

Pojmy a data mining: na rozhraní umělé inteligence, logiky a psychologie

Radim Bělohlávek



Univerzita Palackého v Olomouci



- nahlédnout do moderní oblasti analýzy dat
- nahlédnout do výzkumu na katedře
- matematické a infromatické základy
- psychologické aspekty



- pojmy a formální konceptuální analýza
slajdy, podrobně str. 3–7, 12–24, přehledově str. 27–48
případně příslušné části z učebního textu
- faktorová analýza binárních a ordinálních dat jako použití
základy: slajdy, str. 5–15
reálný příklad a další: slajdy, str. 47–51, pro zájemce 52–68
- souvislosti s psychologíí pojmů
jsou lidské pojmy formálními koncepty?: slajdy, str. 3–19

- (1) Necht $\langle X, Y, I \rangle$ je formální kontext, $A \subseteq X$ a $B \subseteq Y$ jsou neprázdné množiny. Dokažte, že $\langle A, B \rangle$ je formální koncept v $\langle X, Y, I \rangle$, právě když $\langle A, B \rangle$ tvoří v příslušné tabulce maximální obdélník plný křížků.

Poznámky: (a) Obdélník znamená obdélník až na permutaci řádků a permutaci sloupců. Viz slajdy 18 a 19. (b) Je třeba vhodně definovat pojem maximální obdélník v tabulce.

(2) Vyřešte úkol (i) nebo úkol (ii):

(i) Necht $\langle X, Y, I \rangle$ je formální kontext, kde $X = Y$ a I je relace nerovnosti na množině X , t.j. $\langle x, x \rangle \in I$, právě když $x \neq x$. Dokažte, že konceptuální svaz $\mathcal{B}(X, X, I)$ obsahuje $2^{|X|}$ formálních konceptů.

(ii) Vytvořte binární matici I popisující relaci mezi X a Y , tak že v ní existují dvě různě velké množiny formálních konceptů, tj. $\mathcal{F}, \mathcal{G} \subseteq \mathcal{B}(X, Y, I)$, a to tak, že každá z nich faktorizuje matici I a obě jsou minimální.

Tedy pro \mathcal{F} (a analogicky pro \mathcal{G}) musí platit:

(a) $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$ (\mathcal{F} faktorizuje I);

(b) pro $\mathcal{F}' = \mathcal{F} - \{\langle C, D \rangle\}$, kde $\langle C, D \rangle \in \mathcal{F}$ je libovolný, platí $I \neq A_{\mathcal{F}'} \circ B_{\mathcal{F}'}$ (\mathcal{F} je minimální s vlastností (a)).

Poznámka: V zadání označujeme pro stručnost jak relaci mezi X a Y , tak jí odpovídající binární matici stejným symbolem (I).