Factorizing Three-Way Binary Data with Triadic Formal Concepts*

Radim Belohlavek and Vilem Vychodil

Dept. Computer Science, Palacky University, Olomouc
17. listopadu 12, CZ-771 46 Olomouc, Czech Republic
{radim.belohlavek,vilem.vychodil}@upol.cz

Abstract. We present a problem of factor analysis of three-way binary data. Such data is described by a 3-dimensional binary matrix I, describing a relationship between objects, attributes, and conditions. The aim is to decompose I into three binary matrices, an object-factor matrix A, an attribute-factor matrix B, and a condition-factor matrix C, with a small number of factors. The difference from the various decomposition-based methods of analysis of three-way data consists in the composition operator and the constraint on A, B, and C to be binary. We present a theoretical analysis of the decompositions and show that optimal factors for such decompositions are provided by triadic concepts developed in formal concept analysis. Moreover, we present an illustrative example, propose a greedy algorithm for computing the decompositions.

1 Introduction

1.1 Problem Description

Recently, there has been a growing interest in methods for analysis of three-way and generally N-way data that are based on various matrix decompositions. [10] provides an up-to-date survey with 244 references, see also [5]. An N-way data is represented by an N-dimensional matrix, called also N-dimensional array, or N-dimensional tensor. 2-dimensional matrices are the ordinary matrices whose entries are indexed by two indices (rows and column), N-dimensional matrices have N-indices. Decompositions of N-dimensional matrices go back as far as to 1920s and have been studied in psychometrics since the 1940s [10].

We are concerned with decompositions of three-way binary data, i.e. data represented by a 3-dimensional matrix which is denoted by I in this paper and whose entries, denoted I_{ijt}, are either 0 or 1. The matrix entries are interpreted as follows (clearly, other interpretations are possible):

$$I_{ijt} = \begin{cases} 1 \text{ if object } i \text{ has attribute } j \text{ under condition } t, \\ 0 \text{ if object } i \text{ does not have attribute } j \text{ under condition } t. \end{cases} \quad (1)$$

* Supported by grant no. P202/10/0262 of the Czech Science Foundation and by grant no. MSM 6198959214.

R. Setchi et al. (Eds.): KES 2010, Part I, LNAI 6276, pp. 471–480, 2010.
© Springer-Verlag Berlin Heidelberg 2010

For example (cf. Section 3), objects correspond to students, attributes to student qualities, and conditions to courses passed by the students.

Our aim is to decompose I in a way similar to the one employed in Boolean factor analysis, see e.g. [3,7]. Recall that in Boolean factor analysis, a decomposition $I = A \circ B$, defined by $I_{ij} = \max_{l=1}^{k} A_{il} \cdot B_{lj}$, of an object-attribute binary matrix I is sought into an object-factor matrix A and a factor-attribute matrix B, with k (number of factors) as small as possible. $\circ$ is the well-known Boolean matrix multiplication. In our scenario, the goal is to decompose a 3-dimensional binary matrix I into a product $\circ(A, B, C)$ of three binary matrices, an object-factor matrix A, an attribute-factor matrix B, and a condition-factor matrix C with the number of factors as small as possible. The operator $\circ(\cdot,\cdot,\cdot)$, defined in Section 2, is a 3-dimensional analogue of Boolean matrix multiplication.

As a main contribution of this paper, we show that optimal decompositions (those with the least number of factors) are attained by using so-called triadic concepts [13,17] as factors. In Sections 3 and 4, we present a detailed illustrative example, basic complexity considerations, and a greedy approximation algorithm to compute the decompositions. Section 5 presents issues for future work. Due to lack of space, proofs are omitted.

1.2 Related Work

Decompositions of (2-dimensional) matrices and the related methods of data analysis, such as factor analysis (FA), principal component analysis (PCA), independent component analysis (ICA), singular value decomposition (SVD), and others have been studied for a long time. Recently, there has been a growing interest in two topics. On one hand, there is a growing interest in the methods for decomposition of N-dimensional matrices, see [5] and in particular [10] for a survey. The reason behind is that N-way data naturally appear in many fields including psychometrics, chemometrics, signal processing, computer vision, neuroscience, numerical analysis, and others. On the other hand, there is an interest in the methods for decomposition of data which is constrained in some way. An example is the nonnegative matrix factorization [12]. Such constraints can be seen as semantic constraints which help us interpret the results of decompositions. Particularly relevant to our paper is the work on decompositions of binary data. Several methods, including modifications of the methods designed originally for real-valued data, have been developed, see [16] for an overview. A particular role among them have the methods which decompose a binary matrix into a Boolean product of binary matrices, see e.g. [3,7,14]. Namely, as reported in [14], Boolean matrix decompositions can be interpreted in a straightforward way. The present paper can be seen as an extension of [3] in which we described optimal decompositions of binary matrices, provided theoretical results on various aspects of Boolean decompositions, and an efficient approximation algorithm. In this paper, we seek to extend these results to three-way data. Such an extension is not obvious because several useful properties from the case of two-way data (such as a simple duality due to Galois connection induced by the input matrix) are no more available in the case of three-way data.

