Basic Level in Formal Concept Analysis: Interesting Concepts and Psychological Ramifications

Radim Belohlavek, Martin Trnecka

DATA ANALYSIS AND MODELING LAB
Palacky University, Olomouc, Czech Republic

european social fund in the czech republic
EUROPEAN UNION
MINISTRY OF EDUCATION, YOUTH AND SPORTS
OP Education for Competitiveness
INVESTMENTS IN EDUCATION DEVELOPMENT

Motivation

- Belohlavek R., Klir G.: Concepts and Fuzzy Logic. MIT Press, 2011.
- Two possible interactions with FCA envisioned:
  - A) FCA benefits from Psychology of Concepts (utilizing phenomena studied by PoC)
  - B) Psychology of Concepts benefits from FCA (simple formal framework)
- Psychology of Concepts: big area in cognitive psychology, empirical study of human concepts
- Belohlavek R., Trnecka M.: Basic Level of Concepts in Formal Concept Analysis. In: F. Domenach, D.I. Ignatov, and J. Poelmans (Eds.): ICFCA 2012, LNAI 7278, Springer, Heidelberg, 2012, pp. 28-44.

Previous Work

- Concept lattice usually contains large number of concepts. Difficult to comprehend for a user.
- Our experience: user finds some concepts important (relevant, interesting), some less important, some even "artificial" and not interesting.
- Goal: Select only important concepts.
- Several approaches have been proposed, e.g.:
  - Indices enabling us to sort concepts according to their relevance.
    Kuznetsov's stability index
  - Taking into account additional user's knowledge (background knowledge) to filter relevant concepts
    Belohlavek et al.: attribute dependency formulas, constrained concept lattices
- Our approach: **important are concepts from basic level**.

Basic Level Phenomenon

- Extensively studied phenomenon in psychology of concepts.
- When people categorize (or name) objects, they prefer to use certain kind of concepts.
- Such concepts are called the concepts of the basic level.
- Definition of basic level concepts?:
  Are cognitively economic to use; "carve the world well".
- One feature: Basic level concepts are a compromise between the most general and most specific ones.
- Several informal definitions proposed.

Q: What is this?

A: Dog

. . . Why dog?

There is a number of other possibilities:

- Animal
- Mammal
- Canine beast
- Retriever
- Golden Retriever
- Marley

. . . So why dog?: Because “dog” is a basic level concept.

Basic Level Definitions

- The psychological literature does not contain a single, robust and uniquely intepretable definition of the notion of basic level.
- Several informal definitions.
- We created a metric (function assigning numbers to concepts).
- Large number is stronger indicates that the concept is basic level concept.
- There exists (semi)formalized metrics from psychologists.

Our paper: first step in B)

- We formalize within FCA five selected approaches to basic level.
- Goal: Explore their relationship.
- Several challenges:
  - Every formal model of basic level will be simplistic (potentially a point of criticism from the psychological standpoint).
  - Various issues which are problematic or not yet fully understood from a psychological viewpoint (less knowledge and extensive domain knowledge).

Basic Level Metrics

- For formal context $\langle X, Y, I \rangle$ which describes all the available information regarding the objects and attributes. For a given approach A to basic level, we define a function BL_{A} mapping every concept $\langle A, B \rangle$ in the concept lattice $\mathcal{B}(X, Y, I)$ to $[0, \infty)$ or to $[0, 1]$
- $BL_{\mathrm{A}}(A, B)$ is interpreted as the degree to which $\langle A, B \rangle$ belongs to the basic level.
- A basic level is thus naturally seen as a graded (fuzzy) set rather than a clear-cut set of concepts.

Similarity approach (S)

- Based on informal (one of the first) definition from Eleanor Rosch (1970s).
- Basic level concept satisfies three conditions:
  1. The objects of this concept are similar to each other.
  2. The objects of the superordinate concepts are significantly less similar.
  3. The objects of the subordinate concepts are only slightly more similar.
- Formalized in our previous paper (Belohlavek, Trnecka, ICFCA 2012).

$$BL_{\mathrm{S}}(A, B) = \alpha_1 \otimes \alpha_2 \otimes \alpha_3,$$

- Where α_i represents the degree of validity of the above condition $i = 1, 2, 3$ and $\otimes$ represents a truth function of many-valued conjunction.
- The definition of degrees α_i involves definitions of appropriate similarity measures.

