Triadic Formal Concept Analysis and Triclustering: Searching for Optimal Patterns

Dmitry I. Ignatov

National Research University Higher School of Economics, Moscow, Russia
Department of Data Analysis and Machine Intelligence, Faculty of Computer Science &
Laboratory of Intelligent Systems and Structural Analysis

Moscow
2016

Outline

A large amount of structured and unstructured data generates triadic data.
E.g. folksonomy is a set of triples (user, object, tag)

Concrete examples:

- Bibsonomy.org (user, bookmark, tag)
- Social networking sites (user, group, interest)
- Delicious (user, link, tag)

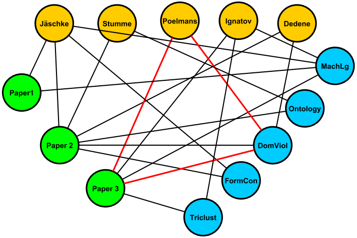


Figure : Folksonomy as a graph.

Main research question

Which triclusters are good approximation of the triconcepts of a given triadic data?

- A.Arnauld, P.Nicole, Logique de Port-Royal (1662)

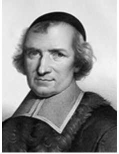


Figure : Antoine Arnauld

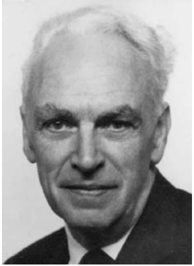
Figure : Pierre Nicole

- A.Arnauld, P.Nicole, Logique de Port-Royal (1662)
- E. Galois (1811-1832)

- A.Arnauld, P.Nicole, Logique de Port-Royal (1662)
- E. Galois (1811-1832)
- G. Birkhoff, since 1930s

Lattices of closed sets. Short history.

- A.Arnauld, P.Nicole, Logique de Port-Royal (1662)
- E. Galois (1811-1832)
- G. Birkhoff, since 1930s
- O. Øre, since 1930s



Lattices of closed sets. Short history.

- A.Arnauld, P.Nicole, Logique de Port-Royal (1662)
- E. Galois (1811-1832)
- G. Birkhoff, since 1930s
- O. Øre, since 1930s
- M. Barbut, B. Monjardet, Ordre et classification, Hachette, Paris, 1970

Bernard Monjardet

- R. Wille, Restructuring lattice theory: An approach based on hierarchies of concepts, 1982

- R. Wille, Restructuring lattice theory: An approach based on hierarchies of concepts, 1982
- B. Ganter, R. Wille, Formale Begriffsanalyse, Springer, 1996
- B. Ganter, R. Wille, Formal Concept Analysis, Springer, 1999
- Chapter in B. Davey, H. Priestly, Introduction to Order and Lattices, 1990.
- Chapter in G. Grätzer (Ed.), General Lattice Theory.
- Concept Data Analysis, C.Carpineto, G. Romano, 2004.
- Galois Connections and Applications, K. Denecke, M. Erné, S. L. Wismath (Eds.), Springer Science & Business Media, 2004

Main conferences on FCA

- International Conference on Conceptual Structures (ICCS), FCA participation starting from 1996 (Proceedings in LNAI, Springer)

- International Conference on Formal Concept Analysis (ICFCA), from 2003 года (Proceedings in LNAI, Springer)

- International Conference on Concept Lattices and Their Applications (CLA), from 2006, special issues

Formal Concept Analysis
[R.Wille, 1982], [B.Ganter and R.Wille, 1999]

- G, a set of objects
- M, a set of attributes
- relation $I \subseteq G \times M$ such that $(g, m) \in I$ if and only if the object g has the attribute m.
- $\mathbb{K} := (G, M, I)$ is a formal context.

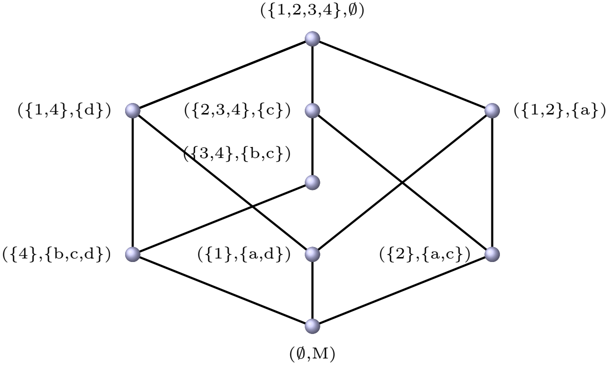
Derivation operators: $A \subseteq G$, $B \subseteq M$

$$A' \stackrel{\text{def}}{=} \{m \in M \mid gIm \text{ for all } g \in A\},\ B' \stackrel{\text{def}}{=} \{g \in G \mid gIm \text{ for all } m \in B\}$$

A formal concept is a pair (A, B): $A \subseteq G,\ B \subseteq M,\ A' = B,$ and $B' = A$.

- A is the extent and B is the intent of the concept (A, B).
- The concepts, ordered by $(A_1, B_1) \geq (A_2, B_2) \iff A_1 \supseteq A_2$ $(B_2 \supseteq B_1)$
  form a complete lattice, called the concept lattice $\underline{\mathfrak{B}}(G, M, I)$.

Example of context of geometrical figures and its concept lattice



	G \ M	a	b	c	d
1		×			×
2		×		×	
3			×	×	
4			×	×	×

a – has exactly 3 vertices,

b – has exactly 4 vertices,

c – has a right angle,

d – is equilateral

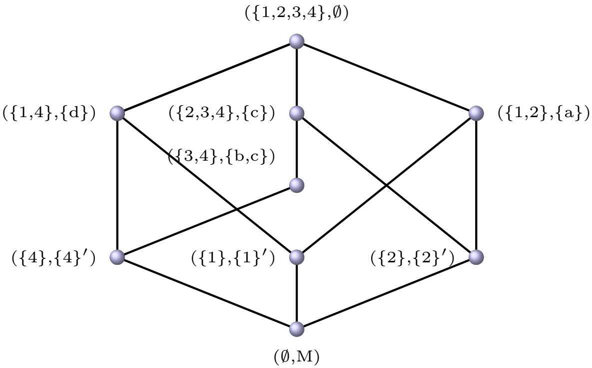
Implications

- Implication $A \to B$ for $A, B \subseteq M$ holds if $A' \subseteq B'$, i.e., every object that has all attributes from A also has all attributes from B.
- Armstrong rules:

$$\frac{}{A \to A}, \quad \frac{A \to B}{A \cup C \to B}, \quad \frac{A \to B, D \cup B \to C}{A \to C}$$

- A Minimal implication base:
  A base with the minimum number of implications [Duquenne, Guigues 1986]
  or
  the stem base, its premises can be given (Ganter 1987) by pseudointents:
  - A set $P \subseteq M$ is a pseudointent if
    $P \neq P''$ and
    $Q'' \subset P$ for every pseudointent $Q \subset P$.

Concept lattice and implications



	G \ M	a	b	c	d
1		×			×
2		×		×	
3			×	×	
4			×	×	×

a – exactly 3 vertices,
b – exactly 4 vertices,
c – has a direct angle,
d – equilateral

Implications:

$abc \to d$

$b \to c$

$cd \to b$

Partial implications or association rules [Luxenburger M., 1991], [Agrawal R. et al., 1993]

- Luxenburger M. Implications partielles dans un contexte. Mathématiques, Informatique et Sciences Humaines, 113 (29) : 35-55, 1991.
- Agrawal R., Imielinski T., Swami A. Mining association rules between sets of items in large databases, Proceedings, ACM SIGMOD Conference on Management of Data, pp. 207-216, 1993.

Let $\mathbb{K} = (G, M, I)$ be a formal context.

Definition 1

Association rule of the context $\mathbb{K}$ is an attribute dependency $A \to B$, where $A, B \subseteq M$.

Definition 2

Support of the association rules $A \to B$ is $supp(A \to B) = \frac{|(A \cup B)'|}{|G|}$.

Definition 3

Confidence of the association rule $A \to B$ is $conf(A \to B) = \frac{|(A \cup B)'|}{|A'|}$.

Example

An object-attribute table (context) of transactions

Clients/goods	Beer	Cakes	Milk	Muesli	Chips
C_1	1	0	0	0	1
C_2	0	1	1	1	0
C_3	1	0	1	1	1
C_4	1	1	1	0	1
C_5	0	1	1	1	1

- $supp(\{\text{Beer, Chips}\}) = 3/5$
- $supp(\{\text{Cakes, Muesli }\} \rightarrow \{\text{ Milk }\}) =$
  $= \frac{|(\{\text{Cakes, Muesli}\} \cup \{\text{Milk}\})'|}{|G|} = \frac{|\{C2, C5\}|}{5} = 2/5$
- $conf(\{\text{Cakes, Muesli }\} \rightarrow \{\text{ Milk }\}) =$
  $= \frac{|(\{\text{Cakes, Muesli}\} \cup \{\text{Milk}\})'|}{|\{\text{Cakes, Muesli}\}'|} = \frac{|\{C2, C5\}|}{|\{C2, C5\}|} = 1$

General task of finding association rules

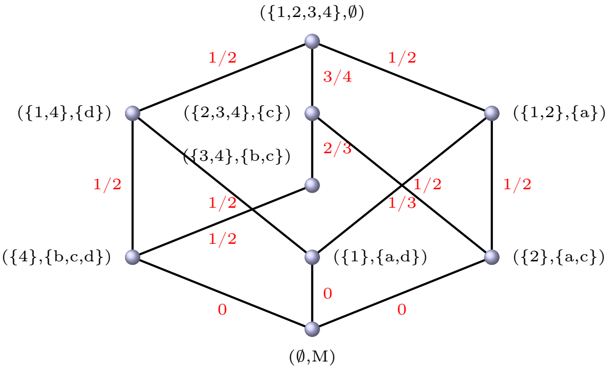
Find all “frequent” (with support greater than a threshold) association rules with confidence greater than a threshold.
Solution stages

- Find all frequent "closed itemsets" (frequent intents)
- For each frequent intent B find all its maximal subintents $A_1, \ldots, A_n$
- Retain only those A_i for which $\text{conf}(A_i \to B) \geq \theta$, where θ is confidence threshold
- Find minimal generators of the remaining A_i, compose rules of the form $mingen(A_i) \to B$.

