Discovery of factors in binary data triangular decomposition of matrices

Radim Belohlavek
State University of New York at Binghamton, U. S. A.
e-mail: rbelohla@binghamton.edu
and
Palacký University, Olomouc, Czech Republic

Abstract

We present new methods of decomposition of an $n \times m$ binary matrix I into a product $A * B$ of an $n \times k$ binary matrix A and a $k \times m$ binary matrix B. These decompositions are alternative to the usual one which is sought in Boolean factor analysis (BFA), where $*$ is a Boolean product of matrices. In the new decompositions, $*$ are the left and the right triangular products of Boolean matrices. In BFA, I is interpreted as object $\times$ attribute matrix, A and B are interpreted as object $\times$ factor and factor $\times$ attribute matrices, and the aim is to find a decomposition with the number k of factors as small as possible. The new decompositions have different semantics from the one with Boolean matrix product. The presented methods are optimal in that they provide the smallest number k possible. We provide a geometric insight showing that the factors correspond to I-beam- and H-beam-shaped patterns in the input matrix. We present an approximation algorithm for computing the optimal decomposition. The algorithm is based on a greedy approximation algorithm for set covering optimization problem. We demonstrate our results by illustrative examples.

Keywords: Binary Data, Matrix Decomposition, Triangular Product, Galois Connection.

1 Introduction

Let I be an $n \times m$ binary matrix, i.e. $I_{ij} \in \{0, 1\}$. Can we decompose I into a triangular product $I = A \triangleleft B$ of an $n \times k$ binary matrix A and a $k \times m$ binary matrix B? What are the best such decompositions, i.e. the decompositions with the least k possible? In this paper, we present answers to these questions.

Recall, see e.g. [1, 10], that the $\triangleleft$-product $A \triangleleft B$ of A and B is defined by

$$(A \triangleleft B)_{ij} = \bigwedge_{l=1}^{k} A_{jl} \rightarrow B_{lj}, \qquad (1)$$

where $\bigwedge$ denotes the truth function of logical conjunction (minimum), and $\rightarrow$ denotes the truth function of logical implication (i.e., $0 \rightarrow 0 = 0 \rightarrow 1 = 1 \rightarrow 1 = 1$ and $1 \rightarrow 0 = 0$). Alternatively, one could look for a decomposition of I into an $\triangleright$-product $I = A \triangleright B$, defined by

$$(A \triangleright B)_{ij} = \bigwedge_{l=1}^{k} B_{lj} \rightarrow A_{jl}. \qquad (2)$$

Looking for such decompositions can be seen as looking for factors in binary data. Namely, if I is interpreted as an object $\times$ attribute matrix, A can be interpreted as an object $\times$ factor matrix, and B can be interpreted as a factor $\times$ attribute matrix. For $I = A \triangleleft B$, the relationship between objects and attributes provided by I is then described via the object $\times$ factor relationship and the factor $\times$ attribute relationship provided by A and B

L. Magdalena, M. Ojeda-Aciego, J.L. Verdegay (eds): Proceedings of IPMU'08, pp. 47–54
Torremolinos (Málaga), June 22–27, 2008

as: An object i has an attribute j ($I_{ij} = 1$) if and only if for every factor l, if l applies to i ($A_{il} = 1$) then j is one of the manifestations of l ($B_{lj} = 1$). For $I = A \triangleright B$, the situation is similar. Since $A \triangleright B = (B^T \triangleleft A^T)^T$, with C^T denoting the transposition of matrix C, we can restrict ourselves to $\triangleleft$-decomposition.

This paper is a continuation of [4]. In [4], we considered a particular type of product $*$ in the decomposition $I = A * B$, namely, the Boolean matrix product $A \circ B$ of A and B, called also logical product, max-min product, or $\circ$-product. Recall that the $\circ$-product is defined by

$$(A \circ B)_{ij} = \bigvee_{l=1}^{k} A_{jl} \cdot B_{lj} \qquad (3)$$

where $\bigvee$ denotes maximum (truth function of logical disjunction) and $\cdot$ is the usual product (truth function of logical conjunction). We described optimal $\circ$-decompositions and an efficient approximation algorithm for finding such decompositions. This type of decomposition of binary matrices is considered in Boolean factor analysis [6]. Early work in this area was done by Markowsky et al., see e.g. [13] which already include complexity results showing hardness of problems related to such decompositions. Recently, the $\circ$-decomposition problem has been discussed in data mining, see e.g. [8, 12, 15]. Note also that various methods for decompositon of binary matrices into matrices which are not binary in general have been reported but we do not discuss this topic in our paper.

Our paper is organized as follows. Section 2 presents the main results. In Section 2.1, we provide a key insight—a geometric way of looking at the problem of triangular decompositions. The decompositions of binary matrices are described and their optimality is proved in Section 2.2. Section 2.3 presents an approximation algorithm for finding optimal decompositions. Section 2.4 contains an illustrative example. Section 3 presents a summary and an outline of future research.

