Dobývání znalostí

Doc. RNDr. Iveta Mrázová, CSc.

Katedra teoretické informatiky
Matematicko-fyzikální fakulta
Univerzity Karlovy v Praze

Dobývání znalostí

– Bayesovské modely –

Doc. RNDr. Iveta Mrázová, CSc.

Katedra teoretické informatiky
Matematicko-fyzikální fakulta
Univerzity Karlovy v Praze

Bayesovská klasifikace

- Bayesův vztah pro výpočet podmíněné pravděpodobnosti
  - ~ pravděpodobnost, že platí hypotéza ***H***, pokud pozorujeme ***E***

$$P(H \mid E) = \frac{P(E \mid H)\,P(H)}{P(E)}$$

- ***P(H)*** ……. apriorní pravděpodobnost hypotézy ***H***
- ***P(H|E)*** ….. aposteriorní (podmíněná) pravděpodobnost
- ***P(E)*** …….. pravděpodobnost pozorování ***E***

Bayesovská klasifikace (2)

→ **nejpravděpodobnější hypotéza** $\boldsymbol{H_{MAP}}$

~ největší aposteriorní pravděpodobnost

$$H_{MAP} = H_J \quad ; \qquad P(H_J \mid E) = \max_t \frac{P(E \mid H_t)\, P(H_t)}{P(E)}$$

→ zanedbat jmenovatele: $P(E) = \sum_t P(E \mid H_t)\, P(H_t)$

$$H_{MAP} = H_J \quad \Leftrightarrow \quad P(E \mid H_J)\, P(H_J) = \max_t \left(P(E \mid H_t)\, P(H_t) \right)$$

Bayesovská klasifikace (3)

Příklad:

- *Poskytnout úvěr klientovi s vysokým příjmem?*
  - $P(PUJCIT) = 0.667$
  - $P(NEPUJCIT) = 0.333$
  - $P(VYS\_PRIJEM|PUJCIT) = 0.91$
  - $P(VYS\_PRIJEM|NEPUJCIT) = 0.12$
  - $P(NIZ\_PRIJEM|PUJCIT) = 0.09$
  - $P(NIZ\_PRIJEM|NEPUJCIT) = 0.88$

  - $P(VYS\_PRIJEM|PUJCIT)\ P(PUJCIT) = 0.607$
  - $P(VYS\_PRIJEM|NEPUJCIT)\ P(NEPUJCIT) = 0.040$

$\rightarrow\ H_{MAP} = PUJCIT$

Naivní Bayesovský klasifikátor

- **Předpoklad:**
  - jednotlivá pozorování $E_1, \ldots, E_K$ jsou **podmíněně nezávislá** při platnosti hypotézy H
  - tento předpoklad bývá v reálných úlohách jen málokdy splněn → proto označení „**naivní**“

→ výpočet aposteriorní pravděpodobnosti hypotézy při platnosti všech $E_1, \ldots, E_K$

$$P(H \mid E_1, \ldots, E_K) = \frac{P(E_1, \ldots, E_K \mid H)\, P(H)}{P(E_1, \ldots, E_K)}$$

Naivní Bayesovský klasifikátor (2)

→ výpočet aposteriorní pravděpodobnosti hypotézy při platnosti všech $\boldsymbol{E_1, \ldots, E_k}$ jako

$$P(H \mid E_1, \ldots, E_K) = \frac{P(H)}{P(E_1, \ldots, E_K)} \prod_{k=1}^{K} P(E_k \mid H)$$

→ při klasifikaci pomocí naivního Bayesovského klasifikátoru budeme hledat hypotézu s největší aposteriorní pravděpodobností $\boldsymbol{H_{MAP}}$

$$H_{MAP} = H_J \quad \Leftrightarrow \quad P(H_J) \prod_{k=1}^{K} P(E_k \mid H_J) = \max_t \left(P(H_t) \prod_{k=1}^{K} P(E_k \mid H_t) \right)$$

Naivní Bayesovský klasifikátor (3)

- $P(H_t) = P(t\check{r}\acute{\imath}dy\_t) =$

  $= Po\check{c}et\_vzor\mathring{u}\_z\_t \,/\, Po\check{c}et\_v\check{s}ech\_vzor\mathring{u}$

- $P(E_k|H_t) = P(vzor\mathring{u}\_z\_t\check{r}\acute{\imath}dy\_t\_vyhovuj\acute{\imath}c\acute{\imath}ch\_E_k =$

$$= \frac{Po\check{c}et\_vzor\mathring{u}\_z\_t\_spl\check{n}uj\acute{\imath}c\acute{\imath}ch\_E_k}{Po\check{c}et\_vzor\mathring{u}\_z\_t}$$

