Optimal factorization of three-way binary data*

Radim Belohlavek and Vilem Vychodil
Department of 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, i.e. data described by a 3-dimensional binary matrix I, describing a relationship between objects, attributes, and conditions. The problem consists in finding a decomposition of I into 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 scenario is similar to that of decomposition-based methods of analysis of three-way data but the difference 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 and propose a greedy approximation algorithm for computing the decompositions.

1. Introduction

Problem Description Methods of analysis of two-way data, i.e. data described by matrices, based on various types of matrix decomposition methods represent an extensive field with applications in many domains. Recently, there has been a growing interest in analogous methods for analysis of three-way and generally N-way data. [7] provides an up-to-date survey with 244 references, see also [2, 8, 16]. Recall that 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. Interestingly, decompositions of N-dimensional matrices go back as far as to 1920s and have been studied in psychometrics since the 1940s (see [7] for historical account).

In this paper, we are concerned with decompositions of three-way binary data. Such data is 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, the presented results are not limited by this particular interpretation, other interpretations are possible):

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

In our illustrative example provided in Section 3, objects correspond to students, attributes to student qualities, and conditions to courses passed by the students. In general, conditions may represent time instances or events at which the objects and attributes have been observed.

Our aim is to decompose I in a way similar to the one employed in Boolean factor analysis, see e.g. [1, 4, 14]. 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 two binary matrices, an object-factor matrix A and a factor-attribute matrix B, with the inner dimension k (number of factors) as small as possible. The decomposition operator $\circ$ involved 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.

The paper presents basic results on the decompositions, particularly on optimal decompositions. We show that decomposing $I = \circ(A,B,C)$ may be interpreted as covering all the entries with 1s of I by cuboids which are full of 1s. Each such cuboid can be interpreted as a factor, which applies to certain objects, attributes, and conditions. Furthermore, we show that so-called triadic concepts [11, 20] studied in formal concept analysis [5] play a fundamental role for the decompositions. Namely, we show that triadic concepts are universal and optimal factors in that every I may be decom-

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

posed using factors which correspond to triadic concepts of I and that triadic concepts provide us with decompositions with the least number of factors. These results are presented in Section 2. In Section 3, we present an example which illustrates how the decompositions may be used to factor-analyze three-way binary data. Section 4 presents basic complexity considerations on the problem of computing decompositions with the smallest number of factors and a greedy.

Related Work Methods for matrix decompositions and dimensionality reduction are considered of fundamental importance in science and engineering [15, 19]. 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 [2, 8, 16] and in particular [7] 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 [10] in which a matrix with non-negative numbers is decomposed into two matrices, both with non-negative numbers. Such constraints can be seen as semantic constraints which help us interpret the results of decompositions. As an example, one of the advantages of non-negative matrix factorization is that the decomposition describes the original data as additively composed of its easily interpretable parts. 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 [18] for an overview and references. A particular role among them have the methods which decompose a binary matrix into a Boolean product of binary matrices, see e.g. [1, 4, 12, 14]. Namely, as reported in [13], while the methods which decompose a binary matrix into matrices with possibly negative real entries are difficult to interpret, Boolean matrix decompositions can be interpreted in a straightforward way. The present paper can be seen as an extension of [1] 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 are no more available in the case of three-way data.

2. Decomposition and Factors

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

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

Our aim is to find a decomposition $I = \circ(A,B,C)$ with the smallest number k of factors possible.

Remark 1. (1) One can see that if $p = 1$, the problem becomes the problem of decomposition of a binary matrix into a Boolean product of binary matrices with the smallest number of factors possible.

(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, this topic is covered in [7].

We are going to show the role of so-called triadic concepts for the decompositions. Triadic concepts were introduced in formal concept analysis (FCA) and we provide the preliminaries below. For further information we refer to [5] (ordinary, or dyadic, FCA) and [11, 20] (triadic FCA). 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 identify them. In particular, 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 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. $\mathbf{K} = \langle X,Y,I \rangle$ 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 [5] 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; the reader is referred to [5]. 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 $\langle x,y\rangle \in I$. Geometrically, formal concepts of $\langle X,Y,I\rangle$ are just maximal rectangular areas in the corresponding binary matrix which are full of 1s [5].