Cue Validity Approach (CV)

- Proposed by Eleanor Rosch (1976).
- Based on the notion of a cue validity of attribute y for concept c, i.e. the conditional probability $p(c|y)$ that an object belongs to c given that it has y.
- The *total cue validity for* c is defined the sum of cue validities of each of the attributes of c.
- Basic level concepts are those with a high total cue validity.

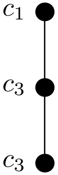
We consider the following probability space: X (objects) are the elementary events, 2^X (sets of objects) are the events, the probability distribution is given by $P(\{x\}) = \frac{1}{|X|}$ for every object $x \in X$. For an event $A \subseteq X$ then, $P(A) = |A|/|X|$. The event corresponding to a set $\{y, \dots\} \subseteq Y$ of attributes is $\{y, \dots\}^{\downarrow}$.

$$\mathit{BL}_{\mathrm{CV}}(A,B) = \sum_{y \in B} P(A|\{y\}^{\downarrow}) = \sum_{y \in B} \frac{|A \cap \{y\}^{\downarrow}|}{|\{y\}^{\downarrow}|}.$$

Cue Validity Remark

Murphy criticizes (Murphy, 1982, 2002) the cue validity approach on the account that total cue validity is monotone w.r.t. inclusion of categories and, hence, achieves its maximum for the most general category.
This claim is wrong. Namely, while it is true that $\langle A_1, B_1\rangle \leq \langle A_2, B_2\rangle$ implies $P(A_1|\{y\}^{\downarrow}) \leq P(A_2|\{y\}^{\downarrow})$, as the author correctly argues, it does not imply that $BL_{\mathrm{CV}}(A_1, B_1) \leq BL_{\mathrm{CV}}(A_2, B_2)$ because the summations run over B_1 and B_2 and we have $B_1 \supseteq B_2$.

	a	b	c
1	$\times$		
2	$\times$	$\times$	
3	$\times$	$\times$	$\times$



	a	b	c	$\sum$
c_1	$\frac{3}{3}$			1
c_2	$\frac{2}{3}$	$\frac{2}{2}$		1.6
c_3	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{1}$	1.8

Cue validity approach is congruent to and under certain assumptions agrees with another view of basic level formulated by Rosch, namely the one based on the number of attributes characterizing the categories, which is often dealt with in experiments.

Category Feature Collocation Approach (CFC)

- Proposed by Jones (1983).
- Defined as product $p(c|y) \cdot p(y|c)$ of the cue validity $p(c|y)$ and the so-called category validity $p(y|c)$.
- The total CFC for c may then be defined as the sum of collocations of c and each attribute.
- Basic level concepts may then again be understood as concepts with a high total CFC.

$$BL_{\mathrm{CFC}}(A,B) = \sum_{y \in Y} (p(c|y) \cdot p(y|c)) = \sum_{y \in Y} \left(\frac{|A \cap \{y\}^{\downarrow}|}{|\{y\}^{\downarrow}|} \cdot \frac{|A \cap \{y\}^{\downarrow}|}{|A|} \right).$$

Category Utility Approach (CU)

- Proposed by Corter and Gluck (1992).
- Utilizes the notation of category utility

$$cu(c) = p(c) \cdot \sum_{y \in Y} [p(y|c)^2 - p(y)^2].$$

- Algorithm COBWEB is base on the CU.

$$\begin{aligned} BL_{\text{CFC}}(A, B) &= P(A) \cdot \sum_{y \in Y} \left[\left(\frac{P(\{y\}^{\downarrow} \cap A)}{P(A)} \right)^2 - P(\{y\}^{\downarrow})^2 \right] = \\ &= \frac{|A|}{|X|} \cdot \sum_{y \in Y} \left[\left(\frac{|\{y\}^{\downarrow} \cap A|}{|A|} \right)^2 - \frac{|\{y\}^{\downarrow}|}{|X|})^2 \right]. \end{aligned}$$

Predictability Approach (P)

- Based on the idea, frequently formulated in the literature.
- Basic level concepts are abstract concepts that still make it possible to predict well the attributes of their objects (enables good prediction, for short).
- This approach has not been formalized in the psychological literature.
- We introduce a graded (fuzzy) predicate $pred$ such that $pred(c) \in [0, 1]$ is naturally interpreted as the truth degree of proposition "concept $c = \langle A, B \rangle$ enables good prediction".
- We use the principles of fuzzy logic to obtain the truth degrees β_1, β_2, and β_3 of the following three propositions:

  1. c has high $pred$;
  2. c has a significantly higher $pred$ than its upper neighbors;
  3. c has only a slightly smaller $pred$ than its lower neighbors.