Naivní Bayesovský klasifikátor (4)

\+ **možnost klasifikovat i neúplně popsané vzory**

\- **nulová aposteriorní pravděpodobnost $P(E_k|H_t)$ při chybějících vzorech v trénovací množině (pro E_k)**

evtl. podhodnocení $P(E_k|H_t)$ při nízké vzájemné četnosti E_k a H_t

Bayesovské sítě
(pravděpodobnostní sítě)

- Jednotlivá pozorování nemusí být navzájem nezávislá

→ Reprezentace znalostí o částečně nezávislých pozorováních → **podmíněná nezávislost**

- Veličiny $\boldsymbol{A}$ a $\boldsymbol{B}$ jsou podmíněně nezávislé při dané veličině $\boldsymbol{C}$, jestliže $\boldsymbol{P(A,B|C) = P(A|C)\,P(B|C)}$
- Označení podmíněné nezávislosti $\boldsymbol{A}$ a $\boldsymbol{B}$ při daném $\boldsymbol{C}$: $\boldsymbol{A \perp B \mid C}$

Bayesovská síť

~ acyklický orientovaný graf zachycující pomocí hran pravděpodobnostní závislosti mezi náhodnými veličinami

- Ke každému uzlu ***u*** (~ náhodné veličině) je přiřazena pravděpodobnostní distribuce tvaru ***P (u | rodiče (u)),*** kde ***rodiče (u)*** jsou uzly, ze kterých vycházejí hrany do uzlu ***u***
- Všechny uzly sítě jsou uspořádány (~ očíslovány) tak, že rodiče jsou před svými dětmi (~ mají nižší pořadové číslo)

Bayesovská síť - pokračování (2)

Potom pro každý uzel u_i platí, že je podmíněně nezávislý na všech uzlech s nižším pořadovým číslem s výjimkou svých rodičů podmíněno rodiči rodičů:

$$u_i \perp \{u_k\};\ k = 1, \ldots, i-1 \setminus rodiče(u_i) \mid rodiče(rodiče(u_i))$$

→ Sdružená pravděpodobnostní distribuce celé sítě:

$$P(u_1, \ldots, u_n) = \prod_{i=1}^{n} P(u_i \mid rodice(u_i))$$

Bayesovská síť - pokračování (3)

Inference (~ pravděpodobnostní odvozování) v Bayesovských sítích

- známe-li strukturu sítě a podmíněné pravděpodob-nostní distribuce přiřazené jednotlivým uzlům, můžeme spočítat aposteriorní pravděpodobnost libovolného uzlu
- **Diagnostická inference (´zdola nahoru´)**
  - Pozorujeme projevy nějakého jevu a zajímá nás, co je příčinou → **Která příčina má maximální aposteriorní pravděpodobnost?**

Bayesovská síť - pokračování (3)

- **Kauzální inference (´shora dolů´)**

  → **Jaká je pravděpodobnost nějakého jevu za předpokladu určité příčiny?**

  - Použití při konstrukci Bayesovské sítě

  X ne každá hrana musí mít kauzální interpretaci

Učení Bayesovských sítí

- Známá struktura (např. od experta)

  → Z dat odvozovat pouze podmíněné pravděpodobnosti

- Z dat odvodit strukturu sítě i pravděpodobnosti

Učení Bayesovských sítí

1. Známá struktura, veličiny plně pozorovatelné (metoda maximálně věrohodného odhadu)

   → spočítat z dat odhady podmíněných pravděpodobnostních distribucí pro jednotlivé uzly sítě

   Cíl: nalézt maximálně věrohodný odhad vzhledem k trénovacím datům (spočítat na základě četností)

$$P(H_t \mid E_1, E_2) = \frac{n(H_t \wedge E_1 \wedge E_2)}{n(E_1 \wedge E_2)}$$

Učení Bayesovských sítí (2)

2. Známá struktura, veličiny částečně pozorovatelné (gradientní metoda nebo algoritmus EM)

- některé veličiny (uzly sítě) nelze pozorovat
  – jsou skryté (~ latentní)
- použít **gradientní metodu**
- maximalizovat pravděpodobnost $P(D \mid h)$ toho, že pozorujeme daná trénovací data D, pokud platí podmíněné pravděpodobnosti

  → výpočet gradientu funkce $\ln P(D \mid h)$

Učení Bayesovských sítí (2)

