From-Below Approximations in Boolean Matrix Factorization: Geometry and New Algorithm I

Radim Belohlavek, Martin Trnecka
radim.belohlavek@acm.org, martin.trnecka@gmail.com

DATA ANALYSIS AND MODELING LAB

Dept. Computer Science
Palacky University, Olomouc, Czech Republic

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

Outline

- Problem Description
- Previous and Related Work
- New Results in From-Below Decompositions
- GreEss: New Algorithm

Problem Description

Basic Problem Given an $n \times m$ Boolean matrix I, find $n \times k$ and $k \times m$ Boolean matrices A and B such that

$$I = A \circ B \quad \text{and} \quad k \text{ is as small as possible.}$$

$A \circ B$ denotes the Boolean matrix product, i.e.

$$(A \circ B)_{ij} = \max_{l=1}^{k} \min(A_{il}, B_{lj}).$$

Example

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

(Our) Motivation—Factor Analysis in Boolean Data

General idea of factor analysis:

– Spearman: General intelligence, objectively determined and measured. Amer. J. Psychology (1904)

– according to Harman:
"The principal concern of factor analysis is the **resolution of a set of variables linearly in terms of (usually) a small number of categories or 'factors'**. … A satisfactory solution will yield factors which convey all the essential information of the original set of variables. Thus, the chief aim is to attain scientific parsimony or **economy of description**."

– factors are new variables "discovered" in data,

- technical aspect:
  - reduction of dimensionality of data
- knowledge discovery aspect:
  - factors = more fundamental aspects of data, original variables are manifestations of (more fundamental) factors

Our situation: Boolean factor analysis

- $I \in \{0,1\}^{n\times m}$ ... object-attribute matrix
- $A \in \{0,1\}^{n\times k}$ ... object-factor matrix
- $B \in \{0,1\}^{k\times m}$ ... factor-attribute matrix

$I = A \circ B$ means:

object i has attribute j iff there exists factor l such that l applies to i and j is one of particular manifestations of l

Small k (no. factors) means that the original attributes are explained by a small no. factors.

Smallest such k is called the **Boolean rank** (or Schein rank) of I:

$$\mathrm{rank_B}(I) = \min\{k \mid I = A \circ B \text{ for some } A \in \{0,1\}^{n\times k}, B \in \{0,1\}^{k\times m}\}$$

Some Equivalent Forms of the Basic Problem

Relations

Given a binary relation $R \subseteq X \times Y$, find smallest Z such that there exist relations $S \subseteq X \times Z$ and $T \subseteq Z \times Y$ whose composition yields R.

Graphs

Given a bipartite graph, find the smallest number of its bicliques that cover all edges of the graph.

Ordered Sets

Find most efficient 2-encoding of the partially ordered set on rows $\cup$ columns given by $\overline{I}$.

Some Equivalent Forms of Boolean Rank

Graphs

Boolean rank of I = bipartite dimension of a bipartite graph.

Formal Concept Analysis

Boolean rank of I = set dimension of the corresponding formal context.

Dimension of Ordered Sets

Boolean rank of I = 2-dimension of the concept lattice of $\overline{I}$.

... some results exist in the related areas, mostly regarding computational complexity, not regarding algorithms for the problem

Beyond the Basic Decomposition Problem

Variants of the basic problem have been considered in the literature.

L_1-norm $||\cdot||$ and the corresponding metric $E(\cdot,\cdot)$ for $C, D \in \{0,1\}^{n\times m}$:

$$||C|| = \textstyle\sum_{i,j=1}^{m,n} |C_{ij}| \quad \text{and} \quad E(C,D) = ||C-D|| = \textstyle\sum_{i,j=1}^{m,n} |C_{ij} - D_{ij}|. \quad (1)$$

- *Discrete Basis Problem* (DBP): given $I \in \{0,1\}^{n\times m}$ and a positive integer k, find $A \in \{0,1\}^{n\times k}$ and $B \in \{0,1\}^{k\times m}$ that minimize $||I - A \circ B||$.
- *Approximate Factorization Problem* (AFP): given I and prescribed error $\varepsilon \geq 0$, find $A \in \{0,1\}^{n\times k}$ and $B \in \{0,1\}^{k\times m}$ with k as small as possible such that $||I - A \circ B|| \leq \varepsilon$.