  This is done in a way analogous to how we approached a similar problem with formalizing the similarity approach to basic level.

Predictability Approach (P)

- Finally, we put

$$BL_{\mathrm{P}}(A,B) = \beta_1 \otimes \beta_2 \otimes \beta_3,$$

  where $\otimes$ is again an appropriate truth function of many-valued conjunction.
- We need to define $pred$.
- For a given concept $c = \langle A, B\rangle$ and attribute $y \in Y$, consider the random variables $V_y : X \to \{0,1\}$ and $V_c : X \to \{0,1\}$ defined by:

$$V_y(x) = 1 \text{ if } \langle x, y\rangle \in I \text{ and } V_y(x) = 0 \text{ if } \langle x, y\rangle \notin I, \text{ and}$$
$$V_c(x) = 1 \text{ if } x \in A \text{ and } V_c(x) = 0 \text{ if } x \notin A.$$

The fact that the value of y is well predictable for objects in c corresponds to the fact that the conditional entropy $E(V_y|V_c = 1)$ is low.

Predictability Approach (P)

$$E(V_y|V_c = 1) = -\frac{|A - \{y\}^{\downarrow}|}{|A|} \cdot \log \frac{|A - \{y\}^{\downarrow}|}{|A|} - \frac{|\{y\}^{\downarrow} \cap A|}{|A|} \cdot \log \frac{|\{y\}^{\downarrow} \cap A|}{|A|}.$$

- Averaging over all the attributes in $Y - B$ (because for $y \in B$ we have $E(V_y|V_c = 1) = 0$), we get quantity:

$$p(c) = \sum_{y \in Y - B} \frac{E(V_y|V_c = 1)}{|Y - B|}.$$

- Since a low value of p corresponds to a good ability to predict by c, i.e. to a high value of $pred(c)$, letting

$$pred(c) = 1 - p(c).$$

Extracting Interesting Concepts - Datasets

- Sports,
  20 objects – sports and 10 attributes (on land, on ice, individual sport . . .), 37 formal concepts.
- Drinks,
  68 objects – drinks and 25 attributes – regarding the composition of drinks (contains alcohol, contains caffeine, contains vitamins, various kinds of minerals), 320 formal concepts.
- DBLP,
  6980 objects – authors of papers in computer science and 19 attributes – conferences in computer science, 2495 formal concepts.

Extracting Interesting Concepts - Experimental Evaluation

Concepts selected for basic level are, for example:

Sports dataset

- "ball games"
- "land sports"
- "individual sports"
- "winter collective sports"
- ...

DBLP dataset

- "Authors publishing in top database conferences"
- "Authors publishing in theory conferences"
- ...

Extracting Interesting Concepts - Experimental Evaluation

Concepts selected for basic level are, for example:

Drinks dataset

- “beers”,
- “energy drinks containing coffein”,
- “liquers”,
- “milk drinks”,
- “mineral waters”,
- “milk drinks”,
- “sweet vitamin drinks”,
- “sparkling drinks”,
- “wines”,
- ...

Comparison of Basic Level Metrics

- Different approaches describe single phenomenon – basic level of concepts.
- For every input data $\langle X, Y, I\rangle$, a given metric BL_{M} (i.e. M being S, CV, CFC, CU, or P) determines a ranking of formal concepts in $\mathcal{B}(X, Y, I)$, i.e. determines the linear quasiorder $\leq_{\mathrm{M}}$ defined by

$$\langle A_1, B_1\rangle \leq_{\mathrm{M}} \langle A_2, B_2\rangle \text{ iff } BL_{\mathrm{M}}(A_1, B_1) \leq BL_{\mathrm{M}}(A_2, B_2).$$

- We examined the pairwise similarities of the rankings $\leq_{\mathrm{S}}$, $\leq_{\mathrm{CV}}$, $\leq_{\mathrm{CU}}$, $\leq_{\mathrm{CFC}}$, and $\leq_{\mathrm{P}}$ for various datasets.
- We used the Kendall tau coefficient to assess the similarities.
- Kendall tau coefficient $\tau(\leq_{\mathrm{M}}, \leq_{\mathrm{N}})$ of rankings $\leq_{\mathrm{M}}$ and $\leq_{\mathrm{N}}$ is a real number in $[-1, 1]$. High values indicate agreement of rankings with 1 in case the rankings coincide. Low values indicate disagreement with -1 in case one ranking is the reverse of the other.
- Even though this approach might seem rather strict, significant patterns were obtained.

