The Discrete Basis Problem and Asso Algorithm for Fuzzy Attributes

Radim Belohlavek, *Senior Member, IEEE*, and Marketa Trneckova

***Abstract*—We present an extension of the discrete basis problem, recently a profoundly studied problem, from the Boolean setting to the setting of fuzzy attributes, i.e., a setting of ordinal data. Our problem consists in finding for a given object-attribute matrix I containing truth degrees and a prescribed number k of factors the best approximate decomposition of I into an object-factor matrix A and a factor-attribute matrix B. Since such matrices represent fuzzy relations, the problem is related to but very different from that of decomposition of fuzzy relations as studied in fuzzy relational equations because neither A nor B are supposed to be known in our problem. We observe that our problem is NP-hard as an optimization problem. Consequently, we provide an approximation algorithm for solving this problem and provide its time complexity in the worst case. The algorithm is inspired by the Asso algorithm, which is known for Boolean attributes and is based on new considerations regarding associations among fuzzy attributes. We provide an experimental evaluation on various datasets and demonstrate that our algorithm is capable of extracting informative factors in data. We conclude with a discussion regarding future research issues.**

***Index Terms*—Decomposition of matrices/relations, factor analysis, fuzzy attribute, fuzzy concept lattice, ordinal data.**

I. Introduction

A. *Problem Description*

CONSIDER an $n \times m$ matrix I whose entries I_{ij} are elements of an ordered scale L (this shall be denoted by $I \in L^{n\times m}$). We assume that L represents a scale of truth degrees and that the matrix I represents a fuzzy (or graded) relation between n objects (matrix rows) and m attributes (columns). The entry I_{ij} is thus interpreted as a degree to which the object i has the attribute j. We are interested in finding for a given number k of factors and a given distance function an approximate decomposition (or, a factorization) of I into a sup-$\otimes$-product $A \circ B$ of an $n \times k$ object-factor matrix A and a $k \times m$ factor-attribute matrix B, i.e., in finding

$$I \approx A \circ B \tag{1}$$

in such a way that I and $A \circ B$ are as similar (i.e., close w.r.t. the given distance) as possible.

The sup-$\otimes$-product $A \circ B$ of A and B is the operation, well known in fuzzy logic, defined by

$$(A \circ B)_{ij} = \bigvee_{l=1}^{k} A_{il} \otimes B_{lj} \tag{2}$$

where $\bigvee$ is the supremum in L (maximum if L is a chain) and $\otimes$ is a suitable many-valued conjunction on L.

In particular, we assume that L is equipped with a partial order $\leq$ with respect to which it forms a complete lattice bounded by 0 and 1, and that the operation $\otimes$ is commutative (i.e., $a \otimes b = b \otimes a$), associative [i.e., $a \otimes (b \otimes c) = (a \otimes b) \otimes c$], has 1 as its neutral element (i.e., $a \otimes 1 = a = 1 \otimes a$), and is distributive over arbitrary suprema [i.e., $a \otimes (\bigvee_{j\in J} b_j) = \bigvee_{j \in J}(a \otimes b_j)$]. Such assumptions are standard in modern fuzzy logic [13], [14]. From a logical point of view, $\otimes$ is considered a truth function of (many-valued) conjunction [13], [14]. Importantly, $\otimes$ induces another operation, $\rightarrow$, called the *residuum of* $\otimes$, which plays the role of the truth function of implication and is defined by

$$a \rightarrow b = \max\{c \in L \mid a \otimes c \leq b\}. \tag{3}$$

Many examples of such scales are known in fuzzy logic [13], [14], among them those where L is the real unit interval $[0, 1]$ or its finite equidistant subinterval $L = \{0, \frac{1}{n}, \ldots, \frac{n-1}{n}, 1\}$, and where $\otimes$ is the Łukasiewicz, minimum, product, or other continuous t-norm. In what follows, we assume that the scale L is equipped with the operation $\otimes$ satisfying the properties mentioned earlier. Moreover, we assume for simplicity that L is a finite linearly ordered scale in the rest of this paper.

Remark 1:

1) An important special case results when L is the two-element scale $\{0, 1\}$ and $\otimes$ represents classical conjunction. Then, the matrices I, A, and B are Boolean matrices and the product (2) is the Boolean matrix product. Our setting then becomes the setting of Boolean matrix factorization/decomposition, which enjoys substantial interest in recent data mining.
2) The interest in our problem from a data analysis viewpoint is explained in the following as part of experimental evaluation (Sections II-B, IV, and V).

To define our problem precisely, we need a notion of closeness of matrices over L. Let $s_L : L \times L \rightarrow [0, 1]$ be an appropriate function measuring closeness of degrees in L (see the following). For matrices $I, J \in L^{n\times m}$, put

$$s(I, J) = \frac{\sum_{i,j=1}^{n,m} s_L(I_{ij}, J_{ij})}{n \cdot m} \tag{4}$$

Manuscript received January 9, 2018; revised October 29, 2018; accepted November 5, 2018. Date of publication November 9, 2018; date of current version June 28, 2019. This work was supported by grant No.15-17899S of the Czech Science Foundation and partially by grant No. IGA_PrF_2018_030. The present work was done during the sustainability period of the ECOP project no. CZ.1.07/2.3.00/20.0059, which is co-financed by the European Social Fund and the state budget of the Czech Republic. *(Corresponding author: Marketa Trneckova.)*

The authors are with the Department of Computer Science, Palacky University, Olomouc 771 47, Czech Republic (e-mail: radim.belohlavek@acm.org; marketa.trneckova@gmail.com).

Digital Object Identifier 10.1109/TFUZZ.2018.2880418

i.e., $s(I,J) \in [0,1]$ is the normalized sum over all matrix entries of the closeness of the corresponding entries in I and J. In general, we require $s_L(a,b) = 1$ if and only if $a = b$, and $s_L(0,1) = s_L(1,0) = 0$, in which case $s(I,J) = 1$ if and only if $I = J$. We furthermore require that $a \leq b \leq c$ implies $s_L(a,c) \leq s_L(b,c)$. For the important case of L being a subchain of $[0,1]$, s_L may be defined by

$$s_L(a,b) = a \leftrightarrow b$$

where $a \leftrightarrow b = \min(a \rightarrow b, b \rightarrow a)$ is the so-called biresiduum (many-valued equivalence) of a and b; here, $\rightarrow$ is the residuum (3) of $\otimes$. For the Łukasiewicz operations, i.e., $a \otimes b = \max(0, a+b-1)$ and $a \rightarrow b = \min(1, 1-a+b)$, which we use in our examples, we obtain $s_L(a,b) = 1 - |a-b|$ this way. Note that we use closeness because of its natural logical interpretation as a many-valued equivalence but, clearly, one could alternatively use the complementary notion of distance instead of closeness. One easily observes that if $s_L(a,b) = a \leftrightarrow b$, the normalized Hamming distance $e(I,J)$ of Boolean matrices I and J, which is defined by $e(I,J) = \frac{\sum_{i,j=1}^{n,m} |I_{ij} - J_{ij}|}{n \cdot m}$ and is often used for Boolean matrices, is just $e(I,J) = 1 - s(I,J)$.

We now present an exact formulation of our problem, the *discrete basis problem* over scale L equipped with $\otimes$, which we denote as *DBP(L)*:

problem *DBP(L)*
input matrix $I \in L^{n \times m}$, positive integer k;
output matrices $A \in L^{n \times k}$ and $B \in L^{k \times m}$ that maximize $s(I, A \circ B)$.

