Zadání

Modely pro práci s neurčitou informací. Pravděpodobnostní přístup, algebraická teorie a její důsledky. Dempster-Shaferův přístup.

1. Modely pro práci s neurčitou informací

1.1. Fuzzy míry

Fuzzy množiny jsou definovány funkcí příslušnosti, která každému prvku uvažovaného světa přiřadí hodnotu z <0;1> vyjadřující, jak silně jsme přesvědčeni, že uvažovaný prvek patří do dané množiny.

Fuzzy míry poukazují, do jaké míry našich pozorování můžeme prvek zařadit do dané množiny. Pro reprezentaci nejistoty přiřazujeme hodnotu neurčitosti (fuzzy míru) každé klasické množině, do níž může daný prvek patřit. Všechny dále popsané kapitoly v tomto dokumentu splňují požadavky kladené na fuzzy míry, proto si nyní tyto požadavky ukážeme. Funkci

$$g : \wp(X) \to \langle 0;1 \rangle \tag{1.1}$$

budeme považovat za fuzzy míru, jsou-li současně splněny požadavky:

(a1) *hraniční podmínky:* $g(\emptyset)=0$, $g(X)=1$;

(a2) *monotónnost:* pro všechny $A,B \in \wp(X) I$ takové, že $A \subseteq B$, platí $g(A) \leq g(B)$

(a3) *spojitost:* pro každou monotónní posloupnost $\{A_1, A_2, ...\}$ podmnožin X (tj. buď $A_1 \subseteq A_2 \subseteq ...$, nebo $A_1 \supseteq A_2 \supseteq ...$) platí:

$$\lim \lim_{i \to \infty} g(A_i) = g\left(\lim_{i \to \infty} A_i\right) \tag{1.2}$$

Pokud je X konečná nemá (a3) smysl a fuzzy mírou bude každé zobrazení g, splňující (a1),(a2). Podmínky (a1),(a2),(a3) neplatí pouze pro fuzzy míry, ale i pro pravděpodobnostní míry (kapitola 1.2.1 Teorie pravděpodobnosti).

1.2. Extenzionální přístup

Přístup ke zpracování neurčité informace, který se snaží skládáním dílčích znalostí napodobit strukturu závislostí existujících ve zkoumané předmětné oblasti.

1.2.1. Teorie pravděpodobnosti (z knihy Umělá inteligence 2)

Každá pravděpodobnostní míra splňuje požadavky na fuzzy míry. Jde tedy o zobrazení:

$$P : \wp(X) \to \langle 0;1 \rangle, \tag{1.3}$$

které kromě axiomů (a1),(a2),(a3) splňuje dále i axiom:

(a4) *aditivnost*: pro každý nejvýše spočetný systém disjunktních podmnožin $X\{A_i\}_{i \in I}$ (tj. pro libovolné dva různé indexy $i, j \in I$ je $A_i \cap A_j = \emptyset$) platí

$$P\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} P(A_i). \tag{1.4}$$

Tento axiom je natolik silný, že z něj je možné odvodit nejen axiomy (a2),(a3), ale i (a5),(a6) (viz kapitola 1.2.8 Dempster-Shaferova teorie). Lze tedy říci, že pravděpodobnostní míra na konečné množině X (přesněji řečeno $\wp(X)$) je speciálním případem fuzzy míry domnění a plausibility (viz kapitola 1.2.8 Dempster-Shaferova teorie). Je to zároveň jediný případ, kdy fuzzy míra splňuje **současně** axiomy (a5),(a6).