2 Optimal decomposition of matrices by triangular products

2.1 Geometric insight: rectangles, I-beams, and H-beams

We are going to show that the problem of decomposing I into $A \circ B$, $A \triangleleft B$, and $A \triangleright B$, can be interpreted geometrically using particular shapes which we call rectangles, I-beams, and H-beams. We call an $n \times m$ Boolean matrix J rectangular-shaped (a rectangle, for short), I-beam-shaped (an I-beam), and H-beam-shaped (an H-beam), if there are Boolean matrices C and D of dimensions $n \times 1$ (column) and $1 \times m$ (row), such that $J = C \circ D$, $J = C \triangleleft D$, and $J = C \triangleright D$, respectively. For instance, for

$$C = \begin{pmatrix} 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \end{pmatrix}^T, \text{ and } D = \begin{pmatrix} 0 & 0 & 1 & 1 & 1 & 0 & 0 \end{pmatrix},$$

the corresponding rectangle, I-beam, and H-beam matrices $C \circ D$, $C \triangleleft D$, and $C \triangleright D$ are

$$\begin{pmatrix} 0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0 \\ 0&0&1&1&1&0&0 \\ 0&0&1&1&1&0&0 \\ 0&0&1&1&1&0&0 \\ 0&0&1&1&1&0&0 \\ 0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0 \end{pmatrix}, \begin{pmatrix} 1&1&1&1&1&1&1 \\ 1&1&1&1&1&1&1 \\ 0&0&1&1&1&0&0 \\ 0&0&1&1&1&0&0 \\ 0&0&1&1&1&0&0 \\ 0&0&1&1&1&0&0 \\ 1&1&1&1&1&1&1 \\ 1&1&1&1&1&1&1 \end{pmatrix}, \text{ and } \begin{pmatrix} 1&1&0&0&0&1&1 \\ 1&1&0&0&0&1&1 \\ 1&1&1&1&1&1&1 \\ 1&1&1&1&1&1&1 \\ 1&1&1&1&1&1&1 \\ 1&1&1&1&1&1&1 \\ 1&1&0&0&0&1&1 \\ 1&1&0&0&0&1&1 \end{pmatrix}.$$

The terms are derived from the geometric shapes of matrices $C \circ D$, $C \triangleleft D$, and $C \triangleright D$. For instance, I-beam matrices are those which can be, after a suitable permutation of rows and columns, brought to a form where entries containing 1 form letter I. Technically speaking, since $C \triangleleft D = \overline{C} \triangleright \overline{D}$, every I-beam is an H-beam, and vice versa. Here, $\overline{M}$ denotes the complement to M, i.e. $(\overline{M})_{ij} = 1 - M_{ij}$. Nevertheless, for conceptual purposes, we will use both I-beams and H-beams. Likewise, one can verify that $A \triangleleft B = \overline{A \circ \overline{B}}$. Therefore, technically, we can reduce the problem of $\triangleleft$-decomposition the that of $\circ$-decomposition. However, we will proceed directly with $\triangleleft$-decomposition because by reducing to $\circ$-decomposition, the natural semantics of $\triangleleft$-product gets lost.

In [4], we observed that if $I = A \circ B$, I is a union ($\bigvee$-superposition) of rectangles J_l, $l = 1, \ldots, k$, where $J_l = A_{\_l} \circ B_{l\_}$, i.e. J_l corresponds to the l-th column of A and to the

l-th row of B. In case of $\triangleleft$- and $\triangleright$-product, we have the following observation which follows easily from definitions:

Given a set

$$\mathcal{F} = \{\langle C_1, D_1\rangle, \ldots, \langle C_k, D_k\rangle\}$$

of $1 \times n$ and $1 \times m$ Boolean vectors C_l and D_l, respectively, define $n \times k$ and $k \times m$ matrices $A_{\mathcal{F}}$ and $B_{\mathcal{F}}$ by

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

That is, the l-th column of $A_{\mathcal{F}}$ is C_l^T and the l-th row of $B_{\mathcal{F}}$ is D_l. Then we have:

Theorem 1 $I = A_{\mathcal{F}} \triangleleft B_{\mathcal{F}}$ *iff* I *is an intersection of* I-*beams* $C_l^T \triangleleft D_l$, $l = 1, \ldots, k$. $I = A_{\mathcal{F}} \triangleright B_{\mathcal{F}}$ *iff* I *is an intersection of* H-*beams* $C_l^T \triangleright D_l$, $l = 1, \ldots, k$.

Due to Theorem 1, factors in the $\triangleleft$- and $\triangleright$-decompositions can be identified with I-beams and H-beams, respectively.