Remark 2:

1) For $L = \{0,1\}$, the problem *DBP(L)* coincides with the well-known DBP for Boolean matrices as defined in [18].
2) Since each $n \times m$ matrix I with entries in L in fact represents a binary L-fuzzy relation R_I between an n-element and an m-element universe, our problem *DBP(L)* may equivalently be described as a problem of finding a decomposition of R_I into a product of fuzzy relations R_A and R_B. Note, however, that *DBP(L)* is very different from that of fuzzy relational equations (FREs) because in FREs, two fuzzy relations represent the input of the problem, either R_I and R_A, or R_I and R_B.

B. Contributions of This Paper

In this paper, we first observe that *DBP(L)* is NP-hard as an optimization problem. Then, we argue that similarly as in the Boolean case, the problem is highly relevant from data analysis viewpoint. In particular, solving the problem corresponds to performing a factor analysis of the data represented by I. In our setting, the analyzed data are of ordinal type, rather than Boolean one. We argue and demonstrate by examples that the extension from Boolean to ordinal data greatly enhances applicability of the decompositions. Since the problem is provably hard, we provide an approximate algorithm inspired by the classic Asso algorithm [18] and discuss the various challenges presented by the extension of this algorithm to ordinal setting. We also present experimental evaluation of our algorithm and observe its computational efficiency as well as its data-analytical effectiveness. We conclude by presenting challenges opened by the current findings.

C. Related Work

Due to the great extent of the various works on matrix decomposition, we limit ourselves to the most relevant work and refer to [7] and [18] for further material and references. A direct predecessor of our work is the previous work in Boolean matrix decomposition. NP-hardness of the basic decomposition problem was established in [21]—one of the earliest papers on this topic. An increase in interest in Boolean matrix decompositions in data mining is due to Miettinen's work, which includes [18] presenting the Boolean version of the DBP and the classic Asso algorithm, as well as several subsequent papers examining Boolean CX and CUR decompositions [16], sparsity issues [17], and selection of the number of factors [19]. Note also the influential papers [11] and [7], which deal with restricted decompositions using so-called tiles and formal concepts in Boolean data.

As regards matrices with entries in scales L, these have been examined in several works, such as works on matrices over semirings and similar algebraic structures [12] and works on binary fuzzy relations [4], [13].

The closest to our problem is the problem of solving FREs; see e.g., [13] as well as [2] and [3] for more recent contributions. Note, however, that the problem of FREs is fundamentally different from the *DBP(L)* problem, because in FREs, we are given two fuzzy relations (or matrices with grades upon obvious identification or fuzzy relations and matrices with grades), I and A (or I and B) and the goal is to determine B (or A) such that I is equal, or approximately equal, to $A \circ B$. This difference reflects itself in that FREs have a very different purpose from that of *DBP(L)* but also in the difficulty of solving: While—as we show in Section II—solving *DBP(L)* is provably hard (NP-hard in technical terms) and one has to resort to approximation algorithms, solving FREs is done in polynomial time and, in fact, very quickly.

Directly relevant to *DBP(L)* are the papers [5] and [8], in which the role of so-called formal concepts of I is examined for exact decompositions $I = A \circ B$, and an algorithm GreCond$_L$ is proposed for this purpose. Even though GreCond$_L$ has primarily been designed for computing exact and almost exact decompositions, it may be easily adopted for computing approximate decompositions as well as for solving the *DBP(L)*. GreCond$_L$ represents the only available algorithm to which our new algorithm, Asso$_L$ may be compared. Let us also mention the possibility of scaling the fuzzy attributes in I to Boolean attributes and using Boolean decomposition algorithms and further insight to eventually obtain an approximate decomposition of I. Such option, which was examined in [6], leads to a considerably worse performance as regards both the quality of decomposition and computation time and we therefore do not consider it here and refer the reader to [6]. The work presented in [6] also provides factor analyses of various ordinal datasets using GreCond$_L$.

Note also that the role of order-theoretic structures in data analysis is extensively examined in [10] (formal concept analysis of Boolean data), [22] (ordered and combinatorial structures), and [4] (closure structures in the setting of fuzzy logic and structures over scales). Ordinal data and the methods for data analysis of such data appear in the literature on mathematical psychology. However, the tools employed there are basically modifications of classical factor analysis methods. In these approaches, grades (truth degrees) are represented by and treated like numbers. Such approaches lead to loss of interpretability, similarly as in the case of Boolean data, which is well documented e.g., in [24]. Note, however, that factor analysis of data with fuzzy attributes is still an open research field: Broadly conceived, our new algorithm represents a particular possible approach. Other approaches shall likely be obtained by combining the ideas from classical factor analysis and its variants on the one hand, and the ideas from fuzzy logic on the other hand, and are worth further examination.

II. Hardness and Significance of DBP(L)

In this section, we observe that *DBP*(L) is provably hard. In addition, we also demonstrate that *DBP*(L) is significant from data analysis point of view, hence designing such approximation algorithms is worth pursuing.

A. NP-Hardness of DBP(L)

Recall that NP-hardness of a computational problem essentially means that the problem cannot be solved with full guarantee by any efficient algorithm, i.e. an algorithm which runs in time polynomial with respect to the size of the input. Consequently, proving that a problem is NP-hard means that when solving this problem one must resort to approximation algorithms. Technically, "NP-hard" means "nondeterministically-polynomial-hard"; for definitions of the concept of NP-hardness of decision problems and optimization problems we refer to [1].

NP-hardness of various problems related to decompositions of Boolean matrices derives from the important result of Stockmeyer [23] claiming NP-hardness of the set basis problem. The following theorem adds to the negative results regarding feasibility of fast exact solvability of decomposition problems; in particular it generalizes the hardness of the Boolean version of DBP [18].

Theorem 1: *DBP*(L) is an NP-hard optimization problem.

Proof: By the definition of NP-hardness of optimization problems, we need to show that the decision version of *DBP*(L), which we denote $\Pi(L)$ in the following, is NP-hard as a decision problem. Note that $\Pi(L)$ consists in deciding for I and k [i.e., the inputs to *DBP*(L)] and an additional input $t \in L$ whether there exist matrices A and B such that $s(I, A \circ B) \geq t$.

One way to establish the claim consists in reducing the set basis problem to $\Pi(L)$. However, since the NP-hardness is already established for the Boolean DBP, we utilize this fact as follows. The Boolean DBP may clearly be identified with our problem for $L = \{0,1\}$, i.e., with *DBP*($\{0,1\}$), taking into account the fact mentioned earlier in this paper that instead of the Hamming distance $E(A \circ B, I)$ used in the Boolean DBP, the *DBP*(L) uses the closeness $s(A \circ B, I)$ and that the correspondence between these two functions is $E(A \circ B, I)/(nm) = 1 - s(A \circ B, I)$ [hence, instead of minimizing E in the Boolean DBP, one maximizes s in *DBP*(L)].

