Quality measure of fuzzy formal concepts

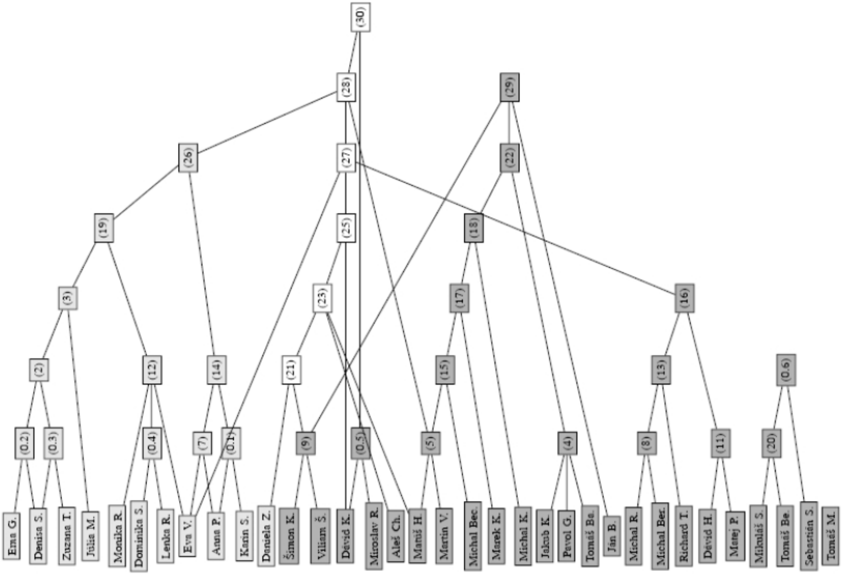
Krajči Stanislav, Antoni Ľubomír
Institute of Computer Science, UPJŠ Košice

FSTA 2012
Liptovský Ján
February 2, 2012

Introduction

- **fuzzy formal concept analysis** – special data-mining method
- method of multi-valued data analysis
- try to discover concepts (i. e. clusters, groups) of similar objects
- [Krajči – Krajčiová, 2008] Social networks and fuzzy formal concept analysis
- special social network: school class
- each student expressed relationships to all schoolmates by values from a given range
- used method:
  Krajči's **one-sided fuzzy concept lattice** including modified **Rice** & **Siff's algorithm**
- obtained results:
  clusters, i.e. groups of pupils sensed by schoolmates in a similar way

- the most of Krajči's clusters:



New experiment

- new data of the same class obtained in 2011
- new approach to gain results
- 29 students (8 girls and 21 boys)
- each student expressed his/her relationship to each schoolmates by 7 values

value	explanation
3	he/she is my very good friend
2	he/she is my friend
1	I tend him/her positively
0	I tend him/her neutrally
-1	I tend him/her negatively
-2	I do not like him/her
-3	I dislike him/her

- the result table
- rows as evaluated schoolmates
- columns as evaluating schoolmates

	gender	name	1	2	3	...	17	18	19	20	21	...	27	28	29
1	M	Ján B.	**3**	0	-1	...	-1	2	—	-1	1	...	1	-1	1
2	M	Tomáš Ba.	2	**3**	3	...	1	2	—	3	1	...	3	2	0
3	M	Michal Bec.	2	3	**3**	...	0	2	—	2	1	...	3	0	1
⋮	⋮	⋮	⋮	⋮	⋮	⋱	⋮	⋮	⋮	⋮	⋮	⋱	⋮	⋮	⋮
17	F	Anna P.	2	2	1	...	**3**	2	—	1	3	...	1	3	1
18	M	Matej P.	2	2	2	...	0	**3**	—	3	2	...	3	1	2
19	M	Miroslav R.	0	-2	-3	...	0	-1	—	-1	0	...	-1	-1	-1
20	M	Michal R.	2	2	2	...	0	3	—	**3**	1	...	3	0	2
21	F	Lenka R.	2	1	2	...	2	2	—	0	**3**	...	2	2	1
⋮	⋮	⋮	⋮	⋮	⋮	⋱	⋮	⋮	⋮	⋮	⋮	⋱	⋮	⋮	⋮
27	M	Richard T.	3	3	3	...	1	3	—	3	2	...	**3**	2	3
28	F	Eva V.	2	3	1	...	3	2	—	1	3	...	3	**3**	1
29	M	Martin V.	2	0	1	...	0	2	—	1	1	...	1	0	**3**

- one of them rejected to participate at the evaluation (19th column)
- he was evaluated only by schoolmates (19th row)
- maximal values on the table diagonale

- using the funkction $x \mapsto \dfrac{x+3}{6}$ can be table values transformed to $\left\{0, \frac{1}{6}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{5}{6}, 1\right\} \subseteq [0,1]$
- **the main purpose:** obtain not only groups of students sensed similar, but give a quality of these groups
- new approach in the experiment:
  fuzzy **context**, fuzzy α-**cut**, fuzzy α-**concept**, **quality** of concept

Fuzzy context

- B – **objects**, $B \neq \emptyset$
- A – **attributes**, $A \neq \emptyset$
- R – **a fuzzy relation on** $B \times A$, **i. e.** $R : B \times A \to [0,1]$

