Relational Factor Analysis with ∘-Matrix Decomposition

Radim Belohlavek
Dept. Systems Science and Industrial Engineering
Binghamton University—SUNY
Binghamton, NY, 13902, U. S. A.
rbelohla@binghamton.edu
and
Palacky University, Olomouc, Czech Republic

Vilem Vychodil
Dept. Computer Science
Palacky University, Olomouc
Tomkova 40, Olomouc 77900, Czech Republic
vilem.vychodil@upol.cz

***Abstract*— The paper presents results on factorization of matrices describing objects and their fuzzy attributes. Entries of the matrices are truth degrees, e.g., numbers from the real unit interval $[0, 1]$. In general, matrix entries can be elements from a complete residuated lattice. We propose a novel method to factorize such matrices which is based on using so-called formal concepts as factors. To factorize an $n \times m$ object-attribute matrix I means to decompose I into a product $A \circ B$ of an $n \times k$ object-factor matrix A and an $k \times m$ factor-attribute matrix B. In addition, we want the number k of factors as small as possible. The product $\circ$ we consider in this paper is the well-known product corresponding to max-t-norm composition of fuzzy relations. We focus on theoretical analysis of the method we propose. We prove several results, e.g., a result which says that our method provides the best factorization in that it leads to the smallest number of factors. In addition, we present an illustrative example.**

I. Problem Setting

Factor analysis (FA) originated with Spearman's monumental development of psychological theory involving a single general factor and a number of specific factors [12]. Nowadays, FA is a well-established branch of statistical data analysis with many applications in numerous fields and with support in several software packages. According to Harman [9, p. 4], "The principal concern of factor analysis is the resolution of a set of variables linearly in terms of (usually) a small number of categories or 'factors'. … A satisfactory solution will yield factors which convey all the essential information of the original set of variables. Thus, the chief aim is to attain scientific parsimony or economy of description." Traditional FA attempts to decompose an input matrix I of dimension $n \times m$ into a product of matrices A of dimension $n \times k$ and B of dimension $k \times m$ such that k is less than m. I represents n objects (rows) and their m variables (columns), A represents description of the objects in terms of new (discovered) k factors and B represents the relationship of the factors to the original variables. Traditionally, the matrix product is the usual product as used in linear algebra. Decades ago, a problem of so-called Boolean factor analysis has been formulated, in which the matrix I is a Boolean matrix (contains 0's and 1's). By and large, the attempts to factorize Boolean matrices using traditional methods of FA failed. New methods have been studied to decompose a Boolean matrix into a Boolean product of two Boolean matrices, A and B. Most of these approaches were heuristics, providing no guarantee that decompositions be found. In [5], we proved that there exists an optimal way to factorize Boolean matrices. Namely, that one can use so-called formal concepts as the factors. In [3], we reformulated the problem of Boolean FA by generalizing it in that the input matrix may contain degrees such as 0.3, 0.6, etc., in addition to 0 and 1. We studied possibilities to factorize I by means of various fuzzy relational products and found optimal solutions in terms of various types of formal concepts associated to the input matrix I. In this paper, we focus in detail on the $\circ$-decomposition (max-t-norm decomposition) of matrix I. Our method heavily relies on an insight provided by results on fuzzy concept lattices and formal concept analysis of data with fuzzy attributes [1], [2], [4], and also [6]. In order to decompose I into $A \circ B$, we employ fuzzy concept lattices with hedges [4]. In particular, we present results which show that in order to decompose I into $A \circ B$, we can use formal concepts of the fuzzy concept lattice with hedges associated to I. Equation

FACTORS = FORMAL CONCEPTS (in sense of FCA)

is the central theme of our approach. The main result of the present paper is optimality of our approach: We prove that for any decomposition of I into $A \circ B$ with k factors there is a decomposition of I into $A' \circ B'$ with $k' \leq k$ with formal concepts as factors. Moreover, since formal concepts are easily interpretable, interpretability of factors is another advantage of our approach. In addition to that, the theoretical insight directly leads to an approach to find formal concept which can be used as factors. This approach is outlined in our paper, too.

In Section II, we provide a survey on fuzzy sets and fuzzy logic and on formal concept analysis of data with fuzzy attributes. In Section III, we present a detailed problem formulation. Section IV describes our main results. Section V contains illustrative examples. In Section VI, we present conclusions, remarks, and an outline of future research.

