Formal Concept Analysis in Boolean Matrix Factorization

Algorithms and Extensions to Ordinal and Fuzzy-Valued Data

Jan Konecny

ICFCA 2023

Formal Concept Analysis

Just basic notation:

- $\langle X, Y, I \rangle$ – formal context
- $(\cdot)^{\uparrow} : 2^X \to 2^Y, (\cdot)^{\downarrow} : 2^Y \to 2^X$ – concept-forming operators
- $\mathcal{B}(X, Y, I)$ set of all formal concepts (also concept lattice)

We unify formal context $\langle X, Y, I \rangle$ with $X = \{1, \ldots, n\}$, $Y = \{1, \ldots, m\}$ with Boolean matrix $\boldsymbol{I} \in \{0, 1\}^{n \times m}$:

$$\boldsymbol{I}_{ij} = 1 \quad \text{iff} \quad \langle i, j \rangle \in I$$

Warning: By abuse of notation, we often do not distinguish between these representations.

Boolean Matrix Factorization

Input: Matrix $\boldsymbol{I}$

- $n \times m$
- contains Boolean values – truth (1) and false (0)

For instance,

$$\begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Goal: to decompose $\boldsymbol{I}$ into $\boldsymbol{A} \circ \boldsymbol{B} \approx \boldsymbol{I}$ where

- $\boldsymbol{A}$ is $n \times k$ matrix object $\times$ factors
- $\boldsymbol{B}$ is $k \times m$ matrix factors $\times$ attributes
- effort $k \ll m$

The symbol $\circ$ in $\boldsymbol{A} \circ \boldsymbol{B}$ is Boolean matrix product:

$$(\boldsymbol{A} \circ \boldsymbol{B})_{ij} = \bigvee_{\ell=1\ldots k} \boldsymbol{A}_{i\ell} \wedge \boldsymbol{B}_{\ell j}$$

For instance:

$$\begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \circ \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Aslo called / equivalent with:

- Boolean matrix decomposition
- Formal context decomposition / factorization
- Problem of covering a bipartite graph by bicliques

J. Orlin
Contentment in graph theory: covering graphs with cliques,
in: Proc. Kon. Neder. Akad. Wet., Amsterdam, ser. A, volume vol. 80, 1977 .

- Problem of finding the 2-dimension of a poset

History

First fundamental results in 70s:

D. S. Nau et al.,
A mathematical analysis of human leukocyte antigen serology,
Math. Biosci. 40(1978), 243–270.

L. Stockmeyer
The set basis problem is NP-complete,
Tech. rep. rc5431, IBM, Yorktown Heights, NY, USA, 1975 .

An increase in interest due to Miettinen's work:

- Boolean version of the discrete basis problem and ASSO

  P. Miettinen, T. Mielikäinen, A. Gionis, G. Das, H. Mannila,
  The discrete basis problem,
  IEEE TKDE 20(2008), 1348–62.

- Boolean CX and CUR decompositions

  P. Miettinen,
  The Boolean column and column-row matrix decompositions,
  Data Mining Knowl. Disc. 17(2008), 39–56.

...

. . .

- sparsity issues

P. Miettinen,
Sparse Boolean matrix factorizations,
Proc. IEEE ICDM 2010, pp. 935–940.

- selection of the number of factors

P. Miettinen, J. Vreeken,
Model order selection for Boolean matrix factorization,
ACM SIGKDD 2011, pp. 51–59.

. . .

...

- restricted decompositions using so-called tiles and formal concepts in Boolean data

F. Geerts, B. Goethals, T. Mielikäinen,
Tiling databases,
Proc. Discovery Science 2004, pp. 278–289.

R. Belohlavek, V. Vychodil,
Discovery of optimal factors in binary data via a novel method of matrix decomposition.
J. Computer and System Sciences 76(1)(2010), 3–20.