A *triadic context* is a quadruple $\langle X,Y,Z,I\rangle$ where X, Y, and Z are non-empty sets, and I is a ternary relation between X, Y, and Z, X, Y, and Z are interpreted as the sets of objects, attributes, and conditions, respectively. 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. In this case, we say that x,y,z (or x,z,y, or the result of listing x,y,z in any other sequence) are related by I. 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 $\langle x_1,\langle x_2,x_3\rangle\rangle \in I^{(1)}$ iff $\langle x_2,\langle x_1,x_3\rangle\rangle \in I^{(2)}$ iff $\langle x_3,\langle x_1,x_2\rangle\rangle \in I^{(3)}$ iff $\langle x_1,x_2,x_3\rangle \in I$ for every $x_1 \in X_1$, $x_2 \in X_2$, $x_3 \in X_3$. Moreover, for $\{i,j,k\} = \{1,2,3\}$ and $D_k \subseteq X_k$, we define a dyadic context

$$\mathbf{K}^{ij}_{D_k} = \langle X_i,X_j,I^{ij}_{D_k}\rangle$$

by $\langle x_i,x_j\rangle \in I^{ij}_{D_k}$ iff for each $x_k \in D_k$: x_i,x_j,x_k are related by I. The concept-forming operators induced by $\mathbf{K}^{(i)}$ are denoted by ${}^{(i)}$. That is, for $C \subseteq X_1$ and $D \subseteq X_2\times X_3$,

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

$$D^{(1)} = \{x_1 \in X_1 \,|\, \text{for each } \langle x_2,x_3\rangle \in D\colon \langle x_1,x_2,x_3\rangle \in I\},$$

and similarly for ${}^{(2)}$ and ${}^{(3)}$. The concept-forming operators induced by $\mathbf{K}^{ij}_{D_k}$ are denoted by ${}^{(i,j,C_k)}$. 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 [20] 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 3-dimensional binary matrices which are cuboidal matrices, i.e. in which after suitable permutations of the slices corresponding to the first, second, and third axis all the 1s form a cuboid. 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 (1).

Theorem 2. *$I = \circ(A,B,C)$ for an $n\times k$ matrix A, $m\times k$ matrix B, and $p\times k$ matrix C, iff I is a* max*-superposition of k cuboids $J_1,\dots,J_k$, i.e.*

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

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

Proof. By easy calculation. □

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 $J_{ijt} \le I_{ijt}$ for all i,j,t. As the following theorem shows, triadic concepts of I correspond to maximal cuboids contained in I.

Theorem 3. *$\langle D_1,D_2,D_3\rangle$ is a triadic concept of I iff $J = \circ(c(D_1),c(D_2),c(D_3))$ 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 $x \in D_i$.*

Proof. Follows from [20, Proposition 1] by a moment's reflection. □

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

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

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,

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

and analogously for $(C_{\mathcal{F}})_{tl}$. 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 [11]. 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 4 (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 5 (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 $I = \circ(A_{\mathcal{F}}, B_{\mathcal{F}}, C_{\mathcal{F}})$.*

Proof. Sketch: The proof is based on Theorem 2 and Theorem 3. Every decomposition may be looked at as a covering of 1s in I by cuboids contained in I. Every such cuboid is contained in a maximal cuboid contained in I. Now, the collection $\mathcal{F}$ of the thus obtained maximal cuboids may be used to form three matrices, $A_{\mathcal{F}}$, $B_{\mathcal{F}}$, and $C_{\mathcal{F}}$, for which $I = \circ(A_{\mathcal{F}}, B_{\mathcal{F}}, C_{\mathcal{F}})$. Clearly, $|\mathcal{F}| \leq k$. □

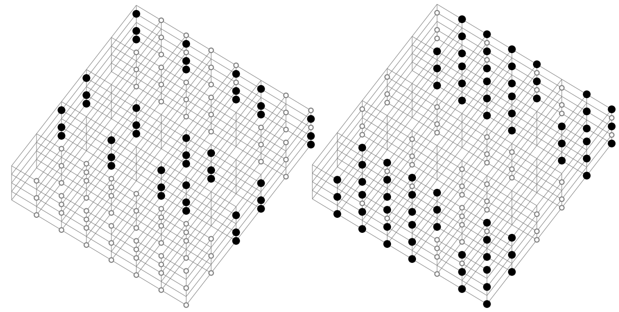
This means that when looking for decompositions of I, one can restrict the search to the set of triadic concepts instead of the set of all possible decompositions.

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 $X = \{\mathsf{a}, \mathsf{b}, \ldots, \mathsf{h}\}$ 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



F_1: "theoretical skills" F_3: "self-confidence"

Figure 1. Geometric meaning of factors.

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 $D_1, \ldots, D_{14}$. 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

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

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:

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

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 $C_{\mathcal{F}}$:

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

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 two of 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. Algorithms and Their Performance

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 $\leq \ln(|U|) + 1$, see [3]. This gives us a simple greedy-approach algorithm for computing all factor concepts.