2 Decomposition and Factors

2.1 Decomposition

Consider an $n \times m \times p$ binary matrix I with entries I_{ijt}. We are interested in decompositions of I into three binary matrices, an $n \times k$ object-factor matrix A with entries A_{ik}, an $m \times k$ attribute-factor matrix B with entries B_{jk}, an $p \times k$ condition-factor matrix A with entries C_{tk}, with respect to a ternary composition $\circ$ defined by

$$\circ(A, B, C)_{ijt} = \max_{l=1}^{k} A_{il} \cdot B_{jl} \cdot C_{tl}. \tag{2}$$

We look for $I = \circ(A, B, C)$ with the smallest number k of factors.

Remark 1. (1) For $p = 1$, the problem becomes the problem of decomposition of a binary matrix into a Boolean product of binary matrices.

(2) Due to lack of space, we do not include observations on the various ways of possible compositions of 3- and lower-dimensional binary matrices. For real-valued matrices, see [10].

2.2 Factors for Decomposition

We are going to show the role of so-called triadic concepts for the decompositions. Triadic concepts were introduced in formal concept analysis (FCA). We provide the preliminaries and refer to [8] (ordinary, or dyadic, FCA) and [13,17] (triadic FCA) for more information. Note that in FCA, one works with relations rather than binary matrices. Since the distinction between relations and binary matrices is only a formal one, we use I to denote both, an $n \times m \times p$ binary matrix and a ternary relation between sets X, Y, and Z, with $|X| = n$, $|Y| = m$, and $|Z| = p$. The correspondence is: $I_{ijt} = 1$ (matrix) iff $\langle x_i, y_j, z_t \rangle \in I$ (relation).

Preliminaries from Dyadic and Triadic FCA. A *formal context* (or *dyadic context*) is a triplet $\langle X, Y, I \rangle$ where X and Y are non-empty sets and I is a binary relation between X and Y, i.e. $I \subseteq X \times Y$. X and Y are interpreted as the sets of objects and attributes, respectively; I is interpreted as the incidence relation ("to have relation"). That is, $\langle x, y \rangle \in I$ is interpreted as: object x has attribute y. A formal context $\mathbf{K} = \langle X, Y, I \rangle$ induces a pair of operators $^{\uparrow} : 2^X \to 2^Y$ and $^{\downarrow} : 2^Y \to 2^X$ defined for $C \subseteq X$ and $D \subseteq Y$ by

$$C^{\uparrow} = \{y \in Y \mid \text{for each } x \in C \colon \langle x, y \rangle \in I\},$$
$$D^{\downarrow} = \{x \in X \mid \text{for each } y \in D \colon \langle x, y \rangle \in I\}.$$

These operators, called *concept-forming operators*, form a Galois connection [8] between X and Y. Usually, there is no danger of misunderstanding and both $^{\uparrow}$ and $^{\downarrow}$ may be denoted by the same symbol, e.g. one uses C' and D' instead of $C^{\uparrow}$ and $D^{\downarrow}$. A *formal concept* (or *dyadic concept*) of $\langle X, Y, I \rangle$ is a pair $\langle C, D \rangle$ consisting of sets $C \subseteq X$ and $D \subseteq Y$ such that $C^{\uparrow} = D$ and $D^{\downarrow} = C$; C and D

are called the *extent* and *intent* of $\langle C, D\rangle$. The collection of all formal concepts of $\langle X, Y, I\rangle$ is denoted by $\mathcal{B}(X, Y, I)$ and is called the *concept lattice* of $\langle X, Y, I\rangle$. That is,

$$\mathcal{B}(X, Y, I) = \{\langle C, D\rangle \mid C^{\uparrow} = D,\ D^{\downarrow} = C\}.$$

A concept lattice equipped with a partial order corresponding to a subconcept-superconcept hierarchy is indeed a complete lattice [8]. A formal context may be visualized by a binary matrix: rows and columns correspond to objects and attributes; an entry corresponding to $x \in X$ and $y \in Y$ equals 1 iff $\langle x, y\rangle \in I$. Formal concepts of $\langle X, Y, I\rangle$ are just maximal rectangular areas in the corresponding binary matrix which are full of 1s [8].