Goal: to decompose $\boldsymbol{I}$ into $\boldsymbol{A} \circ \boldsymbol{B} \approx \boldsymbol{I}$ where

- $\boldsymbol{A}$ is $n \times k$ matrix object $\times$ factors
- $\boldsymbol{B}$ is $k \times m$ matrix factors $\times$ attributes
- effort $k \ll m$

By $\approx$ in $\boldsymbol{A} \circ \boldsymbol{B} \approx \boldsymbol{I}$ we mean, that $\boldsymbol{A} \circ \boldsymbol{B}$ is close to $\boldsymbol{I}$.
We express this as minimalization of reconstruction error:

$$E(\boldsymbol{I}, \boldsymbol{A} \circ \boldsymbol{B}) = E_{\mathrm{u}}(\boldsymbol{I}, \boldsymbol{A} \circ \boldsymbol{B}) + E_{\mathrm{o}}(\boldsymbol{I}, \boldsymbol{A} \circ \boldsymbol{B})$$

where

- $E_{\mathrm{u}}(\boldsymbol{I}, \boldsymbol{A} \circ \boldsymbol{B}) = \{\langle i, j\rangle \ : \ \boldsymbol{I}_{ij} = 1, \ (\boldsymbol{A} \circ \boldsymbol{B})_{ij} = 0\}$
  is undercovering error (uncovering)
- $E_{\mathrm{o}}(\boldsymbol{I}, \boldsymbol{A} \circ \boldsymbol{B}) = \{\langle i, j\rangle \ : \ \boldsymbol{I}_{ij} = 0, \ (\boldsymbol{A} \circ \boldsymbol{B})_{ij} = 1\}$
  is overcovering error

E_u and E_o are not equally serious:

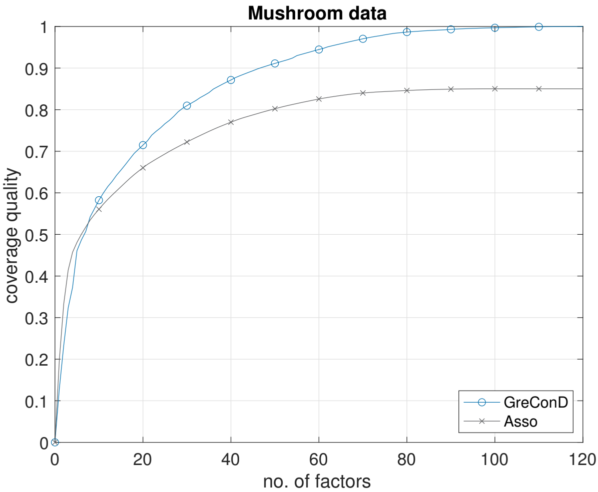
Because the $\circ$-product uses a logical sum

- E_u decreases with adding new factors,
- E_o increases with adding new factors

A sport analogy: long jump

- E_u is an underperformed jump (more strength will fix it)
- E_o is a foul jump (more strength does not fix it)

From-bellow methods – assure $E_\mathrm{o} = 0$.



coverage $= (nm - E)/nm$

- GreConD is from-bellow
- Asso allows for overcovering

Basic variants

Discrete Basis Problem (DBP)

- **Input:** $\boldsymbol{I} \in \{0,1\}^{m \times n}$ and positive integer k
- **Goal:** find $\boldsymbol{A} \in \{0,1\}^{m \times k}$ and $\boldsymbol{B} \in \{0,1\}^{k \times n}$, that minimalize $E(\boldsymbol{I}, \boldsymbol{A} \circ \boldsymbol{B})$

Miettinen P., Mielikainen T., Gionis A., Das G., Mannila H.,
The discrete basis problem,
IEEE Transactional Knowledge and Data Engineering
20(10)(2008), 1348–1362

Approximate Factorization Problem (AFP)

- **Input:** for $\boldsymbol{I} \in \{0,1\}^{m \times n}$ and given error $0 \leq \varepsilon \leq 1$
- **Goal:** find $\boldsymbol{A} \in \{0,1\}^{m \times k}$ and $\boldsymbol{B} \in \{0,1\}^{k \times n}$, s.t. $E(\boldsymbol{I}, \boldsymbol{A} \circ \boldsymbol{B}) \leq \varepsilon$, that minimalize k

Belohlavek R., Vychodil V.
Discovery of optimal factors in binary data via a novel method of matrix decomposition
Journal of Computer and System Sciences 76(1)(2010), 3–20.

Belohlavek, R., Trnecka, M.,
From-below approximations in Boolean matrix factorization: Geometry and new algorithm,
Journal of Computer and System Sciences 81(8)(2015), 1678–1697.

FCA in BMF

The boolean matrix product

$$(\boldsymbol{A} \circ \boldsymbol{B})_{ij} = \bigvee_{\ell=1\ldots k} \boldsymbol{A}_{i\ell} \wedge \boldsymbol{B}_{\ell j}$$

can be written as

$$(\boldsymbol{A} \circ \boldsymbol{B})_{ij} = \bigvee_{\ell=1\ldots k} (\boldsymbol{A}_{\ell\_} \circ \boldsymbol{B}_{\_\ell})$$

where

- $\boldsymbol{A}_{\ell\_}$ is ℓ-th row of $\boldsymbol{A}$,
- $\boldsymbol{B}_{\_\ell}$ is ℓ-th column of $\boldsymbol{B}$.