- R – a **table**
  B – its **rows** (evaluated schoolmates)
  A – its **colums** (evaluating schoolmates)
- $R(b, a)$ – the **degree** to which the object b carries the attribute a

Fuzzy α-cuts

$R : B \times A \to [0, 1]$

$R_\alpha \subseteq B \times A, \quad \alpha \in [0, 1]$

Two different approaches:

- **upper α-cuts**: $R_\alpha = \{\langle b, a\rangle \in B \times A : R(b, a) \geq \alpha\}, \quad \alpha \in [0, 1]$
- **lower α-cuts**: $R_\alpha = \{\langle b, a\rangle \in B \times A : R(b, a) \leq \alpha\}, \quad \alpha \in [0, 1]$

example of fuzzy context

	a_1	a_2	a_3	a_4
b_1	0.4	0.2	0.6	0.3
b_2	0.6	0.8	0.9	0.4
b_3	0.1	0.6	0.7	0.3
b_4	0.2	0.3	0.4	0.2

upper 0.4-cut

	a_1	a_2	a_3	a_4
b_1	×		×	
b_2	×	×	×	×
b_3		×	×	
b_4			×	

Fuzzy α-concept

$X \subseteq B, Y \subseteq A, \alpha \in [0, 1]$

- $X^{\nearrow\alpha} = \{a \in A: (\forall b \in X)\ R(b, a) \geq \alpha\}$
- $Y^{\swarrow\alpha} = \{b \in B: (\forall a \in Y)\ R(b, a) \geq \alpha\}$

If $X = X^{\nearrow\alpha\swarrow\alpha}$ then the pair $\langle X, X^{\nearrow\alpha}\rangle$ is called an **fuzzy α-concept**.
Set of all fuzzy α-concepts is called **α-lattice** (L_α).

example of fuzzy 0.4-concept $\langle\{b_2, b_3\}, \{a_2, a_3\}\rangle$

	a_1	a_2	a_3	a_4
b_1	×		×	
b_2	×	×	×	×
b_3		×	×	
b_4			×	

- generating all α-concepts (e. g.: groups of students):
  - a) by definition to try all subsets of B (complexity: $2^{|B|}$)
  - b) by algorithms (better complexity)

- subset of objects in α-concepts for different $\alpha \in [0,1]$ can be equal
- even for p – positive integer, where $p < n$:
  if table of fuzzy relation R contains $n+1$ different values, then L_α for all $\alpha \in \left(\frac{p}{n}, \frac{p+1}{n}\right]$ are identical
- so it is sufficient to consider α-cuts only for $\alpha \in \left\{\frac{0}{n}, \frac{1}{n}, \dots, \frac{n}{n}\right\}$
- and count how many times every subset of objects appears in all L_α

Quality of fuzzy α-concept

$$q(X) = \frac{\left|\left\{p \in \{0,1,\dots,n\} : \ (\exists Y \subseteq A)\ \langle X, Y\rangle \in L_{\frac{p}{n}}\right\}\right|}{n+1}$$

- the values $q(X)$ are rational numbers
- $q(X) = 0$ X is not α-concept for any α
- $q(X) > 0$ X is α-concept for some α
- a higher number corresponds to a more significant concept

subset	$\frac{0}{n}$-concept	$\frac{1}{n}$-concept	...	...	$\frac{n}{n}$-concept	quality of subset
X_1		✓				$q(X_1)$
X_2						$q(X_2)$
$\vdots$						$\vdots$
$\vdots$						$\vdots$
$X_{2^{\lvert B \rvert}}$						$q(X_{2^{\lvert B \rvert}})$
	$\lvert L_{\frac{0}{n}} \rvert$	$\lvert L_{\frac{1}{n}} \rvert$	...	...	$\lvert L_{\frac{n}{n}} \rvert$	

- subset ordering by quality:

$$q(X_{j_1}) \leq q(X_{j_2}) \leq q(X_{j_3}) \leq \quad \ldots \quad \leq q\left(X_{j_{2^{|B|}-1}}\right) \leq q\left(X_{j_{2^{|B|}}}\right)$$

less significant concepts **significant concepts**

	name	Ján B.	Tomáš Ba.	Michal Bec.	Michal Ber.
1	Ján B.	1	$\frac{1}{2}$	$\frac{1}{3}$	1
2	Tomáš Ba.	$\frac{5}{6}$	1	1	$\frac{2}{3}$
3	Michal Bec.	$\frac{5}{6}$	1	1	$\frac{2}{3}$
4	Michal Ber.	1	$\frac{1}{3}$	$\frac{5}{6}$	1