1.2.2. Teorie pravděpodobnosti (slajdy z přednášek ze SPR)

<u>Podmíněná pravděpodobnost:</u>

$$P(A/B) = P(A,B) / P(B) \tag{1.5}$$

<u>Bayesův vztah:</u>

P(B) = P(B/A) + P(B/¬A)

$$P(A/B) = P(A,B) / (P(B/A) + P(B/\neg A)) \tag{1.6}$$

<u>Sdružená (vzájemná) nezávislost:</u>

$$P(A,B) = P(A) . P(B) \tag{1.7}$$

Dosaďme do definice podmíněné pravděpodobnosti:

P(A,B) = P(A) . P(B)

$$\underline{P(A/B)} = P(A,B) / P(B) = P(A) . P(B) / P(B) = \underline{P(A)} \tag{1.8}$$

Jevy A,B,C vzájemně nezávislé:

$$P(A,B,C) = P(A) \,.\, P(B) \,.\, P(C) \tag{1.9}$$

Jevy A,B,C po dvou nezávislé:

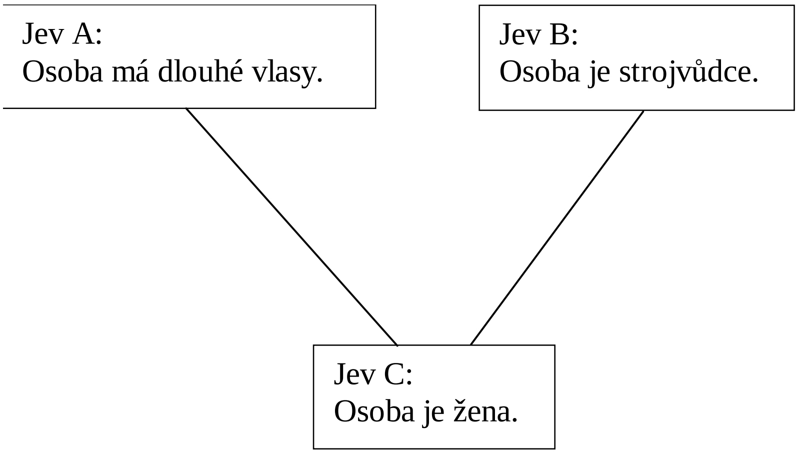
$$P(A,B) = P(A) \,.\, P(B) \;\;\&\; P(B,C) = P(B) \,.\, P(C) \;\&\; P(A,C) = P(A) \,.\, P(C) \tag{1.10}$$

ALE

$$P(A,B,C) \neq P(A) \,.\, P(B) \,.\, P(C) \tag{1.11}$$

Jsou-li jevy A,B,C vzájemně nezávislé, jsou i po dvou nezávislé.
Obrácená věta neplatí.

Podmíněná nezávislost jevů A a B při dané hodnotě jevu C ($A \perp B / C$):



Za předpokladu $A \perp B / C$ je podmíněná pravděpodobnost jevů A, B za podmínky, že nastal jev C, dána vztahem:

$$P(A,B/C) = P(A/C) \,.\, P(B/C) \tag{1.12}$$

Za předpokladu $A \perp B / C$ je sdružená pravděpodobnost jevů A, B, C dána vztahem:

$$P(A,B,C) = P(A,C) \,.\, P(B,C) \,/\, P(C) \tag{1.13}$$

Důkaz:

P(B,C)

$$\mathbf{\underline{P(A,B,C)} = P(A,B/C) \,.\, P(C) = P(A/C) \,.\, P(B/C) \,.\, P(C) = P(A/C) \,.\, P(B,C) = \underline{P(A,C) \,.\, P(B,C) \,/\, P(C)}}$$

P(A,B/C) = P(A/C) . P(B/C)

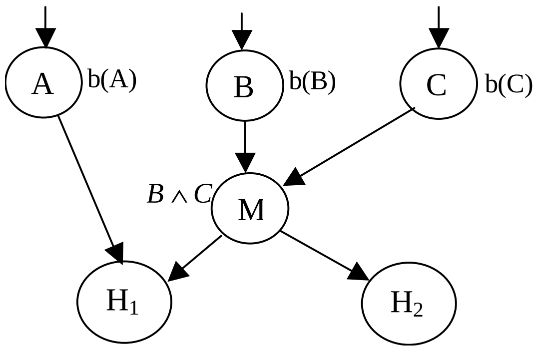
P(A,C) / P(C)