Luxenburger basis

- Spanning tree of the concept lattice diagram
- Duquenne-Guigues implication base

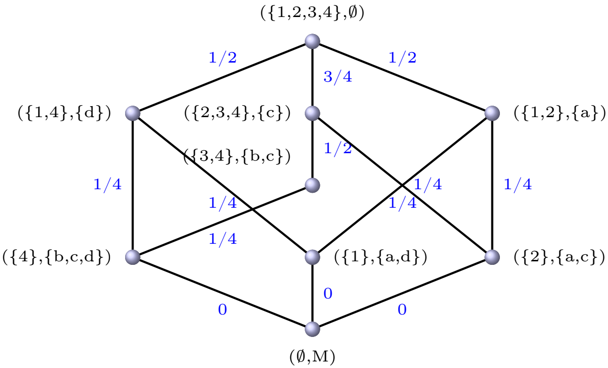
Example. Confidence of association rules

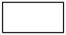


	G \ M	a	b	c	d
1	△	×			×
2	◺	×		×	
3	▭		×	×	
4	□		×	×	×

Good rules with *supp* $\geq 1/2$ and *minconf* $\geq 3/4$

1. $\emptyset \to c$, $\quad \sup(\emptyset \to c) = \text{conf}(\emptyset \to c) = 3/4$;
2. $c \to b$, $\quad \sup(c \to b)= 1/2$, $\text{conf}(c \to b) = 2/3$.



	G \ M	a	b	c	d
1		×			×
2		×		×	
3			×	×	
4			×	×	×

Good rules with *supp* $\geq 1/2$ and *minconf* $\geq 3/4$

1. $\emptyset \to c$, $\sup(\emptyset \to c) = \text{conf}(\emptyset \to c) = 3/4$;
2. $c \to b$, $\sup(c \to b) = 1/2$, $\text{conf}(c \to b) = 2/3$.

Frequent Itemset Mining and FCA

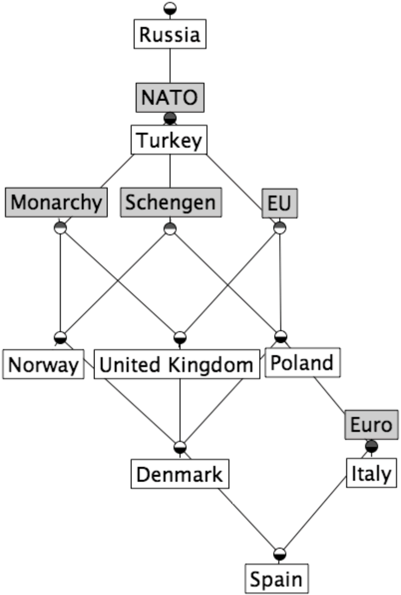
- Agrawal R., RSFDGrC 2011, Moscow

Agrawal et al., Mining Videos from the Web for Electronic Textbooks (2014)

- Start with any (possibly empty) set of objects.
- Generate an implication valid in the current subcontext.
- If the implication is not valid in the entire context, provide an object that violates it.
- Go to the next implication, etc.

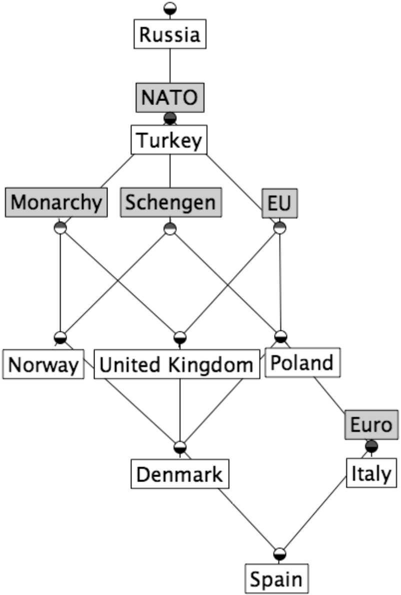
- Start with any (possibly empty) set of objects.
- Generate an implication valid in the current subcontext.
- If the implication is not valid in the entire context, provide an object that violates it.
- Go to the next implication, etc.

Follow the Duquenne-Guigues basis to ask no more questions than is strictly necessary.



Question Is every European monarchy in NATO?

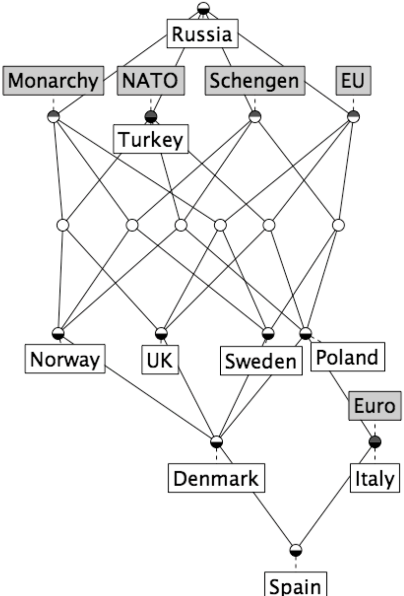
Attribute exploration
European states



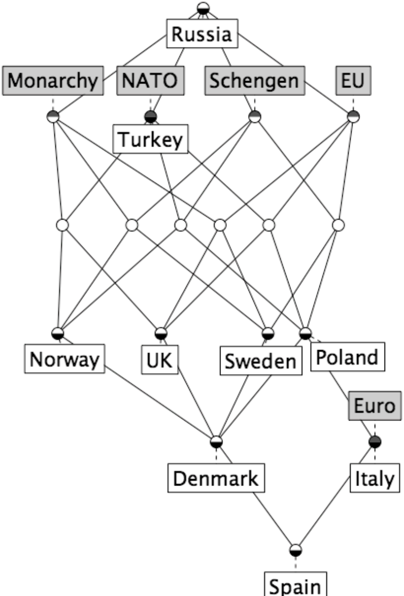
Question Is every European monarchy in NATO?

Answer No: Sweden is not.

	EU	Euro	Schengen	NATO	Monarchy
Sweden	×		×		×



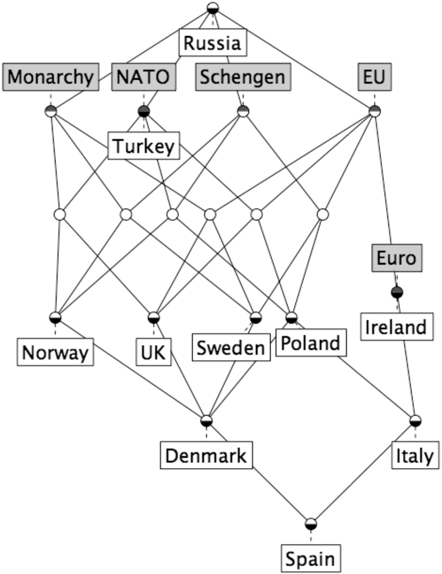
Question: Is every Eurozone country in EU, Schengen, and NATO?



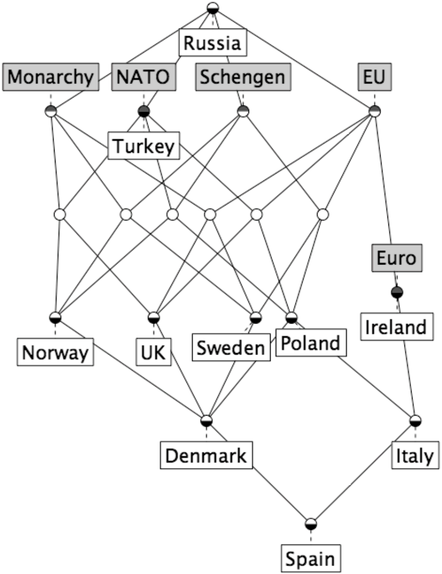
Question Is every Eurozone country in EU, Schengen, and NATO?

Answer No: Ireland is not.

	EU	Euro	Schengen	NATO	Monarchy
Ireland	×	×			



Question Is every Eurozone country in EU?

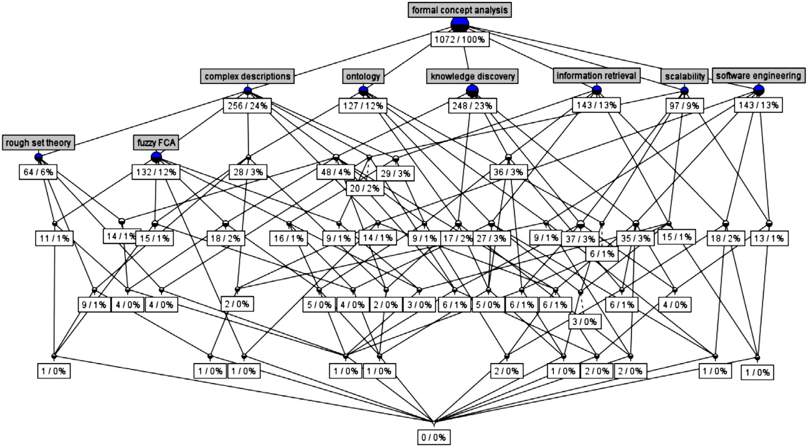


Question Is every Eurozone country in EU?

Answer No: Montenegro is not. . .

FCA in knowledge processing: surveys on models and techniques (a) and applications (b)

Poelmans et al., 2013a,b



Recommendation of advertising terms

Data

Data on purchases of advertising terms. Formal context $\mathbb{K}_{FT} = (F, T, I_{FT} \subseteq F \times T)$, F is the set of advertising companies, T is the set of terms, fIt means that company $f \in F$ bought term $t \in T$. The size of the context is 2000×3000.

Problem statement

Detect markets of advertising terms for making bid recommendations

Solution tools

- FCA: constructing concepts and their generators
- constructing association rules
- association rules + morphology
- association rules + ontology

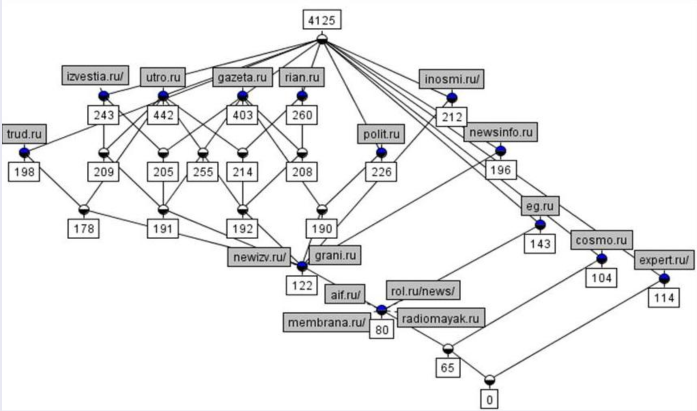
Recommending advertising terms:association rules + morphology

Examples

- $t \xrightarrow{FT} s_i^{ITS}$
- $t \xrightarrow{FT} \bigcup_i s_i^{ITS}$
- $\{mail\ order\ phentermine\} \rightarrow$
  $\{adipex\ online\ order, adipex\ order, adipex\ phentermine, \ldots,$
  $phentermine\ prescription, phentermine\ purchase, phentermine\ sale\}$
  Supp= 19 Conf= 0,95
- $t \xrightarrow{FT} (\bigcup_i s_i)^{ITS}$
- $\{distance\ long\ phone\} \rightarrow \{call\ distance\ long\ phone, carrier\ distance\ long\ phone, \ldots,$
  $distance\ long\ phone\ rate, distance\ long\ phone\ service\}$
  Supp= 37 Conf= 0,88
- $t_1 \xrightarrow{FT} t_2$ such that $t_2^{ITS} \subseteq t_1^{ITS}$
- $\{ink\ jet\} \rightarrow \{ink\}$, Supp= 14 Conf= 0,7