Example 1 Consider the following decomposition $I = A \triangleleft B$ of an 4×5 matrix:

$$\begin{pmatrix} 0&0&1&1&1 \\ 0&0&1&1&0 \\ 0&0&0&0&1 \\ 0&1&1&1&0 \end{pmatrix} = \begin{pmatrix} 1&0&0&1 \\ 1&0&1&0 \\ 1&1&0&0 \\ 0&0&1&0 \end{pmatrix} \triangleleft \begin{pmatrix} 0&0&1&1&1 \\ 1&1&0&0&1 \\ 0&1&1&1&0 \\ 1&0&1&1&1 \end{pmatrix}.$$

According to Theorem 1, this decomposition can be rewritten as a $\bigwedge$-superposition

$$\begin{aligned} \begin{pmatrix} 0&0&1&1&1 \\ 0&0&1&1&0 \\ 0&0&0&0&1 \\ 0&1&1&1&0 \end{pmatrix} &= \begin{pmatrix} 0&0&1&1&1 \\ 0&0&1&1&1 \\ 0&0&1&1&1 \\ 1&1&1&1&1 \end{pmatrix} \wedge \begin{pmatrix} 1&1&1&1&1 \\ 1&1&1&1&1 \\ 1&1&0&0&1 \\ 1&1&1&1&1 \end{pmatrix} \wedge \\ &\wedge \begin{pmatrix} 1&1&1&1&1 \\ 0&1&1&1&0 \\ 1&1&1&1&1 \\ 0&1&1&1&0 \end{pmatrix} \wedge \begin{pmatrix} 1&0&1&1&1 \\ 1&1&1&1&1 \\ 1&1&1&1&1 \\ 1&1&1&1&1 \end{pmatrix} \end{aligned}$$

of I-beams J^1, J^2, J^3, J^4, where J^l results as a $\triangleleft$-product of the l-th column of A and the l-th row of B. Note that the I-beam shape of J^ls becomes apparent after rearrangement (permutation) of rows and columns. Due to small dimensions, the "I"-shape is degenerate in case of J^1 and J^4.

2.2 Optimal decompositions

With respect to the problem of decomposing I into $A \triangleleft B$ or $A \triangleright B$, Theorem 1 says that we need to find a suitable set of I-beams and H-beams, respectively, whose intersection yields I. In this section, we describe optimal ways to decompose I, in that they lead to the least possible numbers of beams, i.e. factors.

The number of I-beams and H-beams which cover a given $n \times m$ Boolean matrix I can be quite large. For instance, there are at least 2^n I-beams covering I since for every $1 \times n$ vector C, we can consider a corresponding $1 \times m$ vector D which has 1s at positions j such that $I_{ij} = 1$ for some i with $C_i = 1$. One can easily check that the corresponding I-beam $C^T \triangleleft D$ covers I, i.e. $I_{ij} \leq (C^T \triangleleft D)_{ij}$. The reason for such a large number of beams is that there is a redundancy involved when we consider all beams. Namely, we can have two beams one inside another. As we shall see, we can restrict ourselves to particular beams which are minimal and thus non-redundant (see below) and which still enable us to decompose I with the least number of beams possible. The non-redundant beams are fixed points of certain operators which we now introduce.

Let $X = \{1, 2, \ldots, n\}$ and $Y = \{1, 2, \ldots, m\}$. This notation will be used from now on. Define operators ${}^{\cap} : 2^X \to 2^Y$ and ${}^{\cup} : 2^Y \to 2^X$, and ${}^{\wedge} : 2^X \to 2^Y$ and ${}^{\vee} : 2^Y \to 2^X$ by letting for $C \subseteq X$ and $D \subseteq Y$,

$$C^{\cap} = \{j \in Y \mid \text{for some } i \in C : I_{ij} = 1\}, \quad (4)$$
$$D^{\cup} = \{i \in X \mid \text{for each } j \in Y\text{: if } I_{ij} = 1 \text{ then } j \in D\}, \quad (5)$$

and

$$C^{\wedge} = \{j \in Y \mid \text{for each } i \in X\text{: if } I_{ij} = 1 \text{ then } i \in C\}, \quad (6)$$
$$D^{\vee} = \{i \in X \mid \text{for some } i \in C : I_{ij} = 1\}. \quad (7)$$

Pairs $\langle{}^{\cap}, {}^{\cup}\rangle$ and $\langle{}^{\wedge}, {}^{\vee}\rangle$ are particular cases of those introduced in [9]. Furthermore, denote by $\mathcal{B}(X^{\cap}, Y^{\cup}, I)$ and $\mathcal{B}(X^{\wedge}, Y^{\vee}, I)$ the sets of fixed points of $\langle{}^{\cap}, {}^{\cup}\rangle$ and $\langle{}^{\wedge}, {}^{\vee}\rangle$, respectively. That is,

$$\mathcal{B}(X^{\cap}, Y^{\cup}, I) = \{\langle C, D\rangle \mid C^{\cap} = D,\, D^{\cup} = C\},$$
$$\mathcal{B}(X^{\wedge}, Y^{\vee}, I) = \{\langle C, D\rangle \mid C^{\wedge} = D,\, D^{\vee} = C\}.$$

Denote the characteristic vectors of $C \subseteq X$ and $D \subseteq Y$ by $c(C)$ and $c(D)$. That is, $c(C)_i = 1$ if $i \in C$ and $c(C)_i = 0$ if $i \notin C$;

the same for D. For $F \subseteq \mathcal{B}(X^\cap, Y^\cup, I)$ we put

$$c(F) = \{\langle c(C), c(D)\rangle \mid \langle C, D\rangle \in F\}. \quad (8)$$