Reflect two important views on BMF. The first one emphasizes the importance of the first k (presumably most important) factors. The second one emphasizes the need to account for (and thus to explain) a prescribed portion of data, which is specified by ε.

Complexity of Decompositions

Theorem

The basic decomposition problem is NP-complete (decision version)/NP-hard (optimization version).

- Proof essentially due to Stockmeyer (1975, set basis problem) and Orlin (1977, graphs).
- Many of related decomposition problems (DBP, AFP, . . .) are provably hard as well.
- Need for approximation algorithms.
- Further complexity results including inapproximability results: Gottlieb, Savage , Yerukhimovich (2005), Gruber, Holzer (2007),

Previous Directly Related Work

- Early work by Markowsky et al. (Math. Biosciences)
- Van Mechelen (psychology, KU Leuven)
- Frolov et al.: Boolean factor analysis by Hopfield-like autoassociative memory. IEEE Trans. Neural Networks 18(2007) 698–707.
- Geerts et al.: Tiling Databases. LNCS 3245(2004), 278–289
- Miettinen et al.: The discrete basis problem. IEEE TKDE 20(10):1348–1362, 2008. (problem, complexity, Asso)
- Further, more recent work (Hyper, Panda, issue of noise),

  overview of algorithms in Part II
- Further heuristic approaches (data mining).

Our Previous Work

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

RB+Vychodil: Factor analysis of incidence data via novel decomposition of matrices. LNAI 5548(2009), 83-97.

RB: Optimal triangular decompositions of matrices with entries from residuated lattices.
Int. J. Approximate Reasoning 50(8)(2009), 1250–1258.

RB: Optimal decompositions of matrices with entries from residuated lattices.
J. Logic and Computation 22(6)(2012), 1405–1425.

Our 2010 (2006) Paper

– Decompositions of I = coverings of 1 in I by rectangles

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

– Best such coverings use maximal rectangles, hence formal concepts.

– Transformations between attribute and factor spaces.

– Algorithm 1 (GreCon) and Algorithm 2 (GreConD).

Optimal Decompositions via Formal Concepts

Given a set $\mathcal{F} = \{\langle C_1, D_1\rangle, \ldots, \langle C_k, D_k\rangle\} \subseteq \mathcal{B}(X, Y, I)$ of formal concepts, define $n \times k$ and $k \times m$ matrices $A_\mathcal{F}$ and $B_\mathcal{F}$ by

$$(A_\mathcal{F})_{il} = (C_l)(i) \quad \text{and} \quad (B_\mathcal{F})_{lj} = (D_l)(j).$$

Theorem (formal concepts of I are optimal factors for I)

Let $I = A \circ B$ for $n \times k$ and $k \times m$ matrices A and B. Then there exists a set $\mathcal{F} \subseteq \mathcal{B}(X, Y, I)$ of formal concepts with

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

such that for the $n \times |\mathcal{F}|$ and $|\mathcal{F}| \times m$ matrices $A_\mathcal{F}$ and $B_\mathcal{F}$ we have

$$I = A_\mathcal{F} \circ B_\mathcal{F}.$$

$\Rightarrow$ Boolean rank achievable by formal concepts

Algorithms from our 2010 paper

- Algorithm 1 (GreCon): direct application of Set Cover algorithm
- Obstacle: Computing and repeatedly going through $\mathcal{B}(X, Y, I)$ slow for large data.
- Algorithm 2 (GreConD): needs not compute $\mathcal{B}(X, Y, I)$
- Is much faster than Algorithm 1 (3 sec vs. 20 min). (Probably fastest existing algorithm.)
- Computes decompositions of the same quality as Algorithm 1.

New Contributions

- Emphasize the role of from-below decompositions.
- New theoretical results, exploiting "geometry of decompositions": some 1s are more important than others.
- New algorithm (GreEss).
- Extensive experimental evaluation (Part II).

From-Below Decompositions

E (metric measuring error) may be seen as being the sum of two components, E_u ("uncovered") and E_o ("overcovered"):

$$E(I, A \circ B) = E_u(I, A \circ B) + E_o(I, A \circ B), \text{ where}$$
$$E_u(I, A \circ B) = |\{\langle i, j\rangle \,;\, I_{ij} = 1, (A \circ B)_{ij} = 0\}|,$$
$$E_o(I, A \circ B) = |\{\langle i, j\rangle \,;\, I_{ij} = 0, (A \circ B)_{ij} = 1\}|.$$

Even though E_u and E_o look symmetric, they have a highly non-symmetric role in BMF.

