Katedra Sztucznej Inteligencji i Matematyki Stosowanej
WI ZUT Szczecin

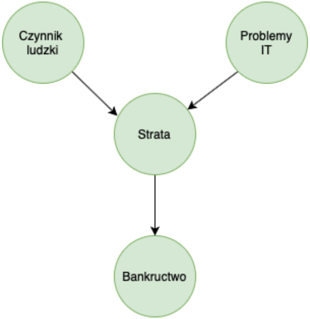
Sztuczna inteligencja i maszynowe uczenie w systemach interaktywnych

Joanna Kolodziejczyk
jkolodziejczyk@zut.edu.pl

November 9, 2020

Plan wykładu

Co prowadzi do bankructwa firmy?



Do czego służą sieci Bayes'a?

1. Używa się je do wyrażania wiedzy o bezpośrednich związkach między zmiennymi. Można uzyskać bardzo skomplikowane relacje.
2. Przechowują dane/prawdopodobieństwa w formie, w której łatwo wprowadza się dowody/obserwacje (evidence) i aktualizacje - ze względu na modułowość.

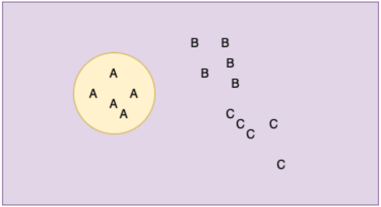
Do czego nie używa się sieci Bayes'a?

1. Nie są strukturami do reprezentacji problemów typu graph traversal search.
2. To nie to samo co sieć (network).
3. Są to graficzne modele probabilistyczne.
4. Można powiązać z modelami Markowa.

Zastosowania sieci Bayes'a

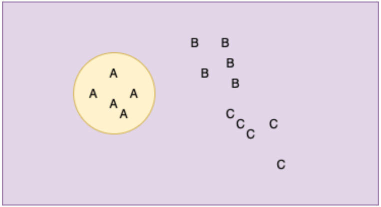
- Diagnoza medyczna już od lat 80tych. Modelują sposób myślenia lekarza.
- Przetwarzanie obrazów - etykietowanie pikseli z informacją o ich sąsiadach.
- Przetwarzanie języka naturalnego - związki pomiędzy elementami zdania, czy wartościami.

Prawdopodobieństwo



Jakie jest prawdopodobieństwo zdarzenia A?

Prawdopodobieństwo



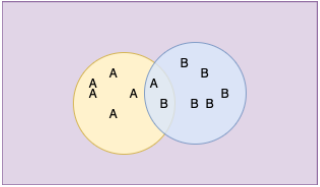
Jakie jest prawdopodobieństwo zdarzenia A?

Podstawowa idea: Policzyć wszystkie A i podzielić przez całkowitą liczbę możliwości:

$$P(A) = \frac{(\#A)}{(\#A + \#B + \#C)}$$

Prawdopodobieństwo warunkowe

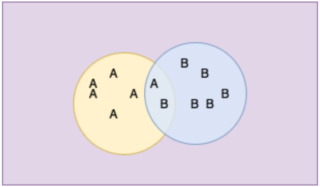
Rzeczywiste problemy posługują się zazwyczaj wieloma zmiennymi, które można zaobserwować jednocześnie.



Jakie jest prawdopodobieństwo zdarzenia A i B?

Prawdopodobieństwo warunkowe

Rzeczywiste problemy posługują się zazwyczaj wieloma zmiennymi, które można zaobserwować jednocześnie.



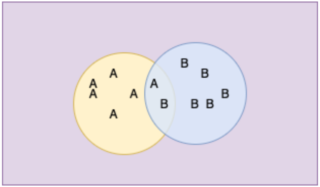
Jakie jest prawdopodobieństwo zdarzenia A i B?