Consider now the proof of NP-hardness of the Boolean DBP [18]. When reducing the set basis problem to the decision version of the Boolean DBP, i.e., to $\Pi(\{0,1\})$ modulo the earlier identification, the instances of the set basis problem are actually assigned instances of the Boolean DBP with $t = 1$ only. These assigned instances therefore consist of a Boolean matrix $I \in \{0,1\}^{n \times m}$, a positive integer k, and $t = 1$ [note that the question $s(A \circ B, I) \geq 1$, i.e., $s(A \circ B, I) = 1$, is equivalent to $E(A \circ B, I) \leq 0$, i.e., to $E(A \circ B, I) = 0$ in terms of Hamming distance]. Clearly, it is now enough to assign to every such instance (in a polynomial time) $\langle I, k, 0\rangle$ an instance $\langle I', k', t'\rangle$ of the decision version $\Pi(L)$ of *DBP*(L) in such a way that the answer to $\langle I, k, 0\rangle$ as an instance of $\Pi(\{0,1\})$ is yes if and only if the answer to $\langle I', k', t'\rangle$ as an instance of $\Pi(L)$ is yes. This is, however, easy as it suffices to take $\langle I', k', t'\rangle = \langle I, k, 0\rangle$. Namely, the answer yes to $\langle I, k, 0\rangle$ as an instance of $\Pi(L)$ means that there exist matrices $A \in L^{n \times k}$ and $B \in L^{k \times m}$ such that $s(A \circ B, I) = 1$, which means $A \circ B = I$. Now, since I_{ij} equals 0 or 1 only and since one always has $a \otimes b = 1$ iff $a = 1$ and $b = 1$, it is not difficult to verify that matrices $A \in L^{n \times k}$ and $B \in L^{k \times m}$ for which $A \circ B = I$ exist if and only if there exist matrices $C \in \{0,1\}^{n \times k}$ and $D \in \{0,1\}^{k \times m}$ for which $C \circ D = I$. That is, the answer to $\langle I, k, 0\rangle$ as an instance of $\Pi(L)$ is positive iff the answer to $\langle I, k, 0\rangle$ as an instance of $\Pi(\{0,1\})$ is positive. This completes the proof.

B. Significance of DBP(L)

The interest in our problem from a data analysis viewpoint derives from the fact that an approximate decomposition (1) provides an explanation of the input data I describing n objects via a possibly large number m of attributes by means of a possibly small number k of factors, which may be considered as newly discovered fundamental attributes. Moreover, the explanation has a simple transparent semantics, which is a significant aspect in the realm of factor analysis. Namely, due to the basic principles of fuzzy logic, our factor model (1) has the following interpretation: The degree I_{ij} to which object i has attribute j equals the degree of the following proposition—there exists factor l such that l applies to i and j is one of the particular manifestations of l.

Note that the kind of data we consider, i.e., data in which objects are described by attributes whose values are in a certain scale L of grades, commonly appear in many fields. The grades in L, such as 0, 0.5, and 1, are often described verbally, such as "not at all," "to some extent," and "fully," respectively. Examples of such data include questionnaires and surveys in which respondents answer questions by selecting grades (e.g., a respondent is asked to select the level of his satisfaction with the company's customer service from "not satisfied at all," "satisfied to some extent," and "fully satisfied"); performance data in which the objects are evaluated by means of attributes expressing the extent to which the object meets a certain criterion; or encyclopedic data (corpus data) in which a collection of objects in a certain domain of interest is described by certain graded attributes such as in the following example.

We now present a small example illustrating that the decompositions of ordinal data involved in *DBP*(L) provide us with

TABLE I
FIVE MOST POPULAR DOG BREEDS

	Energy	Playfulness	Friend. towards dogs	Friend. tow. strangers	Friend. tow. other pets	Protection ability	Exercise	Affection	Ease of training	Watchdog ability	Grooming
Labrador Retrievers	5	6	5	6	6	3	4	6	6	5	3
Golden Retrievers	4	6	6	6	6	3	4	6	6	4	4
Yorkshire terriers	5	5	3	4	3	2	2	4	3	6	5
German shepherds	4	3	2	3	4	6	5	4	6	6	3
Beagles	4	4	6	6	6	2	4	6	2	5	2

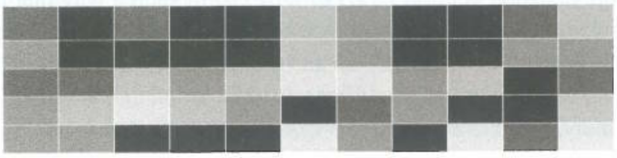


Fig. 1. 5×11 matrix I representing the data in Table I.

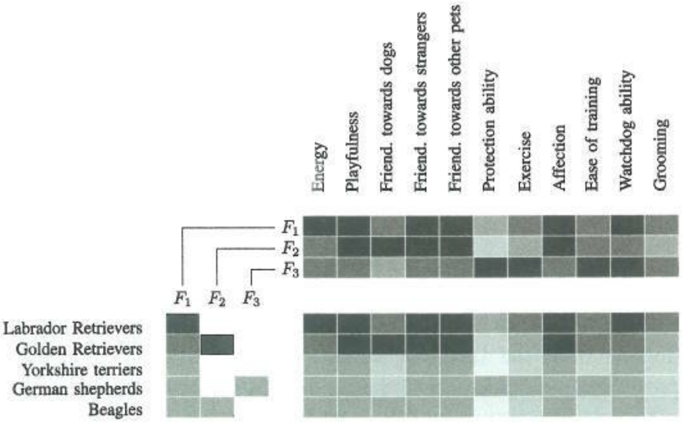
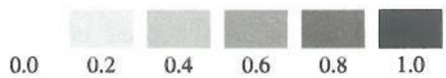


Fig. 2. Approximate decomposition of matrix I into $A_{\mathcal{F}}$ (bottom-left matrix) and $B_{\mathcal{F}}$ (top matrix).

useful information from data analysis point of view. The data in Table I describes 5 most popular dog breeds and their 11 attributes[1] (we analyze the full set of 151 breeds in Section IV).

We transform the original scale $\{1, \ldots, 6\}$ to the six-element chain $L = \{0, 0.2, 0.4, 0.6, 0.8, 1\}$ and use the Łukasiewicz connectives $\otimes$ and $\rightarrow$ (Section I-A). We represent the grades in L by shades of gray as follows:



The 5×11 object-attribute matrix I representing the data in Table I is depicted in Fig. 1.

Fig. 2 presents an approximate decomposition of matrix I using the first three factors obtained from I using our algorithm ASSO_L, which is described in Section III. That is, the 5×11 object-attribute matrix I is decomposed using the set $\mathcal{F}$ consisting of the first three obtained factors into the 5×3 object-factor matrix $A_{\mathcal{F}}$ (bottom-left matrix in Fig. 2) and the 3×11 factor-attribute matrix $B_{\mathcal{F}}$ (top matrix in Fig. 2). The bottom-right

[1]http://www.petfinder.com/

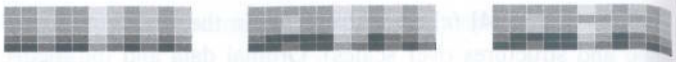


Fig. 3. $\bigvee$-superposition of factor concepts.

matrix in Fig. 2 represents the composition $A_{\mathcal{F}} \circ B_{\mathcal{F}}$ of $A_{\mathcal{F}}$ and $B_{\mathcal{F}}$. One may observe the apparent similarity of the input matrix I and its approximation $A_{\mathcal{F}} \circ B_{\mathcal{F}}$ represented by the first three factors obtained by our algorithm. For completeness, let us note that this particular decomposition was obtained by setting the parameters of ASSO_L to $w_0 = 3$, $w_1 = 1$, and $\tau = 0.85$ (see in the following for the meaning of these parameters).