1.2.3. Teorie pravděpodobnosti (ze cvičení z RPZ)

Protože se mi nelíbí, že p. Kouba nazývá vztah (1.6) na svých slajdech vztahem Bayesovým, přikládám k tomuto dokumentu ještě soubor „Priloha_k_4_1.pdf“ (nechce se mi to opisovat do Wordu), ve kterém jsou uvedeny a vysvětleny na příkladech další vztahy na podmíněnou a úplnou pravděpodobnost a Bayesuův vztah, který je zde uveden ve známém tvaru. Vztah (1.6) se mi totiž nepodařilo na tento „známější“ tvar převést.

1.2.4. Algebraická teorie

Úloha expertních systémů (dále jen ES) spočívá v nalezení správné hypotézy z množiny různých hypotéz $H = \{H_1, H_2, ...H_n\}$.



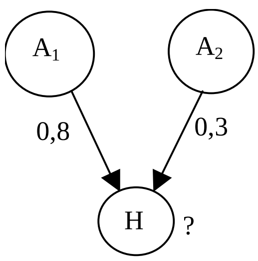
obr. 1.1 – práce expertního systému

A,B,C: náhodné veličiny, u nichž známe pravděpodobnost, s jakou nabývají
M: náhodné veličiny vzniklé složením původních náhodných veličin
b(X): stupeň důvěry v platnost veličiny *X* – zjišťuji od uživatele

ES často pracují s pravidly: $A \rightarrow H(w)$ tj. platí-li *A* (na 100%), pak *H* platí s vahou *w*.
Z výše uvedeného plynou dva problémy, které je třeba řešit:

I. Co když nevím jistě, zda *A* platí, ale mohu *A* ocenit nějakým stupněm důvěry *a* v jeho platnost. Jaký pak bude stupeň důvěry v platnost *H* ? (Problém příspěvku pravidla)

II. Co když mám několik pravidel s týmž závěrem *H* ? Jak určit váhu výroku *H* z příspěvků jednotlivých pravidel ? (Problém sdružování příspěvků)



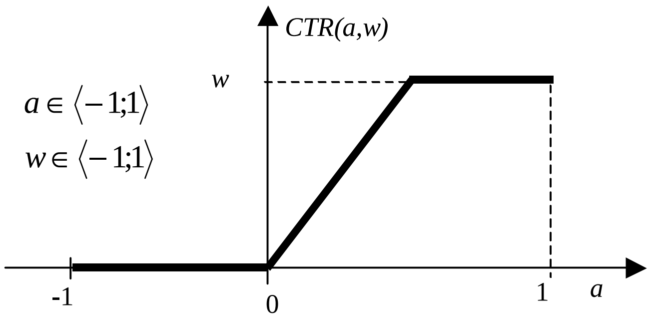
S řešením obou problémů přišel Dr. Hájek ve své algebraické teorii.

ad I. Algebraická teorie předpokládá váhy předpokladů v rozmezí <-1;1>, kde

-1 … znamená určitě ne
0 … znamená nevím
+1 … znamená určitě ano

Dále se zavádí tzv. příspěvek *CTR(a,w)*, který
1) $CTR(1,w)=w$
2) $CTR(a,w)=0$ if $a \leq 0$
3) $CTR(a,w) \approx monotónní$ if $0<a<1$

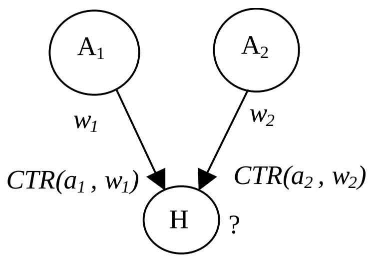
kde *a* je váha předpokladu (tj. stupeň důvěry v *a*) a *w* je váha hypotézy (pravidla)



obr. 1.2 – funkce příspěvků

Funkce příspěvků vrací stupeň důvěry v platnost hypotézy *H*.