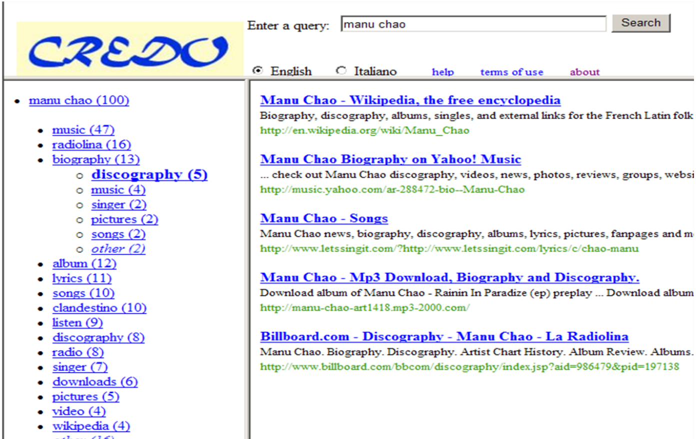
Taxonomy of web-site visitors

Diagram of the ordered set of 25 most stable concepts



Credo

- Metasearch system using concept lattices
- http://credo.fub.it
- Claudio Carpineto, Giovanni Romano. Concept Data Analysis: Theory and Applications



- Cimiano et. al, Automatic acquisition of taxonomies from text: FCA meets NLP, 2003
- Data on touristic business

	bookable	rentable	driveable	rideable	joinable
hotel	x				
apartment	x	x			
car	x	x	x		
bike	x	x	x	x	
excursion	x				x
trip	x				x

KVO

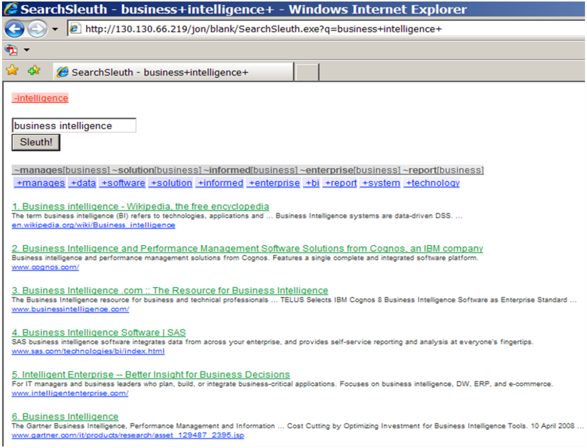
- International Research Group Knowledge, Visualisation and Ordering
- NLP, knowledge representation, information retrieval, data mining, usability knowledge models
- http://www.kvocentral.org/

Software

- Search Sleuth (metasearch system)
- Image Sleuth (search in collections of images)
- Mail Sleuth (plugin for e-mails)
- ToscanaJ (data analysis)

Search Sleuth

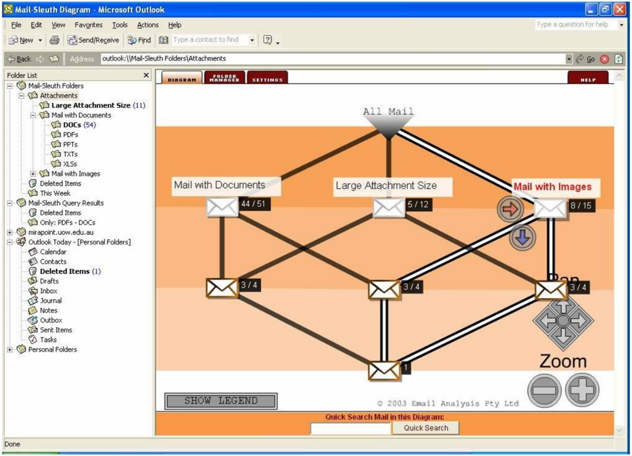
- Processes results of search queries to Yahoo
- Passing to more general (more specific) categories by clicking -term (+term)



- FCA-based system for looking images, navigation and search in their collections

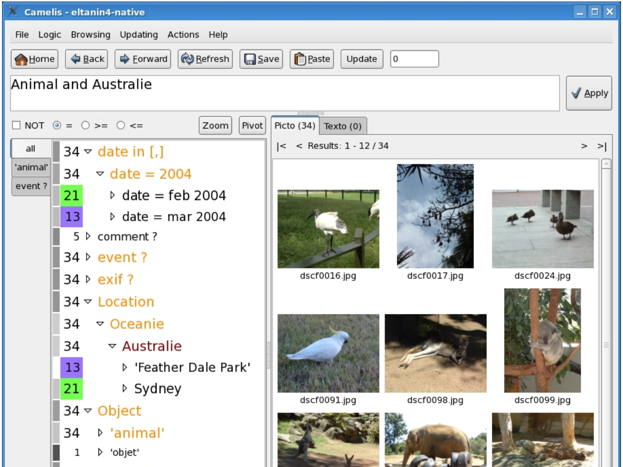
Mail Sleuth

- Plugin for Outlook, using concept lattices as a means of visualization and representing data from an e-mail account



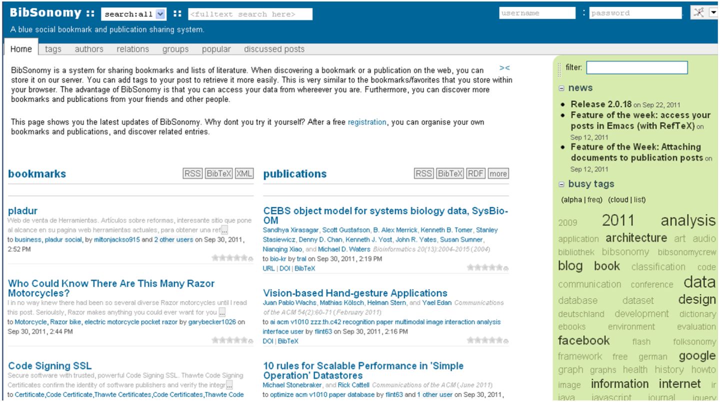
Camelis

- System of automatic indexing and navigation in data using concept lattices
- Sebastien Ferre



Bibsonomy

- http://bibsonomy.org/ a web-service of social bookmarks
- University of Kassel



JSM-method of hypothesis generation
[Kuznetsov 1994], [Ganter, Kuznetsov 2000]

A target attribute $w \notin M$,

- positive examples: Set $G_+ \subseteq G$ of objects known to have w,
- negative examples: Set $G_- \subseteq G$ of objects known not to have w,
- undetermined examples: Set $G_\tau \subseteq G$ of objects for which it is unknown whether they have the target attribute or do not have it.

Three subcontexts of $\mathbb{K} = (G, M, I)$: $\mathbb{K}_\varepsilon := (G_\varepsilon, M, I_\varepsilon)$, $\varepsilon \in \{-, +, \tau\}$ with respective derivation operators $(\cdot)^+$, $(\cdot)^-$, and $(\cdot)^\tau$.

A positive hypothesis $H \subseteq M$ is an intent of $\mathbb{K}_+$ not contained in the intent g^- of any negative example $g \in G_-$: $\forall g \in G_- \quad H \not\subseteq g^-$. Equivalently,

$$H^{++} = H, \quad H' \subseteq G_+ \cup G_\tau.$$

Example of a learning context

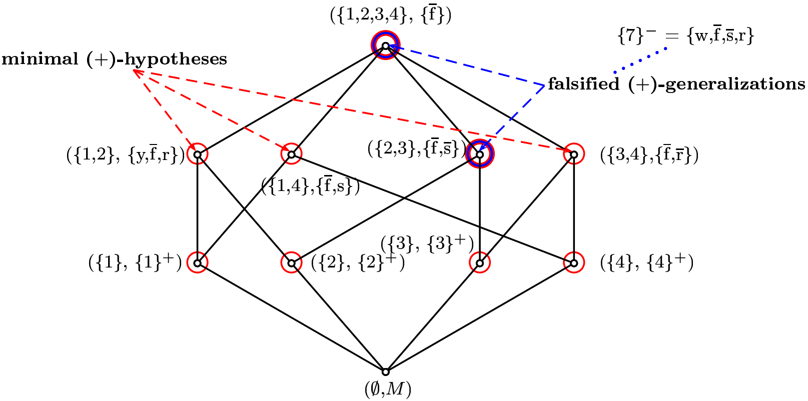
	G \ M	color	firm	smooth	form	fruit
1	apple	yellow	no	yes	round	+
2	grapefruit	yellow	no	no	round	+
3	kiwi	green	no	no	oval	+
4	plum	blue	no	yes	oval	+
5	toy cube	green	yes	yes	cubic	−
6	egg	white	yes	yes	oval	−
7	tennis ball	white	no	no	round	−
8	mango	green	no	yes	oval	τ

Natural scaling of the context

	G \ M	w	y	g	b	f	$\bar{f}$	s	$\bar{s}$	r	$\bar{r}$	fruit
1	apple		$\times$				$\times$	$\times$		$\times$		+
2	grapefruit		$\times$				$\times$		$\times$	$\times$		+
3	kiwi			$\times$			$\times$		$\times$		$\times$	+
4	plum				$\times$		$\times$	$\times$			$\times$	+
5	toy cube			$\times$		$\times$		$\times$			$\times$	−
6	egg	$\times$				$\times$		$\times$			$\times$	−
7	tennis ball	$\times$					$\times$		$\times$	$\times$		−
8	mango			$\times$			$\times$	$\times$			$\times$	τ

Abbreviations:
“g” for green, “y” for yellow, “w” for white, “f” for firm, “$\bar{f}$” for nonfirm,
“s” for smooth, “$\bar{s}$” for non-smooth, “r” for round,
“$\bar{r}$” for non-round.

Positive Concept Lattice



	G \ M	w	y	g	b	f	$\bar{f}$	s	$\bar{s}$	r	$\bar{r}$	**fruit**
1	apple		×				×	×		×		+
2	grapefruit		×				×		×	×		+
3	kiwi			×			×		×		×	+
4	plum				×		×	×			×	+
5	toy cube			×		×		×			×	−
6	egg	×				×		×			×	−
7	tennis ball	×					×		×	×		−
8	mango			×			×	×			×	τ

Classification of undetermined example mango

	G\M	w	y	g	b	f	$\bar{\text{f}}$	s	$\bar{\text{s}}$	r	$\bar{\text{r}}$	fruit
1	apple		×				×	×		×		+
2	grapefruit		×				×		×	×		+
3	qiwi			×			×		×		×	+
4	plum				×		×	×			×	+
5	toy cube			×		×		×			×	−
6	egg	×				×		×			×	−
7	tennis ball	×					×		×	×		−
8	mango		×				×	×			×	τ

Object mango was classified as a positive example since:

- for (+)-hypothesis $\{\bar{\text{r}}, \bar{\text{f}}\}$
  $\{\bar{\text{r}}, \bar{\text{f}}\} \subseteq \text{mango}^{\tau} = \{\text{y}, \bar{\text{f}}, s, \bar{\text{r}}\}$;
- for (−)-hypothesis {w} and $\{\text{f, s, }\bar{\text{r}}\}$:
  $\{\text{w}\} \not\subseteq \text{mango}^{\tau}$,
  $\{\text{f, s, }\bar{\text{r}}\} \not\subseteq \text{,mango}^{\tau}$.