For $C_1, C_2 \subseteq X$ and $D_1, D_2 \subseteq Y$, put

$$\langle C_1, D_1\rangle \leq_{\mathrm{I}} \langle C_2, D_2\rangle \text{ iff } C_1 \supseteq C_2 \,\&\, D_1 \subseteq D_2.$$

In terms of I-beams, $\langle C_1, D_1\rangle \leq_{\mathrm{I}} \langle C_2, D_2\rangle$ means that the I-beam $c(C_1)^T \triangleleft c(D_1)$ corresponding to $\langle C_1, D_1\rangle$ is contained in the I-beam $c(C_2)^T \triangleleft c(D_2)$ corresponding to $\langle C_2, D_2\rangle$, i.e. that $(c(C_1) \triangleleft c(D_1))_{ij} \leq (c(C_2) \triangleleft c(D_2))_{ij}$ for every i and j.

Theorem 2 (fixpoints=minimal I-beams) *$\langle C, D\rangle$ is a fixed point of $\langle^\cap, ^\cup\rangle$ iff the corresponding I-beam is a minimal one which covers I, i.e. iff $\langle C, D\rangle$ is minimal with respect to $\leq_{\mathrm{I}}$ such that $I_{ij} \leq (c(C)^T \triangleleft c(D))_{ij}$ for all i and j.*

Proof. By direct verification; cf. also [9]. □

The following theorem asserts that we can always use fixed points from $\mathcal{B}(X^\cap, Y^\cup, I)$ as factors.

Theorem 3 (universality) *For every I there exists $F \subseteq \mathcal{B}(X^\cap, Y^\cup, I)$ such that $I = A_{c(F)} \triangleleft B_{c(F)}$.*

Proof. From the fact that every I is the intersection of I-beams $c(C)^T \triangleleft c(D)$, for $\langle C, D\rangle \in \mathcal{B}(X^\cap, Y^\cup, I)$; omitted due to lack of space. □

Theorem 3 does not, however, say anything about the number of fixed points which are needed as factors for I. The next theorem says that taking fixed points as factors gives us the least number possible.

Theorem 4 (optimality) *Let $I = A \triangleleft B$ for $n \times k$ and $k \times m$ binary matrices A and B. Then there exists a set $F \subseteq \mathcal{B}(X^\cap, Y^\cup, I)$ of fixed points with*

$$|F| \leq k$$

such that for the $n \times |F|$ and $|F| \times m$ binary matrices $A_{c(F)}$ and $B_{c(F)}$ we have

$$I = A_{c(F)} \triangleleft B_{c(F)}.$$

Proof. Let $I = A \triangleleft B$. Due to Theorem 1, I is an intersection of I-beams $J_1, \ldots, J_k$ which correspond to the columns and rows of A and B, respectively, and cover I. Clearly, every J_l contains some minimal I-beam J'_l. Denote by C_l and D_l the subsets of X and Y for which $J'_l = c(C_l)^T \triangleleft c(D_l)$. By Theorem 2, $\langle C_l, D_l\rangle$s are fixed points, i.e. $\langle C_l, D_l\rangle \in \mathcal{B}(X^\cap, Y^\cup, I)$. Put $F = \{\langle C_l, D_l\rangle \mid 1 \leq l \leq k\}$. Using assumption and Theorem 1 and the fact that I is the intersection of the collection of all I-beams corresponding to fixed points, we get

$$\begin{aligned} I &= \bigwedge_{l=1}^{k} J_l \geq \bigwedge_{l=1}^{k} c(C_l)^T \triangleleft c(D_l) = \\ & A_{c(F)} \triangleleft B_{c(F)} \geq \\ & \bigwedge_{\langle C,D\rangle \in \mathcal{B}(X^\cap, Y^\cup, I)} c(C)^T \triangleleft c(D) = I. \end{aligned}$$

Therefore, $A_{c(F)} \triangleleft B_{c(F)} = I$. Observing that $|F| \leq k$ finishes the proof. □

Example 2 Consider again the $\triangleleft$-decomposition

$$\begin{pmatrix} 0&0&1&1&1\\ 0&0&1&1&0\\ 0&0&0&0&1\\ 0&1&1&1&0 \end{pmatrix} = \begin{pmatrix} 1&0&0&1\\ 1&0&1&0\\ 1&1&0&0\\ 0&0&1&0 \end{pmatrix} \triangleleft \begin{pmatrix} 0&0&1&1&1\\ 1&1&0&0&1\\ 0&1&1&1&0\\ 1&0&1&1&1 \end{pmatrix},$$