Lemma

Let $A' \in \{0,1\}^{n\times(k+1)}$ and $B' \in \{0,1\}^{(k+1)\times m}$ result by adding to A and B a single column and row, respectively. Then

$$E_u(I, A' \circ B') \leq E_u(I, A \circ B) \text{ and } E_o(I, A' \circ B') \geq E_o(I, A \circ B).$$

That is, adding factors, E_u decreases and E_o increases. Important due to hardness of decompositions.

Lemma provides a warning.

Namely, we should be careful with committing E_o error because E_o never decreases by adding further factors.

The most extreme strategy is not to commit E_o error at all, i.e. add the constraint $E_o(I, A \circ B) = 0$.

As the requirement $E_o(I, A \circ B) = 0$ is equivalent to $A \circ B \leq I$, we call a BMF algorithm producing results with zero E_o a **from-below factorization algorithm** and say that A and B provide a from-below factorization of I.

We omit results on optimal approximate from-below factorizations (see our paper).

Role of Entries Containing 1

Reformulation of the basic problem in terms of labeled diagram of $\mathcal{B}(I)$:

Find smallest subset $\mathcal{F}$ of nodes in the diagram such that whenever there exists a path from a node $\gamma(i)$ up to $\mu(i)$, then some such path goes through some $c \in \mathcal{F}$.

Recall:

$$\gamma(i) = \langle \{i\}^{\uparrow\downarrow}, \{i\}^{\uparrow} \rangle \quad \text{and} \quad \mu(j) = \langle \{j\}^{\downarrow}, \{j\}^{\downarrow\uparrow} \rangle.$$

According to the Basic Theorem of FCA, $\gamma(i) \leq \mu(j)$ iff $I_{ij} = 1$. Moreover, the concepts $\langle C, D \rangle$ that cover $\langle i, j \rangle$ are just those in the interval

$$\mathcal{I}_{ij} = [\gamma(i), \mu(j)].$$

Intervals play important role.
For $C \subseteq X$ and $D \subseteq Y$, denote by $\mathcal{I}_{C,D}$ the interval

$$\mathcal{I}_{C,D} = [\gamma(C), \mu(D)] := \{\langle E, F\rangle \in \mathcal{B}(I) \mid \gamma(C) \leq \langle E, F\rangle \leq \mu(D)\}$$

in $\mathcal{B}(I)$ where $\gamma(C) = \langle C^{\uparrow\downarrow}, C^{\uparrow}\rangle$ and $\mu(D) = \langle D^{\downarrow}, D^{\downarrow\uparrow}\rangle$.

We have:

Lemma

(a) *$\mathcal{I}_{C,D}$ is non-empty if and only if $C \times D \subseteq I$. In particular, $\mathcal{I}_{ij}$ is non-empty if and only if $I_{ij} = 1$.*

(b) *$\mathcal{I}_{C,D} = \{\langle E, F\rangle \in \mathcal{B}(I) \mid C \subseteq E, D \subseteq F\} = \{\langle E, F\rangle \in \mathcal{B}(I) \mid C^{\uparrow\downarrow} \subseteq E, D^{\downarrow\uparrow} \subseteq F\}$. In particular, $\mathcal{I}_{ij}$ is the set of all concepts that cover $\langle i, j\rangle$.*

(c) *If $(A_{\mathcal{F}} \circ B_{\mathcal{F}})_{ij} = 1$ then $\mathcal{F}$ contains at least one concept in $\mathcal{I}_{ij}$.*

ε-Essential Entries

Given a prescribed error $\varepsilon \geq 0$, is it possible to identify a matrix $J \leq I$ with a small number $||J||$ of 1s with the property that for any $\mathcal{F} \subseteq \mathcal{B}(I)$, if $J \leq A_{\mathcal{F}} \circ B_{\mathcal{F}}$ then $E(I, A_{\mathcal{F}} \circ B_{\mathcal{F}}) \leq \varepsilon$?

We call any matrix J satisfying the above properties ε-*essential*. In other words, the coverage of 1s in J by any $\mathcal{F} \subseteq \mathcal{B}(I)$ is sufficient for the error of $A_{\mathcal{F}} \circ B_{\mathcal{F}}$ to be at most ε.

In what follows, we restrict to 0-essential matrices, which we simply call *essential*, or essential parts of I, and utilize them in a new BMF algorithm.