Every factor F_l $(l = 1, 2, 3)$ is represented by the lth column in $A_{\mathcal{F}}$ and the lth row in $B_{\mathcal{F}}$. The entries $(A_{\mathcal{F}})_{il}$ indicate the degrees to which factor l applies to breed i, while $(B_{\mathcal{F}})_{lj}$ represents the degree to which attribute j is a particular manifestation (is typical) of factor l. For example, F_1 is manifested by energy, playfulness, two kinds of friendliness, and affection (attributes with degree equal to 1 in the first row of $B_{\mathcal{F}}$) and applies in particular to Labradors and Golden Retrievers (breeds with high degrees in the first column of $A_{\mathcal{F}}$). The factor may hence be termed *suitable for kids*. On the other hand, the four attributes with the highest degree in the row of F_3 (protection ability, exercise, ease of training, and watchdog ability) tell us that this factor is naturally interpreted as *guardian dog*. The corresponding column shows that F_3 applies to German shepherds and separates them clearly from the other breeds.

Every factor may conveniently be represented by a 5×11 matrix. For instance, the matrix representing the first factor results by the crossproduct of the first column of $A_{\mathcal{F}}$ and the first row of $B_{\mathcal{F}}$, and is depicted as the first matrix in Fig. 3.

Comparing this matrix to the input matrix I, one easily observes which parts of the input data are explained by this first factor. Furthermore, by means of $\bigvee$-superposition of factors, one may combine the factors. For instance, adding the second factor to the first, one obtains the second matrix in Fig. 3. Still adding the third factor, one obtains the third matrix in Fig. 3. Note also that the first, the first two, and the first three factors explain 0.81%, 0.84%, and 0.85% of the input data. From this viewpoint, one may conclude that already the first three factors explain the data very well.

III. ASSO_L

ASSO_L—our new algorithm for approximately solving the *DBP*(L)—is inspired by ASSO [18], the classic algorithm for the Boolean DBP. As we shall see, its extension from the Boolean setting to the setting with fuzzy attributes over scales L is not obvious. In the following text, we analyze the conceptual issues involved and provide a description of ASSO_L.

A. Association Matrix

The ordinary ASSO is based on using the rows of the association matrix $\mathcal{A}$ of the input matrix I. These rows are then used as candidate basis vectors, i.e., as rows of the $k \times m$ factor-attribute matrix B. The ordinary association matrix $\mathcal{A}$ is an

$m \times m$ Boolean matrix such that

$$\mathcal{A}_{pq} = 1 \quad \text{iff} \quad c(p,q) \geq \tau \tag{5}$$

where $c(p,q)$ is the confidence of the association rule $\{p\} \Rightarrow \{q\}$ given by I and τ is a user-specified association threshold.

The confidence $c(p,q)$ may be understood as a conditional probability, namely the probability of

object has attribute q provided it has attribute p

given that objects, which represent elementary events, are equally probable. In presence of grades, we consider conditional probabilities $c_a(p,q)$ of

an object has attribute q provided it has attribute p to degree at least a.

In this sense, $c_a(p,q)$ represents our confidence that the presence of p to degree at least a implies the (full, i.e., to degree 1) presence of q. Note that other options to conceive of logical associations are possible, but the one we choose is simple enough and, as we shall see, leads to a good quality of the resulting decompositions.

Unlike the Boolean case, collections of objects sharing certain attributes to prescribed degrees are naturally conceived as fuzzy sets rather than ordinary sets. Thus, according to the common usage, see e.g., [4], we denote by $\{{}^a/p\}^\downarrow$ the collection of all objects sharing attribute p to degree at least a and define it as the fuzzy set to which object $i = 1, \dots, n$ belongs to degree

$$\{{}^a/p\}^\downarrow(i) = a \to I(i,p).$$

Here again, $\to$ is the residuum (many-valued implication) of the particular structure of truth degrees. Likewise, the collection of objects having p to degree at least a and having q is defined by

$$\{{}^a/p, {}^1/q\}^\downarrow(i) = (a \to I(i,p)) \wedge I(i,q)$$

because $\{{}^a/p, {}^1/q\}^\downarrow(i) = [\{{}^a/p\} \cup \{{}^1/q\}]^\downarrow(i) = \{{}^a/p\}^\downarrow(i) \wedge \{{}^1/q\}^\downarrow(i) = (a \to I(i,p)) \wedge (1 \to I(i,q)) = (a \to I(i,p)) \wedge I(i,q)$. These formulas may be obtained from considerations on Galois connections induced by fuzzy relations [4], since these are the mathematical counterparts of assignments of objects sharing a given collection of attributes (the formulas may also be obtained directly on intuitive grounds).

To evaluate the conditional probability that defines $c_a(p,q)$, we deal with so-called fuzzy events and probabilities of fuzzy events in the sense of Zadeh [25]. The probability measure of fuzzy events, which is involved in our situation, is thus a function P assigning to every fuzzy set A of objects a number $P(A) \in [0,1]$—the probability of the fuzzy event A. As in the classical case, we assume that the objects (i.e., the rows of the input matrix I) represent elementary events and are equally probable. Zadeh's formulas [25] for conditional probabilities $P(\cdot|\cdot)$ of fuzzy events then yield that the confidence in question is defined by

$$\begin{aligned} c_a(p,q) &= P(\{{}^1/q\}^\downarrow | \{{}^a/p\}^\downarrow) = \frac{P(\{{}^a/p\}^\downarrow \cap \{{}^1/q\}^\downarrow)}{P(\{{}^a/p\}^\downarrow)} \\ &= \frac{P(\{{}^a/p, {}^1/q\}^\downarrow)}{P(\{{}^a/p\}^\downarrow)} = \frac{|\{{}^a/p, {}^1/q\}^\downarrow|}{|\{{}^a/p\}^\downarrow|} \end{aligned}$$

where $|A|$ denotes the cardinality of a fuzzy set A. With $|A| = \sum_{i=1}^{n} A(i)$, sometimes called the scalar cardinality of fuzzy sets, we thus obtain

$$|\{{}^a/p, {}^1/q\}^\downarrow| = \sum_{i=1}^{n} \{{}^a/p, {}^1/q\}^\downarrow(i), \text{ and}$$

$$|\{{}^a/p\}^\downarrow| = \sum_{i=1}^{n} \{{}^a/p\}^\downarrow(i).$$

Notice that when deriving the formula for $c_a(p,q)$, we utilized $\{{}^a/p\}^\downarrow \cap \{{}^1/q\}^\downarrow = (\{{}^a/p\} \cup \{{}^1/q\})^\downarrow = \{{}^a/p, {}^1/q\}^\downarrow$, which is a basic property of Galois connections [4].

Now, if we want to follow the basic logic of the ordinary, Boolean ASSO algorithm, we need to transform the confidence $c_a(p,q)$ to a truth value $\mathcal{A}_{pq}$ of the association matrix $\mathcal{A}$ by means of a user-defined threshold $\tau \in [0,1]$, and thus obtain the candidate basis vectors $\mathcal{A}_{p\_}$ (rows of $\mathcal{A}$). In the ordinary case, this is simply done by a comparison of the confidence $c(p,q)$—which coincides with $c_1(p,q)$ in our notation in the fuzzy setting—to τ, see (5).