and the corresponding I-beams $J^1, \ldots, J^4$, which are

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

Furthermore, consider fixed points $\langle C_1, D_1\rangle = \langle\{1,2,3\},\{3,4,5\}\rangle$, $\langle C_2, D_2\rangle = \langle\{3\},\{5\}\rangle$, $\langle C_3, D_3\rangle = \langle\{2,4\},\{2,3,4\}\rangle$, of I. Then, each of the I-beams J^ls ($l = 1, \ldots, 4$) contains some of the minimal I-beams corresponding to $\langle C_1, D_1\rangle$, $\langle C_2, D_2\rangle$, or $\langle C_3, D_3\rangle$. Putting now $\mathcal{F} = \{\langle C_1, D_1\rangle, \langle C_2, D_2\rangle, \langle C_3, D_3\rangle\}$, we have $I = A_{\mathcal{F}} \triangleleft B_{\mathcal{F}}$. Denoting by $(A_{\mathcal{F}})_{\_l}$ and $(B_{\mathcal{F}})_{l\_}$ the l-th column of $A_{\mathcal{F}}$ and the l-th row of $B_{\mathcal{F}}$, $I = A_{\mathcal{F}} \triangleleft B_{\mathcal{F}}$ can further be rewritten as $I = (A_{\mathcal{F}})_{\_1} \triangleleft (B_{\mathcal{F}})_{1\_} \wedge (A_{\mathcal{F}})_{\_2} \triangleleft (B_{\mathcal{F}})_{2\_} \wedge (A_{\mathcal{F}})_{\_3} \triangleleft (B_{\mathcal{F}})_{3\_}$, which shows a $\bigwedge$-decomposition of I into minimal I-beams covering I. With our example, we have

$$\begin{pmatrix} 0&0&1&1&1\\ 0&0&1&1&0\\ 0&0&0&0&1\\ 0&1&1&1&0 \end{pmatrix} = \begin{pmatrix} 0&0&1&1&1\\ 0&0&1&1&1\\ 0&0&1&1&1\\ 1&1&1&1&1 \end{pmatrix} \wedge \begin{pmatrix} 1&1&1&1&1\\ 1&1&1&1&1\\ 0&0&0&0&1\\ 1&1&1&1&1 \end{pmatrix} \wedge \begin{pmatrix} 1&1&1&1&1\\ 0&1&1&1&0\\ 1&1&1&1&1\\ 0&1&1&1&0 \end{pmatrix}.$$

Remark 1 Theorems 2, 3, and 4 have their counterparts for the case of $\triangleright$-decomposition. They say that fixed points of $\langle^\wedge, ^\vee\rangle$ are minimal H-beams containing I, that for every I, there is $F \subseteq \mathcal{B}(X^\wedge, Y^\vee, I)$ such that for the corresponding $\mathcal{F} = c(F)$ we have $I = A_\mathcal{F} \triangleright B_\mathcal{F}$, and that for every decomposition $I = A \triangleright B$, there is at least as good a decomposition which uses fixed points from $\mathcal{B}(X^\wedge, Y^\vee, I)$, i.e. minimal H-beams, as factors. We omit details.

Theorem 4 says that the best decomposition $I = A \triangleleft B$ is within the class of decompositions which use fixed points from $\mathcal{B}(X^\cap, Y^\cup, I)$ as factors. In this sense, $\mathcal{B}(X^\cap, Y^\cup, I)$ (the space of all minimal I-beams covering I) provides us with an optimal space of factors. This space is a subspace of the space of all possible factors (the space of all I-beams covering I, cf. Theorem 1).

2.3 Algorithm for optimal decompositions

We are going to present an approximation algorithm for finding an optimal decomposition $I = A \triangleleft B$ which uses minimal I-beams covering I as conditions/factors.

Observe first a connection between the problem of decomposition to a dual version of the well-known set covering optimization problem for which we refer to [5]. The dual problem which is important for our purposes can be formulated as follows. We are given a set $\mathcal{U}$ and a collection $\mathcal{S}$ of supersets of $\mathcal{U}$ whose intersection is $\mathcal{U}$, i.e. $\bigcap \mathcal{S} = \mathcal{U}$. The task is to find the smallest subcollection $\mathcal{C}$ of $\mathcal{S}$ with $\bigcap \mathcal{C} = \mathcal{U}$. It is easily seen that this problem is reducible to the set covering optimization problem and *vice versa*. The set covering problem (hence, the dual problem as well) is NP-hard but there exists an efficient greedy approximation algorithm which achieves an approximation ratio $\leq \ln(|\mathcal{U}|) + 1$, see [5]. A straightforward modification of this algorithm gives us an approximation algorithm for the dual problem with the same approximation ratio. Notice that the problem of finding an optimal set of conditions is reducible to the dual version of the set covering optimization problem. Indeed, put $\mathcal{U} = \{\langle x_i, y_j\rangle \mid I_{ij} = 1\}$ and $\mathcal{S} = \{C \triangleleft D \mid \langle C, D\rangle \in \mathcal{B}(X^\cap, Y^\cup, I)\}$. That is, $\mathcal{U}$ is the set of all pairs for which the corresponding entry I_{ij} is 1, and $\mathcal{S}$ is the set of all I-beams corresponding to fixed points from $\mathcal{B}(X^\cap, Y^\cup, I)$. Then, due to Theorem 1, the solution to the dual problem is the smallest set $F \subseteq \mathcal{B}(X^\cap, Y^\cup, I)$ for which $I = A_{c(F)} \triangleleft B_{c(F)}$.