A *triadic context* is a quadruple $\langle X_1, X_2, X_3, I\rangle$ where X_1, X_2, and X_3 are non-empty sets (interpreted as the sets of objects, attributes, and conditions, respectively), and I is a ternary relation between X_1, X_2, and X_3. I is interpreted as the incidence relation ("to have-under relation"). That is, $\langle x, y, z\rangle \in I$ is interpreted as: object x has attribute y under condition z. For convenience, a triadic context is denoted by $\langle X_1, X_2, X_3, I\rangle$. A triadic context $\mathbf{K} = \langle X_1, X_2, X_3, I\rangle$ induces the following dyadic contexts: $\mathbf{K}^{(1)} = \langle X_1, X_2 \times X_3, I^{(1)}\rangle$, $\mathbf{K}^{(2)} = \langle X_2, X_1 \times X_3, I^{(2)}\rangle$, $\mathbf{K}^{(3)} = \langle X_3, X_1 \times X_2, I^{(3)}\rangle$, with the binary relations $I^{(1)}$, $I^{(2)}$, and $I^{(3)}$ defined by

$$\langle x_1, \langle x_2, x_3\rangle\rangle \in I^{(1)} \text{ iff } \langle x_2, \langle x_1, x_3\rangle\rangle \in I^{(2)} \text{ iff } \langle x_3, \langle x_1, x_2\rangle\rangle \in I^{(3)} \text{ iff } \langle x_1, x_2, x_3\rangle \in I.$$

for every $x_1 \in X_1$, $x_2 \in X_2$, $x_3 \in X_3$. The concept-forming operators induced by $\mathbf{K}^{(i)}$ are denoted by ${}^{(i)}$. That is, for $C \subseteq X_1$ and $D \subseteq X_2 \times X_3$, we have

$$C^{(1)} = \{\langle x_2, x_3\rangle \in X_2 \times X_3 \mid \text{for each } x_1 \in C\colon \langle x_1, x_2, x_3\rangle \in I\},$$
$$D^{(1)} = \{x_1 \in X_1 \mid \text{for each } \langle x_2, x_3\rangle \in D\colon \langle x_1, x_2, x_3\rangle \in I\},$$

and similarly for ${}^{(2)}$ and ${}^{(3)}$. A *triadic concept* of $\langle X_1, X_2, X_3, I\rangle$ is a triplet $\langle D_1, D_2, D_3\rangle$ of $D_1 \subseteq X_1$, $D_2 \subseteq X_2$, and $D_3 \subseteq X_3$, such that for every $\{i, j, k\} = \{1, 2, 3\}$ with $j < k$ we have $D_i = (D_j \times D_k)^{(i)}$; D_1, D_2, and D_3 are called the *extent*, *intent*, and *modus* of $\langle D_1, D_2, D_3\rangle$. The set of all triadic concepts of $\langle X_1, X_2, X_3, I\rangle$ is denoted by $\mathcal{T}(X_1, X_2, X_3, I)$ and is called the concept trilattice of $\langle X_1, X_2, X_3, I\rangle$; the reader is referred to [17] to details on the notion of a trilattice and for the trilattice structure on $\mathcal{T}(X_1, X_2, X_3, I)$.

Triadic concepts can be represented by particular 3-dimensional binary matrices, namely cuboidal matrices. Formally, J is a cuboidal matrix (shortly, a cuboid) if there exist an $n \times 1$ binary vector A, an $m \times 1$ binary vector B, and a $p \times 1$ binary vector C, such that $J = \circ(A, B, C)$.

The following theorem explains the role of cuboids for decompositions (2).

Theorem 1. *$I = \circ(A, B, C)$ for an $n \times k$ matrix A, $m \times k$ matrix B, and $p \times k$ matrix C, if and only if is a* max*-superposition of k cuboids $J_1, \ldots, J_k$, i.e.*

$$I = J_1 \max \cdots \max J_k.$$

For each $l = 1, \ldots, k$, $J_l = \circ(A_{\_l}, B_{\_l}, C_{\_l})$, i.e. each J_l is the product of the l-th columns of A, B, and C.

As a result, to decompose I using a small number of factors, one needs to find a small number of cuboids in I which are full of 1s and cover all the entries of I with 1s.

We say that a cuboid J is contained in I if $J_{ijt} \leq I_{ijt}$ for all i, j, t. As the following theorem shows, triadic concepts of I correspond to maximal cuboids contained in I.

Theorem 2. *$\langle D_1, D_2, D_3\rangle$ is a triadic concept of I if and only if $J = \circ(c(D_1), c(D_2), c(D_3))$ is a maximal cuboid contained in I (i.e., any other cuboid which is contained in I is also contained in J). Here, $c(D_i)$ denotes the characteristic vector of D_i, i.e. $c(D_i)(x) = 1$ iff $x \in D_i$.*