Co musi się wydarzyć, aby A i B zostało zaobserwowane?
Powiedzmy, że najpierw musi się zdarzyć A, a potem gdy A już się dzieje to musi się zdarzyć B.
Zapisujemy:

$$P(A) * P(B|A) \text{ lub } P(B) * P(A|B)$$

Prawdopodobieństwo Bayes'a

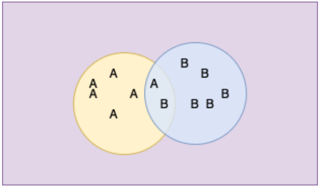
Zasada obserwacji dwóch zdarzeń: P(Jedno wydarzenie | Inne wydarzenie)



Jakie jest prawdopodobieństwo zdarzenia A i B?

Prawdopodobieństwo Bayes'a

Zasada obserwacji dwóch zdarzeń: P(Jedno wydarzenie | Inne wydarzenie)

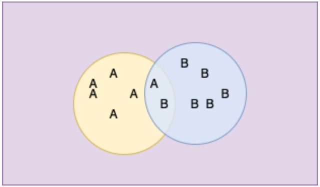


Jakie jest prawdopodobieństwo zdarzenia A i B?

$P(A \wedge B) = P(A) * P(B|A)$ lub $P(A \wedge B) = P(B) * P(A|B)$

Prawdopodobieństwo Bayes'a

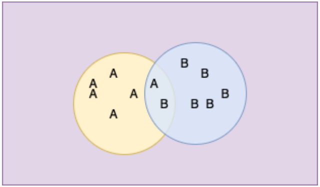
Zasada obserwacji dwóch zdarzeń: P(Jedno wydarzenie | Inne wydarzenie)



Jakie jest prawdopodobieństwo zdarzenia A i B?

$P(A \wedge B) = P(A) * P(B|A)$ lub $P(A \wedge B) = P(B) * P(A|B)$
Możemy wyizolować P(A|B) lub P(B|A) i podstawić do jednego ze wzorów.

Twierdzenie Bayes’a



Twierdzenie

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

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

Rozkłady łączne, warunkowe i brzegowe

Rozkład łączny prawdopodobieństwa

Opisuje, w jaki sposób dwie lub więcej zmiennych przyjmują jednocześnie pewne wartości. Aby uzyskać prawdopodobieństwo rozkładu łącznego zmiennych A i B, należy wziąć pod uwagę P(A = a i B = b).

Rozkład warunkowy prawdopodobieństwa

Analizuje, jak rozkładają się prawdopodobieństwa zmiennej A przy określonej wartości B: $P(A = a | B = b)$

Rozkład brzegowy prawdopodobieństwa

Brzegowy rozkład prawdopodobieństwa jest rozkładem, który wynika z uśredniania jednej zmiennej w celu uzyskania rozkładu prawdopodobieństwa drugiej zmiennej.

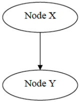
Sieci Bayes'a - definicja

Belief networks

Sieci Bayes'a *Bayesian networks* lub sieci probabilistyczne, sieci przyczynowe to popularna struktura grafu acyklicznego wykorzystywana we wnioskowaniu w warunkach niepewności z wykorzystaniem prawdopodobieństwa.

Zalety

- Umożliwienie zwartą reprezentację łącznego rozkładu reguł opartych na zmiennych.
- Możliwość reprezentacji niezależności relacji pomiędzy krawędziami i zmiennymi losowymi.

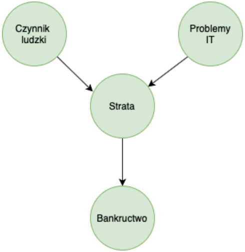


Sieci Bayes'a

Jest to graf, który można zdefiniować jako:

- Zmienne losowe tworzą węzły w sieci. Zmienne losowe mogą być binarne (trzęsienie ziemi: tak lub nie), dyskretne (trzęsienie ziemi: brak, drżenie, wibracje, poważne, i katastrofa) oraz ciągłe (trzęsienie ziemi: w skali Richtera).
- Zbiór skierowanych krawędzi łączących pary węzłów.
- Strzałka z węzła X do Y oznacza, że X jest rodzicem Y i X ma bezpośredni wpływ na Y.
- Każdy węzeł w korzeniu ma tablicę prawdopodobieństw a priori. Pozostałe węzły mają przypisane CPT (conditional probability table) tablicę prawdopodobieństw warunkowych. CTP wyraża liczbowo wpływ jaki mają rodzice na węzły potomne.
- W grafie nie ma cykli, czyli jest to acykliczny graf skierowany DAG.