The algorithm first computes all the fixed points from $\mathcal{B}(X^\cap, Y^\cup, I)$ and then finds a small set $F \subseteq \mathcal{B}(X^\cap, Y^\cup, I)$ of factors, i.e. $I = A_{c(F)} \triangleleft B_{c(F)}$. The algorithm is based on two observations regarding the computation of $\mathcal{B}(X^\cap, Y^\cup, I)$ and selecting F.

Step 1: computing minimal I-beams The problem of computing the fixed points from $\mathcal{B}(X^\cap, Y^\cup, I)$ can be reduced to the problem of computing fixed points of a particular closure operator. Namely, one can check the following observation.

Observation 1 *The operator* $^{\cap\cup} : 2^X \to 2^X$ *sending* $C \subseteq X$ *to* $(C^\cap)^\cup$ *is a closure operator. That is,* $C \subseteq C^{\cap\cup}$*;* $C_1 \subseteq C_2$ *implies* $C_1^{\cap\cup} \subseteq C_2^{\cap\cup}$*; and* $C^{\cap\cup} = (C^{\cap\cup})^{\cap\cup}$.

Furthermore, $\mathcal{B}(X^\cap, Y^\cup, I)$ can be recovered from the fixed points of $^{\cap\cup}$, i.e. from sets $C \subseteq X$ such that $C = C^{\cap\cup}$:

Observation 2 $\mathcal{B}(X^\cap, Y^\cup, I) = \{\langle C, C^\cap\rangle \mid C = C^{\cap\cup}\}$.

Fixed points of closure operators (called also closed sets in the literature) can efficiently be computed by several algorithms, see e.g. [7, 11].

Step 2: selecting factors from minimal I-beams Theorem 1 and the above considerations allow us to see that the problem of finding a smallest set $F \subseteq \mathcal{B}(X^\cap, Y^\cup, I)$ such that $I = A_{cF} \triangleleft B_{cF}$, i.e. an optimal set of factors, is reducible to a problem which is dual to the well-known set covering optimization problem. For the set covering optimization problem we refer to [5]. The dual problem which is important for our purposes can be formulated as follows. We are given a set $\mathcal{U}$

and a collection $\mathcal{S}$ of supersets of $\mathcal{U}$ whose intersection is $\mathcal{U}$, i.e. $\bigcap \mathcal{S} = \mathcal{U}$. The task is to find the smallest subcollection $\mathcal{C}$ of $\mathcal{S}$ with $\bigcap \mathcal{C} = \mathcal{U}$. It is easily seen that this problem is reducible to the set covering optimization problem and *vice versa*. The set covering problem (hence, the dual problem as well) is NP-hard but there exists an efficient greedy approximation algorithm which achieves an approximation ratio $\leq \ln(|\mathcal{U}|) + 1$, see [5]. A straightforward modification of this algorithm gives us an approximation algorithm for the dual problem with the same approximation ratio.

Observe now that the problem of finding an optimal set of factors is reducible to the dual to the set covering optimization problem. Indeed, put $\mathcal{U} = \{\langle x_i, y_j\rangle \mid I_{ij} = 1\}$ and $\mathcal{S} = \{C \lhd D \mid \langle C, D\rangle \in \mathcal{B}(X^\cap, Y^\cup, I)\}$. That is, $\mathcal{U}$ is the set of all pairs for which the corresponding entry I_{ij} is 1, and $\mathcal{S}$ is the set of all I-beams corresponding to fixed points from $\mathcal{B}(X^\cap, Y^\cup, I)$. Then, due to Theorem 1, the solution to the dual problem is the smallest set $F \subseteq \mathcal{B}(X^\cap, Y^\cup, I)$ for which $I = A_{c(F)} \lhd B_{c(F)}$.

Theorem **??** allows us to speed up the algorithm for finding an optimal set F by inserting the mandatory factors from $\mathcal{M}(X, Y, I)$ to F in the beginning. Putting together the above observations results in the following greedy approximation algorithm for computing the factors.

Algorithm 1

INPUT: binary matrix I
OUTPUT: $F \subseteq \mathcal{B}(X^\cap, Y^\cup, I)$ such that $I = A_{c(F)} \lhd B_{c(F)}$
$\mathcal{S} \leftarrow \mathcal{B}(X^\cap, Y^\cup, I)$
$\mathcal{U} \leftarrow \{\langle i, j\rangle \mid I_{ij} = 0\}$
$F \leftarrow \mathcal{M}(X, Y, I)$
remove every $\langle C, D\rangle \in \mathcal{M}(X, Y, I)$ from $\mathcal{S}$
for $\langle C, D\rangle \in \mathcal{M}(X, Y, I)$:
  for each $\langle i, j\rangle \in C \times (Y - D)$:
    remove $\langle i, j\rangle$ from $\mathcal{U}$
while $\mathcal{U} \neq \emptyset$:
  select $\langle C, D\rangle \in \mathcal{S}$ such that $|(C \times (Y - D)) \cap \mathcal{U}|$ is maximal
  add $\langle C, D\rangle$ to F
  remove $\langle C, D\rangle$ from $\mathcal{S}$
  for each $\langle i, j\rangle \in C \times (Y - D)$:
    remove $\langle i, j\rangle$ from $\mathcal{U}$
return F

The previous results prove correctness of Algorithm 1. Experiments are beyond the scope of this paper.