Classification of undetermined example soap

	G\M	w	y	g	b	f	$\bar{f}$	s	$\bar{s}$	r	$\bar{r}$	fruit
1	apple		×				×	×		×		+
2	grapefruit		×				×		×	×		+
3	qiwi			×			×		×		×	+
4	plum				×		×	×			×	+
5	toy cube			×		×		×			×	−
6	egg	×				×		×			×	−
7	tennis ball	×					×		×	×		−
8	soap	×				×		×		×		τ

Object soap was classified as a negative example since:

- for (−)-hypothesis {w} :
  $\{w\} \subseteq soap^{\tau} = \{w, f, s, r\}$,
- but there is no any (+)-hypothesis included in
  $soap^{\tau} = \{w, f, s, r\}$.

Classification of undetermined example shampingon

	G\M	w	y	g	b	f	$\bar{f}$	s	$\bar{s}$	r	$\bar{r}$	fruit
1	apple		×				×	×		×		+
2	grapefruit		×				×		×	×		+
3	qiwi			×			×		×		×	+
4	plum				×		×	×			×	+
5	toy cube			×		×		×			×	−
6	egg	×				×		×			×	−
7	tennis ball	×					×		×	×		−
8	shampingon	×					×	×			×	τ

Object shampingon was classified as a contradictory example since:

- for (+)-hypothesis $\{\bar{f}, s\}$
  $\{\bar{f}, s\} \subseteq \text{shampingon}^{\tau} = \{w, \bar{f}, s, \bar{r}\}$;
- for (−)-hypothesis $\{w\}$
  $\{w\} \subseteq \text{shampingon}^{\tau} = \{w, \bar{f}, s, \bar{r}\}$.

Classification of undetermined example watermelon

	G\M	w	y	g	b	f	$\bar{f}$	s	$\bar{s}$	r	$\bar{r}$	fruit
1	apple		×				×	×		×		+
2	grapefruit		×				×		×	×		+
3	qiwi			×			×		×		×	+
4	plum				×		×	×			×	+
5	toy cube			×		×		×			×	−
6	egg	×				×		×			×	−
7	tennis ball	×					×		×	×		−
8	watermelon			×		×		×		×		τ

Object watermelon was left undetermined since:

- for (+)-hypotheses $\{y, \bar{f}, r\}$, $\{\bar{f}, s\}$ и $\{\bar{f}, \bar{r}\}$:
  $\{y, \bar{f}, r\} \not\subseteq \text{watermelon}^{\tau} = \{g, f, s, r\}$,
  $\{\bar{f}, s\} \not\subseteq \text{watermelon}^{\tau} = \{g, f, s, r\}$, $\{\bar{f}, \bar{r}\} \not\subseteq \text{watermelon}^{\tau} = \{g, f, s, r\}$.
- for (−)-hypotheses $\{w\}$ и $\{f, s, \bar{r}\}$:
  $\{w\} \not\subseteq \text{watermelon}^{\tau} = \{g, f, s, r\}$,
  $\{f, s, \bar{r}\} \not\subseteq \text{watermelon}^{\tau} = \{g, f, s, r\}$.

Hypotheses vs. implications

A positive hypothesis h corresponds to an implication $h \to \{w\}$ in the context $K_+ = (G_+, M \cup \{w\}, I_+ \cup G_+ \times \{w\})$.
A negative hypothesis h corresponds to an implication $h \to \{\bar{w}\}$ in the context $K_- = (G_-, M \cup \{\bar{w}\}, I_- \cup G_- \times \{\bar{w}\})$.
Hypotheses are special implications: their premises are closed (in K_+ or in K_-).

	G \ M	w	y	g	b	f	$\bar{\text{f}}$	s	$\bar{\text{s}}$	r	$\bar{\text{r}}$	fruit	nonfruit
1	apple		×				×	×		×		×	
2	grapefruit		×				×		×	×		×	
3	kiwi			×			×		×		×	×	
4	plum				×		×	×			×	×	
5	toy cube			×		×		×			×		×
6	egg	×				×		×			×		×
7	tennis ball	×					×		×	×			×

Conclusion of the intro part

See recent surveys [Poelmans et al., In Expert Syst. and Appl., 2013a,b]

FCA is a convenient model for

1. construction and visualization of taxonomies of subject domains
2. bimodal clustering of objects in various domains
3. compact representation of dependencies in various domains (by bases of implications and association rules)
4. construction, update, merging, and quality control of ontologies
5. for a variety of applications that need analysis of object-attribute data
6. several problems in Machine Learning and Data Mining

Coinage the term bicluster

The term biclustering was proposed by B. Mirkin in the book Mathematical Classification and Clustering. Kluwer Academic Publishers (1996).

p. 296

> The term biclustering refers to simultaneous clustering of both row and column sets in a data matrix. Biclustering addresses the problems of aggregate representation of the basic features of interrelation between rows and columns as expressed in the data.

- Let $A_{n \times m}$ be a matrix, numeric or boolean
- $X = \{x_1, x_2, \ldots, x_n\}$ is a set of rows
- $Y = \{y_1, y_2, \ldots, y_n\}$ is a set of columns
- $I \subseteq X$ and $J \subseteq Y$ are subsets of rows and columns
- $A_{IJ} = (I, J)$ is a submatrix of A
- Cluster of rows $A_{IY} = (I, Y)$
- Cluster of columns $A_{XJ} = (X, J)$
- Bicluster is a submatrix of A in the form $A_{IJ} = (I, J)$
- $\mathcal{B} = \{B_k = (I_k, J_k)\}$ is a set of biclusters

BiMax algorithm: rediscovery formal concepts in Bioinformatics

Definition [S. Barkow et al, 2006]

Given m genes, n situations and a binary table e such that $e_{ij} = 1$ (gene i is active in situation j) or $e_{ij} = 0$ (gene i is not active in situation j) for all $i \in [1, m]$ and $j \in [1, n]$, the pair $(G, C) \in 2^{\{1,\dots,n\}} \times 2^{\{1,\dots,m\}}$ is called an inclusion-maximal bicluster if and only if (1) $\forall i \in G, j \in C : e_{ij} = 1$ and (2) $\nexists (G_1, C_1) \in 2^{\{1,\dots,n\}} \times 2^{\{1,\dots,m\}}$ with (a) $\forall i_1 \in G_1, \forall j_1 \in C_1$: $e_{i_1 j_1} = 1$ and (b) $G \subseteq G_1 \wedge C \subseteq C_1 \wedge (G_1, C_1) \neq (G, C)$.

BiMax algorithm: rediscovery formal concepts in Bioinformatics

Definition [S. Barkow et al, 2006]

Given m genes, n situations and a binary table e such that $e_{ij} = 1$ (gene i is active in situation j) or $e_{ij} = 0$ (gene i is not active in situation j) for all $i \in [1, m]$ and $j \in [1, n]$, the pair $(G, C) \in 2^{\{1,\dots,n\}} \times 2^{\{1,\dots,m\}}$ is called an inclusion-maximal bicluster if and only if (1) $\forall i \in G, j \in C : e_{ij} = 1$ and (2) $\nexists (G_1, C_1) \in 2^{\{1,\dots,n\}} \times 2^{\{1,\dots,m\}}$ with (a) $\forall i_1 \in G_1, \forall j_1 \in C_1$: $e_{i_1 j_1} = 1$ and (b) $G \subseteq G_1 \wedge C \subseteq C_1 \wedge (G_1, C_1) \neq (G, C)$.

Denote by H the set of genes (objects in general), by S the set of situations (attributes in general), and by $E \subseteq H \times S$ the binary relation given by the binary table e, $|H| = m$, $|S| = n$.

Proposition [Kuznetsov et al., 2009]

For every pair (G, C), $G \subseteq H$, $C \subseteq S$ the following two statements are equivalent.

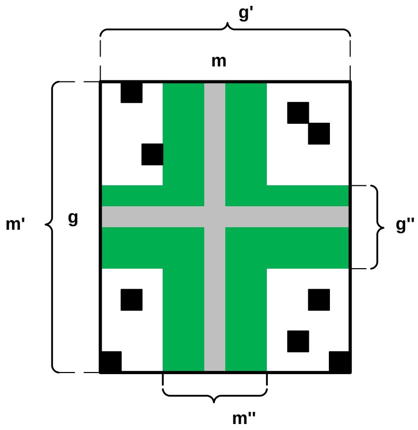
1. (G, C) is an inclusion-maximal bicluster of the table e;
2. (G, C) is a formal concept of the context (H, S, E).

Let $\mathbb{K} = (G, M, I \subseteq G \times M)$ be a formal context.

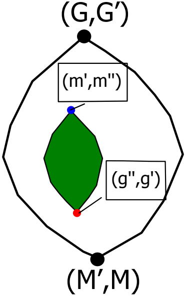
Definition 1

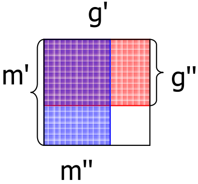
If $(g, m) \in I$, then (m', g') is called an object-attribute or OA-bicluster with density $\rho(m', g') = \frac{|I \cap (m' \times g')|}{|m'| \cdot |g'|}$.

Geometric interpretation of OA-bicluster
[D. Ignatov and S. Kuznetsov, 2010]



“Think of OA-bicusters as latticial intervals” [V. Duquenne’s comment at CLA2013]





OA-biclustering properties
[D. Ignatov and S. Kuznetsov, 2010]

Properties

1. $0 \leq \rho \leq 1$.
2. OA-bicluster (m', g') is a formal concept iff $\rho = 1$.
3. if (m', g') is an OA-bicluster, then $(g'', g') \leq (m', m'')$.

OA-biclustering properties
[D. Ignatov and S. Kuznetsov, 2010]

Property

The constraint $\rho(A, B) \geq \rho_{min}$ is neither monotonic nor anti-monotonic w.r.t. $\sqsubseteq$ relation, where $(A, B) \sqsubseteq (C, D)$ iff $A \subseteq C$ and $B \subseteq D$.

OA-biclustering properties
[D. Ignatov and S. Kuznetsov, 2010]

Property

The constraint $\rho(A, B) \geq \rho_{min}$ is neither monotonic nor anti-monotonic w.r.t. $\sqsubseteq$ relation, where $(A, B) \sqsubseteq (C, D)$ iff $A \subseteq C$ and $B \subseteq D$.

If $\rho_{min} = 0$, this means that we consider the set of all OA-biclusters of the context $\mathbb{K}$.

