Basic Level of Concepts in Formal Concept Analysis

Radim Belohlavek, Martin Trnecka



Palacky University, Olomouc, Czech Republic



INVESTMENTS IN EDUCATION DEVELOPMENT

Motivation

- Belohlavek R., Klir G. J., Lewis H., III, Way E. C.: Concepts and fuzzy logic: misunderstanding, misconceptions, and oversights. Int. J. Approximate Reasoning, 2010.
- Belohlavek R., Klir G. J.: Concepts and Fuzzy Logic. MIT Press, 2011.
- Psychology of Concepts: big area in cognitive psychology, empirical study of human concepts
- 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)
- Our paper: first step in A)
- Utilize in FCA the so-called basic level of concepts.

In Particular ...

- 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.

Q: What is this?



A: Dog

 \ldots 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 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. Murphy G.: The Big Book of Concepts. MIT Press, 2002.
- We use one of the first approaches, due to Eleanor Rosch (1970s): Objects of the basic level concepts are similar to each other, objects of superordinate concepts are significantly less similar, while objects of the subordinate concepts are only slightly more similar to each other.

An Approach to Basic Level in FCA

Formal concept $\langle A,B\rangle$ belongs to the basic level if it satisfies following properties:

- (BL1) $\langle A, B \rangle$ has a high cohesion.
- (BL2) $\langle A, B \rangle$ has a significantly larger cohesion than its upper neighbours.
- (BL3) $\langle A, B \rangle$ has a slightly smaller cohesion than its lower neighbours.

Cohesion of formal concept = measure of mutual similarity of objects.

Upper neighbors of $\langle A, B \rangle$ are the concepts that are more general than $\langle A, B \rangle$ and are directly above $\langle A, B \rangle$ in the hierarchy of concepts.

Lower neighbors of $\langle A, B \rangle$ are the concepts that are more specific than $\langle A, B \rangle$ and are directly below $\langle A, B \rangle$ in the hierarchy of concepts.

Definition

$$\begin{split} \mathcal{UN}(c) &= \{ d \in \mathcal{B}(X,Y,I) \, | \, c < d \text{ and there is no } d' \text{ for which } c < d' < d \}, \\ \mathcal{LN}(c) &= \{ d \in \mathcal{B}(X,Y,I) \, | \, c > d \text{ and there is no } d' \text{ for which } c > d' > d \}. \end{split}$$

Similarity

Similarity of objects x_1 and x_2 on $\langle X, Y, I \rangle$ can be view as similarity of their corresponding intents.

$$sim(x_1, x_2) = sim_Y(\{x_1\}^{\uparrow}, \{x_2\}^{\uparrow}).$$
 (1)

 $sim(x_1, x_2)$ denotes the degree (or index) of similarity of objects x_1 and x_2 .

Definition

For $B_1, B_2 \subseteq Y$

$$sim_{SMC}(B_1, B_2) = \frac{|B_1 \cap B_2| + |Y - (B_1 \cup B_2)|}{|Y|},$$

$$sim_J(B_1, B_2) = \frac{|B_1 \cap B_2|}{|B_1 \cup B_2|}.$$
(2)
(3)

Cohesion

coh(c) denotes the degree (or index) of cohesion of formal concept c.

Basic Level Degree

- We can compute for every formal concepts $\langle A, B \rangle$ of $\langle X, Y, I \rangle$ degree BL(A, B) to which $\langle A, B \rangle$ is a concept from the basic level.
- Concepts from the basic level need to satisfy conditions (BL1), (BL2), and (BL3), it seems natural to construe BL(A, B) as the degree to which a conjunction of the three propositions, (BL1), (BL2), and (BL3), is true.

$$BL(A,B) = \mathcal{C}(\alpha_1(A,B), \alpha_2(A,B), \alpha_3(A,B)), \tag{6}$$

where

- $\alpha_i(A, B)$ is the degree to which condition (BLi) is satisfied, i = 1, 2, 3,
- ${\mathcal C}$ is a "conjunctive" aggregation function

• Simple form of ${\mathcal C}$

$$\mathcal{C}(\alpha_1,\alpha_2,\alpha_3)=\alpha_1\otimes\alpha_2\otimes\alpha_3.$$

• Degrees are numbers in [0,1], we can use product t-norm $a \otimes b = a \cdot b$.

