatrywanym przypadku $\nu_{pr} = 6 \cdot 10^{14}$ Hz. Skoro siatka, mająca potencjał $U = 3$ V zatrzymuje wszystkie fotoelektrony, to ich energia kinetyczna $mv^2/2 \leqslant eU$, gdzie $m$ i $e$ to masa i ładunek elektronu, a $v$ – jego prędkość, przy czym znak równości odpowiada elektronom o maksymalnej prędkości $v_{\max}$. Energia, potrzebna do wykonania pracy wyjścia i nadania elektronowi prędkości, pochodzi od fotonu o częstości $\nu$, więc zgodnie z prawem zachowania energii mamy

$$h\nu = W + \frac{mv_{\max}^2}{2}$$

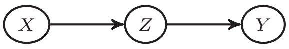
(równanie Einsteina). Stąd, korzystając z otrzymanych zależności, mamy $h\nu = h\nu_{pr} + eU$ i ostatecznie $\nu = (h\nu_{pr} + eU)/h$, a po podstawieniu danych liczbowych $\nu = 13{,}2 \cdot 10^{14}$ Hz.



siecią bayesowską dla naszych zmiennych, to połączenie $(*)$ (zawsze prawdziwe!) i $(**)$ (założenie o faktoryzacji) daje nam $p_{X|Y,Z} = p_{X|Z}$, czyli warunkową niezależność $X$ i $Y$ (oznaczaną jako $X \perp Y|Z$). Z drugiej strony mamy

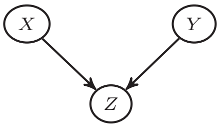
$$p_{X,Y}(x,y) = p_Z(0)p_{Y|Z}(y|0)p_{X|Z}(x|0) + p_Z(1)p_{Y|Z}(y|1)p_{X|Z}(x|1),$$

czego bez dodatkowych założeń nie moglibyśmy zapisać jako iloczynu dwóch funkcji, z których jedna jest zależna tylko od $x$, a druga tylko od $y$. W ogólności nie występuje zatem niezależność $X$ od $Y$, co zostało zaobserwowane przez bywalca baru.



Rys. 2

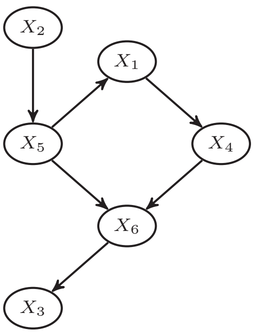
Zastanówmy się, jakie jeszcze dwukrawędziowe sieci bayesowskie mogą rządzić zachowaniem trzech zmiennych. Gdyby $p_{X,Y,Z}$ faktoryzowało się wedle grafu na rysunku 2, mielibyśmy $p_{X,Y,Z} = p_X p_{Z|X} p_{Y|Z}$. Odpowiadałoby to sytuacji, w której $X$ jest oceną ze sprawdzianu z trygonometrii, $Z$ oceną roczną z matematyki, a $Y$ – średnią na świadectwie. Wówczas $X$ i $Y$ są zależne (choć delikatnie), przestaje to jednak mieć miejsce, jeśli rozpatrzymy sytuację pod warunkiem oceny z matematyki (nie mylić z warunkiem z matematyki). Odpowiada to prostej do sprawdzenia (wykorzystując $p_{X,Y,Z} = p_X p_{Z|X} p_{Y|X,Z}$) równości $p_{Y|Z} = p_{Y|X,Z}$.



Rys. 3

Ostatnia możliwość jest przedstawiona na rysunku 3. Wówczas ma miejsce zależność $p_{X,Y,Z} = p_X p_Y p_{Z|X,Y}$. Zmienne $X, Y$ są niezależne (zachodzi bowiem $p_Y = p_{Y|X}$), zależność ta nie występuje jednak w ogólności po „zwarunkowaniu” po $Z$. Ta sieć znajduje odzwierciedlenie w sytuacji, gdy $Z$ oznacza uruchomienie się alarmu przeciwpożarowego, $X$ – ćwiczenia ewakuacji, a $Y$ – faktyczny pożar. Nie ma powodów, aby sądzić, że szansa na wystąpienie pożaru budynku była zależna od planu przeprowadzania w nim ćwiczeń przeciwpożarowych, jeśli jednak słyszymy alarm i wiemy, że na dany dzień nie były zapowiadane ćwiczenia, powinniśmy jak najszybciej (ale bez paniki!) wziąć nogi za pas.