1-4244-1214-5/07/$25.00 ©2007 IEEE

II. PRELIMINARIES

a) Fuzzy sets and fuzzy logic: We assume that the reader is familiar with basic principles of fuzzy logic. We refer to [10] for general overview of fuzzy sets and fuzzy logic and to [1], [7] for the setting we use in this paper. We use complete residuated lattices as structures of truth degrees. A complete residuated lattice is an algebra $\mathbf{L} = \langle L, \wedge, \vee, \otimes, \rightarrow, 0, 1\rangle$ such that $\langle L, \wedge, \vee, 0, 1\rangle$ is a complete lattice with 0 and 1 being the least and greatest element of L, respectively; $\langle L, \otimes, 1\rangle$ is a commutative monoid; $\otimes$ and $\rightarrow$ satisfy the adjointness property, i.e. $a \otimes b \leq c$ iff $a \leq b \rightarrow c$. A truth-stressing hedge on $\mathbf{L}$ (hedge, for short) is a unary operation * on L satisfying (i) $1^* = 1$, (ii) $a^* \leq a$, (iii) $(a \rightarrow b)^* \leq a^* \rightarrow b^*$, (iv) $a^{**} = a^*$, for all $a, b \in L$. Elements a of L are called truth degrees. $\otimes$ and $\rightarrow$ are (truth functions of) "fuzzy conjunction" and "fuzzy implication". Hedge * is a (truth function of) logical connective "very true", see [7], [8] for details. A common choice of $\mathbf{L}$ is a structure with $L = [0, 1]$ (unit interval), $\wedge$ and $\vee$ being minimum and maximum, $\otimes$ being a left-continuous t-norm, such as the Łukasiewicz, product, or minimum, with the corresponding $\rightarrow$, see [7]. Another common choice is to take a finite chain L with appropriate truth functions such as the Łukasiewicz or minimum. A special case of a (finite) complete residuated lattice is a two-element Boolean algebra (structure of truth degrees of classical logic). Two boundary cases of hedges are (i) identity, i.e. $a^* = a$ ($a \in L$); (ii) globalization: $a^* = 1$ if $a = 1$, $a^* = 0$ else. Given a structure $\mathbf{L}$ of truth degrees, we define usual notions: an $\mathbf{L}$-set (fuzzy set) A in universe U is a mapping $A\colon U \rightarrow L$, $A(u)$ being interpreted as "the degree to which u belongs to A". Let $\mathbf{L}^U$ denote the collection of all $\mathbf{L}$-sets in U. The operations with $\mathbf{L}$-sets are defined componentwise. For instance, intersection of $\mathbf{L}$-sets $A, B \in \mathbf{L}^U$ is an $\mathbf{L}$-set $A \cap B$ in U such that $(A \cap B)(u) = A(u) \wedge B(u)$ ($u \in U$). For fuzzy sets $A, B \in \mathbf{L}^U$, we put $A \subseteq B$ (A is contains in B) iff, for each $u \in U$, $A(u) \leq B(u)$. In the following we use well-known properties of residuated lattices and fuzzy structures which can be found in [1], [7].

b) Fuzzy concept lattices with hedges: In this section, we summarize basic notions of formal concept analysis (FCA) of data with fuzzy attributes. FCA is a method of exploratory data analysis. The input data consists of a data table describing a relationship between objects and attributes. The output of FCA consists of a hierarchically ordered collection of clusters. The clusters are called formal concepts and can be seen as natural concepts well-understandable and interpretable by humans. We refer the reader to [6] for FCA with binary attributes and to [1], [2], [4], [11] for FCA of data with fuzzy attributes.

A data table with fuzzy attributes can be represented by a triplet $\langle X, Y, I\rangle$ where X and Y are non-empty sets of objects (table rows) and attributes (table columns), and $I : X \times Y \rightarrow L$ is a fuzzy relation with $I(x, y)$ representing the degree to which object $x \in X$ has attribute $y \in Y$ (table entry corresponding to row x and column y). Let ^{*X} and ^{*Y} be hedges on $\mathbf{L}$. For $A \in \mathbf{L}^X$, $B \in \mathbf{L}^Y$ (i.e. A is a fuzzy set of objects, B is a fuzzy set of attributes), we define fuzzy sets $A^{\uparrow} \in \mathbf{L}^Y$ (fuzzy set of attributes), $B^{\downarrow} \in \mathbf{L}^X$ (fuzzy set of objects) by

$$A^{\uparrow}(y) = \textstyle\bigwedge_{x \in X} \big(A(x)^{*X} \rightarrow I(x, y)\big), \qquad (1)$$

$$B^{\downarrow}(x) = \textstyle\bigwedge_{y \in Y} \big(B(y)^{*Y} \rightarrow I(x, y)\big). \qquad (2)$$

We put

$$\mathcal{B}(X^{*X}, Y^{*Y}, I) = \{\langle A, B\rangle \in \mathbf{L}^X \times \mathbf{L}^Y \mid A^{\uparrow} = B,\, B^{\downarrow} = A\}$$

and define a partial order $\leq$ on $\mathcal{B}(X^{*X}, Y^{*Y}, I)$ by $\langle A_1, B_1\rangle \leq \langle A_2, B_2\rangle$ iff $A_1 \subseteq A_2$ (or, iff $B_2 \subseteq B_1$; both ways are equivalent). Operators $^{\downarrow}, ^{\uparrow}$ form so-called Galois connection with hedges, see [4]. $\langle \mathcal{B}(X^{*X}, Y^{*Y}, I), \leq\rangle$ is called a fuzzy concept lattice associated to $\langle X, Y, I\rangle$. Elements $\langle A, B\rangle$ of $\mathcal{B}(X^{*X}, Y^{*Y}, I)$ are naturally interpreted as concepts (clusters) hidden in the input data represented by I. Namely, $A^{\uparrow} = B$ and $B^{\downarrow} = A$ say that B is the collection of all attributes shared by all objects from A, and A is the collection of all objects sharing all attributes from B. Note that these conditions represent exactly the definition of a concept as developed in the so-called Port-Royal logic; A and B are called the extent and the intent of the concept $\langle A, B\rangle$, respectively, and represent the collection of all objects and all attributes covered by the particular concept. Furthermore, $\leq$ models the natural subconcept-superconcept hierarchy—concept $\langle A_1, B_1\rangle$ is a subconcept of $\langle A_2, B_2\rangle$ iff each object from A_1 belongs to A_2 (dually for attributes).