For example, we can write

$$I = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \circ \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

as

$$\begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix} \circ \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \end{pmatrix} \vee \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \circ \begin{pmatrix} 0 & 0 & 1 & 1 & 0 \end{pmatrix} \vee \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} \circ \begin{pmatrix} 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

i.e.

$$\begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \vee \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \vee \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Belohlavek R., Vychodil V.
Discovery of optimal factors in binary data via a novel method of matrix decomposition
Journal of Computer and System Sciences 76(1)(2010), 3–20.

Observation:

- BMF = covering with rectangles
- Formal concepts (as maximal rectangles) are ideal factors.

This is quite a trivial result.
Important role in showing the DM community this view.

universality
There is $\mathcal{F} = \{\langle A_1, B_1\rangle, \langle A_2, B_2\rangle, \dots, \langle A_k, B_k\rangle\} \subseteq \mathcal{B}(\boldsymbol{I})$ such that

$$\boldsymbol{I} = \boldsymbol{A}_{\mathcal{F}} \circ \boldsymbol{B}_{\mathcal{F}}.$$

where

$$(\boldsymbol{A}_{\mathcal{F}})_{i\ell} = \begin{cases} 1 & \text{if } i \in A_\ell, \\ 0 & \text{otherwise.} \end{cases} \qquad (\boldsymbol{B}_{\mathcal{F}})_{\ell j} = \begin{cases} 1 & \text{if } j \in B_\ell, \\ 0 & \text{otherwise.} \end{cases}$$

optimality
If there are $\boldsymbol{A} \in \{0,1\}^{n\times k}$ and $\boldsymbol{B} \in \{0,1\}^{k\times m}$, then there is $\mathcal{F} \subseteq \mathcal{B}(\boldsymbol{I})$ s.t.

$$\boldsymbol{I} = \boldsymbol{A}_{\mathcal{F}} \circ \boldsymbol{B}_{\mathcal{F}}.$$

and $|\mathcal{F}| \leq k$.

Some Algorithms

GreCon

- Compute $\mathcal{B}(X, Y, I)$
- Iteratively greedily select concepts from $\mathcal{B}(X, Y, I)$ which cover the most (yet uncovered) ones

Efficient implementation in:

Martin Trnecka, Roman Vyjidacek:
Revisiting the GreCon algorithm for Boolean matrix factorization.
Knowl. Based Syst. 249: 108895 (2022)

(uses incidence counters ...)

GreConD

- Does not compute $\mathcal{B}(X, Y, I)$ in advance
- Finds the concepts on demand.

Algorithm 1: GreConD(I)

$\mathcal{U} \leftarrow \{\langle i,j\rangle \mid I_{ij} = 1\}$;
$\mathcal{F} \leftarrow \varnothing$;

while $\mathcal{U} \neq \varnothing$ **do**

$D \leftarrow \varnothing$;
$V \leftarrow 0$;

$$|D \oplus j| = |(D \cup j)^{\downarrow} \times (D \cup j)^{\downarrow\uparrow} \cap \mathcal{U}|$$

select j that maximizes $|D \oplus j|$;
while $|D \oplus j| > V$ **do**

$V \leftarrow |D \oplus j|$;
$D \leftarrow (D \cup j)^{\uparrow\downarrow}$;
select j that maximizes $|D \oplus j|$;

$C \leftarrow D^{\downarrow}$;
$\mathcal{F} \leftarrow \mathcal{F} \cup \{\langle C, D\rangle\}$;

for $\langle i,j\rangle \in \mathcal{U}$ **do**

if $\langle i,j\rangle \in C \circ D$ **then** $\mathcal{U} \leftarrow \mathcal{U} - \{\langle i,j\rangle\}$;

GreConD – Demonstration

$$\boldsymbol{I} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$\mathcal{U}$ is collection of entries with ones in $\boldsymbol{I}$.