Przykład sieci Bayes'a



CzL	P(CzL)
NIE	0.7
TAK	0.3

IT	P(IT)
NIE	0.8
TAK	0.2

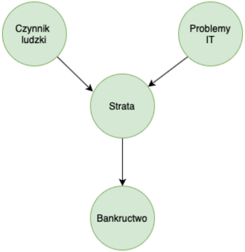
	Strata		
	Mała	Duża	Ups!
CzL i IT	0.5	0.3	0.2
CzL i nie IT	0.8	0.15	0.05
nie CzL i IT	0.8	0.1	0.1
nie CzL i nie IT	0.9	0.08	0.02

	Bankructwo	
	NIE	TAK
Mała strata	0.95	0.05
Duża strata	0.8	0.2
Ups! strata	0.5	0.5

Dyskusja

- Czy bankructwo zależy od czynnika ludzkiego?
- TAK
- Czy bankructwo zależy od czynnika ludzkiego jeżeli wiemy, że są problemy z IT?
- NIE, gdyż IT wpływa na bankructwo niezależnie i tak samo wpływa czynnik ludzki.
- Czy bankructwo zależy od czynnika ludzkiego jeżeli znamy stratę?
- NIE, gdyż mamy bliższego przodka czyli stratę, więc nie interesuje nas czynnik ludzki.

Przykład sieci Bayes’a



CzL	P(CzL)
NIE	0.7
TAK	0.3

IT	P(IT)
NIE	0.8
TAK	0.2

	Strata		
	Mała	Duża	Ups!
CzL i IT	0.5	0.3	0.2
CzL i nie IT	0.8	0.15	0.05
nie CzL i IT	0.8	0.1	0.1
nie CzL i nie IT	0.9	0.08	0.02

	Bankructwo	
	NIE	TAK
Mała strata	0.95	0.05
Duża strata	0.8	0.2
Ups! strata	0.5	0.5

Rozkład warunkowy prawdopodobieństwa

- Każdy węzeł, zmienna ma w sieci powiązany z nią rozkład warunkowy prawdopodobieństwa (CPD).
- Jeżeli węzeł ma rodzica, to powiązane z nim CPD reprezentowane jest jako P(wartość | wartość rodzica)
- Jeżeli węzeł nie ma rodzica, to CPD przedstawia bezwarunkowe prawdopodobieństwo P(wartość).

Dlaczego używać SB?

- Ze względu na budowę modularną, po dodaniu jednego węzła wystarczy dodać CPD w lokalnych węzłach. Propagacja wiedzy będzie wykonana podczas procesu wnioskowania.
- Można wprowadzić wiedzę lokalnie lub intuicyjnie, i pozwolić modelom wnioskować odległe relacje.
- Można obliczyć prawdopodobieństwo z danych, jeśli znamy struktury domeny.
- Można wyodrębnić strukturę i prawdopodobieństwa biorąc pod uwagę dostępną dużą moc obliczeniową, choć jest to nieoptymalne.

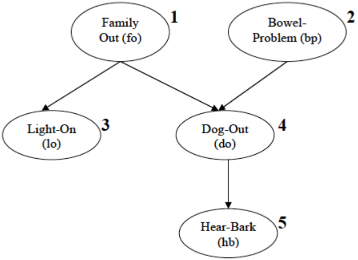
Przykład

Przykład pokazujący próbę zbudowania modelu sytuacji, w której wyodrębnia się związki przyczynowo-skutkowe. Zdarzenie i fakty/dowody nie zawsze są kompletne (niepewność), i możemy opisać je probabilistycznie.