III. PROBLEM FORMULATION IN DETAIL

For our purpose, consider the following (somewhat simplified) formulation of the problem of factor analysis. Suppose Z is an $n \times m$ real-valued matrix with entries $Z_{ij} \in \mathbb{R}$ ($1 \leq i \leq n$, $1 \leq j \leq m$) describing the value of j-th variable on i-th object. The aim is to find a collection of other variables $f_1, \dots, f_k$, so called factors, such that the original variables can be approximately expressed by a linear combination of the factors. In more detail, the aim is to find an $n \times k$ matrix A and a $k \times m$ matrix F such that

$$Z \approx A \circ F,$$

i.e. Z is approximately equal to $A \circ F$ ($\circ$ is the usual matrix product). Each A's entry $A_{il} \in \mathbb{R}$, called a loading, represents value of l-th factor on i-th object; each F's entry $F_{lj} \in \mathbb{R}$ represents how j-th variable is manifested in l-th factor. An important feature of the classical model of factor analysis is its linearity: Z is (approximately equal to) the ordinary matrix product of A and F, i.e.

$$Z_{ij} \approx \sum_{l=1}^{k} A_{il} \cdot F_{lj},$$

i.e. each original variable is considered as a linear combination of new variables (factors). Moreover, one requires that the number of factors is smaller than the number of the original variables, i.e. $k < m$. Then, the collection of all factors can be considered as a collection of a relatively small number of variables representing latent features such that the directly observable features can be expressed in terms of the latent features. Making the latent features explicit is advantageous. First, the dimension of the model reduces (the objects are represented in a k-dimensional space rather than in an m-dimensional one). Second, the factors can be thought of as representing more fundamental (explanatory) features than those represented by the original variables.

Needless to say, the general idea of factor analysis, i.e. finding a relatively small set of latent features, is independent of the linearity of the model. That is to say, the original variables can be expressed by means of the factors also in other ways than by a linear combination. Our formulation of the problem of relational factor analysis follows. The input data is an $n \times m$ matrix I with entries I_{ij} ($1 \leq i \leq n$, $1 \leq j \leq m$) being elements of a support set L of a complete residuated lattice $\mathbf{L}$. We call such matrices L-valued matrices. Thus, in particular, I can be a $[0,1]$-valued matrix such as

$$\begin{pmatrix} 1 & .9 & 0 & .2 & .1 \\ 1 & .9 & .2 & .2 & .8 \\ .8 & .8 & .9 & .9 & .1 \\ 1 & .2 & .1 & .1 & 1 \end{pmatrix}.$$

I describes a relationship between n objects and m attributes. Each entry I_{ij} ($1 \leq i \leq n$, $1 \leq j \leq m$) represents a value to which object i has attribute j (level of expression of attribute j on object i). The aim is to find an $n \times k$ L-valued matrix A (i.e., $A_{ik} \in L$) and a $k \times m$ L-valued matrix B (i.e., $B_{kj} \in L$) such that

$$I \approx A \circ B,$$

where $\approx$ denotes "equal" (or "approximately equal") and $\circ$ denotes a $\circ$-composition based on the multiplication $\otimes$ of $\mathbf{L}$, i.e. $\otimes$ is a (truth function of) "fuzzy conjunction". In our paper, we consider exact decomposition of I, i.e. we are looking for $I = A \circ B$. That is, we want

$$I_{ij} = \bigvee_{l=1}^{k} A_{ik} \otimes B_{kj} \tag{3}$$

where $\bigvee$ denotes logical "or" (truth function of disjunction), i.e. $\bigvee$ denotes supremum in $\mathbf{L}$. Note also that $\bigvee$ is just maximum if truth degrees from L are linearly ordered which is the case, e.g., if we consider $[0,1]$-valued matrices. For example, if we consider $\min$ as our "fuzzy conjunction" (in other words, our structure $\mathbf{L}$ of truth degrees forms the well-known Gödel chain), the above matrix can be decomposed by

$$\begin{pmatrix} 1 & .9 & 0 & .2 & .1 \\ 1 & .9 & .2 & .2 & .8 \\ .8 & .8 & .9 & .9 & .1 \\ 1 & .2 & .1 & .1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & .2 & .8 \\ .8 & .9 & .1 \\ 0 & .1 & 1 \end{pmatrix} \circ \begin{pmatrix} 1 & .9 & 0 & .2 & .1 \\ .2 & .2 & 1 & 1 & 0 \\ 1 & .2 & .1 & .1 & 1 \end{pmatrix}.$$

