Bayesovské rozhodování - kritétium minimální střední ztráty

Lukáš Slánský, Ivana Čapková

6. června 2001

1 Formulace úlohy

JE DÁNO:

$\mathcal{X}$ — množina možných pozorování (příznaků) x

$\mathcal{K}$ — množina hodnot skrytého parametru k

$p(x,k)$ — funkce $p : \mathcal{X} \times \mathcal{K} \to \langle 0{,}1 \rangle$ udává sdruženou pravděpodobnost jevu $p(x,k)$, že objekt je ve stavu k a současně nabývá příznak hodnoty x.

$\mathcal{D}$ — množina možných rozhodnutí$W(k,d)$ — diskriminační fuknce $W : \mathcal{K} \times \mathcal{D} \to \mathcal{R}$, označující ztrátu vyvolanou rozhodnutím d, pokud je objekt ve stavu k.

Poznámky: Hodnotu $p(x,k)$ často počítáme jako součin $p(x,k) = p(x|k) \cdot p(k)$.

Pozorování $x \in \mathcal{X}$ může být číslo, vektor, nečíselný objekt, funkce, graf apod. Stejnou formu mohou mít i množiny $\mathcal{K}$, $\mathcal{D}$.

2 Realizace

Cílem úlohy je nalezení strategie $d : \mathcal{X} \to \mathcal{D}$, která minimalizuje riziko.

Toto riziko lze udat funkcí

$$R(d) = \sum_{x \in \mathcal{X}} \sum_{k \in \mathcal{K}} p(x,k) W(k, d(x))$$

Ve spojitém případě (případ nekonečných množin $\mathcal{X}$, $\mathcal{K}$) má riziko tvar

$$R(d) = \int_{x \in \mathcal{X}} \int_{k \in \mathcal{K}} p(x,k) W(k, d(x))$$

kde symbol p tentokrát neznamená pravděpodobnost, ale hustotu pravděpodobnosti.

V aplikacích je však většinou množina $\mathcal{X}$ konečná, což je dáno omezenou přesností měření.

Nyní hledejme optimální rozhodovací pravidlo tím, že tuto funkci minimalizujeme:

$$\min_{d\in\mathcal{D}} R(d) = \min_{d\in\mathcal{D}} \sum_{x\in\mathcal{X}} \sum_{k\in\mathcal{K}} p(x,k)W(k,d(x))$$

$$\min_{d\in\mathcal{D}} R(d) = \sum_{x\in\mathcal{X}} \min_{d\in\mathcal{D}} \sum_{k\in\mathcal{K}} p(x,k)W(k,d(x))$$

tedy

$$\min_{d\in\mathcal{D}} R(d) = \sum_{x\in\mathcal{X}} \min_{d\in\mathcal{D}} L_x(d)$$

$$L_x(d) = \sum_{k\in\mathcal{K}} p(x,k)W(k,d(x))$$

kde $L_x(d)$ se nazývá ztrátou obrazu x při rozhodnutí $d(x)$.

$$L_x(k_r) = \sum_{k\in\mathcal{K}} p(x,k)W(k,d(x))$$

Optimální rozhodující pravidlo je takové, které zaručuje pro každé x

$$L_x(d^*) = \min_{d\in\mathcal{D}} L_x(d)$$

3 Speciální případ - klasifikace

Úlohu nazýváme klasifikací, jestliže $\mathcal{K} = \mathcal{D}$, tudíž výsledkem našeho rozhodnutí je skrytý stav objektu.

Potom např. $\mathcal{K} = \mathcal{D} = \Omega = \omega_1, \omega_2, ...\omega_N$, $W : \Omega \times \Omega \to \mathcal{R}$. Funkci $W(i,j)$ lze popsat čtvercovou maticí, jejíž prvek (i,j) značí ztrátu, kterou utrpíme rozhodnutím ω_j, je-li objekt ve stavu $_omega_i$.

Jedním ze speciálních případů je rozhodování podle kritéria minimální chyby, kde matice $W(i,j)$ má speciální tvar:

$$W(i,j) = \begin{cases} 0 & \text{pro} \quad i = j \\ 1 & \text{jinak} \end{cases}$$

Potom platí:

$$L_x(d) = \sum_{s=1}^{N} p(x,\omega_s)W(\omega_s,d(x))$$

a jelikož platí Bayesův vzorec $p(x,\omega_s) = p(\omega_s|x)p(x)$,

$$L_x(d) = p(x)\sum_{s=1}^{N} p(\omega_s|x)W(\omega_s,d(x))$$

a vzhledem k podobě matice W

$$L_x(d) = p(x) \sum_{s=1, s\neq d(x)}^{N} p(\omega_s|x)$$

a protože $\sum_{s=1,s\neq d(x)}^{N} p(\omega_s|x) + P(d(x)|x) = 1$,

$$L_x(d) = p(x)(1 - p(d(x)|x))$$

Nyní budeme $L_x(d)$ minimalizovat: Protože p(x) rozhodnutím neovlivníme, lze psát

$$L_x(d^*) \leq L_x(d) \Leftrightarrow (1 - P(d^*(x)|x)) \leq (1 - P(d(x)|x))$$

pro každé $d(x)$.

Pak tedy $d^*(x) = \arg\max_{d(x)\in\Omega} P(d(x)|x)$.