ad II.



obr. 1.3 – problém sdružování příspěvků

Předpokládáme, že funkce *CTR(a,w)* nabývá hodnot z intervalu <-1;1>. Potřebujeme znát:

$$CTR(a_1,a_2,w_1,w_2)=CTR(a_1,w_1) \oplus CTR(a_2,w_2), \quad \textbf{(1.14)}$$

kde $\oplus$ je operace, která sdružuje příspěvky. Pro ni platí:

1.
$$\begin{aligned} x \neq -1 \quad &: \quad x \oplus 1 = 1 \\ x \neq 1 \quad &: \quad x \oplus -1 = -1 \end{aligned} \quad \textbf{(1.15)}$$

2.
$$x, y, z \in \leq (-1;1) \quad \textbf{(1.16)}$$

a) asociativita $(x \oplus y) \oplus z = x \oplus (y \oplus z)$ **(1.17)**

b) komutativita $x \oplus y = y \oplus x$ **(1.18)**

c) $x \oplus 0 = 0$ **(1.19)**

d) $-x \oplus x = 0$ **(1.20)**

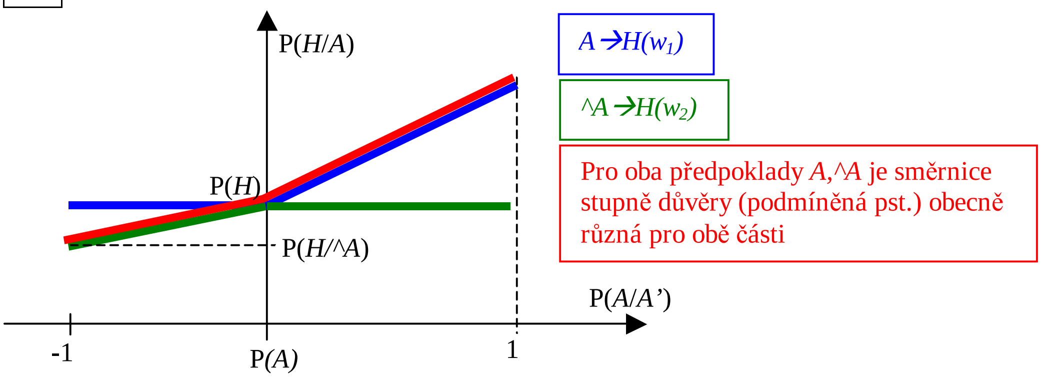
e) $x \leq y \,:\, x \oplus z \leq y \oplus z$ **(1.21)**

Dále musí platit : $w_1 \oplus w_2 \oplus \ldots \oplus w_n = 1$ **(1.22)**

což je tzv. Archimedovská operace.

1.2.5. Řešení problémů algebraické teorie v ES Prospektor (FEL expert)

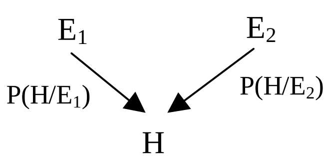
ad I.



obr. 1.4 – funkce příspěvků pro ES PROSPECTOR

P(A) … pravděpodobnost, že A platí
P(^A) … pravděpodobnost, že A neplatí
H(w) … hypotéza a váha v její platnost
A' … provedené pozorování jevu A
P(A) … apriorní pravděpodobnost tvrzení A
P(H) … apriorní pravděpodobnost hypotézy H
P(A/A') … pravděpodobnost jevu A za předpokladu, že jsem provedl jeho pozorování

ad II.



Hledáme

$$P(H/E_1,E_2) = P(H/E_1) \oplus P(H/E_2) \qquad \textbf{(1.23)}$$

Sdružení obou příspěvků je možné pouze tehdy, je-li E_1 a E_2 podmíněně nezávislé při dané hodnotě H (viz kapitola 1.2.2).