2.4 Illustrative example

In this section, we present an illustrative example of a $\lhd$-decomposition. Our main aim is to illustrate the notions and results introduced in previous sections.

We use a generic example related to the $\lhd$-type of decomposition. Suppose we have an $n \times m$ binary matrix I describing n objects and m attributes. The attributes correspond to roles/tasks which the objects can play/perform. Thus, $I_{ij} = 1$ indicates object i can play/perform role/task j, $I_{ij} = 0$ denotes the opposite. We want to see if there is a set of features (factors) using which we can interpret the data the following way: An object can play/perform a given role/task if and only if all the features displayed by the object are compatible with the given role/task. Such a data can result when suitability of objects to play/perform roles/tasks is assessed (by an expert, for instance) and the question arises whether we can find a rationale in the form of factors behind such assessment. In terms of the $\lhd$-product, we are looking for a decomposition of I into a $\lhd$-product $A \lhd B$ of binary matrices A and B with dimensions $n \times k$ and $k \times m$. A_{il} indicates object i has feature l, B_{lj} indicates feature l is compatible with role/task j. The factors to be discovered thus correspond to the features when looking for such interpretation of data.

In particular, consider the following 12×8 binary matrix:

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

We want to find a smallest set F set of fixed points from $\mathcal{B}(X^\cap, Y^\cup, I)$ such that $I =$

Table 1: Fixed points from $\mathcal{B}(X^{\cap}, Y^{\cup}, I)$.

c_i	$\langle C_i, D_i \rangle$
c_0	$\langle \{\}, \{\} \rangle$
c_1	$\langle \{1,5,9,11\}, \{4,6,7,8\} \rangle$
c_2	$\langle \{2,4,12\}, \{3,4,5,7\} \rangle$
c_3	$\langle \{3,6,7\}, \{1,3,4,6,8\} \rangle$
c_4	$\langle \{3,6,7,8,10\}, \{1,2,3,4,5,6,8\} \rangle$
c_5	$\langle \{1,3,5,6,7,9,11\}, \{1,3,4,6,7,8\} \rangle$
c_6	$\langle \{1,2,4,5,9,11,12\}, \{3,4,5,6,7,8\} \rangle$
c_7	$\langle \{1,2,3,4,5,6,7,9,11,12\},$ $\{1,3,4,5,6,7,8\} \rangle$
c_8	$\langle \{1,2,3,4,5,6,7,8,9,10,11,12\},$ $\{1,2,3,4,5,6,7,8\} \rangle$

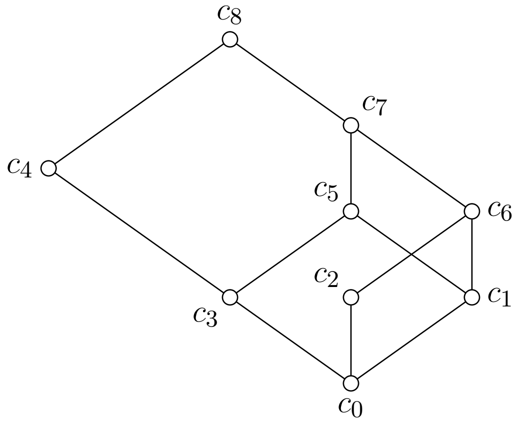

Figure 1: Hase diagram of lattice of fixed points.

$A_{c(F)} \triangleleft B_{c(F)}$. $\mathcal{B}(X^{\cap}, Y^{\cup}, I)$ contains 9 fixed points which we denote by $c_0, \ldots, c_8$. They are shown in Tab. 1. The Hasse diagram of the partially ordered set of fixed points from $\mathcal{B}(X^{\cap}, Y^{\cup}, I)$, ordered by $\langle C, D \rangle \leq \langle E, F \rangle$ iff $C \subseteq E$, is depicted in Fig. 1. According to the previous results, the fixed points correspond to minimal I-beams covering I. Let us now look at the problem of factorization of I. First, one can verify that $\mathcal{O}(X, Y, I) = \{c_1, c_2, c_3, c_4\}$ and $\mathcal{A}(X, Y, I) = \{c_0, c_1, c_2, c_4, c_5, c_6, c_7\}$. Thus, $\mathcal{O}(X, Y, I) \cap \mathcal{A}(X, Y, I) = \{c_1, c_2, c_4\}$. According to Theorem **??**, F needs to contain $\mathcal{O}(X, Y, I) \cap \mathcal{A}(X, Y, I) = \{c_1, c_2, c_4\}$. Then, $F'' = \{c_1, c_2, c_4\}$ is almost a set of factor concepts. One can check that there are just two minimal sets of factor concepts, namely

$$F = \{c_1, c_2, c_3, c_4\} \quad \text{and} \quad F' = \{c_1, c_2, c_4, c_5\}.$$