Opis sytuacji

- Pan X wracając wieczorem do domu chciałby wiedzieć, czy ktoś jest w domu.
- Żona wychodząc z domu zostawia zapalone światło nad drzwiami wejściowymi.
- Jednakże zapala też światło, gdy spodziewa się gości.
- Pan X ma psa. Gdy nie ma nikogo w domu, to pies biega przed domem.
- Pies biega przed domem, gdy cierpi na niestrawność.
- Jeżeli pies jest wypuszczony, to pan X powinien usłyszeć jego warczenie (lub coś co wydaje się być warczeniem).
- Czasami może pomylić swojego psa z innymi psami słysząc ich warczenie.

Diagram do przykładu



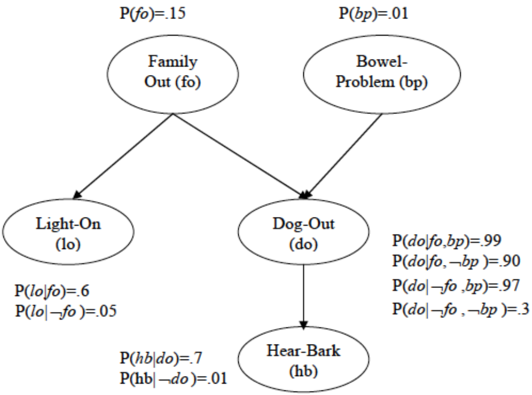
Sieć pokazuje zależność przyczynowo-skutkową i można dzięki niej:

- przewidzieć co się stanie (jeśli rodzina jest poza domem, pies jest wypuszczony)
- wnioskować z obserwowanych przyczyn, efekt (obserwuję, że świeci się światło i pies jest wypuszczony — wnioskuję: rodzina jest poza domem).

Dyskusja przykładu

1. Związki przyczynowe nie są bezwzględne. Bywa, że wszyscy wyjdą z domu i nie zostawią zapalonego światła lub nie wypuszczą psa.
2. W takich wypadkach SB nadal działa. Czy powinno się domniemywać, że nikogo nie ma w domu, gdy światło jest zapalone, ale nie słychać psa? A co w przypadku, gdy słychać psa, ale światło jest zgaszone?
3. Jeżeli znane są odpowiednie prawdopodobieństwa warunkowe to będzie można dokonywać wnioskowania o dowolnych zależnościach w grafie.
4. Niestety, w rzeczywistości nie zawsze będzie możliwe uzyskanie prawdopodobieństw dla wszystkich możliwych kombinacji zdarzeń (rozkład łączny).
5. Sieci Bayes'a pozwalają na wnioskowanie z małego zestawu prawdopodobieństw, odnoszących się jedynie sąsiednich węzłów.

Schemat z CPD



Na czym polega?

Podaje prawdopodobieństwo zdarzenia lub podaje prawdopodobieństwo zdarzenia jeżeli zaobserwowane są inne zdarzenia.

- Można obliczyć (ewaluować SB) prawdopodobieństwa warunkowego węzła przy założeniu, że zdarzenia z innych węzłów zostały zaobserwowane (dowód) (zmienne przyjęła wartość).
- Na przykład, zauważono, że światło się świeci ($lo = true$), ale nie słychać psa ($hb = false$), to $P(fo|lo, \neg hb) = 0.5$)
- W rzeczywistych przypadkach, sieci mają tysiące węzłów, które mogą być ewaluowane wielokrotnie, za każdym razem, gdy pojawia się nowy dowód.

Krawędzie w grafie i ich interpretacja

Niezależność

1. W sieciach Bayes'a krawędzie wyrażają zależności pomiędzy zmiennymi losowymi.
2. Założenie to implikuje jakie rozkłady prawdopodobieństw są niezbędne do wykonania wnioskowania.
3. Zdarzenie jest niezależne od wszystkich innych zdarzeń, które nie są jego potomkami.