OA-biclustering properties
[D. Ignatov and S. Kuznetsov, 2010]

Property

The constraint $\rho(A,B) \geq \rho_{min}$ is neither monotonic nor anti-monotonic w.r.t. $\sqsubseteq$ relation, where $(A,B) \sqsubseteq (C,D)$ iff $A \subseteq C$ and $B \subseteq D$.

If $\rho_{min} = 0$, this means that we consider the set of all OA-biclusters of the context $\mathbb{K}$.
For $\rho_{min} = 0$ every formal concept is "contained" in a OA-bicluster of the context $\mathbb{K}$.

Proposition

For each $(A_c, B_c) \in \mathfrak{B}(G,M,I)$ there exists a OA-bicluster $(A_b, B_b) \in \mathbf{B}$ such that $(A_c, B_c) \sqsubseteq (A_b, B_b)$.

OA-biclustering Algorithm Complexity
[D. Ignatov and S. Kuznetsov, 2010]

Proposition 1

For a given formal context $\mathbb{K} = (G, M, I)$ and $\rho_{min} = 0$ the largest number of OA-biclusters is equal to $|I|$, all OA-biclusters can be generated in time $O(|I| \cdot (|G| + |M|))$.

Proposition 2

For a given formal context $\mathbb{K} = (G, M, I)$ and $\rho_{min} > 0$ the largest number of OA-biclusters is equal to $|I|$, all OA-biclusters can be generated in time $O(|I| \cdot |G| \cdot |M|)$.

OA-biclustering Algorithm Complexity
[D. Ignatov and S. Kuznetsov, 2010]

Proposition 1

For a given formal context $\mathbb{K} = (G, M, I)$ and $\rho_{min} = 0$ the largest number of OA-biclusters is equal to $|I|$, all OA-biclusters can be generated in time $O(|I| \cdot (|G| + |M|))$.

Proposition 2

For a given formal context $\mathbb{K} = (G, M, I)$ and $\rho_{min} > 0$ the largest number of OA-biclusters is equal to $|I|$, all OA-biclusters can be generated in time $O(|I| \cdot |G| \cdot |M|)$.

OA-biclustering versus FCA

$$O(|I| \cdot |G| \cdot |M|) \quad \text{VS} \quad O(|L| \cdot |G|^2 \cdot |M|)$$

Recommendation of an advertisement phrase [Ignatov et al., 2008]

Input data

Data on purchases of bids (advertisement phrases), the formal context $\mathbb{K}_{FT} = (F, T, I_{FT})$, where F is the set of advertisers, T — the set of bids, fIt denotes that advertiser $f \in F$ bought bid $t \in T$. The size of the context is 2000×3000.

Problem statement

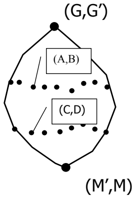
It is required to extract bid markets for further recommendations.

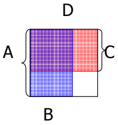
Approaches

- FCA with constraints (D-miner algorithm, [Besson et al., 2004])
- OA-biclustering

Detecting large market sectors with D-miner

[Besson et al, 2004], D-miner, $O(|G|^2|M||L|)$





D-miner results

Minimal extent size	Minimal intent size	Number of concepts
0	0	8 950 740
10	10	3 030 335
15	10	759 963
15	15	150 983
15	20	14 226
20	15	661

Detecting large market sectors with OA-biclustering

[Ignatov et al., 2010], OA-biclustering, $O(|G||M||I|)$

OA-biclustering results

Threshold, ρ_{min}	Number of OA-biclusters
0	92345
0.1	89735
0.2	80893
0.3	65881
0.4	45665
0.5	25921
0.6	10066
0.7	2081
0.8	165
0.9	3
1	0
$\lvert L \rvert$	8 950 740

Examples of market sectors

Hosting market

{affordable hosting web, business hosting web, cheap hosting, cheap hosting site web, cheap hosting web, company hosting web, cost hosting low web, discount hosting web, domain hosting, hosting internet, hosting page web, hosting service, hosting services web, hosting site web, hosting web}

Hotel market

{ angeles hotel los, atlanta hotel, baltimore hotel, dallas hotel, denver hotel, diego hotel san, francisco hotel san, hotel houston, hotel miami, hotel new orleans, hotel new york, hotel orlando, hotel philadelphia, hotel seattle, hotel vancouver }

Experiments settings

- Two algorithms: sequential and parallel
- C#, Microsoft Visual Studio 2008.
- Parallelized by Task Parallel Library из Microsoft .NET Framework 4.0.

Intel Pentuim IV Core 2 Duo, 2 GHz, RAM 3Gb

Data sets

UCI Machine Learning Repository

Dataset	# objects	# attributes	$\|I\|$	Density	$\|\mathfrak{B}(G, M, I)\|$
advertising	2000	3000	92 345	0,015	8 950 740
breast-cancer	286	43	2851	0,232	9918
flare	1389	49	18057	0,265	28742
postoperative	90	26	807	0,345	2378
SPECT	267	23	2042	0,333	21550
vote	435	18	3856	0,492	10644
zoo	101	28	862	0,305	379

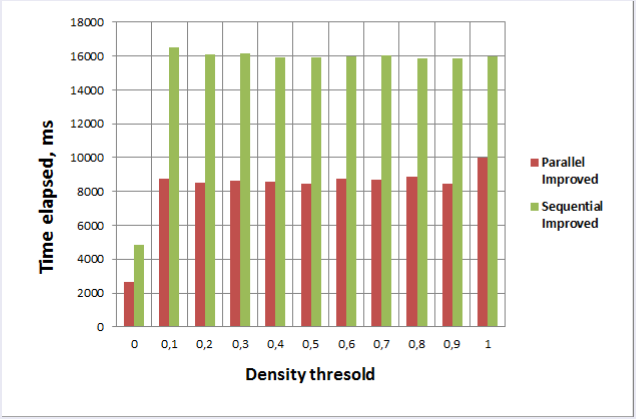
Dependency between the number of biclusters and minimal density threshold ρ_{min}

Dataset	advertising	breast-cancer	flare	postoperative	SPECT	vote
$\|\mathfrak{B}(G,M,I)\|$	8950740	9918	28742	2378	21550	10644
$\rho = 0$	92345	2851	18057	807	2042	3856
$\rho = 0,1$	89735	2851	18057	807	2042	3856
$\rho = 0,2$	80893	2851	18057	807	2042	3856
$\rho = 0,3$	65881	2849	18050	807	2042	3855
$\rho = 0,4$	45665	2678	17988	807	2029	3829
$\rho = 0,5$	25921	1908	17720	725	1753	3527
$\rho = 0,6$	10066	310	16459	402	835	2575
$\rho = 0,7$	2081	17	9353	18	262	1458
$\rho = 0,8$	165	2	1450	2	85	382
$\rho = 0,9$	3	2	293	2	32	33
$\rho = 1$	0	2	3	2	12	1

Fraction of generated concepts and the number of biclusters

Dataset	advertising	breast-cancer	flare	postoperative	SPECT	vote
Reduction	96,9	3,5	1,6	2,9	10,6	2,8

Execution time of OA-biclustering algorithms for advertising dataset



Definition 1

Triadic context $\mathbb{K} = (G, M, B, Y)$ consists of set G (objects), M (attributes), B (conditions) and ternary relations $Y \subseteq G \times M \times B$. Triple $(g, m, b) \in Y$ means that the object g has the attribute m under the condition b.

Definition 2

(Formal) triconcept of $\mathbb{K}$ is a triple (X, Y, Z) which is maximal w.r.t. its components inclusion, i.e. $X \subseteq G$, $Y \subseteq M$, $Z \subseteq B$ и $X \times Y \times Z \subseteq Y$

Primes and double primes operators

Prime operators of 1-sets	Their double prime counterparts
$m' = \{ (g,b) \mid (g,m,b) \in Y \}$	$m'' = \{ \tilde{m} \mid (g,b) \in m' \quad \mathit{and} \quad (g,\tilde{m},b) \in Y \}$
$g' = \{ (m,b) \mid (g,m,b) \in Y \}$	$g'' = \{ \tilde{g} \mid (m,b) \in g' \quad \mathit{and} \quad (\tilde{g},m,b) \in Y \}$
$b' = \{ (g,m) \mid (g,m,b) \in Y \}$	$b'' = \{ \widetilde{b} \mid (g,m) \in b' \quad \mathit{and} \quad (g,m,\tilde{b}) \in Y \}$

Box operators
[D. Ignatov, S. Kuznetsov et al., 2011]

For $\mathbb{K} = (G, M, B, Y \subseteq G \times M \times B)$ and $(g, m, b) \in Y$

$$
\begin{aligned}
g^{\square} &= \{\ g_i\ |\exists b_i\ (g_i, b_i) \in m'\ or\ \exists m_i\ (g_i, m_i) \in b'\} \\
m^{\square} &= \{\ m_i\ |\exists b_i\ (m_i, b_i) \in g'\ or\ \exists g_i\ (g_i, m_i) \in b'\} \\
b^{\square} &= \{\ b_i\ |\exists g_i\ (g_i, b_i) \in m'\ or\ \exists m_i\ (m_i, b_i) \in g'\}
\end{aligned}
$$

Tricluster definition
[D. Ignatov, S. Kuznetsov et al., 2011] appeared in General Systems Journal in 2013

- The paper's copy at Research Gate

Let $\mathbb{K} = (G, M, B, Y)$ be a formal context.

Definition 1

Tricluster is a triple $T = (g^{\square}, m^{\square}, b^{\square})$, where $(g, m, b) \in Y$.

Tricluster definition
[D. Ignatov, S. Kuznetsov et al., 2011] appeared in General Systems Journal in 2013

- The paper's copy at Research Gate

Let $\mathbb{K} = (G, M, B, Y)$ be a formal context.

Definition 1

Tricluster is a triple $T = (g^{\square}, m^{\square}, b^{\square})$, where $(g, m, b) \in Y$.

Definition 2

Tricluster density $\rho(A, B, C)$ is defined as number of from Y in the tricluster (A, B, C), i.e $\rho(A, B, C) = \frac{|I \cap A \times B \times C|}{|A||B||C|}$.

Tricluster definition

[D. Ignatov, S. Kuznetsov et al., 2011] appeared in General Systems Journal in 2013

- The paper's copy at Research Gate

Let $\mathbb{K} = (G, M, B, Y)$ be a formal context.

Definition 1

Tricluster is a triple $T = (g^{\square}, m^{\square}, b^{\square})$, where $(g, m, b) \in Y$.

Definition 2

Tricluster density $\rho(A, B, C)$ is defined as number of from Y in the tricluster (A, B, C), i.e $\rho(A, B, C) = \frac{|I \cap A \times B \times C|}{|A||B||C|}$.

Definition 3

Tricluster $T = (A, B, C)$ is called dense if its density exceeds a predefined minimal threshold, i.e. $\rho(T) \geq \rho_{min}$.

Tricluster definition
[D. Ignatov, S. Kuznetsov et al., 2011] appeared in General Systems Journal in 2013

- The paper's copy at Research Gate

Let $\mathbb{K} = (G, M, B, Y)$ be a formal context.