We are going to use triadic concepts of I for decompositions of I the following way. For a set

$$\mathcal{F} = \{\langle D_{11}, D_{12}, D_{13}\rangle, \ldots, \langle D_{k1}, D_{k2}, D_{k3}\rangle\}$$

of triadic concepts of I, we denote by $A_{\mathcal{F}}$ the $n \times k$ matrix in which the l-th column consists of the characteristic vector $c(D_{l1})$ of the extent D_{l1} of $\langle D_{l1}, D_{l2}, D_{l3}\rangle$, $B_{\mathcal{F}}$ the $m \times k$ matrix in which the l-th column consists of the characteristic vector $c(D_{l2})$ of the intent D_{l2} of $\langle D_{l1}, D_{l2}, D_{l3}\rangle$, $C_{\mathcal{F}}$ the $p \times k$ matrix in which the l-th column consists of the characteristic vector $c(D_{l3})$ of the modus D_{l3} of $\langle D_{l1}, D_{l2}, D_{l3}\rangle$. That is,

$$(A_{\mathcal{F}})_{il} = \begin{cases} 1 \text{ if } i \in (D_{l1}), \\ 0 \text{ if } i \notin (D_{l1}), \end{cases} (B_{\mathcal{F}})_{jl} = \begin{cases} 1 \text{ if } j \in (D_{l2}), \\ 0 \text{ if } j \notin (D_{l2}), \end{cases} (C_{\mathcal{F}})_{tl} = \begin{cases} 1 \text{ if } t \in (D_{l3}), \\ 0 \text{ if } t \notin (D_{l3}). \end{cases}$$

If $I = \circ(A_{\mathcal{F}}, B_{\mathcal{F}}, C_{\mathcal{F}})$, $\mathcal{F}$ can be seen as a set of factors which fully explain the data. In such a case, we call the triadic concepts from $\mathcal{F}$ *factor concepts*. Given I, our aim is to find a small set $\mathcal{F}$ of factor concepts.

Using triadic concepts of I as factors is intuitively appealing because triadic concepts are simple models of human concepts according to traditional logic approach [13]. In fact, factors are often called "(hidden) concepts" in the ordinary factor analysis. In addition, the extents, intents, and modi of the concepts, i.e. columns of $A_{\mathcal{F}}$, $B_{\mathcal{F}}$, and $C_{\mathcal{F}}$, have a straightforward interpretation: they represent the objects, attributes, and conditions to which the factor concept applies (see Section 3 for particular examples).

The next result says that triadic concepts of I are universal factors.

Theorem 3 (universality). *For every I there is $\mathcal{F} \subseteq \mathcal{T}(X, Y, I)$ such that $I = \circ(A_{\mathcal{F}}, B_{\mathcal{F}}, C_{\mathcal{F}})$.*

The following theorem may be considered the main result. It says that, as far as exact decompositions of I are concerned, triadic concepts are optimal factors in that they provide us with decompositions of I with the least number k of factors.

Theorem 4 (optimality). *If $I = \circ(A, B, C)$ for $n \times k$, $m \times k$, and $p \times k$ binary matrices A, B, and C, there exists a set $\mathcal{F} \subseteq \mathcal{T}(X, Y, I)$ of triadic concepts of I with $|\mathcal{F}| \leq k$ such that for the $n \times |\mathcal{F}|$, $m \times |\mathcal{F}|$, and $p \times |\mathcal{F}|$ matrices $A_{\mathcal{F}}$, $B_{\mathcal{F}}$, and $C_{\mathcal{F}}$ we have $I = \circ(A_{\mathcal{F}}, B_{\mathcal{F}}, C_{\mathcal{F}})$.*

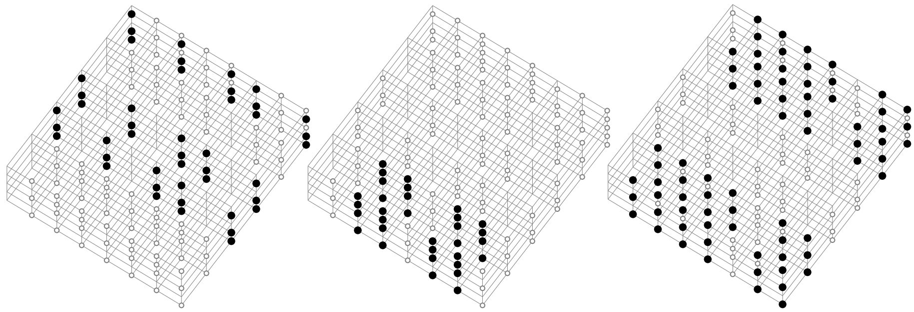

Fig. 1. Geometric meaning of factors as maximal cuboids

This means that when looking for decompositions of I, one can restrict the search to the set of triadic concepts instead of the set of all possible decompositions.

3 Illustrative Example

In this section, we present an illustrative example of factorization. We consider input data containing information about students and their performance in various courses. Such data is frequently obtained from student evaluation and recommendation systems that are used during the process of admission to universities. Factor analysis of this type of data can help reveal important factors describing skills of students under various conditions.