2. Známá struktura, veličiny částečně pozorovatelné

gradientní metoda

p_{ijk} … podmíněná pravděpodobnost toho, že veličina Y_i nabývá hodnoty y_{ij}, pokud rodiče U_i této veličiny nabývají hodnoty u_{ik}

$$\frac{\partial \ln P(D \mid h)}{\partial p_{ijk}} = \eta \sum_{d \in D} \frac{P(Y_i = y_{ij}, U_i = u_{ik} \mid d)}{p_{ijk}}$$

Učení Bayesovských sítí (2)

2. Známá struktura, veličiny částečně pozorovatelné

gradientní metoda

→ aktualizace hodnot p_{ijk} podle:

$$p_{ijk} := p_{ijk} + \eta \sum_{d \in D} \frac{P\left(y_{ij}, u_{ik} \mid d\right)}{p_{ijk}}$$

+ případná další normalizace p_{ijk}

Pro nepozorovatelné veličiny lze potřebné hodnoty dopočítat z pozorovatelných hodnot pomocí běžné inference v Bayesovské síti

Učení Bayesovských sítí (2)

2. Známá struktura, veličiny částečně pozorovatelné

algoritmus EM (expectation maximization)

IDEA: kdybychom znali hodnoty skrytých veličin, mohli bychom přímo spočítat parametry podmíněné pravděpodobnostní distribuce

INICIALIZACE: náhodně zvol parametry distribuce

HLAVNÍ CYKLUS (EXPEKTACE + MAXIMALIZACE:

1) na základě parametrů distribuce spočítej hodnoty skrytých veličin
2) z takto spočítaných hodnot urči maximálně věrohodný odhad parametrů distribuce
3) pokud došlo ke změně parametrů, vrať se ke Kroku 1.

Učení Bayesovských sítí (2)

3. neznámá struktura, veličiny plně pozorovatelné (prohledávání prostoru modelů)

   výpočetně náročné – různých množin rodičů může být 2^k

4. Neznámá struktura, veličiny částečně pozorovatelné

Bayesovská síť - příklady

Příklad: trénovací data pro Bayesovský klasifikátor

KLIENT	PŘÍJEM	KONTO	POHLAVÍ	NEZAMĚSTNANÝ	ÚVĚR
K1	VYSOKÝ	VYSOKÉ	ŽENA	NE	ANO
K2	VYSOKÝ	VYSOKÉ	MUŽ	NE	ANO
K3	NÍZKÝ	NÍZKÉ	MUŽ	NE	NE
K4	NÍZKÝ	VYSOKÉ	ŽENA	ANO	ANO
K5	NÍZKÝ	VYSOKÉ	MUŽ	ANO	ANO
K6	NÍZKÝ	NÍZKÉ	ŽENA	ANO	NE
K7	VYSOKÝ	NÍZKÉ	MUŽ	NE	ANO
K8	VYSOKÝ	NÍZKÉ	ŽENA	ANO	ANO
K9	NÍZKÝ	STŘEDNÍ	MUŽ	ANO	NE
K10	VYSOKÝ	STŘEDNÍ	ŽENA	NE	ANO
K11	NÍZKÝ	STŘEDNÍ	ŽENA	ANO	NE
K12	NÍZKÝ	STŘEDNÍ	MUŽ	NE	ANO

Bayesovská síť - příklady (2)

Příklad (pokračování):

- Apriorní pravděpodobnost různých hodnot cílového atributu ***ÚVĚR*** :

P(ÚVĚR(ANO)) = 8/12 = 0,667
P(ÚVĚR(NE)) = 4/12 = 0,333
P(KONTO(STŘEDNÍ) | ÚVĚR(ANO)) = 2/8 = 0,25
P(KONTO(STŘEDNÍ) | ÚVĚR(NE)) = 2/4 = 0,5
P(NEZAMĚSTNANÝ(NE) | ÚVĚR(ANO)) = 5/8 = 0,625
P(NEZAMĚSTNANÝ(NE) | ÚVĚR(NE)) = 1/4 = 0,25
. . . . . .

Bayesovská síť - příklady (3)

Příklad (pokračování):

→ Pro uchazeče o úvěr, který má středně vysoký zůstatek na kontě a není nezaměstnaný, spočítáme:

$$P(\text{ÚVĚR}(\text{ANO}))\,P(\text{KONTO}(\text{STŘEDNÍ})\mid\text{ÚVĚR}(\text{ANO}))\cdot$$
$$\cdot P(\text{NEZAMĚSTNANÝ}(\text{NE})\mid\text{ÚVĚR}(\text{ANO})) = 0{,}1042$$