	α-concept							
subset	$\alpha = 0$	$\alpha = \frac{1}{6}$	$\alpha = \frac{1}{3}$	$\alpha = \frac{1}{2}$	$\alpha = \frac{2}{3}$	$\alpha = \frac{5}{6}$	$\alpha = 1$	quality
$\emptyset$								0.00
$\{1\}$								0.00
$\{2\}$								0.00
$\{3\}$								0.00
$\{4\}$						✓		**0.14**
$\{1, 2\}$								0.00
$\{1, 3\}$								0.00
$\{1, 4\}$						✓	✓	**0.29**
$\{2, 3\}$				✓	✓	✓	✓	**0.57**
$\{2, 4\}$								0.00
$\{3, 4\}$								0.00
$\{1, 2, 3\}$				✓				**0.14**
$\{1, 2, 4\}$								0.00
$\{1, 3, 4\}$								0.00
$\{2, 3, 4\}$				✓	✓	✓		**0.43**
$\{1, 2, 3, 4\}$	✓	✓	✓	✓	✓	✓	✓	**1.00**

- for 29 students is not effective to try all subsets (2^{29} possibilities)
- **modified Ganter's algorithm**
- original algorithm: computation the next concept from previous one
- our modification: provide α-cuts and quality measure of fuzzy concepts

input B, A, R, n
for $\left(\alpha = \frac{0}{n}, \frac{1}{n}, \ldots, \frac{n}{n}\right)$ do
  $X \leftarrow \emptyset^{\nearrow\alpha\swarrow\alpha}$
  $L_\alpha \leftarrow X$
  $q(X) \leftarrow q(X) + \frac{1}{n+1}$
  while $X \neq B$ do
    for $(i = |B|, |B - 1|, \ldots, 0)$ do
      $W \leftarrow ((X \cap \{1, 2, \ldots, i - 1\}) \cup \{i\})^{\nearrow\alpha\swarrow\alpha}$
      if $((X \cap \{1, 2, \ldots, i - 1\} = W \cap \{1, 2, \ldots, i - 1\})$ and $(i \in X \setminus W))$
        $L_\alpha \leftarrow X$
        $X \leftarrow W$
        $q(X) \leftarrow q(X) + \frac{1}{n+1}$
output L_α, q

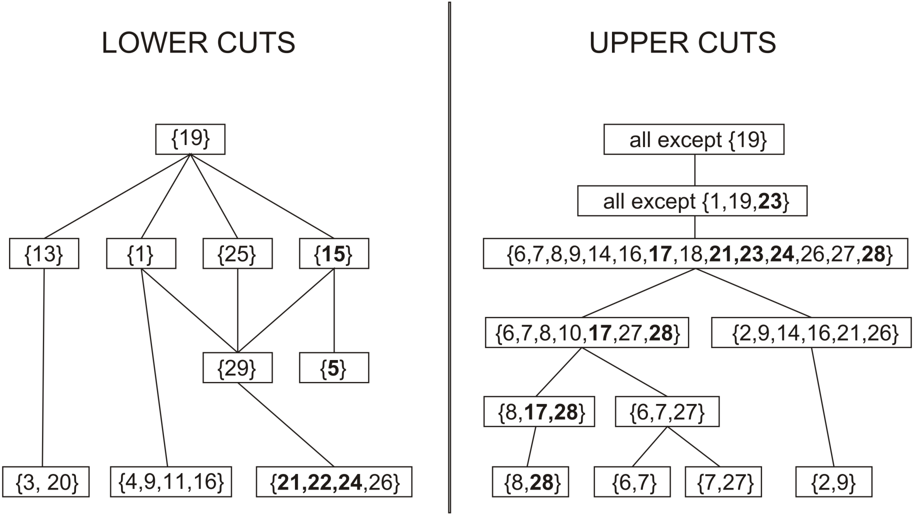
Result of the experiment

- **we obtained totally:**
  about 10 000 concepts in upper cuts (UC) and 50 000 concepts in lower cuts (LC)
- **number of significant concepts**, i. e. $q(X) > 0.25$:
  UC: about 50 concepts, LC: about 70 concepts

quality	upper cuts concepts	quality	lower cuts concepts
0.71	all students except $\{19\}$	0.71	$\{19\}$
0.43	all students except $\{1, 19\}$	0.57	$\{19, 25\}$
0.43	all students except $\{19, 23\}$	0.57	$\{13, 19\}$
0.43	all students except $\{1, 19, 23\}$	0.57	$\{1, 19\}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$

- **the most significant concepts:**
  UC: groups of students sensed by schoolmates positive
  LC: groups of students sensed by schoolmates negative
- **e. g. :** student (19) who rejected to participate at the evaluation:
  UC: **does not occur** in the most significant groups
  LC: **occurs** in the most significant groups
- **gender division** of the groups visible in UC and LC

- the most significant relationships:



Conclusions

I. conclusion:

- this approach give some useful information about structure of selected social network
- can help to class teacher or personal managers to compose teams
- first usage of quality measure of fuzzy concepts directly linked with social networks

II. conclusion:

- it is appropriate to try cuts with lower and upper boundaries

Fuzzy α, β-cuts:

$$R_{\alpha,\beta} = \{\langle b, a\rangle \in B \times A : \ \alpha \leq R(b, a) \leq \beta\}, \quad \alpha, \beta \in [0, 1]$$

- this approach require to execute n^2 cuts and may be more precise

III. conclusion:

- **next aim**: compare results with modified Rice & Siff method for experiment in 2011

Thank you for your attention

stanislav.krajci@upjs.sk
lubomir.antoni@student.upjs.sk