Snadno můžeme vyjádřit:

$$P(E_1, E_2 / H) = P(E_1 / H).P(E_2 / H). \tag{1.24}$$

Po aplikaci Bayesova vztahu dostaneme:

$$P(E_1, E_2 / H) = \frac{P(H / E_1, E_2)P(E_1 E_2)}{P(H)}$$

$$P(H / E_1, E_2) = P(E_1 / H)P(E_2 / H)P(H) = P(E_1 / H)P(E_2, H) = \frac{P(E_1, H) \cdot P(E_2, H)}{P(H)}$$

S vědomím, že $P(A, B) = P(A / B)P(B)$ a s opětovným použitím Bayesova vztahu můžeme čitatel dále upravit:

$$P(H / E_1, E_2) = \frac{P(H / E_1)P(E_1) \cdot P(H / E_2)P(E_2)}{P(H)} \tag{1.25}$$

Zavedeme transformaci – tzv. šanci O

$$O(*) = \frac{P(*)}{1 - P(*)}$$
$$P(*) = \frac{O(*)}{1 + O(*)} \tag{1.26}$$

potom lze psát:

$$O(H / E_1, E_2) = \frac{O(H / E_1)O(H / E_2)}{O(H)}\ , \tag{1.27}$$

a to je právě způsob, jakým ES Prospektor (FEL expert) kombinuje jednotlivé příspěvky. Takže v praxi se příspěvky kombinují podle následujícího schématu:

$$\begin{array}{cc}
E_1 & E_2 \\
\downarrow & \downarrow \\
\mathrm{P}(H/E_1) & \mathrm{P}(H/E_2) \\
\downarrow & \downarrow \\
O(H/E_1) = \dfrac{P(H / E_1)}{1 - P(h / E_1)} & O(H/E_2) = \dfrac{P(H / E_2)}{1 - P(h / E_2)}
\end{array}$$

$$P(H / E_1, E_2) = \frac{O(H / E_1, E_2)}{1 + O(H / E_1, E_2)} \tag{1.28}$$

1.2.6. Intenzionální přístup

Přístup ke zpracování neurčité informace, který se snaží nalézt globální model příslušné struktury závislostí.

1.2.7. Princip maxima entropie

(Integrace znalostí v pravděpodobnostních modelech expertních systémů)

Vstupní znalosti expertního systému zpravidla poskytují dva zdroje: jednak odhady expertů a dále výsledky statistických metod použitých nad soubory reálných dat. Pro využití expertním systémem je potřeba provést *integraci* těchto *znalostí*. Velice často přitom nastává situace, že zadané znalosti jednoznačně neurčují *sdruženou pravděpodobnostní distribuci* (reprezentující bázi znalostí). Se vstupními (částečnými) znalostmi může tedy být konzistentní celá třída pravděpodobnostních distribucí. Ta může obsahovat nekonečně mnoho potenciálně použitelných reprezentací znalostí, ale my se z nich snažíme zvolit tu, která je podle nějakého kritéria optimální.

Informační (shannonovská) **entropie** pravděpodobnostní distribuce se definuje takto:

$$H(P) = -\sum_{x \in X} P(x) \ln(P(x)) \ , \tag{1.29}$$

kde

P ... pravděpodobnostní distribuce
X ... obor hodnot P

Tato veličina (často označovaná jako míra neuspořádanosti) je nepřímo úměrná informaci, kterou v sobě pravděpodobnostní distribuce obsahuje.

Princip maxima entropie: Vybíráme tu distribuci, která maximalizuje informační entropii:

$$H(P^*) = \max_{P^* \in \Pi_K} (H(P)) \ , \tag{1.30}$$

kde

K ... systém zadaných pravidel
Π_K ... třída měr konzistentní s K

Intuitivní zdůvodnění: Pokud bychom vybrali jinou míru než P^*, přidali bychom znalost, která není v K.

1.2.8. Dempster-Shaferova teorie

Nebyla nikde přednášena. Je uvedena v knize Mařík, Štěpánková, Lažanský – Umělá inteligence 2 na str.80-87.