Denote by $\mathcal{E}(I)$ the $n \times m$ Boolean matrix given by

$$(\mathcal{E}(I))_{ij} = 1 \text{ iff } \mathcal{I}_{ij} \text{ is non-empty and minimal w.r.t. } \subseteq,$$

where $\subseteq$ denotes set inclusion. Note that

$$\mathcal{I}_{ij} \subseteq \mathcal{I}_{i'j'} \text{ iff } \gamma(i') \leq \gamma(i) \text{ and } \mu(j) \leq \mu(j') \text{ iff } \{i\}^{\uparrow} \subseteq \{i'\}^{\uparrow} \text{ and } \{j\}^{\downarrow} \subseteq \{j'\}^{\downarrow}$$

and that $\mathcal{E}(I)$ is easy to compute.

Importantly:

Theorem

(a) $\mathcal{E}(I)$ is an essential part of I.
(b) If J is an essential part of a reduced I then $\mathcal{E}(I) \leq J$.

(I is reduced if any two rows (columns) are different.)
Note: Independently, sufficiency to cover minimal intervals (part (a)) was shown in Ganter, Glodeanu (Ordinal factor analysis, ICFCA 2012).

GreEss: New Decomposition Algorithm

Basic idea:

Factorizations of $\mathcal{E}(I)$ may be used to obtain factorizations of I:

Theorem

Let I be a Boolean matrix. If $\mathcal{G} \subseteq \mathcal{B}(\mathcal{E}(I))$ be a set of factor concepts of $\mathcal{E}(I)$, i.e. $\mathcal{E}(I) = A_{\mathcal{G}} \circ B_{\mathcal{G}}$, then every set $\mathcal{F} \subseteq \mathcal{B}(I)$ containing at least one concept from $\mathcal{I}_{C,D}$ for each $\langle C, D \rangle \in \mathcal{G}$ is a set of factor concepts of I, i.e. $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$.

May be generalized to arbitrary factorizations of $\mathcal{E}(I)$.

Implies that the rank of $\mathcal{E}(I)$ provides an upper bound on the rank of I:

Theorem

For every Boolean matrix I we have $\mathrm{rank}_{\mathrm{B}}(I) \leq \mathrm{rank}_{\mathrm{B}}(\mathcal{E}(I))$.

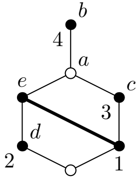
Remark The estimation by Theorem is not tight. Namely, as one may check that for

$$I = \begin{pmatrix} 10111 \\ 01101 \\ 01001 \\ 10110 \end{pmatrix} = \begin{pmatrix} 110 \\ 011 \\ 001 \\ 100 \end{pmatrix} \circ \begin{pmatrix} 10110 \\ 00101 \\ 01001 \end{pmatrix}, \text{ we have } \mathcal{E}(I) = \begin{pmatrix} 00001 \\ 00100 \\ 01000 \\ 10010 \end{pmatrix}.$$

While we see that $\mathrm{rank}_{\mathrm{B}}(I) \leq 3$ (in fact, the rank equals 3), one may easily check that $\mathrm{rank}_{\mathrm{B}}(\mathcal{E}(I)) = 4$.

Example Objects $1, \ldots, 4$ and attributes $a, \ldots, e$. The underlined entries are just the essential 1s, i.e. those with $\mathcal{E}(I)_{ij} = 1$.

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



According to the above results one can easily see that the set $\mathcal{F}$ consisting of the elements of the singleton intervals and either of the two elements of $\mathcal{I}_{1e}$ is the smallest set of factor concepts, i.e. $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$ and $|\mathcal{F}| = \mathrm{rank}_{\mathrm{B}}(I)$.

GreEss: Basic Idea

Pseudocodes:

- Compute intervals in $\mathcal{B}(I)$.
- Search these intervals in a greedy manner for factor concepts of I.

Behind is the following utilization of the previous theorems:

- Compute $\mathcal{E}(I)$.
- Start computing set $\mathcal{G}$ of factors of $\mathcal{E}(I)$ as in GreConD.
- Stop (computing $\mathcal{G}$) when the "conservative estimations" of the counterparts of $\langle C, D\rangle \in \mathcal{G}$ cover I, i.e. stop when $I = \bigcup_{\langle C,D\rangle\in\mathcal{G}} C^{\uparrow_I\downarrow_I} \times D^{\downarrow_I\uparrow_I}$.
- Then, select one concept per interval $\mathcal{I}_{C,D}$ for $\langle C, D\rangle \in \mathcal{G}$ in $\mathcal{B}(I)$ in a greedy manner.