$D = \emptyset$

Find $j \in Y$ that maximizes $|D \oplus j|$:

- $j = 1$:

$$\{1\}^{\downarrow} \circ \{1\}^{\downarrow\uparrow} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

(highlighted entries are those in $\mathcal{U}$).. $|D \oplus j| = 4$

- $j = 2$:

$$\{2\}^{\downarrow} \circ \{2\}^{\downarrow\uparrow} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

  $|D \oplus 2| = 6$

- $j = 3$ and $j = 4$:

$$\{3\}^{\downarrow} \circ \{3\}^{\downarrow\uparrow} = \{4\}^{\downarrow} \circ \{4\}^{\downarrow\uparrow} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

  $|D \oplus 3| = |D \oplus 4| = 4$

- $j = 5$:
  $|D \oplus 5| = 4$

$j = 2$ maximizes $|D \oplus j|$
$D \leftarrow D \cup \{2\}^{\downarrow\uparrow} = \{1, 2\}$

Repeat:
Find $j \in Y$ that maximizes $|D \oplus j|$:

- $j = 1$ and $j = 2$: they are already in D
- $j = 3$ and $j = 4$:

$$\{1,2,3\}^{\downarrow} \circ \{1,2,3\}^{\downarrow\uparrow} = \{1,2,4\}^{\downarrow} \circ \{1,2,4\}^{\downarrow\uparrow} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$|D \oplus 3| = |D \oplus 4| = 4$$

- $j = 5$:

$$\{1,2,5\}^{\downarrow} \circ \{1,2,5\}^{\downarrow\uparrow} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$|D \oplus 5| = 3.$

We did not find a concept with better coverage, therefore we take $\langle D^{\downarrow}, D \rangle$ as the first factor.

We remove covered entries from $\mathcal{U}$

$$\begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

And continue with finding the next factor.

Essential elements

Consider for every entry $\boldsymbol{I}_{ij} = 1$, the interval

$$\mathcal{I}_{ij} = [\gamma(i), \mu(j)] = \{c \in \mathcal{B}(\boldsymbol{I}) \mid \gamma(i) \leq c \leq \mu(j)\},$$

in $\mathcal{B}(\boldsymbol{I})$, where $\gamma(i) = \{i^{\uparrow\downarrow}, i^{\uparrow}\}$ and $\mu(i) = \{j^{\downarrow}, i^{\downarrow\uparrow}\}$.

Entries $\langle i, j \rangle$ for which $\mathcal{I}_{ij}$ is minimal w.r.t. $\subseteq$ are called essential.
It constitutes a new object-attribute relation, $\mathcal{E}(\boldsymbol{I}) \in \{0,1\}^{n \times m}$:

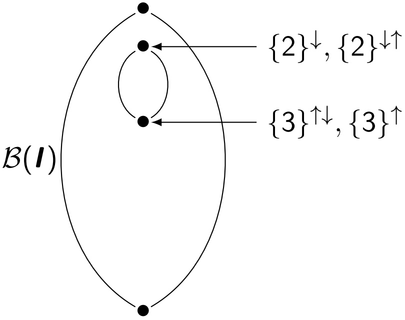
$$(\mathcal{E}(\boldsymbol{I}))_{ij} = 1 \quad \text{iff} \quad \mathcal{I}_{ij} \text{ is nonempty (a) and minimal (b).}$$

(a) – equivalent to $\boldsymbol{I}_{ij} = 1$

(b) – equivalent to: for all objects i' with $\{i'\}^{\uparrow} \subset \{i\}^{\uparrow}$ we have $I_{i'j} = 0$; analogously for attributes.

For any $\mathcal{F} \subseteq \mathcal{B}(\boldsymbol{I})$, if $\boldsymbol{J} \subseteq \boldsymbol{A}_{\mathcal{F}} \circ \boldsymbol{B}_{\mathcal{F}}$ then $\boldsymbol{I} = \boldsymbol{A}_{\mathcal{F}} \circ \boldsymbol{B}_{\mathcal{F}}$.

$$\begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$



$\mathcal{I}_{3,2}$ in not minimal: $\mathcal{I}_{3,3} \subset \mathcal{I}_{3,2}$
$\mathcal{I}_{3,3}$ is minimal.

$$\mathcal{E}(\boldsymbol{I}) = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

- $\mathcal{E}(\boldsymbol{I})$ tends to be significantly smaller than I.
- We may focus on the entries $\langle i, j \rangle \in \mathcal{E}(\boldsymbol{I})$, and ignore the (less important) entries in $\boldsymbol{I}$ that are not in $\mathcal{E}(\boldsymbol{I})$.