Note that if $L = \{0,1\}$, which is a particular case of our setting, our problem becomes a well-known problem of Boolean factor analysis. In this case, one is looking for decompositions of Boolean matrices such as

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

Such a decomposition can be interpreted as follows. In addition to the n original objects and their m original binary attributes, the decomposition employs k new factors. Matrices A and B represent relationships between objects and factors and between factors and the original attributes, respectively. In particular, $A_{il} \in L$ represents a degree to which i-th object has l-th factor, i.e. a degree to which l-th factor is manifested on i-th object; $B_{lj} \in L$ represents a degree to which l-th factor subsumes j-th original attribute, i.e. a degree to which j-th variable is manifested on l-th factor. Therefore, instead of the classical factor analysis, relational factor analysis is suitable for presence/absence data with degrees of expression of attributes on objects. The factors, as in case of classical factor analysis, can be seen as representing latent features using which one can express both the original objects and the original (directly observable) attributes.

IV. Formal concepts as optimal factors

This section presents our approach to factorization of L-valued matrices. Conceptually, the approach is motivated by our earlier papers [3], [5]. We try to utilize formal concepts of particular concept lattices as factors. That is, we attempt to employ as factors certain formal concepts of concept lattices. The concept lattice we consider is a concept lattice associated to the input matrix I. Needless to say, with a slight abuse of notation, matrix I can be obviously identified with a data table $\langle X, Y, I \rangle$ with fuzzy attributes, with a set $X = \{x_1, \dots, x_m\}$ of objects corresponding to matrix rows and a set $Y = \{y_1, \dots, y_n\}$ of attributes corresponding to matrix columns. That is, $I(x_i, y_j)$ coincides with I_{ij} and, as a consequence, we do not distinguish between matrix I and data table $\langle X, Y, I \rangle$. Therefore, we can consider a concept lattice associated to $\langle X, Y, I \rangle$. In this paper, we consider concept lattices which are particular concept lattices with hedges, see Section II. Namely, we consider $\mathcal{B}(X^{*_X}, Y^{*_Y}, I)$ where ^{*_X} is identity and ^{*_Y} is an arbitrary hedge. We denote such concept lattices simply by $\mathcal{B}(X, Y^{*_Y}, I)$. This approach covers, as a particular case, the approach described in [3], where we discussed various types of decompositions of I including the $\circ$-decomposition for which we proposed to use ordinary fuzzy concept lattices $\mathcal{B}(X, Y, I)$, i.e. $\mathcal{B}(X^{*_X}, Y^{*_Y}, I)$ with both ^{*_X} and ^{*_Y} being identities.

The reason for considering $\mathcal{B}(X, Y^{*_Y}, I)$ in this paper is the following. Hedge ^{*_Y} enables us to select only particular formal concepts which we want to use as factor concepts. For instance, with ^{*_Y} being globalization, every concept $\langle A_i, B_i \rangle$ of $\mathcal{B}(X, Y^{*_Y}, I)$ can be interpreted as consisting of a fuzzy set A_i of objects and a (ordinary) set $B^{*_Y} = \{y \mid B(y) = 1\}$ of attributes. Such concepts tend to be more easily interpretable by users, see [4]. Moreover, the number of such concepts can be significantly smaller than the number of "unrestricted" concepts in $\mathcal{B}(X, Y, I)$. This results into a faster computation of factor concepts. In addition to that, we will see that $\mathcal{B}(X, Y^{*_Y}, I)$ provides us with optimal factor concepts restricted by requirement involving ^{*_Y}.

Up to now, we did not say what we mean by "using formal concepts as factors". This is what we are going to explain now.

Suppose $\mathcal{F} = \{\langle A_1, B_1\rangle, \dots, \langle A_k, B_k\rangle\} \subseteq \mathcal{B}(X, Y^{*_Y}, I)$ is a set of formal concepts of $\mathcal{B}(X, Y^{*_Y}, I)$. We define an $n \times k$ matrix $A_{\mathcal{F}}$ and an $k \times m$ matrix $B_{\mathcal{F}}$ by

$$(A_{\mathcal{F}})_{il} = A_l(x_i)$$

and

$$(B_{\mathcal{F}})_{lj} = B_l(y_j).$$

That is, the l-th column of $A_{\mathcal{F}}$ consists of membership degrees $A_l(x_i)$ of the extent of l-th concept $\langle A_l, B_l\rangle$ of $\mathcal{F}$, and the l-th row of $B_{\mathcal{F}}$ consists of membership degrees $B_l(y_j)$ of the intent of l-th concept $\langle A_l, B_l\rangle$ of $\mathcal{F}$. Our aim is to find, given matrix I, a small set $\mathcal{F} \subseteq \mathcal{B}(X, Y^{*_Y}, I)$ such that $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$.