Niezależność węzłów — dyskusja

- Zgodnie z teorią prawdopodobieństwa, by uzyskać pełny rozkład prawdopodobieństwa wymagane jest dla n binarnych zmiennych losowych $2^n - 1$ łącznych rozkładów prawdopodobieństw.
- Zatem pełny rozkład dla rozpatrywanego przykładu wymagałby 31 wartości, a zostało określone tylko 10.
- Wynika to z wbudowanego założenia o niezależności.
- Zatem im większa sieć tym większy zysk, np. dla 10 węzłów pełny rozkład to 1023 wartości, a w sieci trzeba podać tylko 21 (zmienne binarne).

Niezależność zmiennych — przykład

Czy zmienne losowe family-out i hear-bark są niezależne?

Intuicja podpowiada, że NIE są niezależne, bo jeśli rodzina opuszcza dom, to jest bardziej prawdopodobne, że pies jest wypuszczony i usłyszymy jego warczenie.

Czy zmienne losowe family-out i hear-bark są niezależne jeżeli dog-out?

Pytamy czy:

$$P(hb|fo, do) \stackrel{?}{=} P(hb|do)$$

Są niezależne, to, że słychać warczenie zależy tylko od zaobserwowanego faktu, że pies jest wypuszczony i nie ma nic wspólnego z tym czy ktoś jest, czy nie w domu.

Interpretacja krawędzi

Interpretacja przyczynowa krawędzi grafu

Nieobecność rodziny ma bezpośrednie połączenie, związek przyczynowy z byciem psa na zewnątrz, co z kolei ma bezpośredni związek z jego warczeniem, bo w grafie istnieje bezpośrednie połączenie.
Wykorzystuje się założenie niezależności, które narzuca interpretacja przyczynowa.

Uwaga!

Gdyby chcieć uzyskać bezpośrednią zależność pomiędzy słyszeniem warczenia psa i bycia rodziny poza domem (bo pies częściej warczy, gdy rodzina jest poza domem niż jeżeli w nim pozostała), należałoby dodać bezpośrednie połączenie pomiędzy tymi węzłami.

Zasada niezależności

Zasada określająca zależność w sieciach Bayesa

Zmienna a jest zależna od zmiennej b, przy danych dowodach (obserwacjach) $E = \{e_1 \dots e_n\}$, jeśli istnieje ścieżka (d-connecting) z a do b przy danym E.

- E może być puste.
- E nie zawiera a ani b: $a, b \notin E$.

Zmienne niezależne

Jeżeli a nie jest zależne od b przy danym E, to a jest niezależne od b przy danym E.

Dyskusja

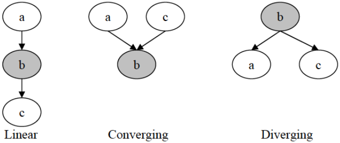
- Jeżeli a i b są niezależne, to mamy na myśli, że:

$$P(a|b) = P(a)$$

- Jednakże, zależność może się zmienić na skutek zaobserwowania nowej zmiennej losowej gdy pojawi się pewien dowód e i wówczas:

$$P(a|b, e) \neq P(a|e)$$

Różne możliwości połączeń węzłów



- liniowe - *b* jest połączone liniowo ze swoimi bezpośrednimi sąsiadami *a* i *c*
- zbieżne - *b*, oba sąsiednie węzły *a* i *c* są rodzicami węzła *b*
- rozbieżne - *b* jest rodzicem dla węzłów *a* i *c*.

Połączenie istnienie ścieżki pomiędzy węzłami

Definicja d-connecting

Ścieżka z węzła q do r jest d-connecting przy danych dowodach E, jeżeli każdy wewnętrzny węzeł n w ścieżce ma następujące własności:

1. jest albo liniowy lub rozbieżny i nie jest elementem w E
2. albo jest zbieżny i n lub jakikolwiek z jego potomków jest w E.

$\forall_{n_i}(((n_i \in Lin \vee n_i \in Div) \wedge n_i \notin E) \vee (n_i \in Con \wedge (n_i \in E \vee dec(n_i) \in E)))$

Definicja d-separate

Dwa węzły są d-separate jeżeli nie istnieje d-connecting ścieżka pomiędzy nimi.