Thus, I can be decomposed by $I = A_{c(F)} \triangleleft B_{c(F)}$ or $I = A_{c(F')} \triangleleft B_{c(F')}$. For instance, we have

$$A_{c(F)} = \begin{pmatrix} 1&0&0&0\\ 0&1&0&0\\ 0&0&1&1\\ 0&1&0&0\\ 1&0&0&0\\ 0&0&1&1\\ 0&0&1&1\\ 0&0&0&1\\ 1&0&0&0\\ 0&0&0&1\\ 1&0&0&0\\ 0&1&0&0 \end{pmatrix}, \quad B_{c(F)} = \begin{pmatrix} 0&0&0&1&0&1&1&1\\ 0&0&1&1&1&0&1&0\\ 1&0&1&1&0&1&0&1\\ 1&1&1&1&1&1&0&1 \end{pmatrix}$$

and similarly for $A_{c(F')}$ and $B_{c(F')}$. Therefore, instead of an 8-dimensional space of roles/tasks, the objects are described in a 4-dimensional space of features with the above interpretation. For instance, object 2 has features 3 and 4, but does not have feature 1; features 2 and 4 are compatible with role/task 5, etc. A particular interpretation of the factors depends on the meaning of objects and roles/tasks.

3 Conclusions, Future Research

We presented a new way to decompose binary matrices into triangular products of binary matrices. The decomposition uses fixed points of particular operators as the elements of the inner space (factors). The geometrical insight provided here enables us to see that such decomposition is optimal in that it provides us with the least number of factors possible. Future research will include the following problems:

– Algorithms. The presented algorithm can be improved. Namely, it is not necessary to compute first the set of all fixed points. Instead, fixed points which are good factors can be computed in a similar way as in the algorithm for $\circ$-decomposition presented in [4].

– Approximate decomposition, i.e. looking for A and B such that I is approximately equal to $A \triangleleft B$.

– Restrictions on $\mathcal{F}$ such a suitably defined independence of factors.

– Can we use factors (i.e., new attributes) for more efficient reasoning and manipulation of objects?

– Extension from binary matrices to matrices containing more general entries, such as numbers from the unit interval $[0, 1]$, using results from formal concept analysis of data with fuzzy attributes, see e.g. [2, 3, 9, 14]. A paper on this topic is in preparation.

Acknowledgements

Supported by grant No. 1ET101370417 of GA AV ČR, by grant No. 201/05/0079 of the Czech Science Foundation, and by institutional support, research plan MSM 6198959214.

References

[1] Bandler W., Kohout L. J.: Mathematical relations. In: M. G. Singh *et al.* (Eds.): *International Encyclopedia of Systems and Control.* Pergamon Press, Oxford, 1985, pp. 4000–4008.

[2] Belohlavek R.: *Fuzzy Relational Systems: Foundations and Principles.* Kluwer, Academic/Plenum Publishers, New York, 2002.

[3] Belohlavek R.: Concept lattices and order in fuzzy logic. *Annals of Pure and Applied Logic* **128**(1–3)(2004), 277–298.

[4] Belohlavek R., Vychodil V.: Discovery of optimal factors in Boolean factor analysis via novel method of binary matrix decomposition (submitted to *J. Computer and System Sciences*). Extension of: On Boolean factor analysis with formal concepts as factors. SCIS & ISIS 2006, Tokyo, Japan, pp. 1054–1059.

[5] Cormen T. H., Leiserson C. E., Rivest R. L., Stein C.: *Introduction to Algorithms, 2nd Ed.* MIT Press, 2001.

[6] Frolov A. A., Húsek D., Muraviev I. P., Polyakov P. A.: Boolean factor analysis by Hopfield-like autoassociative memory. *IEEE Transactions on Neural Networks* Vol. **18**, No. 3, May 2007, pp. 698–707.

[7] Ganter B., Wille R.: *Formal Concept Analysis. Mathematical Foundations.* Springer, Berlin, 1999.

[8] F. Geerts, B. Goethals, T. Mielikäinen. Tiling Databases. DS 2004, LNCS **3245**, pp. 278–289.

[9] Georgescu G., Popescu A.: Non-dual fuzzy connections. *Archive for Mathematical Logic* **43**(2004), 1009–1039.

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

[11] Kuznetsov S., Obiedkov S.: Comparing performance of algorithms for generating concept lattices. *J. Experimental and Theoretical Articial Intelligence* **14**(2–3)(2002), 189–216.

[12] P. Miettinen, T. Mielikäinen, A. Gionis, G. Das, H. Mannila. The Discrete Basis Problem. PKDD 2006, LNAI **4213**, pp. 335–346.

[13] D. S. Nau, G. Markowsky, M. A. Woodbury, D. B. Amos. A Mathematical Analysis of Human Leukocyte Antigen Serology. *Math. Biosciences* **40**(1978), 243–270.

[14] Popescu A.: A general approach to fuzzy concepts. *Mathematical Logic Quarterly* **50**(3)(2004), 1–17.

[15] J. Vaidya, V. Atluri, Q. Guo: The Role Mining Problem: Finding a Minimal Descriptive Set of Roles. *ACM Symposium on Access Control Models and Technologies,* June, 2007, pp. 175–184.