Recall that $\mathcal{B}(X, Y, I)$ denotes $\mathcal{B}(X^{*_X}, Y^{*_Y}, I)$ with both ^{*_X} and ^{*_Y} being identities. Moreover, denote by $\text{fix}(^{*_Y})$ the set of fixed points of ^{*_Y}, i.e.

$$\text{fix}(^{*_Y}) = \{a \in L \mid a^{*_Y} = a\}.$$

We need the following theorem.

Theorem 1 ([4]): $\mathcal{B}(X, Y^{*_Y}, I) = \{\langle A, B\rangle \in \mathcal{B}(X, Y, I) \mid \langle A, B\rangle = \langle D^{\Downarrow}, D^{\Downarrow\Uparrow}\rangle$ for $D \in (\text{fix}(^{*_Y}))^Y\}$.

Here, $^{\Downarrow}$ and $^{\Uparrow}$ are the arrow operators "without hedges", i.e. defined by $A^{\Uparrow}(y) = \bigwedge_{x \in X}(A(x) \to I(x, y))$ and $B^{\Downarrow}(y) = \bigwedge_{y \in Y}(B(y) \to I(x, y))$. The following assertion provides us with the first insight.

Lemma 2: For $\langle X, Y, I\rangle$, we have

$$I(x, y) = \bigvee_{\langle A, B\rangle \in \mathcal{B}(X, Y^{*_Y}, I)} A(x) \otimes B(y).$$

Proof: Use Theorem 1 and direct computation. We omit details due to lack of space. ■

According to Lemma 2, I can be reconstructed from $\mathcal{B}(X, Y^{*_Y}, I)$. The way I is being reconstructed is this. For each formal concept $\langle A, B\rangle$ from $\mathcal{B}(X, Y^{*_Y}, I)$, one considers the rectangular $\mathbf{L}$-relation $A \otimes B$ defined by $(A \otimes B)(x, y) = A(x) \otimes B(y)$. Lemma 2 says that I results by adding all of these rectangular relations together by means of taking suprema (maxima if, e.g. $L = [0, 1]$). Now, any pair $\langle A, B\rangle$ of $A \in \mathbf{L}^X$ and $B \in \mathbf{L}^Y$ can be considered as a rectangle with $A \otimes B$ defined as above being the corresponding rectangular $\mathbf{L}$-relation. For rectangles $\langle A_1, B_1\rangle$ and $\langle A_2, B_2\rangle$ we can put $\langle A_1, B_1\rangle \trianglelefteq \langle A_2, B_2\rangle$ if for each $x \in X$, $y \in Y$ we have $A_1(x) \le A_2(x)$ and $B_1(y) \le B_2(y)$. The following lemma shows that formal concepts from $\mathcal{B}(X, Y^{*_Y}, I)$ correspond to certain maximal rectangles with the corresponding rectangular $\mathbf{L}$-relation contained in I.

Lemma 3: A pair $\langle A, B\rangle$ of $A \in \mathbf{L}^X$ and $B \in \mathbf{L}^Y$ is a formal concept in $\mathcal{B}(X, Y^{*_Y}, I)$ iff $\langle A, B\rangle$ is a maximal rectangle such that $A \otimes B \subseteq I$ and $A = (B^{*^Y})^{\Downarrow}$.

Proof: Similar to the proof of an analogous statement for $\mathcal{B}(X, Y, I)$, see [1]. We omit details. ■

Note that the additional condition requiring $A = (B^{*^Y})^{\Downarrow}$ is automatically satisfied in case ^{*_Y} is identity. Namely, in this case, it follows from the fact that $\langle A, B\rangle$ is a maximal rectangle such that $A \otimes B \subseteq I$. If ^{*_Y} is globalization, the additional condition says that $A(x) = \bigwedge_{B(y)=1} I(x, y)$. In general, the condition can be interpreted as follows. Even if we consider, instead of B, the fuzzy set B^{*_Y}, i.e. the largest fuzzy set which is contained in B and uses as membership degrees only the restricted degrees from $\text{fix}(^{*_Y})$, the collection A of objects which share all attributes from B^{*_Y} does not change.

Definition 1: Call I k-factorizable if $I = A \circ B$ for some $n \times k$ L-valued matrix A and some $k \times m$ L-valued matrix B.

If, moreover, for every $l = 1, \dots, k$, $B_{lj} \in \text{fix}(*_Y)$, we say that I is k-$*_Y$-factorizable.

Therefore, I is k-$*_Y$-factorizable iff I can be decomposed into $A \circ B$ where A is an L-valued matrix and B is a matrix with values restricted to fixpoints of ^{*_Y}. This means that if ^{*_Y} is identity, B is an L-valued matrix (no restriction), if ^{*_Y} is globalization, B is an $\{0, 1\}$-valued matrix (Boolean matrix). The condition of k-$*_Y$-factorizability is reasonable because one might want to have a crisp relationship between factors and original attributes or to see if it is possible to have such factors. Symmetrically, one might want to have a crisp relationship between objects and factors.