Our model data in represented by a triadic context $\langle X, Y, Z, I\rangle$ where $X = \{\mathsf{a}, \mathsf{b}, \dots, \mathsf{h}\}$ is a set of students (objects); $Y = \{\mathsf{co}, \mathsf{cr}, \mathsf{di}, \mathsf{fo}, \mathsf{in}, \mathsf{mo}\}$ is a set of student qualities (attributes): communicative, creative, diligent, focused, independent, motivated; and $Z = \{\mathsf{AL}, \mathsf{CA}, \mathsf{CI}, \mathsf{DA}, \mathsf{NE}\}$ is a set of courses passed by the students (conditions): algorithms, calculus, circuits, databases, and networking. The fact that x is related with y under z is interpreted so that "student x showed quality y in course z". We consider I given by the following table:

	AL						CA						CI						DA						NE					
	co	cr	di	fo	in	mo	co	cr	di	fo	in	mo	co	cr	di	fo	in	mo	co	cr	di	fo	in	mo	co	cr	di	fo	in	mo
a	1	1	1	1	1	1	0	0	1	1	0	1	1	1	0	0	1	1	0	0	1	1	0	1	1	1	0	0	1	1
b	1	1	0	0	1	1	0	0	0	0	0	0	1	1	0	0	1	1	1	1	0	0	0	0	1	1	0	0	1	1
c	1	1	1	1	0	1	0	0	1	1	0	1	1	1	0	0	0	0	1	1	1	1	0	1	1	1	0	0	0	0
d	1	1	1	1	1	1	0	0	1	1	0	1	1	1	0	0	1	1	0	0	1	1	0	1	1	1	0	0	1	1
e	1	1	0	0	1	1	0	0	0	0	0	0	1	1	0	0	1	1	1	1	0	0	0	0	1	1	0	0	1	1
f	1	1	1	1	1	1	0	0	1	1	0	1	1	1	0	0	1	1	1	1	1	1	0	1	1	1	0	0	1	1
g	1	1	0	0	1	1	0	0	0	0	0	0	1	1	0	0	1	1	0	0	0	0	0	0	1	1	0	0	1	1
h	0	0	1	1	0	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0

The rows of the table correspond to students, the columns correspond to attributes under the various conditions (courses). The triadic context $\langle X, Y, Z, I\rangle$ contains 14 triadic concepts:

$$\begin{aligned}
D_1 &= \langle \emptyset, \{\mathsf{co}, \mathsf{cr}, \mathsf{di}, \mathsf{fo}, \mathsf{in}, \mathsf{mo}\}, \{\mathsf{AL}, \mathsf{CA}, \mathsf{CI}, \mathsf{DA}, \mathsf{NE}\}\rangle,\\
D_2 &= \langle \{\mathsf{f}\}, \{\mathsf{co}, \mathsf{cr}, \mathsf{mo}\}, \{\mathsf{AL}, \mathsf{CI}, \mathsf{DA}, \mathsf{NE}\}\rangle,\\
D_3 &= \langle \{\mathsf{c}, \mathsf{f}\}, \{\mathsf{co}, \mathsf{cr}, \mathsf{di}, \mathsf{fo}, \mathsf{mo}\}, \{\mathsf{AL}, \mathsf{DA}\}\rangle,\\
D_4 &= \langle \{\mathsf{b}, \mathsf{c}, \mathsf{e}, \mathsf{f}\}, \{\mathsf{co}, \mathsf{cr}\}, \{\mathsf{AL}, \mathsf{CI}, \mathsf{DA}, \mathsf{NE}\}\rangle,\\
D_5 &= \langle \{\mathsf{a}, \mathsf{d}, \mathsf{f}\}, \{\mathsf{mo}\}, \{\mathsf{AL}, \mathsf{CA}, \mathsf{CI}, \mathsf{DA}, \mathsf{NE}\}\rangle,\\
D_6 &= \langle \{\mathsf{a}, \mathsf{d}, \mathsf{f}\}, \{\mathsf{co}, \mathsf{cr}, \mathsf{di}, \mathsf{fo}, \mathsf{in}, \mathsf{mo}\}, \{\mathsf{AL}\}\rangle,\\
D_7 &= \langle \{\mathsf{a}, \mathsf{c}, \mathsf{d}, \mathsf{f}\}, \{\mathsf{co}, \mathsf{cr}, \mathsf{di}, \mathsf{fo}, \mathsf{mo}\}, \{\mathsf{AL}\}\rangle,\\
D_8 &= \langle \{\mathsf{a}, \mathsf{c}, \mathsf{d}, \mathsf{f}, \mathsf{h}\}, \{\mathsf{di}, \mathsf{fo}, \mathsf{mo}\}, \{\mathsf{AL}, \mathsf{CA}, \mathsf{DA}\}\rangle,\\
D_9 &= \langle \{\mathsf{a}, \mathsf{b}, \mathsf{d}, \mathsf{e}, \mathsf{f}, \mathsf{g}\}, \{\mathsf{co}, \mathsf{cr}, \mathsf{in}, \mathsf{mo}\}, \{\mathsf{AL}, \mathsf{CI}, \mathsf{NE}\}\rangle,\\
D_{10} &= \langle \{\mathsf{a}, \mathsf{b}, \mathsf{c}, \mathsf{d}, \mathsf{e}, \mathsf{f}, \mathsf{g}\}, \{\mathsf{co}, \mathsf{cr}\}, \{\mathsf{AL}, \mathsf{CI}, \mathsf{NE}\}\rangle,\\
D_{11} &= \langle \{\mathsf{a}, \mathsf{b}, \mathsf{c}, \mathsf{d}, \mathsf{e}, \mathsf{f}, \mathsf{g}\}, \{\mathsf{co}, \mathsf{cr}, \mathsf{mo}\}, \{\mathsf{AL}\}\rangle,\\
D_{12} &= \langle \{\mathsf{a}, \mathsf{b}, \mathsf{c}, \mathsf{d}, \mathsf{e}, \mathsf{f}, \mathsf{g}, \mathsf{h}\}, \emptyset, \{\mathsf{AL}, \mathsf{CA}, \mathsf{CI}, \mathsf{DA}, \mathsf{NE}\}\rangle,\\
D_{13} &= \langle \{\mathsf{a}, \mathsf{b}, \mathsf{c}, \mathsf{d}, \mathsf{e}, \mathsf{f}, \mathsf{g}, \mathsf{h}\}, \{\mathsf{mo}\}, \{\mathsf{AL}\}\rangle,\\
D_{14} &= \langle \{\mathsf{a}, \mathsf{b}, \mathsf{c}, \mathsf{d}, \mathsf{e}, \mathsf{f}, \mathsf{g}, \mathsf{h}\}, \{\mathsf{co}, \mathsf{cr}, \mathsf{di}, \mathsf{fo}, \mathsf{in}, \mathsf{mo}\}, \emptyset\rangle.
\end{aligned}$$