R. Belohlavek, M. Trnecka
From-below approximations in Boolean matrix factorization: Geometry and new algorithm.
J. Comput. Syst. Sci. 81(8): 1678-1697 (2015)

- GreEss algorithm

Essential elements (called tight) with part of the theory also appears in:

C.V. Glodeanu, B. Ganter
Applications of Ordinal Factor Analysis.
ICFCA 2013: 109-124

The 8M algorithm

R. Belohlavek, M. Trnecka
The 8M Algorithm from Today's Perspective.
ACM Transactions on Knowledge Discovery from Data 15 (2)(2021), article 22.

8M

- first algorithm for BMF (1983)
- part of BMDP (statistics package; now non-existent)
- does not use the geometric view, uses classical linear algebra
- still it gives good results

8M+

- improvement of 8M using FCA
- "two steps forward, one step back" idea: included in GreConD and Asso $\rightarrow$ significant improvement

Extension to a setting with fuzzy attributes

Factorization of Matrices with Truth Degrees

Input: Matrix $\boldsymbol{I}$

- $n \times m$
- contains truth degrees – truth (1) and false (0), and intermediate degrees

For instance,

$$\begin{pmatrix} 0.8 & 1 & 0 & 0 & 0 \\ 1 & 0.8 & 0 & 0 & 0.2 \\ 0.4 & 1 & 0.8 & 0.4 & 0 \\ 0.2 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Goal: to decompose $\boldsymbol{I}$ into $\boldsymbol{A} \circ \boldsymbol{B} \approx \boldsymbol{I}$ where

- $\boldsymbol{A}$ is $n \times k$ matrix object $\times$ factors
- $\boldsymbol{B}$ is $k \times m$ matrix factors $\times$ attributes
- effort $k \ll m$

The structure of truth degrees = complete residuated lattice

Definition

Complete residuated lattice is a structure $\mathbf{L} = \langle L, \wedge, \vee, \otimes, \rightarrow, 0, 1\rangle$ s.t.

- $\langle L, \wedge, \vee, 0, 1\rangle$ is a complete lattice, i.e. poset where arbitrary infima and suprema exist (the lattice order of $\mathbf{L}$ is denoted $\leq$);
- $\langle L, \otimes, 1\rangle$ is a commutative monoid, i.e., $\otimes$ is a commutative, associative binary operation with $a \otimes 1 = a$ for all $a \in L$;
- $\otimes$ and $\rightarrow$ satisfy the adjoint property, i.e.,

$$a \otimes b \leq c \qquad \Longleftrightarrow \qquad a \leq b \rightarrow c.$$

Joseph A. Goguen
The logic of inexact concepts.
Synthese (1969): 325-373.

Example

Typical examples, $L = [0,1]$ and $\otimes$ and $\rightarrow$ given as:

- Łukasiewicz
$$a \otimes b = \max(a + b - 1, 0),$$
$$a \rightarrow b = \min(1 - a + b, 1),$$

- Gödel
$$a \otimes b = \min(a, b),$$
$$a \rightarrow b = \begin{cases} 1 & \text{if } a \leq b, \\ b & \text{otherwise,} \end{cases}$$

- Goguen (product)
$$a \otimes b = a \cdot b,$$
$$a \rightarrow b = \begin{cases} 1 & \text{if } a \leq b, \\ \frac{b}{a} & \text{otherwise.} \end{cases}$$

Factorization of Matrices with Truth Degrees

Goal: to decompose $\boldsymbol{I}$ into $\boldsymbol{A} \circ \boldsymbol{B} \approx \boldsymbol{I}$ where

- $\boldsymbol{A}$ is $n \times k$ matrix object $\times$ factors
- $\boldsymbol{B}$ is $k \times m$ matrix factors $\times$ attributes
- effort $k \ll m$

The $\circ$-product is defined as:

$$(\boldsymbol{A} \circ \boldsymbol{B})_{ij} = \bigwedge_{\ell=1}^{k} \boldsymbol{A}_{i\ell} \otimes \boldsymbol{B}_{\ell j}$$

For matrices $\boldsymbol{I}, \boldsymbol{J} \in L^{n \times m}$

$$s(\boldsymbol{I}, \boldsymbol{J}) = \frac{\sum_{i,j=1}^{m,n} s_L(\boldsymbol{I}_{ij}, \boldsymbol{J}_{ij})}{n \cdot m}$$

i.e. $s(\boldsymbol{I}, \boldsymbol{J}) \in [0, 1]$ is the normalized sum over all matrix entries of the closeness of the corresponding entries in I and J.