Definition 1

Tricluster is a triple $T = (g^{\Box}, m^{\Box}, b^{\Box})$, where $(g, m, b) \in Y$.

Definition 2

Tricluster density $\rho(A, B, C)$ is defined as number of from Y in the tricluster (A, B, C), i.e $\rho(A, B, C) = \frac{|I \cap A \times B \times C|}{|A||B||C|}$.

Definition 3

Tricluster $T = (A, B, C)$ is called dense if its density exceeds a predefined minimal threshold, i.e. $\rho(T) \geq \rho_{min}$.

For a given formal context $\mathbb{K} = (G, M, B, Y)$ denote by $\mathbf{T}(G, M, B, Y)$ the set of all its (dense) triclusters.

Geometric interpretation: a cross-like structure [D. Gnatyshak et al., 2011]

Suppose $\mathbb{K} = (G, M, B, Y)$ is a triadic context, and consider the triple $(\widetilde{g}, \widetilde{m}, \widetilde{b}) \in Y$.
Then object $\overline{g}$ will be added to $\widetilde{g}^{\square}$ iff

$$\{(\overline{g}, \widetilde{m}, b) \,|\, b \in B \wedge (\overline{g}, \widetilde{m}, b) \in Y\} \neq \emptyset \text{ or}$$

$$\{(\overline{g}, m, \widetilde{b}) \,|\, m \in M \wedge (\overline{g}, m, \widetilde{b}) \in Y\} \neq \emptyset.$$

If at least one of the elements from "grey" cells is an element of Y, then $\overline{g}$ is added to $\widetilde{g}^{\square}$.

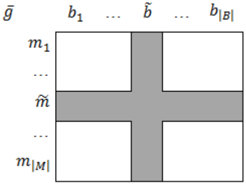


Figure : $\overline{g}$ addition condition

Tricluster properties
[D. Ignatov, S. Kuznetsov et al., 2011]

Property 1

For every triconcept (A, B, C) of the triadic context $\mathbb{K} = (G, M, B, Y)$ and non-empty sets A, B и C it holds that $\rho(A, B, C) = 1$.

Property 2

For every tricluster (A, B, C) of the triadic context $\mathbb{K} = (G, M, B, Y)$ it holds that $0 \leq \rho(A, B, C) \leq 1$.

Useful property
[D. Ignatov and S. Kuznetsov et al., 2011]

Proposition

Let $\mathbb{K} = (G, M, B, Y)$ be a triadic context and $\rho_{min} = 0$. For every triconcept $T_c = (A_c, B_c, C_c) \in \mathfrak{T}(G, M, B, Y)$ there exist a tricluster $T = (A, B, C) \in \mathbf{T}(G, M, B, Y)$ such that $A_c \subseteq A, B_c \subseteq B, C_c \subseteq C$.

Folksonomy
[T. van der Wal, 2004]

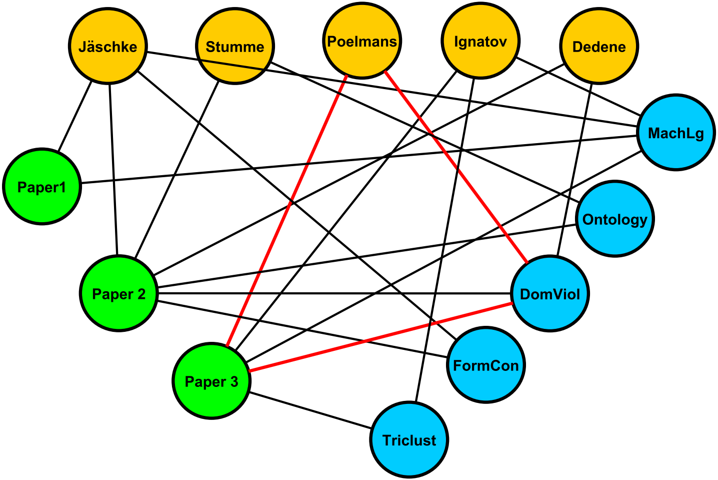
Definition

A quadruple (U, T, R, Y) is called folksonomy, where U is a set of users, T is a set of tags, R is a set of resources, and $Y \subseteq U \times T \times R$.

A triple $(u, t, r) \in Y$ denotes that the user u assigns the tag t to the resource r.

Folksonomy example

It is inspired by Bibsonomy (http://bibsonomy.org).



Example: triconcepts VS triclusters.

Folksonomy example

r_1:

	t_1	t_2	t_3
u_1		×	×
u_2	×	×	×
u_3	×	×	×

r_2:

	t_1	t_2	t_3
u_1	×	×	×
u_2	×		×
u_3	×	×	×

r_3:

	t_1	t_2	t_3
u_1	×	×	×
u_2	×	×	×
u_3	×	×	

- $\mathfrak{T} = \{(\emptyset, \{t_1, t_2, t_3\}, \{r_1, r_2, r_3\}), (\{u_1\}, \{t_2, t_3\}, \{r_1, r_2, r_3\}), \ldots$ $(\{u_1, u_2, u_3\}, \{t_1, t_2\}, \{r_3\})\}$
- $|\mathfrak{T}| = 3^3 = 27$
- $T = (\{u_1, u_2, u_3\}, \{t_1, t_2, t_3\}, \{r_1, r_2, r_3\})$, $\rho = 0.89$

Bibsonomy.org (ECML PKDD Discovery Challenge, 2008)

- 1. user (number, no user names available)
- 2. tag
- 3. content_id (matches bookmark.content_id or bibtex.content_id)
- 4. content_type (1 = bookmark, 2 = bibtex)
- 5. date

- The folksonomy size: $|U| = 2\ 337$ users, $|T| = 67\ 464$ tags, $|R| = 28\ 920$ resources that related by $|Y| = 816\ 197$ triples. The size of the cuboid 4 559 624 602 560 cells. $\rho(U, T, R) \approx 1,8 \cdot 10^{-7}$
- Implementation: Python 2.7.1
- System parameters: Pentium Core Duo, 2 GHz, 2 GB RAM.

Bibsonomy Data. Top-11 tags

Tag	Frequency of assignments (user, document)
imported	66636
public	15666
system:imported	11294
nn	9147
video	7610
books	6214
software	5021
tools	4423
web2.0	4215
web	4071
blog	3439

Results for k first triples of data set tas with $\rho_{min} = 0$

k	$\|U\|$	$\|T\|$	$\|R\|$	$\|\mathfrak{T}\|$	$\|\mathbf{T}\|$	Trias, s	TriclEx,s	TriclProb,s
100	1	47	52	57	1	0.2	0.2	0.2
1000	1	248	482	368	1	1	1	1
10000	1	444	5193	733	1	2	46,7	47
100000	59	5823	28920	22804	4462	3386	10311	976
200000	340	14982	61568	-	19053	$>$ 24 h	$>$ 24h	3417

Density of conceptual triclusters distribution for 200 000 first triples of tas data set with $\rho_{min} = 0$

Lower bound ρ	Upper bound ρ	Number of triclusters
0	0,05	18617
0,05	0,1	195
0,1	0,2	112
0,2	0,3	40
0,3	0,4	20
0,4	0,5	10
0,5	0,6	8
0,6	0,7	1
0,7	0,8	1
0,8	0,9	0
0,9	1	49

Algorithms

- TRIAS, [R. Jäschke et al., (2006)]
- OAC triclustering based on
  (a) box and
  (b) prime operators, [D. Ignatov et al. (2011, 2012)]
- TriBox, [B. Mirkin and A. Kramarenko (2011)]
- SpecTric, [D. Ignatov and Z. Sekinaeva (2011)]

Let $\mathbb{K} = (G, M, B, Y)$ be a triadic context.
Prime operators:

$$(g, m)' = \{ b \mid (g, m, b) \in Y \}$$
$$(g, b)' = \{ m \mid (g, m, b) \in Y \}$$
$$(m, b)' = \{ g \mid (g, m, b) \in Y \}$$

Definition

For $(g, m, b) \in I$ an OAC-tricluster based on prime operators is a triple $T = ((m, b)', (g, b)', (g, m)')$.

Geometric interpretaion: a cross-like structure again
[D. Gnatyshak et al., 2011]

Prime based OAC-triclusters are more dense than box operator based ones. Every element corresponding to the "grey" cell is an element of Y.

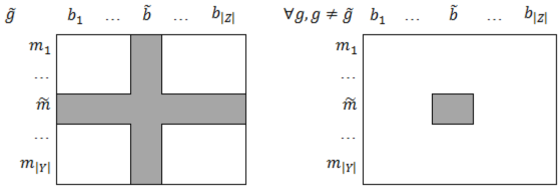


Figure : Prime operator based tricluster structure

OAC-prime triclustering algorithm
[Dmitry Gnatyshak et al., 2012]

Require: $\mathbb{K} = (G, M, B, Y)$ is a tricontext;
ρ_{min} is a density threshold
Ensure: $TSet = \{(X, Y, Z)\}$

for all (g, m): $g \in G$,$m \in M$ do
  $PrOA[g, m] = (g, m)'$
for all (g, b): $g \in G$,$b \in B$ do
  $PrOC[g, b] = (g, b)'$
for all (m, b): $m \in M$,$b \in B$ do
  $PrAC[m, b] = (m, b)'$
for all $(g, m, b) \in I$ do
  $T = (PrAC[m, b], PrOC[g, b], PrOA[g, m])$
  $Tkey = hash(T)$
  if $Tkey \notin Tset.keys \wedge \rho(T) \geq \rho_{min}$ then
    $Tset[Tkey] = T$

A set of triclusters $\mathcal{T} = \{T = (X, Y, Z)\}$ forms the following model of data:

$$r_{ijk} = \max_{t=1,\ldots,|\mathcal{T}|} \lambda_t [(g_{it}, m_{jt}, b_{kt}) \in X_t \times Y_t \times Z_t] + \lambda_0 + \epsilon_{ijk}, \quad (1)$$

where λ_t is a parameter, λ_0 is a constant, ϵ_{ijk} is a residual. TriBox aims at minimizing residuals (or maximizing f) in the model with one tricluster T:

$$L^2 = \sum_{ijk} (r_{ijk} - \lambda_0 - \lambda[(g_i, m_j, b_k) \in X \times Y \times Z])^2 \qquad (2)$$

$$f(T) = \lambda^2 |X||Y||Z| = (\rho(T) - \lambda_0)^2 |X||Y||Z| \qquad (3)$$

Why TriBox?
[Mirkin and Kramarenko, 2011]

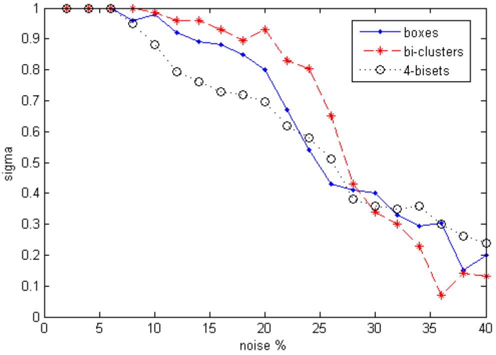


Figure : Graphs of σ measure between the original three concepts and results of *Box* and *Dual* bicluster algorithms (Mirkin, 2008) applied to the binary R_p matrix at different levels of random noise, $p = 1, 2, .., 40\%$. The third graph represents the σ values at 4-bisets by Pensa and Boulicaut, 2005.