Similarity of Rankings Basic Level Concepts

	S	CV	CFC	CU	P
S	1.000	-0.093	-0.215	-0.139	-0.175
CV	-0.093	1.000	0.737	0.754	0.170
CFC	-0.215	0.737	1.000	0.789	0.229
CU	-0.139	0.754	0.789	1.000	0.161
P	-0.175	0.170	0.229	0.161	1.000

Table : Kendall tau coefficients for Sports data.

	S	CV	CFC	CU	P
S	1.000	0.103	0.151	0.067	-0.036
CV	0.103	1.000	0.555	0.581	-0.232
CFC	0.151	0.555	1.000	0.811	-0.061
CU	0,067	0.581	0.811	1.000	-0.064
P	-0.036	-0.232	-0.061	-0.064	1.000

Table : Kendall tau coefficients for Drinks data.

Similarity of Rankings Basic Level Concepts

	S	CV	CFC	CU	P
S	1.000	0.051	0.050	0.039	-0.104
CV	0.051	1.000	0.701	0.699	-0.211
CFC	0.050	0.701	1.000	0.791	-0.164
CU	0.039	0.699	0.791	1.000	-0.107
P	-0.104	-0.211	-0.164	-0.107	1.000

Table : Kendall tau coefficients for synthetic 75×25 datasets.

	S	CV	CFC	CU	P
S	1.000	0.245	0.303	0.350	-0.136
CV	0.245	1.000	0.636	0.631	-0.540
CFC	0.303	0.636	1.000	0.727	-0.645
CU	0.350	0.631	0.727	1.000	-0.458
P	-0.136	-0.540	-0.645	-0.458	1.000

Table : Kendall tau coefficients for synthetic 100×50 datasets.

Similarity of sets of top $\mathbf{r}$ basic level concepts

- Rank correlation may be too strict a criterion.
- Namely, instead of basic-level ranking, one is arguably more interested in the set consisting of the top r concepts of $\mathcal{B}(X, Y, I)$ according to the ranking $\leq_{\mathrm{M}}$ for a given metric M. We denote such set by

$$Top_r^{\mathrm{M}}$$

  with the provision that

  a) if the $(r+1)$-st, ..., $(r+k)$-th concepts are tied with the r-th one the ranking, we add the k concepts to Top_r^{M}.

  b) we do not include concepts to which the metric assigns 0.

Similarity of sets of top $\mathbf{r}$ basic level concepts

- Given metrics M and N, we are interested in whether and to what extent are the sets Top_r^{M} and Top_r^{N} similar.
- For this purpose, we propose the following measure of similarity.
- For formal concepts $\langle C, D\rangle, \langle E, F\rangle \in \mathcal{B}(X, Y, I)$, denote by $s(\langle C, D\rangle, \langle E, F\rangle)$ an appropriately defined degree of similarity. We present results utilizing the similarity based on the simple matching coefficient but other options such as the Jaccard coefficient yield similar results. For two metrics M and N, and a given $r = 1, 2, 3, \ldots,$ we define:

$$S(Top_r^{\mathrm{M}}, Top_r^{\mathrm{N}}) = \min(I_{\mathrm{MN}}, I_{\mathrm{NM}})$$

where

$$I_{\mathrm{MN}} = \frac{\sum_{\langle C,D\rangle \in Top_r^{\mathrm{M}}} \max_{\langle E,F\rangle \in Top_r^{\mathrm{N}}} s(\langle C, D\rangle, \langle E, F\rangle)}{|\,Top_r^{\mathrm{M}}|}$$

and

$$I_{\mathrm{NM}} = \frac{\sum_{\langle E,F\rangle \in Top_r^{\mathrm{N}}} \max_{\langle C,D\rangle \in Top_r^{\mathrm{M}}} s(\langle C, D\rangle, \langle E, F\rangle)}{|\,Top_r^{\mathrm{N}}|}.$$

Similarity of sets of top $\mathbf{r}$ basic level concepts

- $S(Top_r^{\mathrm{M}}, Top_r^{\mathrm{N}})$ may naturally be interpreted as the truth degree of the proposition "for most concepts in Top_r^{M} there is a similar concept in Top_r^{N} and vice versa".
- Because of this interpretation and because S is a reflexive and symmetric fuzzy relation with suitable furhter properties, S is a good candidate for measuring similarity.
- Clearly, high values of S indicate high similarity and $S(Top_r^{\mathrm{M}}, Top_r^{\mathrm{N}}) = 1$ iff $Top_r^{\mathrm{M}} = Top_r^{\mathrm{N}}$.