We require

- $s_L(a, b) = 1$ if and only if $a = b$,
- $s_L(0, 1) = s_L(1, 0) = 0$,

(in which case $s(\boldsymbol{I}, \boldsymbol{J}) = 1$ iff $\boldsymbol{I} = \boldsymbol{J}$.)

Addressing two issues (I)

- Ordinal data and the methods for data analysis of such data appear in the literature on mathematical psychology.
- The tools employed there are basically modifications of classical factor analysis methods: grades (truth degrees) are represented by and treated like numbers.

This leads to loss of interpretability, demonstrated in:

N. Tatti, T. Mielikäinen, A. Gionis, H. Mannila,
What is the dimension of your binary data?
Proc. IEEE ICDM 2006, pp. 603–612.

Addressing two issues (II)

- Ordinal scaling + Boolean algorithms – brings considerably worse performace as regards both the quality of decomposition and computation time.
  Decathlon example at the panel discussion:

  R. Belohlavek, M. Krmelova:
  Factor Analysis of Sports Data via Decomposition of Matrices with Grades.
  CLA 2012: 293-304

- Double scaling – is (mostly) equivalent.

The decathlon example

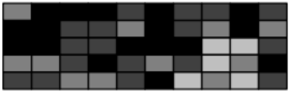
Scores of Top 5 Athletes

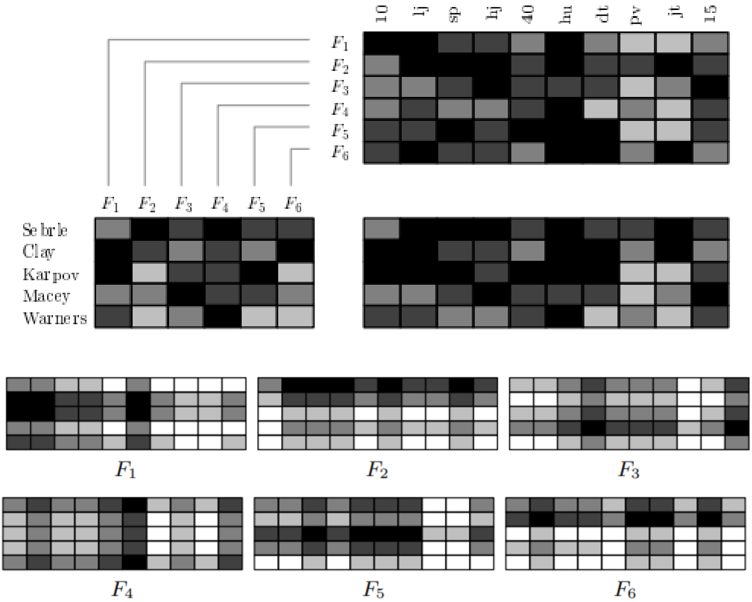
	10	*lj*	*sp*	*hj*	40	*hu*	*di*	*pv*	*ja*	15
Sebrle	894	1020	873	915	892	968	844	910	897	680
Clay	989	1050	804	859	852	958	873	880	885	668
Karpov	975	1012	847	887	968	978	905	790	671	692
Macey	885	927	835	944	863	903	836	731	715	775
Warners	947	995	758	776	911	973	741	880	669	693

Matrix I with Graded Attributes (input to the method)

	10	*lj*	*sp*	*hj*	40	*hu*	*di*	*pv*	*ja*	15
Sebrle	0.50	1.00	1.00	1.00	0.75	1.00	0.75	0.75	1.00	0.75
Clay	1.00	1.00	0.75	0.75	0.50	1.00	1.00	0.50	1.00	0.50
Karpov	1.00	1.00	1.00	0.75	1.00	1.00	1.00	0.25	0.25	0.75
Macey	0.50	0.50	0.75	1.00	0.75	0.75	0.75	0.25	0.50	1.00
Warners	0.75	0.75	0.50	0.50	0.75	1.00	0.25	0.50	0.25	0.75

Graphical Representation of Matrix I





10 lj sp hj 40 hu dt pv jt 15
F_1 F_2 F_3 F_4 F_5 F_6
Sebrle
Clay
Karpov
Macey
Warners
F_1 F_2 F_3 F_4 F_5 F_6

Formal Concept Analysis for Inexact Data

Formal L-concept analysis in sense of:

- Ana Burusco Juandeaburre and Ramón Fuentes-González.
  The study of the L-fuzzy concept lattice.
  Mathware & soft computing. 1994 Vol. 1 Num. 3 p. 209-218 (1994)
- Silke Pollandt.
  Fuzzy Begriffe: Formale Begriffsanalyse von unscharfen Daten.
  Springer–Verlag, Berlin–Heidelberg, 1997.
- Radim Belohlavek.
  Lattices generated by binary fuzzy relations.
  Abstracts of FSTA 1998, Liptovský Ján, Slovakia, p. 11 (1998)

Formal Concept Analysis with Fuzzy Attributes

Input: formal context

attributes ↓ / objects →

	α	β	γ
a	$\times$		
b	$\times$	$\times$	
c			$\times$

$\Longrightarrow$

attributes ↓ / objects →

	α	β	γ
a	1	0	0.1
b	0.9	0.8	0.1
c	0	0	0.6

object b has attribute β

object b has attribute β (at least) in degree 0.8

L-sets and L-relations

Definition

L-set A in universe U is a mapping $A : U \to L$.

- The set of all L-sets in U is denoted L^U.
- If all $u \in U$ different from $u_1, u_2, \ldots, u_n$ satisfy $A(u) = 0$, we can also write A as

$$\{^{A(u_1)}/u_1, ^{A(u_2)}/u_2, \ldots, ^{A(u_n)}/u_n\}.$$

Operations on L-sets defined component-wise:
For instance intersection $A \cap B$ of $A, B \in L^U$ is defined by

$$(A \cap B)(u) = A(u) \wedge B(u) \qquad \text{for all } u \in U.$$

Concept-forming operators

Ordinary formal context $\langle X, Y, I\rangle$ induces operators $\Uparrow: 2^X \to 2^Y$ and $\Downarrow: 2^Y \to 2^X$:

$$y \in A^{\Uparrow} \quad \text{iff} \quad \text{for all } x \in X\text{: } x \in A \text{ implies } \langle x, y\rangle \in I$$

$$x \in B^{\Downarrow} \quad \text{iff} \quad \text{for all } y \in Y\text{: } y \in B \text{ implies } \langle x, y\rangle \in I$$

$$\Downarrow$$

For formal L-context $\langle X, Y, I\rangle$ (I is L-relation between X and Y): induces operators $\uparrow: L^X \to L^Y$ and $\downarrow: L^Y \to L^X$:

$$A^{\uparrow}(y) = \bigwedge_{x \in X} A(x) \to I(x, y)$$

$$B^{\downarrow}(x) = \bigwedge_{y \in Y} B(y) \to I(x, y)$$

Formal (L-)concept, ...

Formal concept: $\langle A, B\rangle$ where $A^{\Uparrow} = B, B^{\Downarrow} = A$.

Formal L-concept: $\langle A, B\rangle$ where $A^{\uparrow} = B, B^{\downarrow} = A$.

A – extent, B – intent.

Set of all concepts:

$$\mathcal{B}(X, Y, I) = \{\langle A, B\rangle \mid \ \langle A, B\rangle \text{ is a formal (L-)concept}\}$$

Sets of extents and intents:

$$\begin{aligned}\mathrm{Ext}(X, Y, I) &= \{A \mid \langle A, B\rangle \in \mathcal{B}(X, Y, I)\}\\ \mathrm{Int}(X, Y, I) &= \{B \mid \langle A, B\rangle \in \mathcal{B}(X, Y, I)\}\end{aligned}$$

Concept lattice: $\mathcal{B}(X, Y, I) + \leq$, where

$$\langle A, B\rangle \leq \langle C, D\rangle \quad \text{iff} \quad A \subseteq C.$$

Factorization of Matrices with Truth Degrees

R. Belohlavek,
Optimal decompositions of matrices with entries from residuated lattices,
J. Logic Comput. 22(2012), 1405–1425.

R. Belohlavek, V. Vychodil
Factorization of matrices with grades.
Fuzzy Sets and Systems 292(1)(2016), 85–97.

R. Belohlavek, J. Konecny
Operators and Spaces Associated to Matrices with Grades and Their Decompositions I,II.
NAFIPS 2008, CLA 2010: 60-69

E. Bartl, R. Belohlavek, J. Konecny
Optimal decompositions of matrices with grades into binary and graded matrices. Ann. Math.
Artif. Intell. 59(2): 151-167 (2010)

Intermezzo: General framework

Triangular products:

$$(\boldsymbol{A} \triangleleft \boldsymbol{B})_{ij} = \bigwedge_{\ell=1}^{k} \boldsymbol{A}_{i\ell} \rightarrow \boldsymbol{B}_{\ell j}$$

$$(\boldsymbol{A} \triangleright \boldsymbol{B})_{ij} = \bigwedge_{\ell=1}^{k} \boldsymbol{B}_{\ell j} \rightarrow \boldsymbol{A}_{i\ell}$$

Studied by Bandler and Kohout:

L. J. Kohout and W. Bandler.
Relational-product architectures for information processing.
Information Sciences, 37(1-3):25–37, 1985.