When truth degrees are involved, we use again a user-defined threshold $\tau \in [0,1]$ to transform the confidence $c_a(p,q) \in [0,1]$ to a truth value in L using a user-defined threshold $\tau \in [0,1]$. However, the thresholding process is more involved compared to the Boolean case, because the result of the thresholding is no longer either 0 or 1 as in the Boolean case, but rather any of the multiple truth values in the scale L. We propose to use for this purpose the following rounding function round_τ.

Put first for $r \in (0,1]$

$$\begin{aligned} r_+ &= \min\{a \in L \mid a \geq r\} \\ r_- &= \max\{a \in L \mid a < r\}. \end{aligned}$$

That is, r^+ is the least truth value in L greater than or equal to r and r^- is the largest truth value in L smaller than r. Furthermore, define a rounding function round_τ as follows:

$$\text{for } r = 0\text{:}\ \ \text{round}_\tau(r) = 0$$

$$\text{for } r > 0\text{:}\ \ \text{round}_\tau(r) = \begin{cases} r_+ & \text{if } \frac{1-|r_+ - r|}{|r_+ - r^-|} \geq \tau. \\ r_- & \text{otherwise.} \end{cases}$$

The purpose of round_τ is to round off the confidence degree r to its appropriate neighboring truth degree in L. Now, as it is easy to see, r is in the interval $[r^-, r^+]$ and round_τ rounds off r to r^+ or r^-, depending on whether r is in $[t, r^+]$ or $[r^-, t)$, respectively, where t splits the interval $[r^-, r^+]$ into its left and right parts that comprise $(1-\tau) \cdot 100\%$ and $\tau \cdot 100\%$ of the size of the interval.

Observe that this generalizes the Boolean version involved in the classical ASSO, because there, trivially, $[r^-, r^+] = [0,1]$ and, therefore, the interval $[0,1]$ is split by τ into $[0,\tau)=[r^-,t)$ and $[\tau,1]=[t,r^+]$. Then, the entry $\mathcal{A}_{pq}$ of the association matrix $\mathcal{A}$ is set according to where the confidence $r=c(p,q)$ sits in $[0,1]$: one puts $\mathcal{A}_{pq}=1$ if $c(p,q)\in[\tau,1]$ while $\mathcal{A}_{pq}=0$ if $c(p,q)\in[0,\tau)$, cf., (5).

Since we deal—in our setting of associations between fuzzy attributes—with confidence degrees $c_a(p,q)$, we may define for every attribute p and every *suitable* grade $a\in L-\{0\}$ a candidate basis vector, i.e., a row $\mathcal{A}_{(p,a),\_}$ of a prospective association matrix $\mathcal{A}$, by

$$\mathcal{A}_{(i,a),j}=\text{round}_\tau(c_a(p,q)).$$

Let us now pick a set $K\subseteq L-\{0\}$ of *suitable* grades. We then obtain an association matrix of dimension $(m\cdot|K|)\times m$, i.e., $\mathcal{A}\in L^{(m\cdot|K|)\times m}$, because one entry, $\mathcal{A}_{(p,a),q}$, is obtained for every pair $1\leq p,q\leq m$ of objects, and every truth value $a\in K$. One may verify that if $L=\{0,1\}$ and $K=\{1\}$, then $\mathcal{A}$ is just the ordinary $m\times m$ association matrix defined in [18]. The presence of intermediate grades allows us to broaden the set of candidate basis vectors. Namely, in addition to the possible choice $K=\{1\}$, we may pick K containing more grades, e.g., $K=L-\{0\}$ which seems a natural choice, and thus enlarge the search space for factorization.

B. Procedures Cover and Asso$_L$

The basic idea of our new algorithm, Asso$_L$, may be described as follows. The algorithm iteratively computes k factors one by one, with the provision that it stops with less than k factors if the addition of any new factor would only worsen the error function, i.e., would decrease the value of the function s, defined in (4), in our case. Let A and B denote the object-factor and factor-attribute matrices computed so far. The next factor, which is described by a new column and a new row to be added to A and B, is computed as follows. For every candidate row of B, i.e., the row of the association matrix $\mathcal{A}$, one determines the best corresponding candidate column of A. "Best" means that the value of a function Cover (see later) is maximized. The candidate row of B and column of A with the highest value of Cover are then added as a new factor to B and A.

The purpose of the function Cover is to yield a high value for factors whose addition is likely to lead to good resulting matrices A and B, i.e., with high value of s. In the Boolean case, this is relatively straightforward: we want a high number c of entries $\langle i,j\rangle$ for which $I_{ij}=1$ and $(A\circ B)_{ij}=1$, i.e., 1s in I that are "covered" by the factors, and a small number o of entries for which $I_{ij}=0$ and $(A\circ B)_{ij}=1$, i.e., are "overcovered" by the factors. This reasoning leads to the formula

$$w^+\cdot c-w^-\cdot o$$

as the definition of Cover in the Boolean case. The weights reflect relative importance of c and o. Since "overcovering" cannot be undone by adding further factors, the benefit of covering an entry with 1 [i.e., having $I_{ij}=1$ and $(A\circ B)_{ij}=1$] reflected by the weight w^+ is smaller than the drawback due to overcovering an entry with 0 [i.e., having $I_{ij}=0$ and $(A\circ B)_{ij}=1$). As a consequence, one should set w^+ larger than w^- as a rule in practice.

In a fuzzy setting, the design of an appropriate Cover function is more delicate compared to the Boolean case. One reason is that the coverage of entry $\langle i,j\rangle$ of I is a matter of degree. We therefore need to account for a partial coverage and a partial overcoverage. For instance, if $I_{ij}=0.5$ and $(A\circ B)_{ij}=0.4$, then one may consider $\langle i,j\rangle$ almost covered and thus consider

$$I_{ij}\leftrightarrow(A\circ B)_{ij}=0.5\leftrightarrow 0.4=0.9$$

as the degree to which $\langle i,j\rangle$ is covered. Likewise, if $I_{ij}=0.5$ and $(A\circ B)_{ij}=0.6$, then $\langle i,j\rangle$ is slightly overcovered and

$$\neg(I_{ij}\leftrightarrow(A\circ B)_{ij})=\neg(0.5\leftrightarrow 0.6)=0.1$$

may be thought of as a degree to which $\langle i,j\rangle$ is overcovered. Analogously to the Boolean case, one could obtain the value of Cover by adding the degrees corresponding to the first type of entries, multiply them with w^+ and subtract from this number the w^--multiple of the sum of the degrees corresponding to the second type of entries.

This, however, would not yet be an appropriate approach. Consider a situation in which $I_{ij}=0.5$, w^- is even five times larger than w^+, and the so-far computed matrices A and B yield $(A\circ B)_{ij}=0.3$. Suppose we now have two options. First, adding a factor resulting in A_1 and B_1 with $(A_1\circ B_1)_{ij}=0.4$; second, adding a factor resulting in A_2 and B_2 with $(A_2\circ B_2)_{ij}=0.52$. Intuitively, the second choice is preferable because the factor commits only a slight overcovering of $I_{ij}=0.5$. However, the function Cover described earlier would lead to the selection of the first factor. Namely (for simplicity, we disregard entries other than $\langle i,j\rangle$), the first factor contributes by $w^+\cdot(I_{ij}\leftrightarrow(A_1\circ B_1)_{ij})=w^+\cdot 0.9$, while the second one contributes by $-w^-\cdot\neg(I_{ij}\leftrightarrow(A_2\circ B_2)_{ij})=-w^-\cdot 0.02$, i.e., even represents a decrease in value of Cover.