Following the observations from Section 2, it suffices to take $\mathcal{F} = \{D_1, \ldots, D_{14}\}$ as the set of factor concepts which then yields a factorization of I into an 8×14 object-factor matrix $A_{\mathcal{F}}$, a 6×14 attribute-factor matrix $B_{\mathcal{F}}$, and a 5×14 conditions-factor matrix $C_{\mathcal{F}}$. However, there exists a smaller set $\mathcal{F}$ of factor concepts consisting of

$$\begin{aligned}
F_1 = D_8 &= \langle \{\mathsf{a}, \mathsf{c}, \mathsf{d}, \mathsf{f}, \mathsf{h}\}, \{\mathsf{di}, \mathsf{fo}, \mathsf{mo}\}, \{\mathsf{AL}, \mathsf{CA}, \mathsf{DA}\}\rangle,\\
F_2 = D_4 &= \langle \{\mathsf{b}, \mathsf{c}, \mathsf{e}, \mathsf{f}\}, \{\mathsf{co}, \mathsf{cr}\}, \{\mathsf{AL}, \mathsf{CI}, \mathsf{DA}, \mathsf{NE}\}\rangle,\\
F_3 = D_9 &= \langle \{\mathsf{a}, \mathsf{b}, \mathsf{d}, \mathsf{e}, \mathsf{f}, \mathsf{g}\}, \{\mathsf{co}, \mathsf{cr}, \mathsf{in}, \mathsf{mo}\}, \{\mathsf{AL}, \mathsf{CI}, \mathsf{NE}\}\rangle.
\end{aligned}$$

If we fix the order of objects, attributes, and conditions in sets X, Y, and Z, respectively, we can uniquely represent subsets of object, attributes, and conditions by characteristic vectors. For instance, we let $\mathsf{a} < \mathsf{b} < \cdots < \mathsf{h}$, i.e., a has index 1, b has index 2, etc. Similarly, we assume $\mathsf{co} < \mathsf{cr} < \cdots < \mathsf{mo}$ and $\mathsf{AL} < \mathsf{CA} < \cdots < \mathsf{NE}$. As a consequence, extents, intents, and modi of F_1, F_2, F_3 can be represented by characteristic vectors as follows:

$$F_1 = \langle 10110101, 001101, 11010\rangle, \quad F_2 = \langle 01101100, 110000, 10111\rangle,$$
$$F_3 = \langle 11011110, 110011, 10101\rangle.$$

Using $\mathcal{F} = \{F_1, F_2, F_3\}$, we obtain the following 8×3 object-factor matrix $A_{\mathcal{F}}$, 6×3 attribute-factor matrix $B_{\mathcal{F}}$, and 5×3 conditions-factor matrix $C_{\mathcal{F}}$:

$$A_{\mathcal{F}} = \begin{pmatrix} 1\,0\,1\\ 0\,1\,1\\ 1\,1\,0\\ 1\,0\,1\\ 0\,1\,1\\ 1\,1\,1\\ 0\,0\,1\\ 1\,0\,0 \end{pmatrix}, \qquad B_{\mathcal{F}} = \begin{pmatrix} 0\,1\,1\\ 0\,1\,1\\ 1\,0\,0\\ 1\,0\,0\\ 0\,0\,1\\ 1\,0\,1 \end{pmatrix}, \qquad C_{\mathcal{F}} = \begin{pmatrix} 1\,1\,1\\ 1\,0\,0\\ 0\,1\,1\\ 1\,1\,0\\ 0\,1\,1 \end{pmatrix}.$$

