

## Použití Bayesovských sítí

Data mining - hledá se model pro dana' data, který bude odražet závislosti mezi jevy (proměnnými) a způsob, jak jeden ovlivňuje druhý'

### Podpora rozhodování

na základě vypočtených pravděpodobností se rozhodujeme buď k nejakej akci nebo k hledání nových evidencí, které by mely vstoupit do modelu

Prí: při podezření na chřipku si budeme rovnou vezmeme aspirin (akce), nebo si ještě předem změříme teplotu (nova' evidence, test)

Také si můžeme spočítat pravděpodobnost, jestliž nejakeho rozhodnutí bude mít spis' výhodu nebo škodu. %

### Budování rozhodovacích stromů

tj. způsob reprezentace scénářů pro několik možných rozhodnutí

a t. d. na jeho základě se vybírá optimální rozhodnutí

## Algoritmy - musí být efektivní

jak vypadají algoritmy pro proti se sbutečným B. S.

(stejně postupy, které jsme ukazovali na příkladech, bychom se v praxi asi moc daleko nedostali.)

%

Příklad: Farmář pěstuje krávy na mléko

Budeme uvažovat nejjednodušší případ pro 1 den

$$\textcircled{1} \rightarrow \textcircled{1}$$

$I$  - infikované mléko ano - ne

$T$  - výsledek testu poz - neg

$$P(I=\text{ano}) = 0,0007 ; P(I=\text{ne}) = 0,9993$$

Má celkem 50 krav, za mléko od každé krávy dostane 2 \$.

Předpokládejme, že 49 krav má mléko v paralelu, a jedná se o výsledek pozitivní. Máto mléko rovinou vylít (v tom případě získa 98 \$) nebo přidat k ostatnímu (v tom případě získa bud 100 \$ nebo nic).

kdyby nedělal test  $\checkmark$  bude očekávaná hodnota zisku

$$0,9993 \cdot 100 + 0,0007 \cdot 0 = 99,93$$

tabuže by to asi risknul.

Protože na jeho rozhodování má vliv jenom pozitivní výsledek testu, vypočteme

$$P(I=\text{ne} | T=\text{poz}) = 0,9351$$

když foto podzrcelé mléko přidá k ostatnímu, bude očekávaná hodnota zisku  $0,9351 \cdot 100 = 93,51$ , což je méně než 98, tabuže ho rádi vylije.

Dalším kritériem pro rozhodování je cena testu - počet test stojí 6 centů.

$$P(I = \text{ne} \wedge T = \text{neg}) = 0,9893 \quad \text{mléko se prodalo}$$

$$P(T = \text{pos}) = 0,0107 \quad \text{mléko se vylilo}$$

Očekávaný zisk je

$$0,9893 \cdot 100 + 0,0107 \cdot 98 = 99,98$$

To je sice vše, než kdyby test nedělal, ale vzhledem k cenie toho testu se to opravdu nevyplatí.

Jedna z možných aplikací Bayesovských sítí -

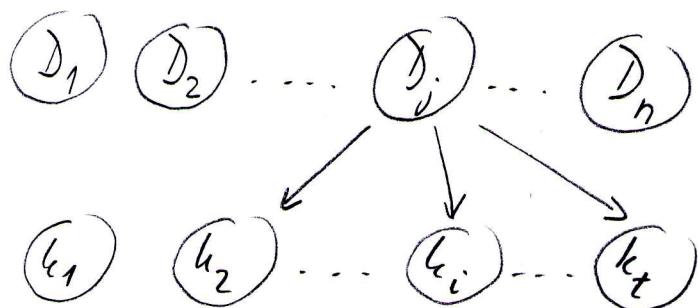
### Dokumentografické informační systémy

1) pro modelování závislostí mezi termíny v dokumentech.

je to podobné jako v příkladu, kdy jsme modelovali  
závislost znaku ve slově

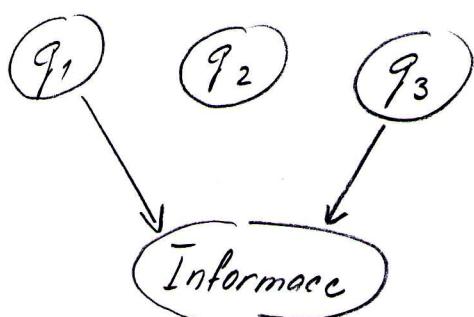
učí se to na databázi dokumentů

2) modelování závislostí termů na dokumentech



říká, že pozorovaný dokument  $D_j$  má termíny  $l_2, l_i, l_t$

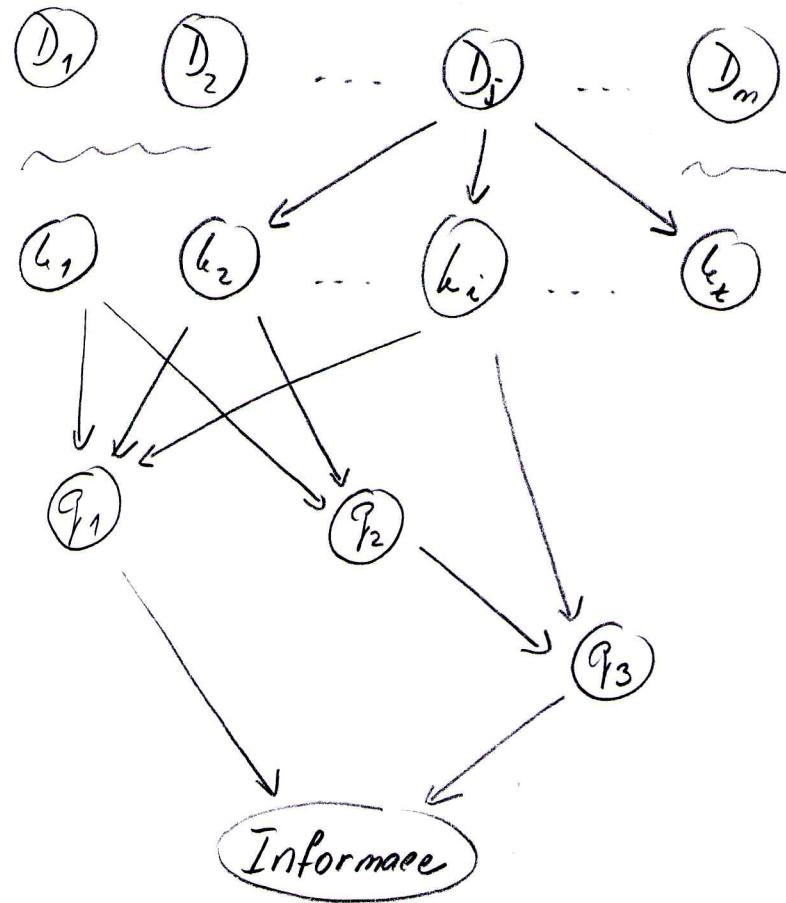
3) modelování závislosti získané informace na dotazech



říká, že informaci získanou z dotazu  $q_1$  a  $q_3$

(41)

4) kombinace 2)a 3)



dodatek je tvorený termími  $k_1, k_2, k_i$