- We could be interested in factorization into these products.

These, together with $\circ$, are covered by a unifying framework proposed in:

E. Bartl, R. Belohlavek
Sup-t-norm and inf-residuum are a single type of relational equations.
Int. J. Gen. Syst. 40(6): 599-609 (2011)

R. Belohlavek
Sup-t-norm and inf-residuum are one type of relational product: Unifying framework and consequences.
Fuzzy Sets Syst. 197: 45-58 (2012)

FCA in BMF

The graded matrix product

$$(\boldsymbol{A} \circ \boldsymbol{B})_{ij} = \bigvee_{\ell=1\ldots k} \boldsymbol{A}_{i\ell} \otimes \boldsymbol{B}_{\ell j}$$

can be written as

$$(\boldsymbol{A} \circ \boldsymbol{B})_{ij} = \bigvee_{\ell=1\ldots k} (\boldsymbol{A}_{\ell\_} \circ \boldsymbol{B}_{\_\ell})$$

where

- $\boldsymbol{A}_{\ell\_}$ is ℓ-th row of $\boldsymbol{A}$,
- $\boldsymbol{B}_{\_\ell}$ is ℓ-th column of $\boldsymbol{B}$.

GreConD$_L$

- designed for computing exact and almost exact decompositions, it may be easily adopted for computing approximate decompositions as well as for solving the DBP(L).

R. Belohlavek,
Optimal decompositions of matrices with entries from residuated lattices,
J. Logic Comput. 22(2012), 1405–1425.

R. Belohlavek, V. Vychodil,
Factorization of matrices with grades.
Fuzzy Sets and Systems 292(1)(2016), 85–97 (preliminary version in LNAI 5548(2009), 83–97).

Algorithm 2: $\mathsf{GreCond}_L(\boldsymbol{I})$

$\mathcal{U} \leftarrow \{\langle i,j\rangle \mid I_{ij} \neq 0\}$;
$\mathcal{F} \leftarrow \emptyset$;

while $\mathcal{U} \neq \emptyset$ **do**

  $D \leftarrow \emptyset$;
  $V \leftarrow 0$; $|D \oplus_a j| = |(D \cup \{{}^a/j\})^{\downarrow} \times (D \cup \{{}^a/j\})^{\downarrow\uparrow} \cap \mathcal{U}|$

  select $\langle j, a\rangle$ that maximizes $|D \oplus_a j|$;
  while $|D \oplus_a j| > V$ **do**
    $V \leftarrow |D \oplus_a j|$;
    $D \leftarrow (D \cup \{{}^a/j\})^{\uparrow\downarrow}$;
    select $\langle j, a\rangle$ that maximizes $|D \oplus_a j|$;

  $C \leftarrow D^{\downarrow}$;
  $\mathcal{F} \leftarrow \mathcal{F} \cup \{\langle C, D\rangle\}$;

  for $\langle i,j\rangle \in \mathcal{U}$ **do**
    if $I_{ij} \leq C(i) \circ D(j)$ **then** $\mathcal{U} \leftarrow \mathcal{U} - \{\langle i,j\rangle\}$;

Other algorithms

Asso for Matrices with Truth Degrees

R. Belohlavek, M. Trneckova
The Discrete Basis Problem and Asso Algorithm for Fuzzy Attributes
IEEE Trans. Fuzzy Syst. 27(7): 1417-1427 (2019)

GreEss for Matrices with Truth Degrees

R. Belohlavek, M. Trneckova
Factorization of matrices with grades via essential entries
Fuzzy Sets Syst. 360: 97-116 (2019)

Future research

- Other algebraic sctructures.
- Significance of BMF for DM + ML (reduction of dimensionality)

R. Belohlavek, J. Outrata, M. Trnecka
Impact of Boolean factorization as preprocessing methods for classification of Boolean data.
Ann. Math. Artif. Intell. 72(1-2): 3-22 (2014)

- Complexity issues
- Novel approaches to factorization

J. Konecny, M. Trnecka
Boolean Matrix Factorization for Data with Symmetric Variables.
ICDM 2022: 1011-1016