Simple modification (stop when error $\leq \varepsilon$) delivers approximate from-below decompositions.

Algorithm **GreEss**

Input: Boolean matrix I and $\varepsilon \geq 0$
Output: set $\mathcal{F}$ of factors for which $||I - A_{\mathcal{F}} \circ B_{\mathcal{F}}|| \leq \varepsilon$

$\mathcal{G} \leftarrow \mathrm{ComputeIntervals}(I)$ $\;U \leftarrow \{\langle i,j\rangle \,|\, I_{ij} = 1\}$; $\mathcal{F} \leftarrow \emptyset$
while $|U| > \varepsilon$ **do**
  $s \leftarrow 0$ **foreach** $\langle C, D\rangle \in \mathcal{G}$ **do**
    $J = D^{\downarrow_I} \times C^{\uparrow_I}$ $\;F \leftarrow \emptyset$; $s_{\langle C,D\rangle} \leftarrow 0$ **while** *exists* $j \in C_f^{\uparrow_I} - F$ *s.t.* $|(F \cup \{j\})^{\downarrow_J} \times (F \cup \{j\})^{\downarrow_J \uparrow_J} \cap U| > s_{\langle C,D\rangle}$ **do**
      select j maximizing $|(F \cup \{j\})^{\downarrow_J} \times (F \cup \{j\})^{\downarrow_J \uparrow_J} \cap U|$
      $F \leftarrow (F \cup \{j\})^{\downarrow_J \uparrow_J}$; $E \leftarrow (F \cup \{j\})^{\downarrow_J}$ $\;s_{\langle C,D\rangle} \leftarrow |E \times F \cap U|$
    end
    if $s_{\langle C,D\rangle} > s$ **then** $\langle E', F'\rangle \leftarrow \langle E, F\rangle$; $\langle C', D'\rangle \leftarrow \langle C, D\rangle$
  end
  add $\langle E', F'\rangle$ **to** $\mathcal{F}$; **remove** $\langle C', D'\rangle$ **from** $\mathcal{G}$
  $U \leftarrow U - E' \times F'$
end
return $\mathcal{F}$

Algorithm **ComputeIntervals**

Input: Boolean matrix I
Output: Set $\mathcal{G} \subseteq \mathcal{B}(\mathcal{E}(I))$

$\mathcal{E} \leftarrow \mathcal{E}(I)$; $U \leftarrow \{\langle i,j\rangle \,|\mathcal{E}_{ij} = 1\}$
while $U \neq \emptyset$ **do**
  $D \leftarrow \emptyset$; $s \leftarrow 0$
  while *exists* $j \notin d$ *with* $|((D \cup \{j\})^{\downarrow_{\mathcal{E}}})^{\uparrow_I \downarrow_I} \times ((D \cup \{j\})^{\downarrow_{\mathcal{E}} \uparrow_{\mathcal{E}}})^{\downarrow_I \uparrow_I} \cap U| > s$ **do**
    select j which maximize $|((D \cup \{j\})^{\downarrow_{\mathcal{E}}})^{\uparrow_I \downarrow_I} \times ((D \cup \{j\})^{\downarrow_{\mathcal{E}} \uparrow_{\mathcal{E}}})^{\downarrow_I \uparrow_I} \cap U|$
    $D \leftarrow (D \cup \{j\})^{\downarrow_{\mathcal{E}} \uparrow_{\mathcal{E}}}$; $C \leftarrow (D \cup \{j\})^{\downarrow_{\mathcal{E}}}$ $s \leftarrow |C^{\uparrow_I \downarrow_I} \times D^{\downarrow_I \uparrow_I} \cap U|$
  end
  add $\langle C, D\rangle$ **to** $\mathcal{G}$
  $U \leftarrow U - C^{\uparrow_I \downarrow_I} \times D^{\downarrow_I \uparrow_I}$
end
return C, D

Theorem

GreEss is correct and provides a from-below approximation of I.

Further Issues

Experiments . . . see Part II

Future Research

- Understand better $\mathcal{E}(I)$ and its concept lattice.
- Exploit $\mathcal{E}(I)^n$.
- Approximate decompositions with $E_o \geq 0$