

COBWEB - zhlukovanie

10. víťaz

- hierarchické zhlukovanie - inkrementálne
 generuje hierarchický strom, kt. sa aktualizuje
 po príchode každého nového príkladu

nový príklad → zaradi do existujúceho zhluku

→ vytvorí nový zhluk

→ modifikuje hierarchiu

→ spojenie 2 zhlukov a zaradenie nového do nich
 → rozdelenie zhluku a zaradenie nového do jedného z nich

$$U = \sum_k \sum_i \sum_j p(a_i = v_{ij} | c_k) p(c_k | a_i = v_{ij}) p(a_i = v_{ij})$$

predpodobnosť
 (čím ľ, tým stôr majú príklady v jednom zhluku či isté hodnoty)

predikčnosť
 (čím ľ, tým menej platí, že príklady iných zhlukov budú mať rovnaké hodnoty)

zácha (hodnoty, kt. sa častejšie spolu vyskytujú, budú mať silnejší vplyv na rozhodnutie kvality)

- tabuľková doména



→ jediná rovná kružka

prejaví sa aj v strome

i sa hodnota jedného atribútu

⊙ $P(c_k) = 1$

každý zhluk má svoju sadu pravdepodobností

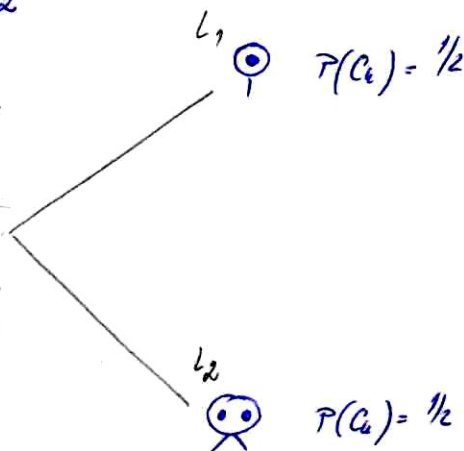
v_{ij}	$P(v_{ij} c_k)$
1B	1
2B	0
1C	1
2C	0
1J	1
2J	0
3J	0

} Bičiky
 } vnútro < číre reálne
 } Jadrá



$P(c_k) = 2/2$ pravdepodobnosť zhluku

v_{ij}	$P(v_{ij} c_k)$
1B	1/2
2B	1/2
1C	2/2 → 1
2C	0
1J	1/2
2J	1/2
3J	0



v_{ij}	$P(v_{ij} c_k)$
1B	1
2B	0
1C	1
2C	0
1J	1
2J	0
3J	0

v_{ij}	$P(v_{ij} c_k)$
1B	0
2B	1
1C	1
2C	0
1J	0
2J	1
3J	0

5 možností rozdelenia

$l=1$ a) k_1 \odot
 $l=2$ b) k_2 \odot \odot
 $l=3$ c) k_3 \odot \odot \odot
 $l=4$ d) k_4 \odot \odot \odot \odot
 $l=5$ e) k_5 \odot \odot \odot \odot \odot

-> počíta sa ÚŽITOČNOSŤ na rybek sypámej reky

$$U = \sum_k \sum_i \sum_j P(V_{ij}/C_k) \cdot P(C_k/V_{ij}) \cdot P(V_{ij})$$

$P(V_{ij})$ -> pravdepodobnosť
 dávajúca mášiu samu
 rasko sa rybkymicim
 celkovú vlastnosťam

V_{ij}	celk. počet rybkymicim	$P(V_{ij})$	$P(V_{ij}/C_k)$ v-tiede					$P(C_k/V_{ij})$ celkov				
			a	b	c	d	e	a	b	c	d	e
1B	1	1/3	1	0	0	1/2	0	1	0	0	1	0
2B	2	2/3	0	1	1	1/2	1	0	1/2	1/2	1/2	2/2
1C	2	2/3	1	1	0	1/2	1/2	1/2	1/2	0	1/2	1/2
1E	1	1/3	0	0	1	1/2	1/2	0	0	1	1	1
1J	1	1/3	1	0	0	1/2	0	1	0	0	1	0
2J	2	2/3	0	1	1	1/2	1	0	1/2	1/2	1/2	2/2
3J	0	0	0	0	0	0	0	0	0	0	0	0