Formulas

$$\begin{aligned}
\alpha_{1}^{*}(A,B) &= coh^{*}(A,B), \\
\alpha_{2}^{\varnothing*}(A,B) &= 1 - \frac{\sum_{c \in \mathcal{UN}(A,B)} coh^{*}(c)/coh^{*}(A,B)}{|\mathcal{UN}(A,B)|}, \\
\alpha_{2}^{m*}(A,B) &= 1 - \max_{c \in \mathcal{UN}(A,B)} coh^{*}(c)/coh^{*}(A,B), \\
\alpha_{3}^{\varnothing*}(A,B) &= \frac{\sum_{c \in \mathcal{LN}(A,B)} coh^{*}(A,B)/coh^{*}(c)}{|\mathcal{LN}(A,B)|}, \\
\alpha_{3}^{m*}(A,B) &= \min_{c \in \mathcal{LN}(A,B)} coh^{*}(A,B)/coh^{*}(c). \end{aligned}$$
(10)

• * means \emptyset or m

• Values of $\alpha_1(A, B)$, $\alpha_2(A, B)$ and $\alpha_3(A, B)$ (and their variants) may naturally be interpreted as the truth degrees to which the propositions in (BL1), (BL2) and (BL3) are true.

Meaning of Formulas

- If $coh^*(c_1) \leq coh^*(c_2)$, then $\frac{coh^*(c_1)}{coh^*(c_2)} \in [0, 1]$ may be interpreted as the truth degree of " $coh^*(c_1)$ is only slightly smaller than $coh^*(c_2)$ ".
- $1 \frac{coh^*(c_1)}{coh^*(c_2)} \in [0, 1]$ may be interpreted as the truth degree of proposition " $coh^*(c_1)$ is significantly smaller than $coh^*(c_2)$ ".

Lemma

If $\langle A_1, B_1 \rangle \leq \langle A_2, B_2 \rangle$ then $coh^m(A_2, B_2) \leq coh^m(A_1, B_1)$.

However, for coh^{\emptyset} such property no longer holds.

Solution of Problem

Instead of considering $\mathcal{UN}(A, B)$, (all upper neighbors of $\langle A, B \rangle$), we consider only

$$\mathcal{UN}^{\leq}(A,B) = \{ c \in \mathcal{UN}(A,B) \, | \, coh^{\varnothing}(c) \leq coh^{\varnothing}(A,B) \},$$

i.e. only the upper neighbors with a smaller cohesion. It seems natural to disregard $\langle A, B \rangle$ as a candidate for a basic level concept if the number of "wrong upper neighbors" is relatively large, i.e. if $\frac{|\mathcal{UN}^{\leq}(A,B)|}{|\mathcal{UN}(A,B)|} < \theta$ for some parameter θ .

Analogous, instead of considering $\mathcal{LN}(A, B)$, we consider only

$$\mathcal{LN}^{\geq}(A,B) = \{ c \in \mathcal{LN}(A,B) \, | \, coh^{\varnothing}(c) \geq coh^{\varnothing}(A,B) \}$$

and similar condition for the number of "wrong lower neighbors" given by θ .

Experiments

- We performed several experiments.
- We used relative small datasets.
- Subjectivity factor plays a significant role.
- Datasets describing commonly known objects, for which most people would probably agree with selected basic level concepts.
- For every dataset $\langle X, Y, I \rangle$ we compute the basic level degree of all concepts of the concept lattice $\mathcal{B}(X, Y, I)$.
- $BL_{\rm s}^{\rm c,a}(A,B)$: s is SMC or J and indicates whether $sim_{\rm SMC}$ or $sim_{\rm J}$ was used; c is \emptyset or m and indicates whether coh^{\emptyset} or $coh^{\rm m}$ was used; a is \emptyset or m and indicates whether $\alpha_2^{\emptyset*}$ and $\alpha_3^{\emptyset*}$, or $\alpha_2^{\rm m*}$ and $\alpha_3^{\rm m*}$ was used.