Definition 2: An $n \times m$ L-valued matrix I is called $*_Y$-concept-factorizable if there is a set $\mathcal{F}$ of formal concepts of $\mathcal{B}(X, Y^{*_Y}, I)$ such that for the corresponding matrices $A_{\mathcal{F}}$ and $B_{\mathcal{F}}$ we have $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$.

Therefore, if I is $*_Y$-concept-factorizable with $\mathcal{F}$, we call the formal concepts of $\mathcal{F}$ factors (or factor concepts).

The following theorem shows that each I is concept-factorizable. In fact, one can take $\mathcal{F} = \mathcal{B}(X, Y^{*_Y}, I)$, i.e. the set of all formal concepts of $\mathcal{B}(X, Y^{*_Y}, I)$.

Theorem 4 (universality of $*_Y$-concept-factorizability): For $\mathcal{F} = \mathcal{B}(X, Y^{*_Y}, I)$ we have $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$. Therefore, every binary matrix I is $*_Y$-concept-factorizable.

Proof: The assertion is a direct consequence of Lemma 2. We omit details. ■

Although Theorem 4 gives a way to factorize I, the obvious disadvantage is that the number $|\mathcal{B}(X, Y^{*_Y}, I)|$ of all concepts of $\mathcal{B}(X, Y^{*_Y}, I)$ (and thus the number of factors suggested by Theorem 4) is usually much larger than m, i.e. than the number of original attributes. It can even happen to be infinite. The next theorem shows that m, in fact, serves as the upper bound of the number of factor-concepts we can use. Consider the following set of formal concepts:

$$\mathcal{A}(X, Y^{*_Y}, I) = \{\langle \{1/y_j\}^{\downarrow}, \{1/y_j\}^{\downarrow\uparrow}\rangle \mid 1 \le j \le m\} \quad (4)$$

Formal concepts from $\mathcal{A}(X, Y^{*_Y}, I)$ are called attribute concepts. Using results from [4], it can be shown that elements of $\mathcal{A}(X, Y^{*_Y}, I)$ are indeed formal concepts from $\mathcal{B}(X, Y^{*_Y}, I)$. The following theorem shows that $\mathcal{A}(X, Y^{*_Y}, I)$ can be taken as a set of factor-concepts.

Theorem 5 (factorizability by attribute concepts): Let $\mathcal{F} = \mathcal{A}(X, Y^{*_Y}, I)$. Then I is concept-factorizable with $\mathcal{F}$ being the set of factor-concepts.

Proof: The proof follows from the fact that for every $\langle A, B\rangle \in \mathcal{A}(X, Y^{*_Y}, I)$ we have $A(x_i) = I(x_i, y_j)$ and $B(y_j) = 1$. The rest is routine and we omit details. ■

We are now going to show our main results, namely, the optimality of $*_Y$-concept-factorizability.

Theorem 6 (optimality of $*_Y$-concept-factorizability): Let I be k-$*_Y$-factorizable. Then there is $\mathcal{F} \subseteq \mathcal{B}(X, Y^{*_Y}, I)$ such that $|\mathcal{F}| \leq k$ and I is $*_Y$-concept-factorizable using $\mathcal{F}$ as the set of factor-concepts.

Proof: Sketch (details to be presented in full version): Let $I = A \circ B$ such that A and B are $n \times k$ and $k \times m$ L-valued matrices satisfying $B_{lj} \in \text{fix}(^{*_Y})$, i.e. I is k-$*_Y$-factorizable using A and B. One can see that for every $l = 1, \ldots, k$, the pair $\langle A_i, B_j\rangle$ consisting of $\mathbf{L}$-sets A_i and B_i corresponding to l-th column of A and l-th row of B are rectangles with the corresponding rectangular $\mathbf{L}$-relation $A_i \otimes B_j$ contained in I. Furthermore, it can be shown that $\langle A_i, B_j\rangle$ is contained in some maximal rectangle $\langle C_i, D_j\rangle$ which satisfies $C_i = (D_j^{*_Y})^{\Downarrow}$. According to Lemma 2, $\langle C_i, D_j\rangle$ is a formal concept from $\mathcal{B}(X, Y^{*_Y}, I)$. If we denote by $\mathcal{F}$ the collection of all such formal concepts $\langle C_i, D_j\rangle$, it is clear that $|\mathcal{F}| \leq k$ and that $A_{\mathcal{F}} \circ B_{\mathcal{F}} = I$. This proves the theorem. ■

Remark 7: Note that, under the notation from the proof of Theorem 6, we might have $|\mathcal{F}| < k$ (the number of factor-concepts may be strictly smaller than the number of the original factors). Namely, it may happen that two different rectangles $\langle A_l, B_l\rangle$ and $\langle A_{l'}, B_{l'}\rangle$ are contained in the same maximal rectangle, i.e. one can take $\langle C_l, D_l\rangle = \langle C_{l'}, D_{l'}\rangle$.

In fact, as shown by Theorem 6, if the number of factors is taken as a quality of factorization, $*_Y$-concept-factorization is the best way to factorize if we restrict our attention to $*_Y$-factorizability. Obviously, with ^{*_Y} being identity, there is no restriction involved and we obtain, as a corollary, the theorem from [3] saying that formal concepts from $\mathcal{B}(X, Y, I)$ are optimal factors.