One can check that $I = \circ(A_{\mathcal{F}}, B_{\mathcal{F}}, C_{\mathcal{F}})$, i.e., I decomposes into three (two-dimensional) matrices using three factors. Note that the meaning of the factors can be seen from the extents, intents, and modi of the factor concepts. For instance, F_1 applies to students $\mathsf{a}, \mathsf{c}, \mathsf{d}, \mathsf{f}, \mathsf{h}$ who are diligent, focused, and motivated in algorithms, calculus, and databases. This suggests that F_1 can be interpreted as "having good background in theory / formal methods". In addition, F_2 applies to students who are communicative and creative in algorithms, circuits, databases, and networking. This may indicate interests and skills in "practical subjects". Finally, F_3 can be interpreted as a factor close to "self-confidence" because it is manifestated by being communicative, creative, independent, and motivated. As a result, by finding the factors set $\mathcal{F} = \{F_1, F_2, F_3\}$, we have explained the structure of the input data set I using three factors which describe the abilities of student applicants in terms of their skills in various subjects.

Let us recall that the factor concepts $\mathcal{F} = \{F_1, F_2, F_3\}$ can be seen as maximal cuboids in I. Indeed, I itself can be depicted as three-dimensional box where the axes correspond to students, their qualities, and courses. Figure 1 shows the three factors depicted as cuboids. White and black circlets in Figure 1 correspond to elements in I. Namely, a circlet is present on the intersection of $x \in X$, $y \in Y$, and $z \in Z$ in the diagram iff $\langle x, y, z\rangle \in I$. Furthermore, the circlet is black iff $\langle x, y, z\rangle \in I$ belongs to the factor F_i which is iff x belongs to the extent of F_i, y belongs to the intent of F_i, and z belongs to the modus of F_i.

4 Algorithm

We now outline an algorithm for computing the matrix decompositions described in Section 2. Since the problem of finding a minimal decomposition of $\langle X, Y, Z, I\rangle$ is reduced to a problem of finding a minimal subset $\mathcal{F} \subseteq \mathcal{T}(X, Y, Z, I)$ of formal concepts which cover the whole set I, we can reduce the problem of finding a matrix decomposition to the set-covering problem. The universe U that should be covered corresponds to $I \subseteq X \times Y \times Z$. The family $\mathcal{S}$ of subsets of the universe U that is used for finding a cover is, in fact, the set of all triadic concepts $\mathcal{T}(X, Y, Z, I)$. More precisely, $\mathcal{S} = \{A \times B \times C \mid \langle A, B, C\rangle \in \mathcal{T}(X, Y, Z, I)\}$. In this setting, we are looking for $\mathcal{C} \subseteq \mathcal{S}$ as small as possible such that $\bigcup \mathcal{C} = U$. Thus, finding factor concepts is indeed an instance of the set-covering problem. The set covering optimization problem is NP-hard and the corresponding decision problem is NP-complete. However, there exists an efficient greedy approximation algorithm for the set covering optimization problem which achieves an approximation ratio $\leq \ln(|\mathcal{U}|) + 1$, see [6].

Algorithm 1, implementing the above-mentioned greedy approach in our setting, computes a set of factor concepts by first computing the set of all triadic concepts which are stored in $\mathcal{S}$, see lines 1–8, and then iteratively selecting formal concepts from $\mathcal{S}$, maximizing their overlap with the remaining tuples in U, see lines 9–17. Notice that the triadic concepts are computed by a reduction to the dyadic case [9]. In line 2, we iterate over all concepts in $\mathcal{B}\big(X, Y \times Z, I^X\big)$

Algorithm 1. ComputeFactors(X, Y, Z, I)