Spectral Triclustering: SpecTric
[Sekinaeva and Ignatov, 2011]

1. The formal tricontext $\mathbb{K} = (G, M, B, Y)$ as a hypergraph $\Gamma = \langle G \sqcup M \sqcup B, E \rangle$
2. Transformation of the hypergraph into an undirected non-weighted graph $\widetilde{\Gamma} = \langle G \sqcup M \sqcup B, \widetilde{E} \rangle$ with information loss
3. Finding the eigenvector for the second smallest eigenvalue of the Laplacian matrix L for the graph

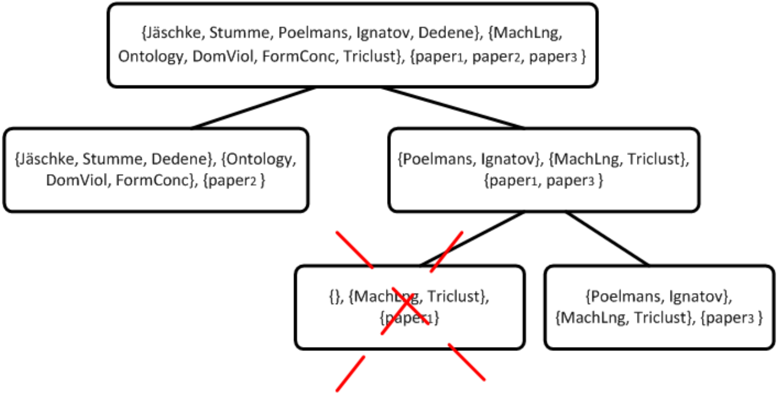
Let $\{i, g, k\} = \{G, M, B\}$, then E_{ij} is an adjacency submatrix and D_k is a diagonal vertex degree matrix.

$$L = \begin{pmatrix} D_G & -E_{GM} & -E_{GB} \\ -E_{MG} & D_M & -E_{MB} \\ -E_{BG} & -E_{BM} & D_B \end{pmatrix}$$

A matrix equation for Min-cut problem:

$$Lv = \lambda D v$$

A SpecTric execution tree for the Bibsonomy example:



TRIAS
[Jäschke et al., 2006]

Trias is a method for finding triadic formal concepts, that are closed 3-sets. Triadic formal concepts can be interpreted as absolutely dense triclusters.

- NextClosure algorithm enumerates all formal concepts of the dyadic context in their lexicographical order
- Trias is a NextClosure extension to the triadic case
- Minimal support constraints are added (triclusters with too small extent, intent or modus are skipped)

Road Map

What is a good tricluster or a collection of triclusters?
[D. Gnatyshak et al., 2012]

- Reasonably low number of found patterns
- High fault-tolerance [Besson et al., 2005]
- High density (average density of triclusters)
- High coverage of an initial data by the generated triclusters
- High diversity of triclusters
- Good triclusters interpretability w.r.t. the data domain and common sense

For a good triclustering algorithm we rather expect not so high time complexity and ability to parallelization.

We used the Jaccard similarity coefficient to find the most similar tricluster t for a given cuboid c. Total similarity with an initial collection of cuboids C has been defined as follows:

$$\sigma(\mathcal{C},\mathcal{T}) = \frac{1}{|\mathcal{C}|}\sum_{c=c_1}^{c_C}\max_{t=t_1,\ldots,t_T}\frac{|G_c \cap G_t|}{|G_c \cup G_t|}\frac{|M_c \cap M_t|}{|M_c \cup M_t|}\frac{|B_c \cap B_t|}{|B_c \cup B_t|} \quad (4)$$

Coverage
[D. Gnatyshak et al., 2011]

Coverage is defined as a fraction of the triples of the context (alternatively, objects, attributes or conditions) included in at least one of the triclusters of the resulting tricluster collection.

Diversity
[D. Gnatyshak et al., 2011]

To define diversity we will use a binary function of 2 triclusters:

$$intersect(\mathcal{T}_i, \mathcal{T}_j) = \begin{cases} 1, & (G_{\mathcal{T}_i} \cap G_{\mathcal{T}_j} \neq \emptyset) \wedge (M_{\mathcal{T}_i} \cap M_{\mathcal{T}_j} \neq \emptyset) \wedge (B_{\mathcal{T}_i} \cap B_{\mathcal{T}_j} \neq \emptyset) \\ 0, & \text{otherwise} \end{cases}$$

where $\mathcal{T}$ is a tricluster set.
The diversity of the tricluster set $\mathcal{T}$:

$$diversity(\mathcal{T}) = 1 - \frac{\sum_j \sum_{i<j} intersect(\mathcal{T}_i, \mathcal{T}_j)}{\frac{|\mathcal{T}|(|\mathcal{T}|-1)}{2}}$$

It is possible to define *intersect* for the sets of objects:

$$intersect_G(\mathcal{T}_i, \mathcal{T}_j) = \begin{cases} 1, & G_{\mathcal{T}_i} \cap G_{\mathcal{T}_j} \neq \emptyset \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

$$diversity_G(\mathcal{T}) = 1 - \frac{\sum_j \sum_{i<j} intersect_G(\mathcal{T}_i, \mathcal{T}_j)}{\frac{|\mathcal{T}|(|\mathcal{T}|-1)}{2}} \tag{6}$$

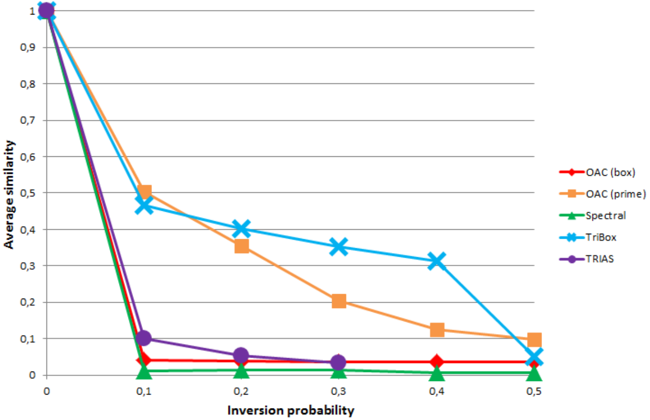
The diversity for the sets of attributes or conditions is similarly defined.

- Fault-tolerance test: contexts $30 \times 30 \times 30$ with three absolutely dense cuboids $10 \times 10 \times 10$ on the main diagonal with a different noise probability

Table : Contexts for experiments on time, number of triclusters, density, coverage and diversity.

Context	$\|G\|$	$\|M\|$	$\|B\|$	# triples	Density
Uniform	30	30	30	2660	0.0985
IMDB	250	795	22	3818	0.00087
BibSonomy	51	924	2844	3000	0.000022

Given a tricontext with three dense $10 \times 10 \times 10$ cuboids on its main diagonal.



Noise-tolerance vs probability of a triple inversion

Experiments

Table : Experiments: time, number of triclusters, density, coverage and diversity.

Algorithm	t, ms	t_{par}, ms	n	ρ_{av}, %	Cov, %	Div, %	Div_G, %	Div_M, %	Div_B, %
	Uniform random context								
OAC ($\square$)	407	196	73	9.88	100.00	0.00	0.00	0.00	0.00
OAC ($\prime$)	312	877	2659	32.23	100.00	92.51	60.07	59.80	59.45
SpecTric	277	-	5	8.74	8.84	100.00	100.00	100.00	100.00
TriBox	6218	1722	1011	74.00	96.02	97.42	66.25	79.53	84.80
Trias	29367	-	38356	100.00	100.00	99.99	99.93	4.07	3.51
	IMDB								
OAC ($\square$)	2314	1573	1500	1.84	100.00	15.65	9.67	0.70	7.87
OAC ($\prime$)	547	2376	1274	53.85	100.00	96.55	94.56	92.14	28.52
SpecTric	98799	-	21	17.07	20.88	100.00	100.00	100.00	100.00
TriBox	197136	55079	328	91.65	98.90	98.89	98.46	95.21	30.94
Trias	102554	-	1956	100.00	100.00	99.89	99.69	52.52	26.18

Table : Experiments: time, number of triclusters, density, coverage and diversity.

Algorithm	t, ms	t_{par}, ms	n	ρ_{av}, %	Cov, %	Div, %	Div_G, %	Div_M, %	Div_B, %
	BibSonomy								
OAC (□)	19297	6803	398	4.16	100.00	79.59	67.28	42.83	79.54
OAC (/)	13556	9400	1289	9466	100.00	99.74	88.58	99.51	99.53
SpecTric	5906563	-	2	50	100.00	100.00	100.00	100.00	100.00
TriBox	> 24 hrs								
Trias	110554	-	1305	100.00	100.00	99.98	91.70	99.78	99.92

Experiments: Pareto optimality

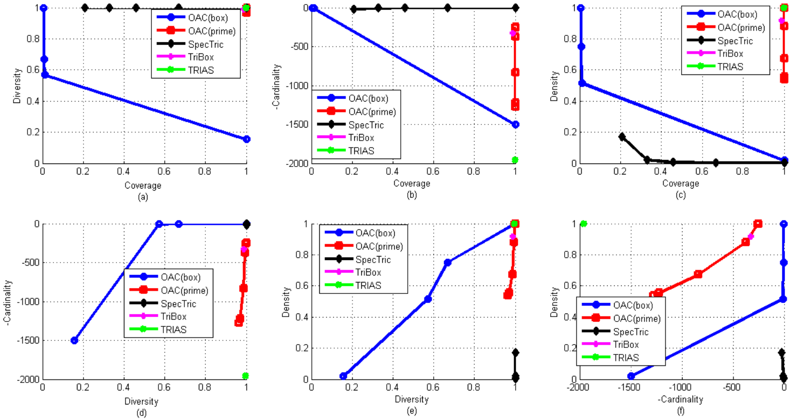


Figure : Pairwise criterion graphs for IMDB dataset

- High elapsed time
- Too large number of small well-interpreted triclusters (triconcepts)

Examples of the triconcepts for the IMDB context:

1. {The Princess Bride (1987), Pirates of the Caribbean: The Curse of the Black Pearl (2003)}, {Pirate}, {Fantasy, Adventure}
2. {Platoon (1986), Letters from Iwo Jima (2006)}, {Battle}, {Drama,War}
3. {V for Vendetta (2005)}, {Fascist, Terrorist, Government, Secret Police ,
   Fight}, {Action, Sci-Fi, Thriller}

- High computational speed on small contexts
- Well-interpreted triclusters, but of the low density
- Diversity is always equal to 1, but it causes low coverage