Similarity of sets of top $\mathbf{r}$ basic level concepts

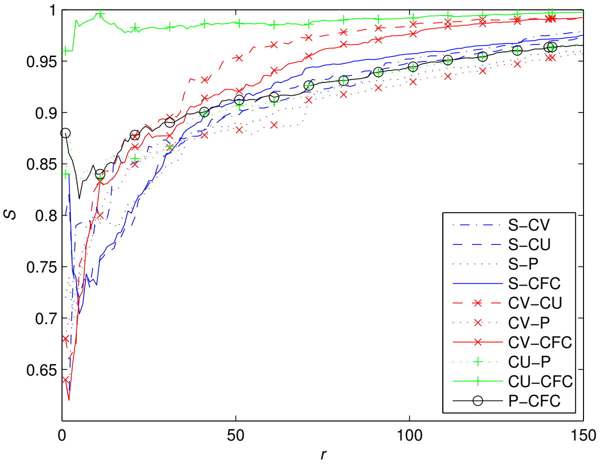


Figure : Similarities S of sets of top r concepts for Drinks data.

Similarity of sets of top $\mathbf{r}$ basic level concepts

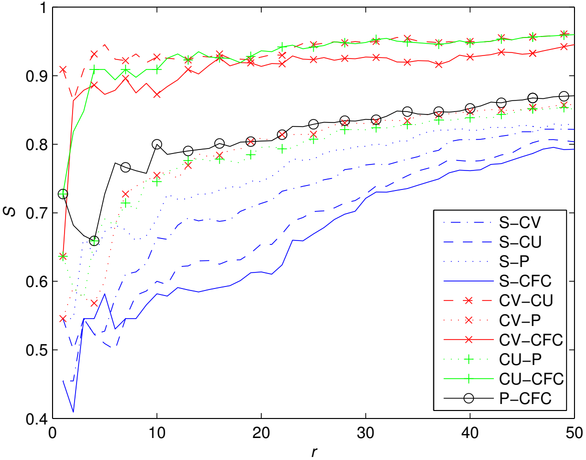


Figure : Similarities S of sets of top r concepts for 75×25 random datasets.

Similarity of sets of top r basic level concepts

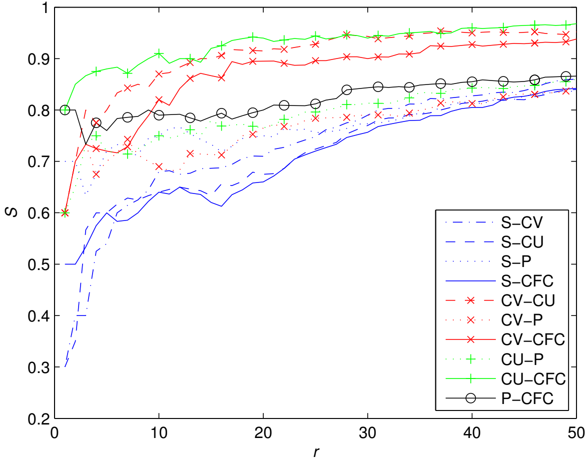


Figure : Similarities S of sets of top r concepts for 100×50 random datasets.

Results

- CU, CFC, and CV may naturally be considered as a group of metrics with significantly similar behavior, while S and P represent separate, singleton groups.
- This observation contradicts the current psychological knowledge. Namely, the (informal) descriptions of S, P, and CU are traditionally considered as essentially equivalent descriptions of the notion of basic level in the psychological literature. On the other hand, CFC has been proposed by psychologists as a supposedly significant improvement of CV and the same can be said of CU versus CFC.

Conclusions

- We formalized within FCA five approaches to basic level.
- All approaches tend deliver informative, natural concepts.
- The experiments performed indicate a mutual consistency of the proposed basic level metrics but also an interesting pattern.

Future Research

- Psychological experiments:
  - Test our methods against respondents' opinion. So far, we used our judgment.
  - Careful design of psychological experiments.
- Present the work to the community of the psychology of concepts.
  - Contrary to rather informal treatment of the basic level in the psychology, we provide simple formal (exact) framework.
  - Several challenges, e.g.:
    a) psychologists usually consider only a tree of concepts;
    b) is basic level a horizontal cut in a concept lattice?