Grafy przedstawione na rysunkach 1, 2, 3 zwykło się nazywać w specjalistycznym żargonie odpowiednio „widelcem” (*fork*), „łańcuchem” (*chain*) i „zderzaczem” (*collider*).



Rys. 4

Z powyższych rozważań wynika, że znajomość sieci bayesowskiej odpowiadającej funkcji $p_{X,Y,Z}$ pozwala na sformułowanie wniosków o (warunkowej) niezależności. Oczywiście, rozpatrywaliśmy trywialny przypadek trzech zmiennych – co się dzieje, kiedy jest ich więcej? Kandydat na sieć bayesowską dla zmiennych $X_1, \ldots, X_n$ musi być grafem skierowanym, i to nie byle jakim: wymagamy od niego, aby nie miał żadnych cykli (*Directed Acyclic Graph*, czyli DAG). Dla przykładu, DAG przedstawiony na rysunku 4 jest siecią bayesowską dla funkcji prawdopodobieństwa $p_{X_1,X_2,X_3,X_4,X_5,X_6}$, jeśli zachodzi

$$p_{X_1,X_2,X_3,X_4,X_5,X_6} = p_{X_2} \cdot p_{X_5|X_2} \cdot p_{X_1|X_5} \cdot p_{X_4|X_1} \cdot p_{X_6|X_5,X_4} \cdot p_{X_3|X_6}$$

i nie można usunąć z powyższego napisu żadnej ze zmiennych „warunkowych”. Jakie (warunkowe) niezależności spełniają wówczas zmienne $X_1, \ldots, X_6$? Okazuje się (i jest to jedno z podstawowych twierdzeń teorii sieci bayesowskich; jego dowód jest bardziej skomplikowany, niż może się na pierwszy rzut oka wydawać), że (warunkowa) niezależność wynika wówczas z występujących w sieci bayesowskiej *d-rozdzielności*. Mówimy, że wierzchołki $X$ i $Y$ w grafie $G$ są *d-połączone* przez zbiór wierzchołków $\mathcal{Z}$, jeśli istnieje pomiędzy nimi ścieżka $X = W_0 - W_1 - \ldots - W_k - W_{k+1} = Y$ (nie bierzemy pod uwagę kierunku krawędzi) taka, że dla dowolnego $1 \leqslant i \leqslant k$ zachodzi

- jeśli $W_{i-1} \to W_i \leftarrow W_{i+1}$, to $W_i$ lub któryś z jego potomków (dzieci, wnuków, prawnuków...) należy do $\mathcal{Z}$
- w przeciwnym przypadku $W_i$ nie należy do $\mathcal{Z}$.

Jeśli $X$, $Y$ nie są $d$-połączone przez $\mathcal{Z}$, to mówimy, że są przez $\mathcal{Z}$ $d$-rozdzielone. W sieci przedstawionej na rysunku 4 wierzchołki $X_2$ i $X_4$ są $d$-rozdzielone przez $X_1$, nie są jednak $d$-rozdzielone przez $\{X_1, X_3\}$. Zgodnie ze wspomnianym

**Rozwiązanie zadania F 953.**
Ponieważ $\sigma_{\mathrm{Ag}} = e\, n_{\mathrm{Ag}}\, \mu_{\mathrm{Ag}}$, a $\sigma_{\mathrm{Cu}} = e\, n_{\mathrm{Cu}}\, \mu_{\mathrm{Cu}}$ to

$$\frac{\sigma_{\mathrm{Ag}}}{\sigma_{\mathrm{Cu}}} = \frac{n_{\mathrm{Ag}}\, \mu_{\mathrm{Ag}}}{n_{\mathrm{Cu}}\, \mu_{\mathrm{Cu}}}.$$

Korzystając z tego, że $n = \rho N/M$, gdzie $\rho$ to gęstość, $M$ – ciężar atomowy, a $N$ – liczba Avogadro, otrzymujemy:

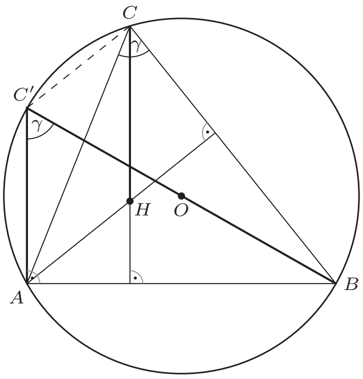
$$\frac{\sigma_{\mathrm{Ag}}}{\sigma_{\mathrm{Cu}}} = \frac{\rho_{\mathrm{Ag}}}{M_{\mathrm{Ag}}} \frac{M_{\mathrm{Cu}}}{\rho_{\mathrm{Cu}}} \frac{\mu_{\mathrm{Ag}}}{\mu_{\mathrm{Cu}}},$$

a stąd

$$\frac{\mu_{\mathrm{Ag}}}{\mu_{\mathrm{Cu}}} = \frac{\sigma_{\mathrm{Ag}}}{\sigma_{\mathrm{Cu}}} \frac{M_{\mathrm{Ag}}}{M_{\mathrm{Cu}}} \frac{\rho_{\mathrm{Cu}}}{\rho_{\mathrm{Ag}}}.$$

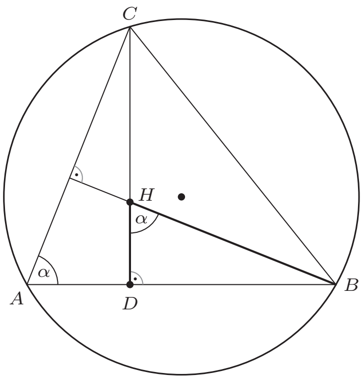
Podstawiając tablicowe wartości gęstości i ciężarów atomowych, dostajemy $\mu_{\mathrm{Ag}}/\mu_{\mathrm{Cu}} \approx 1{,}5$.

Podpowiedź do rysunku 2 z *Małej Delty*




Czworokąt $AHCC'$ jest równoległobokiem $|HC| = |AC'|$, $|\sphericalangle AC'B| = |\sphericalangle ACB| = \gamma$, bo są to kąty oparte na tym samym łuku.

Podpowiedź do rysunku 4 z *Małej Delty*




Z $\triangle HDB$: $\dfrac{|HD|}{\cos\beta} = \cos\alpha$.

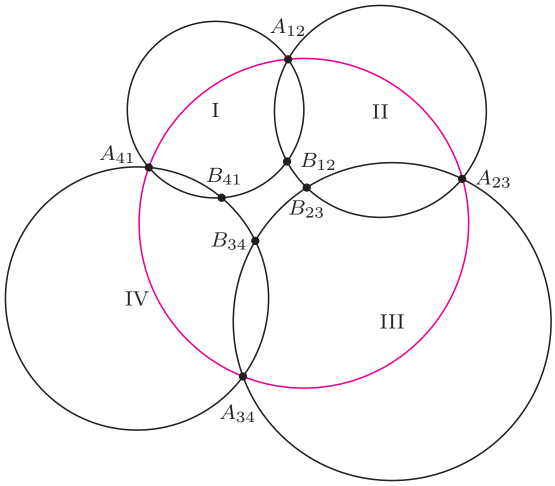
twierdzeniem oznacza to, że w tej sytuacji, jeśli znamy wartość $X_1$, to wiedza o $X_2$ nie dostarcza nam informacji o $X_4$; byłoby jednak inaczej, gdybyśmy na początku poznali jeszcze $X_3$. Odpowiada to „przepływowi informacji" w sieci: między $X_2$ a $X_4$ są dwie ścieżki: $\mathcal{S}_1 = X_2 \to X_5 \to X_1 \to X_4$ oraz $\mathcal{S}_2 = X_2 \to X_5 \to X_6 \leftarrow X_4$. Informacja nie przepływa przez $\mathcal{S}_2$, przepływa jednak przez $\mathcal{S}_1$. Zmieni się to, kiedy dowiemy się czegoś o $X_1$ – wówczas obie ścieżki są zablokowane. Gdy jednak uzyskamy również informację na temat $X_3$, to dowiemy się czegoś o $X_6$, przez co „odblokujemy" ścieżkę $\mathcal{S}_2$ (na tej samej zasadzie, jak w przykładzie o alarmie przeciwpożarowym).