Innymi słowy węzły są połączone jeżeli istniej pomiędzy nimi ścieżka w grafie lub jeżeli istnieją dowody, które sprawiają, że te dwa węzły są skorelowane.

Interpretacja d-connecting

1. Niezależność liniowa lub rozbieżna lub inaczej blokowanie przez dowód ($fo \rightarrow do \rightarrow hb$), gdy $do \in E$.
2. Niezależność połączenie zbieżne: ($fo \wedge bp \rightarrow do$) (jeśli dwa zdarzenia mogą spowodować ten sam stan rzeczy i brak innych połączeń pomiędzy nimi, to są one niezależne).
3. Zależność zbieżna: (zakłada się że $do \in E$ zatem istnieje d-connecting pomiędzy *fo* i *bp*) (np. wiedząc, że rodzina jest w domu powinno to nieznacznie zwiększyć prawdopodobieństwo problemów gastrycznych psa, gdyż wyeliminowaliśmy najbardziej prawdopodobne zdarzenie wpływające na wypuszczenie psa.)

Spójność wartości

Załóżmy, że podano następujące wartości:

$$P(a|b) = .7, P(b|a) = 3, \text{ i } P(b) = .5$$

Spójność wartości

Załóżmy, że podano następujące wartości:

$$P(a|b) = .7, P(b|a) = 3, \text{ i } P(b) = .5$$

Obliczamy teraz regułę Bayesa:

$$P(a|b) = \frac{P(a)P(b|a)}{P(b)} =$$

$$P(a) = \frac{P(b)P(b|a)}{P(b|a)} = \frac{0.5 \cdot 0.7}{0.3} = \frac{0.35}{0.3} > 1$$

Spójność rozkładu w sieci

Spójność

Dobrą własnością sieci Bayes'a jest to, że jeśli określi się wymagane wartości (prawdopodobieństwa każdego węzła biorąc pod uwagę wszystkie możliwe kombinacje jego rodziców), to

1. wartości będą spójne
2. sieć jednoznacznie określi rozkład.

Własność

Spójność w sieci bayesowskiej jest zagwarantowana, gdy zapewnia się spójność dla lokalnych rozkładów.

Łączny rozkład prawdopodobieństwa

Definicja

Łączny rozkład prawdopodobieństwa zbioru zmiennych losowych $v_1 \dots v_n$ definiuje się jako $P(v_1, \dots, v_n)$.
Czyli dla zmiennych boolowskich (a, b) trzeba podać prawdopodobieństwa $P(a, b)$, $P(\neg a, b)$, $P(a, \neg b)$ i $P(\neg a, \neg b)$.

Przykład

Dany jest łączny rozkład prawdopodobieństwa dwóch zmiennych losowych a, b

	a	$\neg a$
b	0.04	0.06
$\neg b$	0.01	0.89

Suma wszystkich łącznych prawdopodobieństw musi wynosić 1, ponieważ prawdopodobieństwo wszystkich możliwych wyników musi wynosić 1. Aby mieć pełen łączny rozkład trzeba podać $2^n - 1$ wartości.

Sieć Bayes'a jako reprezentacja łącznego rozkładu prawdopodobieństwa

Dla sieci N składającej się z węzłów $v_1 \ldots v_n$ prawdziwe jest, że

$$P(v_1 \ldots v_n) = P(v_1)P(v_2|v_1)\ldots P(v_n|v_1 \ldots v_{n-1})$$

Rozkład łączny dla przykładu

Prawdziwy jest łączny rozkład prawdopodobieństwa w sieci Bayes'a, przy założeniu niezależności zmiennych:

$$P(fo, bp, lo, do, hp) = P(fo)P(bp)P(lo|fo)P(do|fo\, bp)P(hb|do)$$

Obliczanie sieci

Cel:

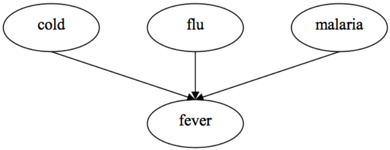
Znaleźć przekonanie (prawdopodobieństwo warunkowe), gdy dane są dowody (zaobserwowane zdarzenia). (Zadanie NP-trudne)

Przykład

Skąd bierzemy wartości?