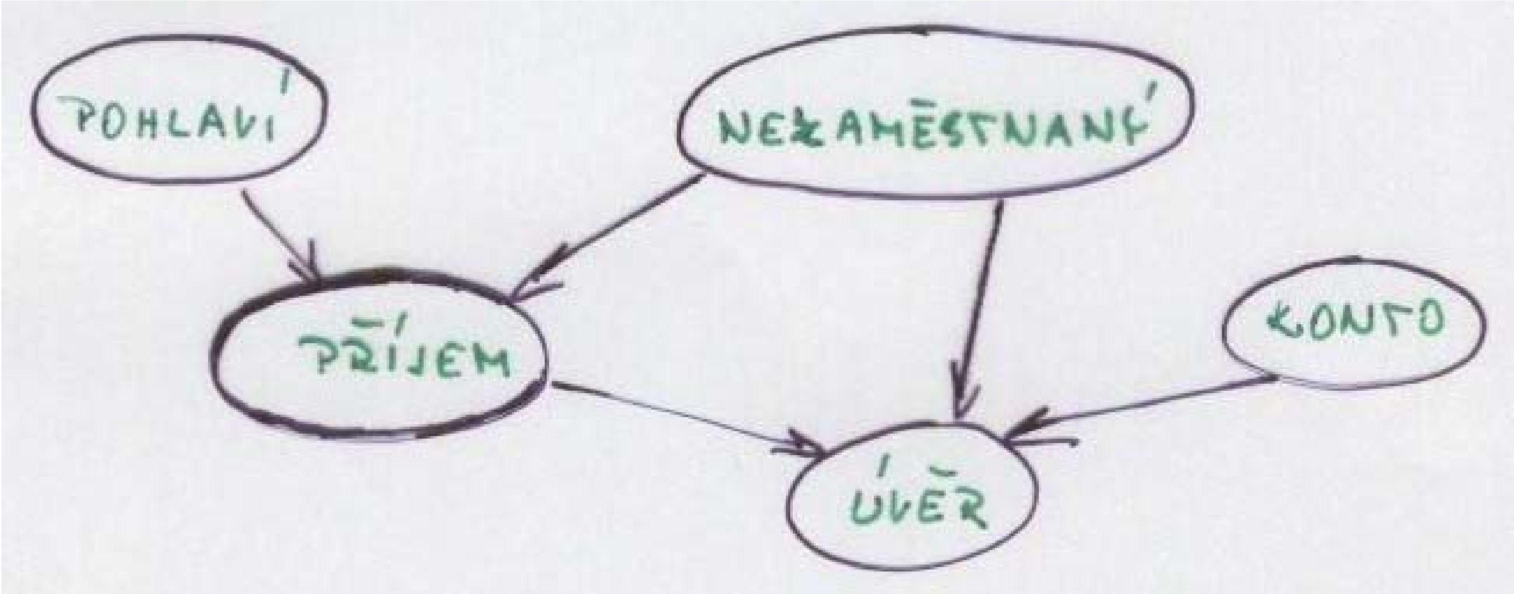
$$P(\text{ÚVĚR}(\text{NE}))\,P(\text{KONTO}(\text{STŘEDNÍ})\mid\text{ÚVĚR}(\text{NE}))\cdot$$
$$\cdot P(\text{NEZAMĚSTNANÝ}(\text{NE})\mid\text{ÚVĚR}(\text{NE})) = 0{,}0416$$

→ **Uchazeč bude zařazen do třídy** ***ÚVĚR (ANO)***

Bayesovská síť - příklady (4)

Příklad (pokračování):

- Bayesovská síť a výpočet podmíněných pravděpodobností v případě známé struktury sítě:



Bayesovská síť - příklady (5)

Příklad (pokračování):

→ sdružená pravděpodobnostní distribuce pro tuto síť:

P(PŘÍJEM, KONTO, POHLAVÍ, NEZAMĚSTNANÝ, ÚVĚR) =
= P(POHLAVÍ) P(NEZAMĚSTNANÝ) P(KONTO) ·
· P(PŘÍJEM | POHLAVÍ, NEZAMĚSTNANÝ) ·
· P(ÚVĚR | PŘÍJEM, NEZAMĚSTNANÝ, KONTO)

Bayesovská síť - příklady (6)

Příklad (pokračování):

→ odhad dílčích pravděpodobností z trénovacích dat:

$$P(\text{POHLAVÍ} = \text{ŽENA}) = n(\text{POHLAVÍ}(\text{ŽENA}))/n = 6/12 = 0.5$$