Należy podkreślić, że choć z sieci bayesowskiej możemy odczytać występujące między zmiennymi warunkowe niezależności, nie możemy na jej podstawie wnioskować o ich braku. Przykładem może być sytuacja, w której $X, Y$ to wyniki dwóch niezależnych rzutów symetryczną monetą (gdzie orzeł kodowany jest przez 0, a reszka przez 1), natomiast $Z$ to odpowiedź na pytanie o równość $X$ i $Y$ (0 – fałsz, 1 – prawda). Sieć bayesowska odpowiadająca tym zmiennym była przedstawiona na rysunku 3. Pomimo tego, że wierzchołki $X$ i $Z$ są $d$-połączone, są one niezależne. Wedle tej zasady działa każdy godny szacunku kryminał: pojedyncza wskazówka nic nie mówi o przestępcy, ale uwzględnienie ich wszystkich pozwala na doprowadzenie go przed oblicze sprawiedliwości.

Czytelnik Pragmatyczny zapewne zaczął się już niecierpliwić – wprowadziliśmy mnóstwo nowych pojęć, a sformułowaliśmy tylko jedno twierdzenie, którego nawet nie udowodniliśmy; po co więc cały ten ambaras? Otóż sieci bayesowskie są bardzo poręcznym matematycznym wynalazkiem, pozwalającym w czytelny sposób przedstawiać zależności między zmiennymi losowymi. Są one użyteczne zwłaszcza wtedy, gdy chcemy zilustrować bezpośrednie związki między nimi. Działa to również w drugą stronę – choć na podstawie samych obserwacji nigdy nie jesteśmy w stanie zidentyfikować związków przyczynowo-skutkowych (jest to zasada, którą każda osoba przeprowadzająca analizę statystyczną powinna mieć na uwadze), zidentyfikowanie najprostszej sieci bayesowskiej można traktować jako krok w dobrym kierunku. Zauważmy bowiem, na przykład, że jeśli wśród zmiennych $X, Y, Z$ zaobserwujemy niezależność $X$ od $Y$, jednak zmienne te są zależne pod warunkiem $Z$, to wiemy, że mamy wówczas do czynienia ze „zderzaczem". Stanowi to pewną mglistą przesłankę (nie zawsze słuszną), że $X$ i $Y$ „składają się" na $Z$; wiemy jednak z całą pewnością, że $Z$ nie może być przyczyną ani $X$, ani $Y$ (wówczas bowiem mielibyśmy do czynienia z inną strukturą niezależności). Sprawę często komplikują tutaj tak zwane „ukryte zmienne" – jak, na przykład, wizyta w barze ostatniej nocy, która nie została uwzględniona podczas analizy...

## Jeśli...

Drogi Czytelniku, narysuj takie cztery okręgi, że I przecina się z II, II z III, III z IV i IV z I. Powstanie osiem punktów – wspólne punkty I i II nazwij $A_{12}$ i $B_{12}$ i podobnie pozostałe. Jeśli trafi Ci się tak, że – jak na rysunku – punkty $A_{12}$, $A_{23}$, $A_{34}$, $A_{41}$ leżą na jednym okręgu, to wówczas punkty $B_{12}$, $B_{23}$, $B_{34}$, $B_{41}$ też będą leżały na jednym okręgu (jak tutaj) lub na jednej prostej. Udowodnij to!




Podobnie gdyby punkty $A_{12}$, $B_{12}$, $A_{34}$, $B_{34}$ leżały na jednej prostej lub jednym okręgu, to wówczas pozostałe punkty też leżałyby na jednej prostej lub na jednym okręgu.

A czy te prawidłowości będą miały miejsce, gdy niektóre punkty $A_{ij}$ zamienisz z punktami $B_{ij}$?

*M. K.*