Examples of the triclusters for the IMDB context:

1. $\rho = 23.08\%$, {Alien (1979), The Shining (1980), The Thing (1982), The Exorcist (1973)}, {Spaceship, Egg, Parasite, Creature, Caretaker, Colorado, Actress, Blood, Helicopter, Scientist, Priest, Washington D.C., Faith}, {Horror}
2. $\rho = 2.09\%$, {The Shawshank Redemption (1994), The Godfather (1972), The Godfather: Part II (1974), . . . , Bonnie and Clyde (1967), Arsenic and Old Lace (1944)}, {Prison, Cuba, Business, 1920s, . . . , Texas, Cellar}, {Crime, Thriller }

TriBox

- A moderate number of well-interpreted triclusters
- High elapsed time
- An efficient parallelization
- Reasonably high coverage and diversity

Examples of the triclusters for the IMDB context:

1. 100%, {Million Dollar Baby (2004), Rocky (1976), Raging Bull (1980)}, {Boxer, Boxing}, {Drama, Sport}
2. 83.33%, {The Sixth Sense (1999), The Exorcist (1973), The Silence of the Lambs (1991)}, {Psychiatrist}, {Drama, Thriller}
3. 33.33%, {Platoon (1986), All Quiet on the Western Front (1930), Glory (1989), Apocalypse Now (1979), Lawrence of Arabia (1962), Saving Private Ryan (1998), Paths of Glory (1957), Full Metal Jacket (1987)}, {Army, General, Jungle, Vietnam, Soldier, Recruit}, {Drama, Action, War}

OAC-box

- Large triclusters of low density
- High coverage, small diversity
- An efficient parallelization
- Difficult to interpret unlike SpecTric

In many cases extent size is small. Examples are given below:

1. 0.9%, {The Shawshank Redemption (1994), The Godfather (1972), Ladri di biciclette (1948), Unforgiven (1992), Batman Begins (2005), Die Hard (1988), . . . , The Green Mile (1999), Sin City (2005), The Sting (1973)}, {Prison, Murder, Cuba, FBI, Serial Killer, Agent, Psychiatrist,. . . , Window, Suspect, Organized Crime , Revenge, Explosion, Assassin, Widow}, {Crime, Drama, Sci-Fi, Fantasy, Thriller, Mystery}
2. 1.07%, {The Great Escape (1963), Star Wars: Episode VI - Return of the Jedi (1983), Jaws (1975), Batman Begins (2005), Blade Runner (1982), Die Hard (1988),. . . , Metropolis (1927), Sin City (2005), Rebecca (1940)}, {Prison, Murder, Cuba, FBI, Serial Killer, Agent, Psychiatrist,. . . , Shower, Alimony, Phoenix Arizona, Assassin, Widow}, {Drama, Thriller, War}

OAC-prime

- It is one of the fastest algorithms
- Moderately large number of dense well-interpreted triclusters
- For $\rho_{min} = 0$ coverage is equal to 1, but remains high for different ρ_{min}
- Diversities are also rather high
- Low efficiency of parallelization

Examples of the triclusters for the IMDB context:

1. 36%, {The Shawshank Redemption (1994), Cool Hand Luke (1967), American History X (1998), A Clockwork Orange (1971), The Green Mile (1999)}, {Prison, Murder, Friend, Shawshank, Banker}, {Crime, Drama}
2. $56,67\%$, {The Godfather: Part II (1974), The Usual Suspects (1995)}, {Cuba, New York, Business, 1920s, 1950s}, {Crime, Drama, Thriller}
3. 60%, {Toy Story (1995), Toy Story 2 (1999)}, {Jealousy, Toy, Spaceman, Little Boy, Fight}, {Fantasy, Comedy, Animation, Family, Adventure}

- There is no a winner according to the comparison criteria
- Method TriBox shows best results but it takes huge computational time
- OAC-triclustering based on prime operators gives the second best results and it is sufficiently fast
- Proposed OAC-triclustering technique seems to be effective and efficient means of reducing the numbers of patterns in comparison with triconcepts (execution time and output size are polynomial)

- We are developing a multimodal clustering framework for relational data – MMC Toolbox.
- Conducting more experiments with triclustering for making recommendations and finding tricommunities.
- Implementing more triclustering methods to comparison (e.g. adding Fenster [Cerf & Boulicaut, 2013])
- We are looking for a Bayesian view on the problem (the starting point is e.g. LDA [Blei et al.])

Road Map

Finding a good tricluster collection: an optimization task

Ignatov et al., Machine Learning 101(1-3): 271-302 (2015)

- The paper's copy at Research Gate

For a given tricontext $\mathbb{K} = (G, M, B, I \subseteq G \times M \times B)$, minimal density $\rho_{min} \in [0, 1]$ and coverage level $\alpha \in [0, 1]$ find

$$\mathcal{T}_{opt} \in Arg \min_{\mathcal{T}_{cov} \subseteq \mathcal{T}} (|\mathcal{T}_{cov}|, -Diversity(\mathcal{T}_{cov}))$$

subject to constraints

(1) $\forall T \in \mathcal{T}_{cov} : \rho(T) \geq \rho_{min}$,

(2) $\forall (g, m, b) \in I \quad \exists (X, Y, Z) \in \mathcal{T}_{cov} : (g, m, b) \in X \times Y \times Z$

or

(2′) $\dfrac{\sum\limits_{(g,m,b) \in I} \left[(g,m,b) \in \bigcup\limits_{(X,Y,Z) \in \mathcal{T}_{cov}} X \times Y \times Z \right]}{|I|} \geq \alpha$

(3) $\forall (X, Y, Z) \in \mathcal{T}_{cov} : |X| \geq minsup_G, |Y| \geq minsup_M, |Z| \geq minsup_B$

What else has been proposed so far?

- Monte-Carlo strategies for tricluster density calculation
- n-ary generalizations of OAC-triclustering
- Usage of sorting on dimensions and various optimizations
- Mixing several constraint-based approaches to triclustering (e.g., mining dense triclusters first and then frequent tri-sets in them)
- Greedy approach to triclustering by coverage of an input context
- Time-aware multimodal clustering

What else has been proposed so far?

- Monte-Carlo strategies for tricluster density calculation
- n-ary generalizations of OAC-triclustering
- Usage of sorting on dimensions and various optimizations
- Mixing several constraint-based approaches to triclustering (e.g., mining dense triclusters first and then frequent tri-sets in them)
- Greedy approach to triclustering by coverage of an input context
- Time-aware multimodal clustering

What else has been done?

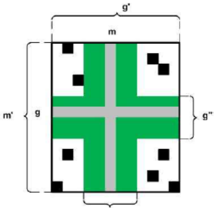
- Extraction communities of criminals operating in Amsterdam-Amstelland police region from unstructured observational police reports [Poelmans at al., 2012]
- Finding tricommunities in the massive amount of unstructured texts resulting from brainstorm sessions (in collaboration with the Witology company) [Ignatov et al., 2013]
- OC-biclustering based recommender system for crowdsourcing platforms [Kaminskaya, 2013]
- Triclustering based perfume recommender system [Venzhega et al., 2012]
- Automatically identifying suitable descriptors for groups on social network sites based on the interests which users indicated on their profile [Gnatyshak et al., 2012]
- Triadic JSM-like classification [Zhuk, 2013]

Multi-valued context: $\mathbb{K} = (G, M, W, I), W \subseteq \mathbb{R}$.
Modified prime operators: δ-operators which generate numeric biclusters of similar values:

$$g^{\delta=x} = \{\tilde{m} \mid (g, \tilde{m}, \tilde{w}) \in I \text{ and } |\tilde{m}(g) - m(g)| \leq x\},$$
$$m^{\delta=y} = \{\tilde{g} \mid (\tilde{g}, m, \tilde{w}) \in I \text{ and } |m(\tilde{g}) - m(g)| \leq y\}$$

$x, y \in \mathbb{R}$

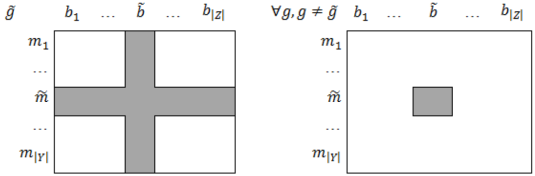
OAN-bicluster for the pair (g, m) with parameters x, y:
$(m^{\delta=y}, g^{\delta=x})$.



Triadic multi-valued context: $\mathbb{K} = (G, M, B, W, I), W \subseteq \mathbb{R}$.

$$(g, m)^{\delta=x} = \{\tilde{b} | (g, m, \tilde{b}, \tilde{w}) \in I \text{ and } |v(g, m, \tilde{b}) - v(g, m, b)| \leq x\},$$
$$(m, b)^{\delta=y} = \{\tilde{g} | (\tilde{g}, m, b, \tilde{w}) \in I \text{ and } |v(\tilde{g}, m, b) - v(g, m, b)| \leq y\},$$
$$(g, b)^{\delta=z} = \{\tilde{m} | (g, \tilde{m}, b, \tilde{w}) \in I \text{ and } |v(g, \tilde{m}, b) - v(g, m, b)| \leq z\}.$$
$$x, y \in \mathbb{R}$$

OACN-tricluster based on δ-operators for the triple (g, m, b) with parameters x, y, z: $((m, b)^{\delta=y}, (g, b)^{\delta=z}, (g, m)^{\delta=x})$.



Modified box operators: β-operators for triclusters of similar values:

$$g^{\beta(Æ,y,z,\Diamond)} = \{\tilde{g} | Æ\tilde{m}(\tilde{g},\tilde{m}) \in b^{\delta=z} \Diamond Æ\tilde{b}(\tilde{g},\tilde{b}) \in m^{\delta=y}\},$$

$$m^{\beta(Æ,x,z,\Diamond)} = \{\tilde{m} | Æ\tilde{g}(\tilde{g},\tilde{m}) \in b^{\delta=z} \Diamond Æ\tilde{m}(\tilde{m},\tilde{b}) \in g^{\delta=x}\},$$

$$b^{\beta(Æ,x,y,\Diamond)} = \{\tilde{b} | Æ\tilde{g}(\tilde{g},\tilde{b}) \in m^{\delta=y} \Diamond Æ\tilde{m}(\tilde{m},\tilde{b}) \in g^{\delta=x}\},$$

$Æ \in \{\forall, \exists\}, x, y, z \in \mathbb{R}, \Diamond \in \{\vee, \wedge\}$.
OACN-tricluster based on β-operators for the triple (g, m, b) with x, y, z:
$T(g, m, b | \beta(\exists, \cdot, \cdot, \vee)) = (g^{\beta(\exists,y,z,\vee)}, m^{\beta(\exists,x,z,\vee)}, b^{\beta(\exists,x,y,\vee)})$.

1. Dmitry I. Ignatov: Towards a closure operator for enumeration of maximal tricliques in tripartite hypergraphs. CoRR abs/1602.07267 (2016)

Thank you!