$$U_1 = 1/3 \cdot 1 \cdot 1 + 2/3 \cdot 0 \cdot 0 + 2/3 \cdot 1 \cdot 1/2 + 1/3 \cdot 0 \cdot 0 + 1/3 \cdot 1 \cdot 1 + 2/3 \cdot 0 \cdot 0 + 0 = 1/3 + 0 + 1/3 + 0 + 1/3 + 0 + 0 = 1$$

$$U_2 = 0 + 1/3 + 1/3 + 0 + 0 + 1/3 + 0 = 1$$

$$U_3 = 0 + 1/3 + 0 + 1/3 + 0 + 1/3 + 0 = 1$$

$$U_4 = 1/6 + 1/6 + 1/6 + 1/6 + 1/6 + 1/6 + 0 = 1$$

$$U_5 = 0 + 2/3 + 1/6 + 1/6 + 0 + 2/3 + 0 = 10/6 = 5/3$$

normalizácia

na počet akulor

\odot \odot \odot

$$U = U_1 + U_2 + U_3 = 3$$

$$U' = U/P = 3/3 = 1$$

\odot \odot \odot

$$U = U_4 + U_2 = 2$$

$$U' = U/P = 2/2 = 1$$

\odot \odot \odot

$$U = U_1 + U_5 = 1 + 5/3 = 8/3$$

$$U' = U/P = \frac{8/3}{2} = 8/6 = \frac{4}{3} \checkmark = 1 \frac{1}{3}$$

\odot \odot \odot $P(C_k) = 3/3$

V_{ij}	$P(V_{ij}/C_k)$
1B	1/3
2B	2/3
1C	2/3
1E	1/3
1J	1/3
2J	2/3
3J	0

\odot $P(C_k) = 1/3$

V_{ij}	$P(V_{ij}/C_k)$
1B	0
2B	0
1C	1
1E	0
1J	1
2J	0
3J	0

\odot $P(C_k) = 1/3$

V_{ij}	$P(V_{ij}/C_k)$
1B	0
2B	1
1C	1
1E	0
1J	0
2J	1
3J	0

\odot \odot $P(C_k) = 2/3$

V_{ij}	$P(V_{ij}/C_k)$
1B	0
2B	1
1C	1/2
1E	1/2
1J	0
2J	1
3J	0

\odot $P(C_k) = 1/3$

V_{ij}	$P(V_{ij})$
1B	0
2B	1
1C	0
1E	1
1J	0
2J	1
3J	0

V_{ij}	cell. poc. pr.	$P(V_{ij})$	$P(V_{ij}/C_k)$					$P(C_k/V_{ij})$				
			a	b	c	d	e	a	b	c	d	e
1B	1	$1/4$	1	0	0	$1/2$	0	1	0	0	1	0
2B	3	$3/4$	0	1	1	$1/2$	1	0	$2/3$	$1/3$	$1/3$	1
1C	2	$2/4 = 1/2$	1	$1/2$	0	$1/2$	$1/3$	$1/2$	$1/2$	0	$1/2$	$1/2$
2C	2	$2/4 = 1/2$	0	$1/2$	1	$1/2$	$2/3$	0	$1/2$	$1/2$	$1/2$	1
1J	1	$1/4$	1	0	0	$1/2$	0	1	0	0	1	0
2J	2	$2/4 = 1/2$	0	1	0	0	$2/3$	0	1	0	0	1
3J	1	$1/4$	0	0	1	$1/2$	$1/3$	0	0	1	1	1