The algorithm can be further modified to get a better performance in terms of the computation time. The main drawback of the algorithm is that it first computes a large set of triadic concepts (e.g., using an algorithm described in [6]) and then it selects (usually) a small subset of it being the set of factor concepts. Algorithm 1 overcomes this problem and computes a set of factor concepts directly without

Algorithm 1: COMPUTEFACTORS(X, Y, Z, I)

set $\mathcal{F}$ **to** $\emptyset$;
set U **to** I;
while $U \neq \emptyset$ **do**
  set A, A', B, B', C, C' **to** $\emptyset$;
  set *removed* **to** 0;
  repeat
    set $\langle A, B, C \rangle$ **to** $\langle A', B', C' \rangle$;
    set *removed* **to** $\mathrm{Size}(U, B, C)$;
    select $\langle B', C' \rangle \in \mathrm{Upd}(B, C)$ **which maximizes** $\mathrm{Size}(U, B', C')$;
    set $\langle A', B', C' \rangle$ **to** $\mathrm{Factor}(B', C')$;
  until *removed* $\geq \mathrm{Size}(U, B', C')$;
  add $\langle A, B, C \rangle$ **to** $\mathcal{F}$;
  set U **to** $U \setminus (A \times B \times C)$;
end
return $\mathcal{F}$

the need to compute all triadic concepts. The algorithm is based on the idea of incremental modification of a promising triadic concept by extending its intent and modus so that the concept covers as much of the remaining values in U as possible.

Algorithm 1 uses the following auxiliary sets. First, a set $\mathrm{Upd}(B, C) \subseteq 2^Y \times 2^Z$ which represents a set of pairs of attributes and conditions that result from B and C by adding a new attribute, a new condition, or both. Formally,

$$\mathrm{Upd}(B, C) = \mathrm{Upd}_Y(B, C) \cup \mathrm{Upd}_Z(B, C) \cup \mathrm{Upd}_{YZ}(B, C),$$

where

$$\begin{aligned}
\mathrm{Upd}_Y(B, C) &= \{B \cup \{y\} \mid y \in Y \setminus B\} \times \{C\}, \\
\mathrm{Upd}_Z(B, C) &= \{B\} \times \{C \cup \{z\} \mid z \in Z \setminus C\}, \\
\mathrm{Upd}_{YZ}(B, C) &= \{B \cup \{y\} \mid y \in Y \setminus B\} \times \{C \cup \{z\} \mid z \in Z \setminus C\}.
\end{aligned}$$

Each element $\langle B', C' \rangle \in \mathrm{Upd}(B, C)$ is used as an enlarged candidate for replacing B and C if a factor concept that corresponds to $\langle B', C' \rangle$ covers larger part of U that the current factor corresponding to $\langle B, C \rangle$. In order to give a precise meaning of the covered part of U, we introduce a number

$$\mathrm{Size}(U, B, C) = \max\bigl(\mathrm{Size}_{Y,Z}(U, B, C), \mathrm{Size}_{Z,Y}(U, B, C)\bigr),$$

where

$$\begin{aligned}
\mathrm{Size}_{Y,Z}(U, B, C) &= |U \cap (A^{Y,Z} \times B^{Y,Z} \times C^{Y,Z})|, \\
\mathrm{Size}_{Z,Y}(U, B, C) &= |U \cap (A^{Z,Y} \times B^{Z,Y} \times C^{Z,Y})|,
\end{aligned}$$

and

$$\begin{aligned}
A^{Y,Z} &= (B \times C)^{(X)}, & A^{Z,Y} &= (B \times C)^{(X)}, \\
B^{Y,Z} &= \bigl(A^{Y,Z} \times C\bigr)^{(Y)}, & B^{Z,Y} &= \bigl(A^{Z,Y} \times C^{Z,Y}\bigr)^{(Y)}, \\
C^{Y,Z} &= \bigl(A^{Y,Z} \times B^{Y,Z}\bigr)^{(Z)}, & C^{Z,Y} &= \bigl(A^{Z,Y} \times B\bigr)^{(Z)}.
\end{aligned}$$

It follows from [11, 20] that both $\langle A^{Y,Z}, B^{Y,Z}, C^{Y,Z} \rangle$ and

factors	50% approximation	95% approximation
4	$(1.766, 1.778)$	$(3.933, 3.940)$
6	$(1.995, 2.003)$	$(5.489, 5.504)$
8	$(2.274, 2.288)$	$(6.878, 6.891)$
10	$(2.524, 2.539)$	$(8.268, 8.287)$
12	$(2.693, 2.707)$	$(9.736, 9.760)$

Figure 2. Mean numbers of factors.

$\langle A^{Z,Y}, B^{Z,Y}, C^{Z,Y}\rangle$ represent triadic concepts which are, in our case, considered to be candidates for a factor concept. Thus, $\mathrm{Size}(U,B,C)$ measures the maximum overlap of such concepts with U. Finally, we introduce $\mathrm{Factor}(B,C)$ to denote one of the latter concepts having the greater overlap with U:

$$\mathrm{Factor}(B,C) = \begin{cases} \langle A^{Y,Z}, B^{Y,Z}, C^{Y,Z}\rangle, \\ \qquad \text{if } \mathrm{Size}(U,B,C) = \mathrm{Size}_{Y,Z}(U,B,C), \\ \langle A^{Z,Y}, B^{Z,Y}, C^{Z,Y}\rangle, \text{otherwise.} \end{cases}$$

It can be shown that Algorithm 1 is sound and produces a set of factor concepts.

