

# Zovšeobecnený jednostranne fuzzy konceptový zväz a jeho ekvivalencia s konceptuálnym škálovaním

Peter Butka

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky,  
Technická univerzita v Košiciach, Letná 9, 042 00 Košice, Slovenská republika  
peter.butka@tuke.sk

**Abstrakt.** Tento príspevok sumarizuje práce súvisiace s návrhom a vlastnosťami modelu zovšeobecného jednostranne fuzzy konceptového zväzu, ktorý predstavuje metódu analýzy objekt-atribútových modelov pomocou tzv. formálnej konceptovej analýzy. Konceptový zväz je hierarchická štruktúra organizujúca objekty do skupín podľa podobnosti na základe nimi zdieľaných atribútov. Metódy na báze konceptových zväzov boli ako konceptuálna metóda dolovania dát úspešne aplikované v rôznych oblastiach ako dolovania v textových dokumentoch, vyhľadávanie informácií, systémy pre podporu rozhodovania, znalostný manažment, atď. V rámci tohto príspevku okrem stručného popisu práce definujúcej spomínaný zovšeobecný typ konceptového zväzu, umožňujúceho spracovávať heterogénne vstupné tabuľky dát, spomenieme prácu na tému ekvivalencie nového prístupu a metódy spracovania heterogénnych vstupov pomocou konceptuálneho škálovania.

**Kľúčové slová:** formálna konceptová analýza, konceptový zväz, konceptuálne škálovanie, analýza dát

## 1 Úvod

Jednou z oblastí konceptuálnych metód dolovania v dátach je tzv. formálna konceptová analýza (Formal Concept Analysis - FCA, [4]). Ide o metódu na báze konceptových zväzov identifikujúcu konceptuálnu štruktúru nad vstupnou množinou (tabuľkou) dát definovanou ako objekt-atribútový model. Konkrétne metódy z tejto oblasti si nachádzajú uplatnenie v dolovaní dát z databáz a textov, vyhľadávaní informácií, znalostnom manažmente, podpore rozhodovania a podnikovej inteligencie. Tento príspevok sumarizuje hlavné myšlienky článkov [1,2].

Konceptový zväz, ako výstup analýzy pomocou metód FCA, môže byť chápaný ako znalostný model alebo hierarchia konceptov. Štandardný (alebo klasický) rámec FCA pracuje s tzv. crisp (binárnym) prípadom, ktorý je založený na použití vstupných tabuliek binárnych dát vyjadrujúcich fakt, či objekt má alebo nemá daný atribút. V praxi však existuje potreba spracúvať rôzne typy atribútov, preto sa objavilo viacero prístupov k fuzzifikácii klasického prístupu. Zovšeobecnenie týchto prístupov (vrátane napríklad aj odkazov na jednotlivé predchádzajúce prístupy) založené na dynamickom použití lokálnych Galoisových konexií je možné nájsť v [5]. Tento model je aj základom pre prístup označený ako zovšeobecný jednostranný fuzzy

konceptový zväz, ktorý tvorí základ referovaných článkov v tomto príspevku stručne popísaných v nasledujúcich kapitolách.

## 2 Zovšeobecný jednostranne fuzzy konceptový zväz (GOSCL)

Aplikačne zaujímavé modely v rámci FCA predstavujú modely využívajúce jednostrannú fuzzifikáciu, kde jedna strana zostáva binárna (crisp) a druhá je popísaná fuzzy množinou (napr. objekty sa interpretujú klasicky, avšak atribúty sú fuzzifikované). V rámci článku [1] sme predstavili zovšeobecný model takejto fuzzifikácie umožňujúcej spracovať heterogénne vstupné kontexty (tabuľky dát obsahujúce rôzne typy atribútov). Pre potreby tohto referujúceho príspevku si tento model tzv. zovšeobecného jednostranne fuzzy konceptového zväzu (angl. Generalized One-Sided Concept Lattice – skr. GOSCL) stručne predstavíme. Referovaný článok obsahuje odkazy na predchádzajúce práce, ako aj podrobnejší matematický popis, dôkazy a príklady.