The point is that the entries which are overcovered, i.e., $I_{ij}<(A\circ B)_{ij}$, need to be looked at as follows: They need to be penalized for overcovering by $w^-\cdot\neg(I_{ij}\leftrightarrow(A\circ B)_{ij})$ but at the same time rewarded for full covering by $w^+\cdot 1$. This type of problem is degenerated in the Boolean case. Namely, this reward may be ignored because it would pertain to all entries with $I_{ij}=0$, would be equal for all such entries, and would hence have no influence on the choice of factors. This explains why the function Cover for the ordinary Asso algorithm does not contain any rewarding term for the overcovered entries.

The aforementioned reasoning leads to the following definition of Cover. Let $\mathcal{F}=\{\langle C_1,D_1\rangle,\ldots,\langle C_k,D_k\rangle,\}$ denote a set of factors (with a fixed ordering of its elements), i.e., pairs $\langle C_l,D_l\rangle$ where $C_l\in L^{1\times n}$ and $D_l\in L^{1\times m}$. Let $A_\mathcal{F}$ and $B_\mathcal{F}$ denote the corresponding matrices obtained from the factors. That is,

$$(A_\mathcal{F})_{il}=(C_l)_i\quad\text{and}\quad(B_\mathcal{F})_{lj}=(D_l)_j \tag{6}$$

where C_l forms the lth column of $A_\mathcal{F}$ and D_l forms the lth row of $B_\mathcal{F}$ for each $l=1,\ldots,k$. In accordance to the earlier

Algorithm 1 COVER.

Input: matrix $I \in L^{n\times m}$, set $\mathcal{F}$ of factors, w^+, w^-
Output: number $\text{COVER}(\mathcal{F}, I, w^+, w^-)$

$J \leftarrow A_\mathcal{F} \circ B_\mathcal{F}$; $c \leftarrow 0$
for $i = 1, \dots, n$ **do**
  for $j = 1, \dots, m$ **do**
    if $I_{ij} \geq J_{ij}$ **then**
      $c \leftarrow c + w^+ \cdot (I_{ij} \leftrightarrow J_{ij})$
    else
      $c \leftarrow c + w^+ - w^- \cdot (1 - (I_{ij} \leftrightarrow J_{ij}))$
    end
  end
end
return c

Algorithm 2 ASSO_L.

Input: matrix $I \in L^{n\times m}$, $k \geq 1$, w^+, w^-, τ, $K \subseteq L - \{0\}$
Output: set $\mathcal{F}$ of factors

1 compute association matrix $\mathcal{A}$
2 $\mathcal{F} \leftarrow \emptyset$
3 **for** $l = 1, \dots, k$ **do**
4   select $\langle C, \mathcal{A}_{(i,a)\_}\rangle$ maximizing $\text{COVER}(\mathcal{F} \cup \{\langle C, \mathcal{A}_{(i,a)\_}\rangle\}, I, w^+, w^-)$
5   add $\langle C, \mathcal{A}_{(i,a)\_}\rangle$ to $\mathcal{F}$
6 **end**
7 **return** $\mathcal{F}$

considerations, we put

$$\begin{aligned}
&\text{COVER}(\mathcal{F}, I, w^+, w^-) \\
&= w^+ \cdot \sum\{I_{ij} \leftrightarrow (A_\mathcal{F} \circ B_\mathcal{F})_{ij}\,;\, I_{ij} \geq (A_\mathcal{F} \circ B_\mathcal{F})_{ij}\} \\
&\quad + w^+ \cdot |\{\langle i,j\rangle\,;\, I_{ij} < (A_\mathcal{F} \circ B_\mathcal{F})_{ij}\}| \\
&\quad - w^- \cdot \sum\{1 - (I_{ij} \leftrightarrow (A_\mathcal{F} \circ B_\mathcal{F})_{ij})\,;\, I_{ij} < (A_\mathcal{F} \circ B_\mathcal{F})_{ij}\}.
\end{aligned}$$

Algorithm 1 provides the pseudocode for computing COVER $(\mathcal{F}, I, w^+, w^-)$.

The aforementioned procedure for computing a set $\mathcal{F}$ of factors, and hence due to (6) for computing matrices $A_\mathcal{F}$ and $B_\mathcal{F}$, is described by Algorithm 2.

Note that the selection in line 4 proceeds by finding for every row $\mathcal{A}_{(i,a)\_}$ of $\mathcal{A}$ the best C w.r.t. COVER and then selecting the best found pair $\langle C, \mathcal{A}_{(i,a)\_}\rangle$. Due to the properties of COVER, the best C for a given $\mathcal{A}_{(i,a)\_}$ is found efficiently in a componentwise manner, i.e., by finding the best entry C_p for every $p = 1, \dots, n$ independently of the other entries C_q of C).

C. Time Complexity

We first derive the worst case of time complexity of ASSO_L. Constructing the association matrix takes $O(n \cdot m^2 \cdot |K|)$ steps: The matrix has $m^2 \cdot |K|$ entries and to compute the entry $\mathcal{A}_{(p,a)q}$, one needs to go over the n objects to evaluate $c_a(p,q)$ and apply the rounding function round_τ. One then needs to obtain k basis vectors. One may observe that a selection of one such vector takes $O(n \cdot m^2 \cdot |K|)$ steps: One has to process $n \cdot |K|$ rows of $\mathcal{A}$; for each row, one finds the corresponding C in $O(n \times m)$ steps. Altogether, the algorithm ASSO_L is of time complexity $O(k \cdot n \cdot m^2 \cdot |K|)$ in the worst case. Note that according to the work presented in [18], the time complexity of the classical ASSO algorithm is $O(k \cdot n \cdot m^2)$, i.e., in the fuzzy setting, the complexity is essentially taking into account that the association matrix has $|K|$-times more rows compared to the Boolean case. Note also that the formula $O(k \cdot n \cdot m^2)$ derives from our formula in the fuzzy setting because in Boolean setting, $K = \{1\}$.

In concrete terms, we ran our experiments on a PC with the Intel Core i7-3517U processor and 8GB RAM. Computation of each of the datasets involved in our experiments was completed in the order of seconds. The actual running time of ASSO_L turns out to be 0.45 of the time needed by GRECOND_L on average over the data used in our experiments.

TABLE II
REAL DATA

dataset	size	$\lvert L\rvert$	$\lVert I\rVert$
Breeds	151×11	6	1963
Decathlon	28×10	5	266
IPAQ	4510×16	3	41624
Music	900×26	7	20377
MovieLens	1000×1000	6	361426

IV. ANALYZING REAL DATA

In this section, we present results of selected analyses of real data which we used in our evaluation. Our primary aim is to illustrate that the presented decomposition method makes it possible to extract natural and easy-to-understand factors from ordinal data. The datasets and their characteristics are described in Table II, in which $|L|$ denotes the number of truth degrees in the scale L and $||I||$ denotes the number of nonzero entries in the input matrix I.