Another important issue in factor analysis is *approximate factorizability,* i.e. the ability to find a set of concept factors $\mathcal{F}$ such that $\circ(A_{\mathcal{F}}, B_{\mathcal{F}}, C_{\mathcal{F}})$ equals I at least to a specified degree of approximation. By a degree of approximation we mean a ratio (usually specified in percents) given by

$$\frac{|A_{\mathcal{F}} \times B_{\mathcal{F}} \times C_{\mathcal{F}}|}{|I|}. \qquad (2)$$

As one can see, the ratio equals 1 (or 100%) iff $\mathcal{F}$ is a set of concept factors such that $I = \circ(A_{\mathcal{F}}, B_{\mathcal{F}}, C_{\mathcal{F}})$ (an exact decomposition). From the point of view of approximation, we might be interested in finding, given a ratio r, a set $\mathcal{F}$ of triadic concepts such that degree of approximation given by (2) is at least r. In other words, we are interested in factors which explain at least r% of the input data. Notice that Algorithm 1 can be easily modified to compute approximate factorizations by adding an additional parameter r and a new halting condition which stops looking for further factor concepts whenever the threshold value r has been reached.

We have performed a series of empirical experiments in order to explore the behavior of the exact and approximate decompositions using both algorithms. In this paper, we present an excerpt of our observations. We have taken a sample of 900 000 randomly generated triadic contexts with density 0.5 (exactly one half of matrix entries are ones) which were known to be decomposable using $4, 6, 7, 10$, and 12 factors. We were interested in getting means values of the numbers of factors that are needed to reach 50% and 95% degree of approximation, see (2). The results are shown in the tables in Fig. 2. Rows in the table correspond to matrices with known numbers of factors and columns correspond to approximation degrees. The table entries are 98% confidence intervals for the mean values of numbers of factors, each of them computed from a sample of 180 000 matrices. The tables show several interesting trends. For instance, the mean numbers of factors needed to explain 50% of the data are surprisingly low compared to the number of all factors, e.g., the mean number of factors in case of 12-decomposable matrices lies in the interval $(2.693, 2.707)$. Hence, if 12 factors explain the whole data, just 2.7 factors are needed (on average) to explain 50% of the data. Further performance evaluation of the algorithms is postponed to a full version of the paper.

References

[1] 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)(2010), 3–20.

[2] Cichocki A. *et al.*: *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation.* J. Wiley, 2009.

[3] Cormen T. H. *et al.*: *Introduction to Algorithms, 2nd Ed.* MIT Press, 2001.

[4] Frolov A. A. *et al.*: Boolean factor analysis by Hopfield-like autoassociative memory. *IEEE Trans. Neural Networks* **18**(3)(2007), 698–707.

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

[6] Jäschke R., Hotho A., Schmitz C., Ganter B., Stumme G.: TRIAS – An Algorithm for Mining Iceberg Tri-Lattices. *Proc. ICDM 2006*, pp. 907–911.

[7] Kolda T. G., Bader B. W.: Tensor decompositions and applications. *SIAM Review* **51**(3)(2009), 455–500.

[8] Kroonenberg P. M.: *Applied Multiway Data Analysis.* J. Wiley, 2008.

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

[10] Lee D., Seung H.: Learning the parts of objects by non-negative matrix factorization. *Nature* **401**(1999), 788–791.

[11] Lehmann F., Wille R.: A triadic approach to formal concept analysis. *Lecture Notes in Computer Science* **954**(1995), 32–43.

[12] Mickey M. R., Mundle P., Engelman L.: Boolean factor analysis. In: W.J. Dixon (Ed.), BMDP statistical software manual, vol. 2, 849–860, Berkeley, CA: University of California Press, 1990.

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

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

[15] Roweis S. T., Saul L. K.: Nonlinear dimensionality reduction by locally linear embedding. *Science* **290**(2000), 2323–2326.

[16] Smilde A., Bro R., Geladi P.: *Multi-way Analysis: Applications in the Chemical Sciences.* J. Wiley, 2004.

[17] Stockmeyer L. J.: The set basis problem is NP-complete. IBM Research Report RC5431, Yorktown Heights, NY, 1975.

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

[19] Tenenbaum J. B., de Silva V., Langford J. C.: A global geometric framework for nonlinear dimensionality reduction. *Science* **290**(2000), 2319–2323.

[20] Wille R.: The basic theorem of triadic concept analysis. *Order* **12**(1995), 149–158.