Tento vztah můžeme přímo přepsat na vztah pro klasifikaci podle kritéria střední ztráty

$$\omega^* = \arg\max_{\omega\in\Omega} P(\omega|x)$$

V některých případech je výpočtově nevýhodné rozhodovat se podle ztráty obrazu $L_x(d(x))$. Vzhledem k tomu, že pouze porovnáváme, můžeme použít jako diskriminační funkci jakoukoli **monotónní** funkci $L_x(d)$. Nejčastěji se pro zjednodušení výpočtů používá logaritmus.

4 Rozhodování s možností nevím - reject option

V některých aplikacích je rozumné uvažovat také možnost, kdy nejsem schopen (nebo ochoten) odpovědět na otázku do které třídy daný objekt patří. To může nastat například pokud by ztráta vyvolaná případným špatným rozhodnutím tak velká, že by převážila nad ztrátou, která vznikne ústupem z rozhodnutí.

V tomto případě rozšíříme množinu identifikátorů tříd o další třídu 'nevím' označenou #. Ztrátová funkce Bayesovského rozhodování podle kritéria minimální chyby má tedy tvar:

$$W(i,j) = \begin{cases} 0 & \text{pro} \quad i = j \\ 1 & \text{pro} \quad i \neq j i \neq \# \\ \epsilon & \text{pro} \quad i = \# \end{cases}$$

5 Dichotomie

Dalším speciálním případem je klasifikace do dvou tříd (dichotomie). Zde lze rozhodovat dle znaménka funkce $g(x)$

$$g(x) = P(\omega_1|x) - P(\omega_2|x)$$

K tomuto rozpoznávání se používá například perceptron.

6 Příklad

Předpokládejme, že lékař určuje pouze podle tělesné teploty, zda je pacient zdravý, má chřipku, zápal plic či jinou nemoc (odpovídá rozhodnutí nevím). Pravděpodobnost určité nemoci při dané teplotě dle statistik je zapsána v následující tabulce:

Teplota	36,4	36,6	36,8	37,0	37,2	37,4	37,6	37,8	38,0	38,2	38,4	38,6	38,8	39,0
Zdravý	0,990	0,958	0,950	0,720	0,510	0,210	0,110	0,010	0,001	0,000	0,000	0,000	0,000	0,000
Chřipka	0,001	0,008	0,010	0,230	0,410	0,680	0,690	0,620	0,580	0,430	0,210	0,090	0,030	0,010
Zápal plic	0,001	0,002	0,002	0,003	0,020	0,040	0,080	0,220	0,270	0,420	0,580	0,430	0,250	0,180

Dlouhodobé zkušenosti dávají dohromady také následující skutečnosti:

Poplatky za vyšetření jsou pro chřipku 300 Kč, pro zápal plic 400 Kč a komplexní vyšetření včetně ostatních nemocí stojí 1000 Kč. Léčení včas rozpoznané chřipky stojí 500 Kč, zápalu plic 1500 Kč, ostatních nemocí 4000 Kč. Léčení pozdě rozpoznané chřipky včetně komplikací přijde namocnici na 1000 Kč, zápalu plic 3500 Kč a ostatních nemocí 5000 Kč.

Při správně rozpoznané chorobě se platí vyšetření a její léčba. Při špatně rozpoznané chorobě nemocnice zaplatí vyšetření na podezřelou chorobu, vyšetření na správnou chorobu a její léčbu. Prohlásí-li doktor, že je pacient zdravý a on přitom trpí nějakou chorobou, zaplatí poté nemocnice vyšetření na danou chorobu a její léčení včetně komplikací.

Otázkou je, jak se lékař rozhodne, dovede-li využívat teorie Bayesovského rozpoznávání.

Nejprve určíme diskriminační funkci:

$$W(i,j) = \begin{bmatrix} 0,00K & 1300,00K & 3900,00K & 6000,00K \\ 300,00K & 800,00K & 2200,00K & 5300,00K \\ 400,00K & 1200,00K & 1900,00K & 5400,00K \\ 1000,00K & 1800,00K & 2900,00K & 5000,00K \end{bmatrix}$$

Nyní již můžeme vyčíslit ztrátovou funkci $L_x(d)$:

Teplota	36,4	36,6	36,8	37,0	37,2	37,4	37,6	37,8	38,0	38,2	38,4	38,6	38,8	39,0
Zdravý	53,20 Kč	210,20 Kč	248,80 Kč	592,70 Kč	971,00 Kč	1460,00 Kč	1929,00 Kč	2564,00 Kč	2701,00 Kč	3097,00 Kč	3795,00 Kč	4674,00 Kč	5334,00 Kč	5 575,00 Kč
Chřipka	342,40 Kč	467,80 Kč498,80 Kč	655,70 Kč	843,00 Kč	1066,00 Kč	1397,00 Kč	1778,00 Kč	1848,00 Kč	2063,00 Kč	2557,00 Kč	3562,00 Kč	4390,00 Kč	4697,00 Kč	
Zápal plic	442,30 Kč	569,40 Kč	601,00 Kč	823,50 Kč	1058,00 Kč	1354,00 Kč	1672,00 Kč	1 976,00 Kč	2014,00 Kč	2124,00 Kč	2488,00 Kč	3517,00 Kč	4399,00 Kč	4728,00 Kč
Jiná nemoc	1034,70 Kč	1 138,20 Kč	1163,80 Kč	1377,70 Kč	1606,00 Kč	1900,00 Kč	2184,00 Kč	2514,00 Kč	2573,00 Kč	2742,00 Kč	3110,00 Kč	3809,00 Kč	4379,00 Kč	4590,00 Kč

Z této tabulky již můžeme bez problémů zjistit, jak se rozhodnout, aby byla ztráta nemocnice minimální.