V. Examples

In this section, we present illustrative examples. Our main aim is to illustrate the notions and results introduced in previous sections.

Take $L = \{\frac{n}{10} \,|\, 0 \leq n \leq 10\}$ equipped with Łukasiewicz operations and identities for hedges as our structure of truth degrees and consider the following input L-valued matrix:

$$I = \begin{pmatrix} 0.9 & 0.3 & 0.5 & 0.8 & 0.9 & 0.5 \\ 0.8 & 0.3 & 0.1 & 0.8 & 0.5 & 0.5 \\ 0.5 & 0.2 & 0.1 & 0.4 & 0.5 & 0.1 \\ 0.2 & 0.0 & 0.0 & 0.2 & 0.0 & 0.0 \\ 0.6 & 0.0 & 0.2 & 0.3 & 0.6 & 0.0 \end{pmatrix}.$$

The input matrix I is 3-factorizable. Indeed, we have $I = A \circ B$ for the following L-valued matrices:

$$A = \begin{pmatrix} 1.0 & 0.8 & 0.0 \\ 0.4 & 0.8 & 0.7 \\ 0.6 & 0.4 & 0.6 \\ 0.0 & 0.2 & 0.4 \\ 0.7 & 0.3 & 0.0 \end{pmatrix}, \quad B = \begin{pmatrix} 0.9 & 0.3 & 0.5 & 0.3 & 0.9 & 0.0 \\ 1.0 & 0.4 & 0.3 & 1.0 & 0.7 & 0.7 \\ 0.1 & 0.6 & 0.2 & 0.4 & 0.0 & 0.0 \end{pmatrix}.$$

Matrices A and B describing relationships between object $\times$ factors and factors $\times$ attributes induce three matrices F_1, F_2, and F_3 such that, for each $l \in 1, 2, 3$, $(F_l)_{ij} = A_{il} \otimes B_{lj}$:

$$F_1 = \begin{pmatrix} 0.9 & 0.3 & 0.5 & 0.3 & 0.9 & 0.0 \\ 0.3 & 0.0 & 0.0 & 0.0 & 0.3 & 0.0 \\ 0.5 & 0.0 & 0.1 & 0.0 & 0.5 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.6 & 0.0 & 0.2 & 0.0 & 0.6 & 0.0 \end{pmatrix},$$

$$F_2 = \begin{pmatrix} 0.8 & 0.2 & 0.1 & 0.8 & 0.5 & 0.5 \\ 0.8 & 0.2 & 0.1 & 0.8 & 0.5 & 0.5 \\ 0.4 & 0.0 & 0.0 & 0.4 & 0.1 & 0.1 \\ 0.2 & 0.0 & 0.0 & 0.2 & 0.0 & 0.0 \\ 0.3 & 0.0 & 0.0 & 0.3 & 0.0 & 0.0 \end{pmatrix},$$

$$F_3 = \begin{pmatrix} 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.3 & 0.0 & 0.1 & 0.0 & 0.0 \\ 0.0 & 0.2 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \end{pmatrix}.$$

The original matrix I is in fact a union of F_1, F_2, and F_3, cf. Lemma 2. Matrices F_1, F_2, and F_3 represent rectangles present in the original matrix I. These rectangles are, however, not maximal in sense of FCA, i.e. they do not correspond to formal concepts in I. Note that $\mathcal{B}(X, Y, I)$ induced by I contains 436 concepts. Matrices F_1, F_2, and F_3 representing rectangles in I can be extended to maximal rectangles representing formal concepts. For instance, we can take the following set of three formal concepts which will play the role of factor-concepts:

$$\begin{aligned} \mathcal{F} = \{ &\langle \{^{1.0}/x_1, {}^{0.6}/x_2, {}^{0.6}/x_3, {}^{0.1}/x_4, {}^{0.7}/x_5\}, \\ &\quad \{^{0.9}/y_1, {}^{0.3}/y_2, {}^{0.5}/y_3, {}^{0.6}/y_4, {}^{0.9}/y_5, {}^{0.3}/y_6\}\rangle, \\ &\langle \{^{0.8}/x_1, {}^{0.8}/x_2, {}^{0.4}/x_3, {}^{0.2}/x_4, {}^{0.3}/x_5\}, \\ &\quad \{^{1.0}/y_1, {}^{0.5}/y_2, {}^{0.3}/y_3, {}^{1.0}/y_4, {}^{0.7}/y_5, {}^{0.7}/y_6\}\rangle, \\ &\langle \{^{0.7}/x_1, {}^{0.7}/x_2, {}^{0.6}/x_3, {}^{0.4}/x_4, {}^{0.4}/x_5\}, \\ &\quad \{^{0.8}/y_1, {}^{0.6}/y_2, {}^{0.4}/y_3, {}^{0.8}/y_4, {}^{0.6}/y_5, {}^{0.5}/y_6\}\rangle\}. \end{aligned}$$