Dog Breeds[2] extends the dataset from Section II-B to 151 breeds. ASSO_L obtained ten factors from this data. These factors explain 83% of data. The most important ones could be interpreted as *excel in sports* or *hound* since this factor contains the attributes "playfulness," "energy," and "affection" to degree 1 and also "ease of train" to a high degree. This factor applies to high degree to breeds such as American Cocker Spaniel, American Foxhound, Border Collie, Irish Setter, Brittany, or Pointer. The second significant factor contains the attributes "protection ability," "exercise," and "watchdog ability" to degree 1. Such a factor may therefore be interpreted as the ability to serve as a *guardian dog* and applies, e.g., to Australian Shepherd, Belgian Sheepdog, and German Shepherd Dog. The other extracted factors may be also interpreted by looking at the attributes present in the fuzzy sets of attributes of the respective factor but they are less significant in terms of the added explanatory power.

Decathlon[3] extends the dataset from [8] and represents data regarding performance of 28 athletes in the ten disciplines of decathlon via a 28×10 matrix I over a five-element scale L. ASSO_L computed a set $\mathcal{F}$ of five factors from these data. These

[2] http://www.petfinder.com/
[3] http://www.sports-reference.com/

factors reconstruct 79% of the input data, i.e., $s(I, A_{\mathcal{F}} \circ B_{\mathcal{F}}) = 0.79$. The most interesting of them is the most important one in terms of coverage (explanatory power). This factor contains "1500 m," "100 m," "400 m," and "hurdles," and may thus naturally be termed *running capability*. The second most important factor in terms of coverage may naturally be interpreted as *explosiveness*, since it contains the attributes "100 m," "shot put," "hurdles," and "discus throw" to high degrees. The factors found were consulted with a decathlon coach who confirmed that they indeed represent natural factors determining good performance in decathlon.

IPAQ Data[4] consists of international questionnaire data regarding physical activity of population and involves 4510 respondents answering 16 questions using a three-element scale. This questionnaire is considered important from the health management point of view, particularly as a source for making government decisions regarding health policy. The questions include those regarding age, sex, body-mass-index, health, to what extent the person bicycles, walks, and other factors. ASSO_L returned a set $\mathcal{F}$ of ten factors from these data. These factors explain 78% of the input data and are naturally interpretable. For instance, the first factor corresponds to and thus may be interpreted as *healthy people who cycle or walk on a regular basis*. The second factor may be interpreted as *healthy people with good education who cycle on a regular basis.*

The *Music Data* comes from [9] and consist of results of a study inquiring how people perceive a speed of song depending of various characteristics of the songs. The data were collected by questionnaires involving 30 participants who were presented 30 samples (29 complex music samples and one simple tone of 528 Hz). The participants recorded their emotional experience using 26 attributes (such as "pleasant," "happy," "exciting," "restful," "intelligible," "ugly," "valuable," "interesting," "slow," "meaningful," "active," "tense," "predictable," "closed," "violent," "strong," "known," "variable," or "like it"), each using a six-element scale L, along with a retrospective time duration and time passage judgement. The data are then represented by a 900×26 matrix with entries in L.

ASSO_L obtained ten factors from the data. These factors explain 83% of the data. The most significant are the first two factors. The first factor contains songs rated as "meaningful" to degree 1, and contains other positive attributes (such as "pleasant," "interesting," or "satisfied") to high degree. This factor may be termed *pleasing songs.* The second one groups songs rated as "ugly," "tense," or "violent," and may thus be termed *unpleasant songs.*

MovieLens[5] is a well-known dataset in the data mining community. It consists of two data tables. The first one represents a set of users and their attributes such as "gender," "age," "sex," "occupation," while the second one represents a set of movies and their attributes such as "production year" or "genre." The last part of this dataset is a relation between these datasets. This relation contains anonymous ratings of approximately 3900 movies made by 6040 MovieLens users who joined MovieLens in 2000. Ratings are made by means of a five-element (five-star) scale. The values are 1 to 5 with 1 indicating that the user does not like the movie and 5 indicating that the user likes a movie a lot. We analyzed the middle-sized version of the MovieLens dataset.

ASSO_L found some interesting factors in these data. The most significant in terms of explaining the data are the first two factors. One of them is a factor that may be interpreted as *recent action movies* as it contains to high degrees particularly action movies from the period 1991–2000. Another significant factor, whose extent applies to a large group of users, may be interpreted as *comedy and romance movies.*

With respect to the experiments with the earlier mentioned data, we performed their analyses with the GRECOND_L algorithm [6], [8] (see Section I-C). Our findings reveal (cf., also the quantitative evaluation in Section V) that often, the first factors found by ASSO_L are similar to those found by GRECOND_L. On some datasets, however, one of the algorithms finds natural factors that are not found by the other algorithm but we do not know whether this may be told in advance. Nevertheless, ASSO_L in most cases has a better coverage by the first couple of factors compared to GRECOND_L, which is clearly useful when only a small number of highly significant factors is used. On the other hand, GRECOND_L is capable of very accurate factorization in most cases, but it needs a higher number of factors to achieve it. These features are congruent with the purpose of both algorithms: To achieve high coverage by a small number of factors in the case of ASSO_L, and to achieve a highly accurate factorization in the case of GRECOND_L.

V. Evaluation of Algorithms

We now evaluate, on both real and synthetic data, the ability of ASSO_L to extract factors from data and to explain, i.e., reconstruct, the input data. The ability to explain data is measured by the degree of similarity $s(I, A_{\mathcal{F}} \circ B_{\mathcal{F}})$, which is introduced by (4), where $\mathcal{F}$ is the examined set of factors; $\mathcal{F}$ usually consists of the first k factors obtained. Due to the nature of the problem, we also speak of coverage of data by factors. In addition to ASSO_L, we also include in our evaluation the algorithm GRECOND_L for comparison purposes. The purpose of GRECOND_L is different from that of ASSO_L (see Section I-C) and to evaluate the distinct features of these algorithms is a part of our aim in this section.

A. Real Data

For the real data, described in Section IV, Table III provides the numbers of factors obtained by ASSO_L and GRECOND_L that achieve a given prescribed coverage. This means that we observe the least l for which $s(I, A_{\mathcal{F}} \circ B_{\mathcal{F}})$ exceeds the prescribed coverage, where $\mathcal{F}$ is the set of the first l factors produced by the given algorithm. For instance, the second row corresponding to Decathlon shows that one needs two factors in order to have $s(I, A_{\mathcal{F}} \circ B_{\mathcal{F}}) \geq 0.85$ with ASSO_L, while for GRECOND_L, one needs four factors to account for 85% of Decathlon data. In this table, we use "NA" to indicate that it is not possible to achieve the prescribed coverage by the factors produced by ASSO_L, which often happens for higher levels of coverage—similarly as with

[4] `http://www.ipaq.ki.se/`, Belohlavek *et al.*, Inf. Sciences 181(2011), 1774–1786.