Usporiadaná štvorica  $c = (B, A, L, R)$  sa nazýva *zovšeobecný jednostranný formálny kontext* ak sú splnené nasledujúce podmienky:

1.  $B$  je neprázdna množina objektov a  $A$  je neprázdna množina atribútov.
2.  $L: A \rightarrow CL$  je zobrazenie z množiny atribútov do triedy všetkých kompletných zväzov  $CL$ . Potom, pre každý atribút  $a$ ,  $L(a)$  predstavuje štruktúru pravdivostných hodnôt atribútu  $a$ .
3.  $R$  je zovšeobecná incidenčná relácia, t.j.,  $R(b, a) \in L(a)$  pre všetky  $b \in B$  a  $a \in A$ . Preto,  $R(b, a)$  reprezentuje stupeň z  $L(a)$  ktorý má objekt  $b \in B$  v atribúte  $a$ .

Relácia  $R$  reprezentuje tabuľku dát pre analýzu. Hlavný rozdiel oproti predchádzajúcim prístupom je v možnosti vytvárať konceptový zväz z tabuliek obsahujúcich atribúty rôzneho typu, t.j. s rôznymi štruktúrami pravdivostných hodnôt.

Ak  $(B, A, L, R)$  je zovšeobecný jednostranný formálny kontext, potom môžeme definovať dvojicu zobrazení  $\uparrow: 2^B \rightarrow \prod_{a \in A} L(a)$  and  $\downarrow: \prod_{a \in A} L(a) \rightarrow 2^B$  takto:

$$\uparrow(X)(a) = \inf_{b \in X} (R(b, a)),$$

$$\downarrow(g) = \{ b \in B : \forall a \in A, g(a) \leq R(b, a) \}.$$

Pár zobrazení  $(\uparrow, \downarrow)$  tvorí Galoisovu konexiu medzi  $2^B$  and  $\prod_{a \in A} L(a)$ . Potom máme  $C(B, A, L, R)$  ako množinu všetkých dvojíc  $(X, g)$ , kde  $X \subseteq B$ ,  $g \in \prod_{a \in A} L(a)$ , spĺňajúcich  $\uparrow(X) = g$  a  $\downarrow(g) = X$ . Množina  $X$  je potom označená *extent* a  $g$  *intent* konceptu  $(X, g)$ . Ak dodefinujeme čiastočné usporiadanie na  $C(B, A, L, R)$  ako  $(X_1, g_1) \leq (X_2, g_2)$  iff  $X_1 \subseteq X_2$  iff  $g_1 \geq g_2$ , potom  $C(B, A, L, R)$  spolu s týmto usporiadaním tvorí kompletný zväz, ktorý nazývame *zovšeobecný jednostranne fuzzy konceptový zväz*.

Následne môžeme definovať inkrementálny algoritmus pre vytvorenie GOSCL, t.j. vstupom je kontext definovaný  $(B, A, L, R)$ . Pre  $b \in B$  je  $R(b)$  element zo súčiny zväzov  $\prod_{a \in A} L(a)$  taký že  $R(b)(a) = R(b, a)$ , t.j.,  $R(b)$  je  $b$ -tý riadok tabuľky dát  $R$ . Nech  $1_L$  definuje najväčší element z  $L = \prod_{a \in A} L(a)$ , t.j.,  $1_L(a) = 1_{L(a)}$  pre všetky  $a \in A$ . Potom pseudokód algoritmu pre tvorbu GOSCL je uvedený nižšie. Praktické príklady

na rôznych dátach je možné nájsť v referovanom článku, prípadne v príbuzných článkoch (napr. [2,3], prípadne iné naše práce dosiahnuteľné v databázach vedeckých článkov).

### Algoritmus (Generalized One-Sided Concept Lattice)

Input:  $(B, A, L, R)$  – zovšeobecnený jednostranný formálny kontext

begin

create lattice  $L := \prod_{a \in A} L(a)$

$C := \{1_L\}, C \subseteq L$  (množina všetkých intentov)