The concepts in $\mathcal{F}$ correspond to the following maximal rectangles which extend the rectangles F_1, F_2, and F_3:

$$C_1 = \begin{pmatrix} 0.9 & 0.3 & 0.5 & 0.6 & 0.9 & 0.3 \\ 0.5 & 0.0 & 0.1 & 0.2 & 0.5 & 0.0 \\ 0.5 & 0.0 & 0.1 & 0.2 & 0.5 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.6 & 0.0 & 0.2 & 0.3 & 0.6 & 0.0 \end{pmatrix},$$

$$C_2 = \begin{pmatrix} 0.8 & 0.3 & 0.1 & 0.8 & 0.5 & 0.5 \\ 0.8 & 0.3 & 0.1 & 0.8 & 0.5 & 0.5 \\ 0.4 & 0.0 & 0.0 & 0.4 & 0.1 & 0.1 \\ 0.2 & 0.0 & 0.0 & 0.2 & 0.0 & 0.0 \\ 0.3 & 0.0 & 0.0 & 0.3 & 0.0 & 0.0 \end{pmatrix},$$

$$C_3 = \begin{pmatrix} 0.5 & 0.3 & 0.1 & 0.5 & 0.3 & 0.2 \\ 0.5 & 0.3 & 0.1 & 0.5 & 0.3 & 0.2 \\ 0.4 & 0.2 & 0.0 & 0.4 & 0.2 & 0.1 \\ 0.2 & 0.0 & 0.0 & 0.2 & 0.0 & 0.0 \\ 0.2 & 0.0 & 0.0 & 0.2 & 0.0 & 0.0 \end{pmatrix}.$$

It is easily seen that I is a union of C_1, C_2, and C_3. The matrices $A_{\mathcal{F}}$ and $B_{\mathcal{F}}$ induced by $\mathcal{F}$ are the following:

$$A_{\mathcal{F}} = \begin{pmatrix} 1.0 & 0.8 & 0.7 \\ 0.6 & 0.8 & 0.7 \\ 0.6 & 0.4 & 0.6 \\ 0.1 & 0.2 & 0.4 \\ 0.7 & 0.3 & 0.4 \end{pmatrix}, \; B_{\mathcal{F}} = \begin{pmatrix} 0.9 & 0.3 & 0.5 & 0.6 & 0.9 & 0.3 \\ 1.0 & 0.5 & 0.3 & 1.0 & 0.7 & 0.7 \\ 0.8 & 0.6 & 0.4 & 0.8 & 0.6 & 0.5 \end{pmatrix},$$

and due to Theorem 6, we have $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$.

VI. Further Remarks and Further Research

Future research will focus on the following topics:

- Approximate factorizability. Extension of the results so that one requires a weaker condition I approximately equals $A_{\mathcal{F}} \circ B_{\mathcal{F}}$.
- Further criteria of what a "good set of factors" means. In addition to the number of factors, one may consider further criteria such as the generality of formal concepts used as factors or independence of formal concepts.
- Algorithms. A further study of algorithms is crucial for efficient implementation of our approach. Research will focus on studies of many-valued variants of set-covering problem, on utilizing theoretical insight provided by available results on fuzzy concept lattices, as well as on heuristic approaches.
- Comparison with classical and non-linear factor analysis. Namely, in a particular case when the entries of matrix I are numbers from $[0, 1]$, one can try to employ classical factor analysis as well as methods of non-linear factor analysis. Both experimental and theoretical comparison are needed here.

Acknowledgment

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

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

[3] Belohlavek R.: Relational factor analysis. Preprint, 2007.

[4] Belohlavek R., Vychodil V.: Reducing the size of fuzzy concept lattices by hedges. In: FUZZ-IEEE 2005, The IEEE International Conference on Fuzzy Systems, May 22–25, 2005, Reno (Nevada, USA), pp. 663–668, ISBN 0–7803–9158–6. Extended version to appear in *J. Advanced Comput. Intelligence and Intelligent Informatics*.

[5] Belohlavek R., Vychodil V.: On Boolean factor analysis with formal concepts as factors. SCIS & ISIS 2006, Int. Conf. Soft Computing and Intelligent Systems & Int. Symposium on Intelligent Systems, Sep 20–24, 2006, Tokyo, Japan, pp. 1054–1059.

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

[7] Hájek P.: *Metamathematics of Fuzzy Logic.* Kluwer, Dordrecht, 1998.

[8] Hájek P.: On very true. *Fuzzy Sets and Systems* **124**(2001), 329–333.

[9] Harman H. H.: *Modern Factor Analysis, 2nd Ed.* The Univ. Chicago Press, Chicago, 1970.

[10] Klir G. J., Yuan B.: *Fuzzy Sets and Fuzzy Logic. Theory and Applications.* Prentice-Hall, 1995.

[11] Pollandt S.: *Fuzzy Begriffe.* Springer-Verlag, Berlin/Heidelberg, 1997.

[12] Spearman C.: General intelligence, objectively determined and measured. *Amer. J. Psychology* **15**(1904), 201–293.