$$U_1 = 1/4 + 0 + 1/4 + 0 + 1/4 + 0 + 0 = 3/4$$

$$U_2 = 0 + 1/2 + 1/8 + 1/8 + 0 + 1/2 + 0 = 10/8 = 5/4$$

$$U_3 = 0 + 1/4 + 0 + 1/4 + 0 + 0 + 1/4 = 3/4$$

$$U_4 = 1/2 + 1/2 + 1/8 + 1/8 + 1/2 + 0 + 1/2 = 6/8 = 3/4$$

$$U_5 = 0 + 3/4 + 1/2 + 1/3 + 0 + 1/3 + 1/2 = 19/12$$

$$U = U_1 + U_2 + U_3 = 3/4 + 5/4 + 3/4 = 11/4$$

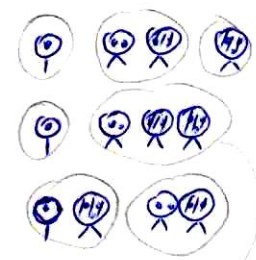
$$U/P = \frac{11}{3} = 11/12 = 0.92$$

$$U = U_1 + U_5 = 3/4 + 19/12 = 22/12 = 7/3$$

$$U/P = \frac{7}{3} = 7/6 = 1.17 \checkmark$$

$$U = U_4 + U_2 = 3/4 + 5/4 = 8/4 = 2$$

$$U/P = \frac{2}{2} = 1$$



V_{ij}	cell. poc. pr.	$P(V_{ij})$	$P(V_{ij}/C_k)$					$P(C_k/V_{ij})$				
			a	b	c	d	e	a	b	c	d	e
1B	0	0	0	0	0	0	0	0	0	0	0	0
2B	3	1	1	1	1	1	1	$1/3$	$1/3$	$1/3$	$2/3$	$2/3$
1C	1	$1/3$	1	0	0	$1/2$	0	1	0	0	1	0
2C	2	$2/3$	0	1	1	$1/2$	1	0	$1/2$	$1/2$	$1/2$	1
1J	0	0	0	0	0	0	0	0	0	0	0	0
2J	2	$2/3$	1	1	0	$1/2$	$1/2$	$1/2$	$1/2$	0	$1/2$	$1/2$
3J	1	$1/3$	0	0	1	$1/2$	$1/2$	0	0	1	1	1

$$U_1 = 0 + 1/3 + 1/3 + 0 + 0 + 1/3 + 0 = 1$$

$$U_2 = 0 + 1/3 + 0 + 1/3 + 0 + 1/3 + 0 = 1$$

$$U_3 = 0 + 1/3 + 0 + 1/3 + 0 + 0 + 1/3 = 1$$

$$U_4 = 0 + 2/3 + 1/6 + 1/6 + 0 + 1/6 + 1/6 = 8/6 = 4/3$$

$$U_5 = 0 + 2/3 + 0 + 2/3 + 0 + 1/6 + 1/6 = 10/6 = 5/3$$

$$U = U_1 + U_2 + U_3 = 3$$

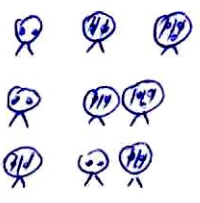
$$U/P = 3/3 = 1$$

$$U = U_1 + U_5 = 1 + 5/3 = 8/3$$

$$U/P = \frac{8}{3} = 8/6 = 1.33 \checkmark$$

$$U = U_2 + U_4 = 1 + 4/3 = 7/3$$

$$U/P = \frac{7}{3} = 7/6 = 1.17$$





$$P(C_6) = 1/4$$

V_{ij}	$P(V_{ij}/C_6)$
1B	1/4
2B	3/4
V \bar{C}	1/2
V \bar{Z}	1/2
1J	1/4
2J	1/2
3J	1/4

Stick figure with a single dot in its eye.

$$P(C_6) = 1/4$$

V_{ij}	$P(V_{ij}/C_6)$
1B	0
2B	1
V \bar{C}	0
V \bar{Z}	0
1J	1
2J	0
3J	0



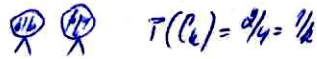
$$P(C_6) = 3/4$$

V_{ij}	$P(V_{ij}/C_6)$
1B	0
2B	1
V \bar{C}	1/3
V \bar{Z}	2/3
1J	0
2J	2/3
3J	1/3

Stick figure with two dots in its eye.

$$P(C_6) = 1/4$$

V_{ij}	$P(V_{ij}/C_6)$
1B	0
2B	1
V \bar{C}	1
V \bar{Z}	0
1J	0
2J	1
3J	0



$$P(C_6) = 2/4 = 1/2$$

V_{ij}	$P(V_{ij}/C_6)$
1B	0
2B	1
V \bar{C}	0
V \bar{Z}	1
1J	0
2J	1/2
3J	1/2



$$P(C_6) = 1/4$$

V_{ij}	$P(V_{ij}/C_6)$
1B	0
2B	1
V \bar{C}	0
V \bar{Z}	1
1J	0
2J	1
3J	0



$$P(C_6) = 1/4$$

V_{ij}	$P(V_{ij}/C_6)$
1B	0
2B	1
V \bar{C}	0
V \bar{Z}	1
1J	0
2J	0
3J	1