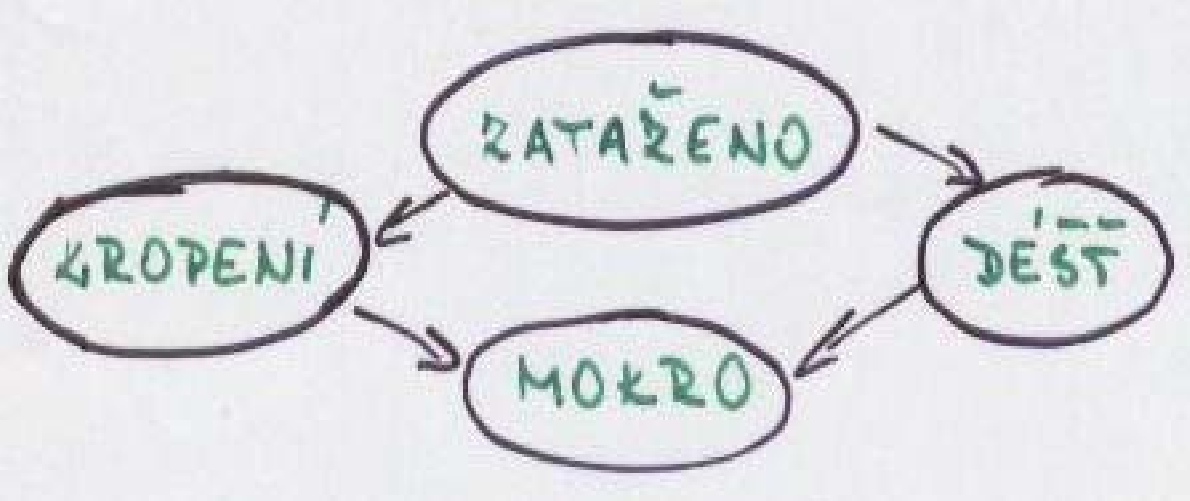
$$P(\text{PŘÍJEM} = \text{VYSOKÝ} \mid \text{POHLAVÍ} = \text{ŽENA}, \text{NEZAM} = \text{ANO}) =$$

$$= \frac{n(\text{PŘÍJEM}(\text{VYSOKÝ}) \wedge \text{POHLAVÍ}(\text{ŽENA}) \wedge \text{NEZAM}(\text{ANO}))}{n(\text{POHLAVÍ}(\text{ŽENA}) \wedge \text{NEZAM}(\text{ANO}))} =$$

$$= 1/4$$

Bayesovská síť - příklady (7)

Příklad: Bayesovská síť a diagnostická inference



- Sdružená distribuce:

$$\boldsymbol{P(Z,K,D,M) = P(Z)\ P(K|Z)\ P(D|Z)\ P(M|K,D)}$$

Bayesovská síť - příklady (8)

Příklad (pokračování):

- Podmíněné pravděpodobnosti uzlů - ***P(Z), P(K|Z), P(D|Z)*** a ***P(M|K,D)***:

P(Z=0)	P(Z=1)
0.5	0.5

Z	P(K=0)	P(K=1)
0	0.5	0.5
1	0.9	0.1

Z	P(D=0)	P(D=1)
0	0.8	0.2
1	0.2	0.8

K	D	P(M=0)	P(M=1)
0	0	1,0	0,0
0	1	0,1	0,9
1	0	0,1	0,9
1	1	0,01	0,99

Bayesovská síť - příklady (9)

Příklad (pokračování):

- Podmíněné pravděpodobnosti - ***P(K=1|M=1)***

$$P(K=1\,|\,M=1) = \frac{P(K=1, M=1)}{P(M=1)} =$$

$$= \frac{\sum_{z,d} P(Z=z, K=1, D=d, M=1)}{P(M=1)} =$$

$$= \frac{\sum_{z,d} P(Z=z)\,P(K=1\,|\,Z=z)\,P(D=d\,|\,Z=z)\cdot P(M=1\,|\,K=1, D=d)}{P(M=1)} = \frac{0{,}2781}{P(M=1)}$$

a ***P(D=1|M=1)***:

$$P(D=1\,|\,M=1) = \frac{0{,}4581}{P(M=1)}$$

Bayesovská síť - příklady (10)

Příklad (pokračování):

a $P(D=1 \mid M=1)$:

$$P(D=1 \mid M=1) = \frac{0{,}4581}{P(M=1)}$$

$\rightarrow P(D=1 \mid M=1) > P(K=1 \mid M=1)$

$\rightarrow$ **Příčinou** ***MOKRO (=1)*** **je** ***DÉŠŤ***

Kauzální inference (´shora dolů´)

- Pravděpodobnost, že bude mokro, pokud je zataženo:

$$P(M=1 \mid Z=1) = \frac{P(M=1, Z=1)}{P(Z=1)} = 0.7452 / P(Z=1)$$