- wymyślane przez ekspertów
- obliczane z danych.

Przykład schematu danego przez ekspertów:



Przykład obliczeniowy

This can be solved by using the Noisy-OR-random variables as follows –

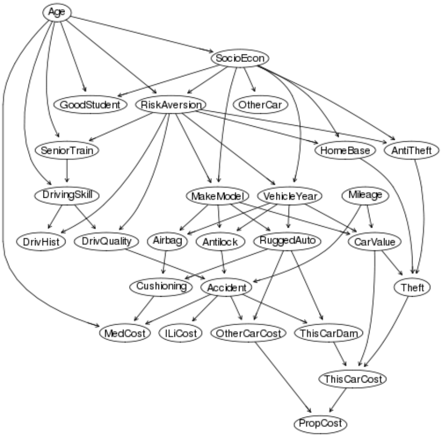
Suppose,

$P(fever \mid cold) = 0.4$ Noisy Parameter = 0.6
$P(fever \mid flu) = 0.8$ Noisy Parameter = 0.2
$P(fever \mid malaria) = 0.9$ Noisy Parameter = 0.1

Then

Cold	*Flu*	*Malaria*	$P(fever)$	$P(\neg fever)$
F	F	F	0.0	1.0
F	F	T	0.9	0.1
F	T	F	0.8	0.2
F	T	T	0.98	$0.02=0.2\times0.1$
T	F	F	0.4	0.6
T	F	T	0.94	$0.06=0.6\times0.1$
T	T	F	0.88	$0.12=0.6\times0.2$
T	T	T	0.988	$0.012=0.6\times0.2\times0.1$

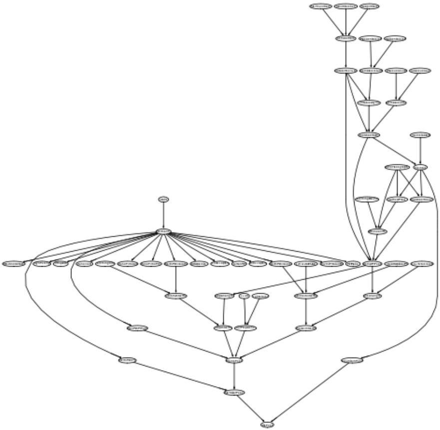
Ubezpieczenia - ryzyko polisy



Ubezpieczenie jest siecią służącą do oceny ryzyka związanego z ubezpieczeniem samochodu.

- Number of nodes: 27
- Number of arcs: 52
- Number of parameters: 984

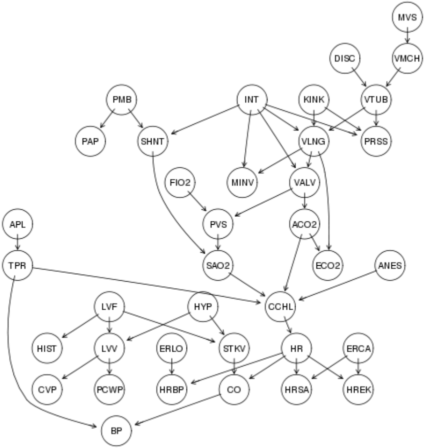
HailFinder - Prognozowanie pogody



Prognoza silnego letniego gradobicia w północno-wschodnim Kolorado

- Number of nodes: 56
- Number of arcs: 66
- Number of parameters: 2656

ALARM - monitorowanie pacjenta



ALARM ("Logiczny Mechanizm Redukcji Alarmów") jest siecią Bayes’a zaprojektowaną w celu zapewnienia systemu komunikatów alarmowych do monitorowania pacjentów.

- Number of nodes: 37
- Number of arcs: 46
- Number of parameters: 509

Dziękuję za uwagę
Czas na pytania ????