Experiment (sports)

		on land	on ice	in water	collective sport	individual sport	using ball	needs opponent	multiple disciplines	points	time
		1	2	3	4	5	6	7	8	9	10
Run	1	×				×					×
Orienteering	2	×				×					\times
Gymnastics	3	×				\times			\times	×	
Triathlon	4	×		×		\times			\times		×
Football	5	×			×		×	×		×	
Inline Hockey	6	×			×		×	×		×	
Tennis	7	×			×		×	×		×	
Baseball	8	×			×		×	×		×	
Ice Hockey	9		×		×			×		×	
Curling	10		×		×					×	
:											

		on land	on ice	in water	collective sport	individual sport	using ball	needs opponent	multiple disciplines	points	time
		1	2	3	4	5	6	7	8	9	10
:											
Cross-country Skiing	11		Х			\times					×
Synchronized Skating	12		Х		Х					×	
Alpine Skiing	13		Х			\times					\times
Biathlon	14		Х			\times			\times		×
Speed Skating	15		×			×					×
Synchronized Swimming	16			×	×				×	×	
Diving	17			×		\times				×	
Water Polo	18			×	×		×	×		×	
Underwater Diving	19			×		×				×	
Rowing	20			×	×						×

intent of concept $\langle A, B \rangle$							basic level degree of $\langle A,B angle$										
on land	on ice	in water	collective sport	individual sport	using ball	needs opponent	multiple disciplines	points	time	$BL_{ m SMC}^{arphi arphi}$	$BL_{ m SMC}^{m{lpha}}$	$BL_{ m SMC}^{moldsymbol{lpha}}$	$BL_{ m SMC}^{ m mm}$	$BL_{J}^{\not {m eta} \not {m eta}}$	$BL_{\mathrm{J}}^{\mathrm{eta}\mathrm{m}}$	$BL_{J}^{m} oldsymbol{arsigma}$	$BL_{ m J}^{ m mm}$
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1	0	0	0.33	0.33	0	0	0.08	0.08
0	0	0	0	0	0	0	0	1	0	0	0	0.11	0.09	0	0	0.05	0.04
0	0	0	0	0	0	0	1	0	0	0	0	0.21	0.20	0	0	0.06	0.06
0	0	0	0	1	0	0	0	0	1	0	0	0.08	0.07	0	0	0.09	0.07
0	0	0	0	1	0	0	0	1	0	0.10	0.07	0.21	0.14	0.10	0.10	0.10	0.09
0	0	0	0	1	0	0	1	0	0	0	0	0.13	0.09	0.05	0.01	0.12	0.11
0	0	0	0	1	0	0	1	0	1	0	0	0.07	0.07	0	0	0.08	0.08
0	0	0	1	0	0	0	0	0	0	0	0	0.22	0.20	0	0	0.07	0.06
0	0	1	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0.14	0.14	0.13	0.13	0.11	0.10	0.03	0.03
0	1	0	0	1	0	0	0	0	1	0.18	0.13	0.36	0.27	0.29	0.23	0.38	0.31
0	1	0	1	0	0	0	0	1	0	0.20	0.19	0.41	0.36	0.27	0.24	0.40	0.35
0	1	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0
									÷								

Experiments



Figure : Concept lattice of Sports. Red square: $BL_{SMC}^{\emptyset\emptyset} > 0$.

Experiment Evaluation

Concepts selected for basic level are, for example:

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

• . . .

Arguably, all of them are likely to be considered natural, basic level concepts.

Concepts not selected for basic level are, for example:

- "individual land sports with multiple disciplines"
- "sports performed in water with multiple disciplines"
- "individual winter sports with multiple disciplines evaluated by time"
- "collective winter sports with opponent evaluated by points"
- "sports evaluated by time"

• . . .

Notes on Experiments

- We consider the method promising and giving reasonable results already at this stage.
- We were not checking the results of our method for a given dataset against a psychological experiment.
- An important observation, basic level depends on the dataset and the selected attributes in particular. Typically, a human expert tends to take into account other information (not only the attributes present in the dataset)
- It seems not to matter very much whether $\alpha_2^{\emptyset^*}$ and $\alpha_3^{\emptyset^*}$, or $\alpha_2^{m^*}$ and $\alpha_3^{m^*}$ is used. On the other hand, it matters significantly whether coh^{\emptyset} or coh^m is used. According to our intuition and the results of this and other experiments we performed, we hypothesize that coh^{\emptyset} is better to use than coh^m .
- More detailed study is needed to support this claim.

Conclusions

- Method utilizing basic level of concepts to select possibility important concepts.
- First results and experience obtained from experiments.
- Method seems to deliver reasonable basic level concepts.
- Simple, FCA-based formal framework for basic level study.

Future Research

- Further experiments (drinks dataset, ...).
- Utilizing other approaches to the basic level developed in the psychology of concepts.
- Algorithmic aspects: Compute efficiently basic level concepts.
- Psychological experiments:
 - Test our method 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?