while  $(B \neq \emptyset)$

{

choose  $b \in B$

$C^* := C$

for each  $c \in C^*$

$C := C \cup \{c \wedge R(b)\}$

$B := B \setminus \{b\}$

}

$C(B, A, L, R) := \emptyset$

for each  $c \in C$

$C(B, A, L, R) := C(B, A, L, R) \cup \{(\downarrow(c), c)\}$

end

Output:  $C(B, A, L, R)$  – množina všetkých konceptov

## 3 Ekvivalencia GOSCL a konceptuálneho škálovania

V ďalšom článku [3] sme sa podrobnejšie zaoberali modelovaním atribútov a riešili detailne ekvivalenciu medzi použitím GOSCL a využitím metódy konceptuálneho škálovania pre spracovania zložitejších atribútov crisp prístupom. Konceptuálne škálovanie reprezentuje iný prístup k vytváraniu konceptových vzťahov z fuzzy modelov. Jedná sa o transformáciu daného objektovo-atribútového modelu na binárny prípad. Je pridané množstvo nových atribútov a nový model je omnoho väčší. Následne je použitá klasická FCA.

V uvedenom článku je dokázané, že prístupy konceptuálneho škálovania a zovšeobecneného jednostranne fuzzy konceptového vzťahu vytvárajú na rovnakom viachodnotovom formálnom kontexte rovnaké uzáverové systémy na množine objektov. Preto sú aj im odpovedajúce konceptové vzťahy izomorfné. Znamená to, že každý zovšeobecnený jednostranný kontext môže byť jednoducho transformovaný do binárneho kontextu práve metódou škálovania. Následne, použitím štandardných metód môže byť príslušný binárny (crisp) kontext skúmaný, pričom výsledky sú aplikovateľné aj pre prípad GOSCL.

Ako aplikácia dokázanej ekvivalencie je v článku prezentovaná metóda redukcie atribútov zovšeobecneného jednostranného kontextu na jednoduchšie, samozrejme pri zachovaní izomorfizmu konceptových vzťahov (pred a po redukcii).

## 4 Záver

Tento príspevok stručne referuje o dvoch výsledkoch súvisiacich s oblasťou formálnej konceptovej analýzy. Prvým výsledkom je návrh modelu a algoritmu tvorby tzv. zovšeobecneného jednostranne fuzzy konceptového zväzu, ktorý umožňuje spracovávať tabuľky dát z rôznymi typmi atribútov. Druhý výsledok sa zaoberá ekvivalenciou medzi spomínaným modelom a metódou spracovania zložitejších atribútov používanou v klasickom prípade konceptovej analýzy, ktorá sa nazýva konceptuálne škálovanie.

**PodĎakovanie.** Tento príspevok vznikol najmä za podpory agentúry VEGA v rámci projektu č. 1/1147/12 a za podpory agentúry APVV v rámci projektu č. APVV-0208-10. PodĎakovanie patrí aj spoluautorom článkov, ktoré tento príspevok sumarizuje, ako aj ďalším projektom spomínaným v daných prácach.

### English Summary

This work summarizes published journal papers related to the model and properties of generalized one-sided concept lattice, which represents method for the analysis of object-attribute models using Formal Concept Analysis approach. Concept lattice is hierarchical structure organizing the objects into groups based on their similarity according to their shared attributes. Methods based on concept lattices were used as conceptual data analysis technique applicable in different areas like text-mining, information retrieval, decision support systems, knowledge management, etc. In this paper we shortly define some mentioned generalized concept lattice model, which is able to process heterogeneous input data tables, and also we provide information on work related to equivalence of our new approach and classical method for heterogeneous inputs based on the conceptual scaling approach.

### Referencie

1. Butka, P., Pócs, J.: Generalization of one-sided concept lattices. *Computing and Informatics*, (2013), vol. 32, no. 2, pp. 355–370.
2. Butka, P., Pócs, J., Pócsová, J.: Use of concept lattices for data tables with different types of attributes. *Journal of Information and Organizational Sciences*, (2012), vol. 36, no. 2, pp. 1–12.
3. Butka, P., Pócs, J., Pócsová, J.: On equivalence of conceptual scaling and generalized one-sided concept lattices. *Information Sciences*, (2014), vol. 259, 20 February 2014, pp. 57-70.
4. Ganter, B., Wille, R.: *Formal Concept Analysis*. Springer, Berlin, (1999).
5. Pócs, J.: Note on Generating Fuzzy Concept Lattices via Galois Connections. *Information Sciences*, (2012), vol. 185, no. 1, pp. 128–136.