/* compute a set $\mathcal{S}$ of all triadic concepts */
1 **set** $\mathcal{S}$ **to** $\emptyset$;
2 **foreach** $\langle A, J\rangle \in \mathcal{B}\big(X, Y \times Z, I^X\big)$ **do**
3   **foreach** $\langle B, C\rangle \in \mathcal{B}(Y, Z, J)$ **do**
4   **if** $A = (B \times C)^{(X)}$ **then**
5   **add** $\langle A, B, C\rangle$ **to** $\mathcal{S}$;
6   **end**
7   **end**
8 **end**
/* compute a set $\mathcal{F}$ of factor concepts */
9 **set** $\mathcal{F}$ **to** $\emptyset$;
10 **set** U **to** I;
11 **while** $U \neq \emptyset$ **do**
12   **select** $\langle A, B, C\rangle \in \mathcal{S}$ **which maximizes** $|U \cap (A \times B \times C)|$;
13   **add** $\langle A, B, C\rangle$ **to** $\mathcal{F}$;
14   **set** U **to** $U \setminus (A \times B \times C)$;
15   **remove** $\langle A, B, C\rangle$ **from** $\mathcal{S}$;
16 **end**
17 **return** $\mathcal{F}$

where $I^X = \{\langle x, \langle y, z\rangle\rangle \mid \langle x, y, z\rangle \in I\}$, cf. $\mathbf{K}^{(1)}$ in Section 2. In line 3, we iterate over all concepts in $\mathcal{B}(Y, Z, J)$ where J was obtained as an intent in the previous line. The condition in line 4 is needed to check whether A is maximal, i.e., whether $\langle A, B, C\rangle$ is a triadic concept. Notice that [9] computes triadic concepts by an analogous reduction which utilizes two nested NextClosure algorithms, however, arbitrary algorithm for computing dyadic formal concepts can do the job (e.g., CbO, Lindig's algorithm), see [11] for a survey and comparison.

We have observed by experiments that Algorithm 1 computes nearly optimal sets of factor concepts in terms of their sizes. Because of the limited scope of the paper, we postpone detailed performance evaluation of the algorithm to a full version of the paper.

5 Further Issues

Future work will include the following topics: Algorithms and experiments (we proposed a greedy algorithm, based on the idea from our [3]; the algorithm need not compute all triadic concepts, instead it computes good factor concepts one by one directly from the data, resulting in a high speed-up; a paper is in preparation); approximate factorization (decompositions for which $\circ(A, B, C)$ approximately equals I); complexity and approximability of the problem of finding decompositions; extension to ordinal data (see [1,2,4] for the case of two-way data).

References

1. Belohlavek, R.: Optimal decompositions of matrices with grades. In: IEEE IS 2008, pp. 15-2–15-7 (2008)
2. Belohlavek, R.: Optimal triangular decompositions of matrices with entries from residuated lattices. Int. J. of Approximate Reasoning 50(8), 1250–1258 (2009)
3. Belohlavek, R., Vychodil, V.: Discovery of optimal factors in binary data via a novel method of matrix decomposition. J. Computer and System Sci. 76(1), 3–20 (2010)
4. Belohlavek, R., Vychodil, V.: Factor analysis of incidence data via novel decomposition of matrices. LNCS (LNAI), vol. 5548, pp. 83–97. Springer, Heidelberg (2009)
5. Cichocki, A., et al.: Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation. J. Wiley, Chichester (2009)
6. Cormen, T.H., et al.: Introduction to Algorithms, 2nd edn. MIT Press, Cambridge (2001)
7. Frolov, A.A., et al.: Boolean factor analysis by Hopfield-like autoassociative memory. IEEE Trans. Neural Networks 18(3), 698–707 (2007)
8. Ganter, B., Wille, R.: Formal Concept Analysis. Mathematical Foundations. Springer, Berlin (1999)
9. Jäschke, R., Hotho, A., Schmitz, C., Ganter, B., Stumme, G.: TRIAS – An Algorithm for Mining Iceberg Tri-Lattices. In: Proc. ICDM 2006, pp. 907–911 (2006)
10. Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. SIAM Review 51(3), 455–500 (2009)
11. Kuznetsov, S., Obiedkov, S.: Comparing performance of algorithms for generating concept lattices. J. Exp. Theor. Artif. Intelligence 14(2-3), 189–216 (2002)
12. Lee, D., Seung, H.: Learning the parts of objects by non-negative matrix factorization. Nature 401, 788–791 (1999)
13. Lehmann, F., Wille, R.: A triadic approach to formal concept analysis. In: Ellis, G., Rich, W., Levinson, R., Sowa, J.F. (eds.) ICCS 1995. LNCS, vol. 954, pp. 32–43. Springer, Heidelberg (1995)
14. Miettinen, P., Mielikäinen, T., Gionis, A., Das, G., Mannila, H.: The Discrete Basis Problem. In: Fürnkranz, J., Scheffer, T., Spiliopoulou, M. (eds.) PKDD 2006. LNCS (LNAI), vol. 4213, pp. 335–346. Springer, Heidelberg (2006)
15. Stockmeyer, L.J.: The set basis problem is NP-complete. IBM Research Report RC5431, Yorktown Heights, NY (1975)
16. Tatti, N., Mielikäinen, T., Gionis, A., Mannila, H.: What is the dimension of your binary data? In: Perner, P. (ed.) ICDM 2006. LNCS (LNAI), vol. 4065, pp. 603–612. Springer, Heidelberg (2006)
17. Wille, R.: The basic theorem of triadic concept analysis. Order 12, 149–158 (1995)