[5] `http://grouplens.org/datasets/movielens/`

TABLE III
QUALITY OF DECOMPOSITIONS (REAL DATA)

		number of factors needed	
dataset	s	ASSO$_L$	GRECOND$_L$
Breeds	0.75	2	3
	0.85	3	5
	0.95	NA	9
	1	NA	16
Decathlon	0.75	1	1
	0.85	2	4
	0.95	NA	8
	1	NA	15
IPAQ	0.75	1	8
	0.85	1	12
	0.95	NA	18
	1	NA	32
Music	0.75	1	7
	0.85	NA	13
	0.95	NA	24
	1	NA	36
MovieLens	0.75	1	419
	0.85	2	727
	0.95	NA	954
	1	NA	1 123

TABLE IV
SYNTHETIC DATA

dataset	size	$\lvert L\rvert$	k	distribution on L in A and B
Set 1	50×50	3	10	$[\frac{1}{3}\ \frac{1}{3}\ \frac{1}{3}]$
Set 2	50×50	5	10	$[\frac{1}{8}\ \frac{1}{8}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}]$
Set 3	100×50	5	25	$[\frac{1}{8}\ \frac{1}{8}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}]$
Set 4	100×100	5	20	$[\frac{1}{8}\ \frac{1}{8}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}]$

the classical ASSO algorithm for Boolean data. Note that "NA" never appears for GRECOND$_L$ because GRECOND$_L$ eventually stops with an exact decomposition.

The results reveal that ASSO$_L$ has a good ability to achieve a reasonably high coverage using the first few factors produced. This ability is considerably better compared to GRECOND$_L$. On the other hand, GRECOND$_L$ is capable to achieve a very precise decomposition, in most cases still with a reasonably small number of factors (the only exception to this is the MovieLens data, for reasons not known to us).

B. Synthetic Data

The synthetic data we used are grouped in collections Set 1–4. Each collection contains 500 $n \times m$ data matrices I and the characteristics of these matrices are given in Table IV. As usual, every I is obtained as a matrix product of randomly generated matrices A and B of dimensions $n \times k$ and $k \times m$. The entries in A and B are the truth degrees in scale L and are selected to follow a prescribed probability distribution. For example, in Set 3 we employed a five-element scale $L = \{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\}$ with the probabilities $p(a)$ of the degrees $a \in L$ in A and B being $p(0) = p(\frac{1}{4}) = \frac{1}{8}$ and $p(\frac{1}{2}) = p(\frac{3}{4}) = p(1) = \frac{1}{4}$. Observe that this scenario generalizes that from the Boolean setting, i.e., the usually considered densities of Boolean matrices. For instance,

TABLE V
COVERAGE s AND $s_=$ BY THE FIRST k FACTORS

		coverage $s/s_=$ by the first k factors	
dataset	k	ASSO$_L$	GRECOND$_L$
Set 1	1	0.864/0.729	0.525/0.309
	4	0.905/0.810	0.866/0.744
	8	0.912/0.826	0.968/0.937
	11	–	1/1
Set 2	1	0.861/0.525	0.620/0.253
	4	0.865/0.539	0.893/0.666
	8	–	0.981/0.929
	10	–	1/1
Set 3	1	0.946/0.790	0.684/0.188
	5	0.955/0.825	0.871/0.509
	10	0.955/0.825	0.916/0.670
	20	–	0.970/0.883
	48	–	1/1
Set 4	1	0.943/0.829	0.648/0.205
	5	0.961/0.850	0.881/0.580
	10	0.961/0.850	0.947/0.795
	20	–	0.991/0.979
	37	–	1/1

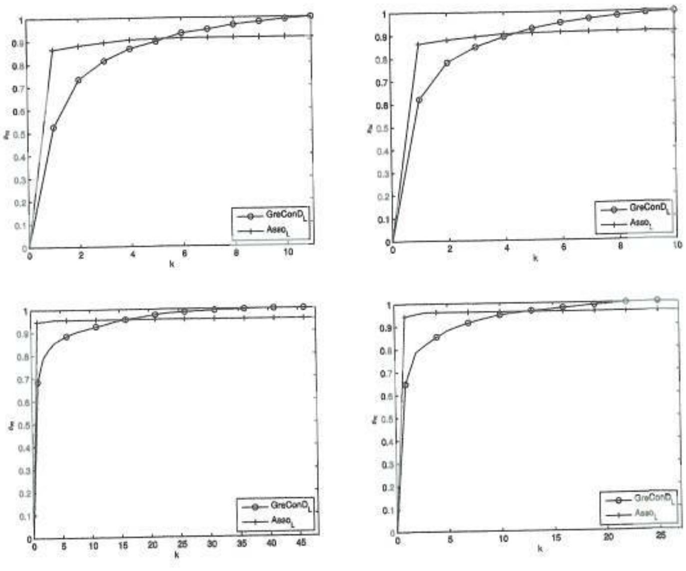


Fig. 4. Coverage s by the first k factors.

for $L = \{0, 1\}$, the distribution $[\frac{1}{4}\ \frac{3}{4}]$ corresponds to density 0.75.

Selected results in terms of the coverage s, see (4), by the first k factors for the datasets are provided by Table V and Fig. 4. In addition, we include the percentage $s_=$ of entries $\langle i, j\rangle$ for which I_{ij} equals $(A_{\mathcal{F}_k} \circ B_{\mathcal{F}_k})_{ij}$. Here, $\mathcal{F}_k$ denotes the set of the first k factors. Therefore, $s_=$ is a measure that is more strict than s: it only considers equal entries, and does not take into account closeness of the corresponding entries in I and $A_{\mathcal{F}_k} \circ B_{\mathcal{F}_k}$, respectively. The character "–" for ASSO$_L$ indicates that no additional factors have been produced with increasing k.

As one may see, the results for synthetic data confirm the findings observed for real data: ASSO$_L$ is good at achieving a reasonable coverage with the first few factors, while GRECOND$_L$ outperforms ASSO$_L$ in achieving almost exact decompositions.

TABLE VI
SYNTHETIC DATA

dataset	size	$\lvert L\rvert$	k	distribution on L in A and B
Set 1	150×150	3	10	$[\frac{1}{3}\ \frac{1}{3}\ \frac{1}{3}]$
Set 2	150×100	5	10	$[\frac{1}{8}\ \frac{1}{8}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}]$
Set 3	100×150	11	10	$[\frac{1}{16}\ \frac{1}{16}\ \frac{1}{16}\ \frac{1}{16}\ \frac{1}{16}\ \frac{1}{16}\ \frac{1}{8}\ \frac{1}{8}\ \frac{1}{8}\ \frac{1}{8}\ \frac{1}{8}]$
Set 4	100×100	21	10	$\frac{1}{21}$ for all
Set 5	100×100	101	10	$\frac{1}{101}$ for all

TABLE VII
COVERAGE s BY THE FIRST FACTOR/BY ALL FACTORS OBTAINED FOR DIFFERENT VALUES OF τ

dataset	$\tau = 0.85$	$\tau = 0.9$	$\tau = 0.95$
Set 1	0.86/0.88	0.85/0.88	0.83/0.90
Set 2	0.77/0.78	0.78/0.79	0.79/0.80
Set 3	0.765/0.766	0.765/0.766	0.757/0.758
Set 4	0.774/0.775	0.774/0.775	0.774/0.775
Set 5	0.791/0.793	0.791/0.793	0.791/0.793

Let us mention that s tends to be large even for a small number of factors computed. The values of s are higher than what one observes for Boolean data. The reason of this interesting behavior is provided in Section V-D.

C. Role of τ in ASSO$_L$ Algorithm

Interestingly, the setting of multiple truth values (that is to say, instead of only 0 and 1, L involves several other truth values) presents us with a new phenomenon, which is advantageous from the data analysis viewpoint. For Boolean data, as is well known, the particular choice of the threshold τ influences grossly the performance of the classic ASSO. A clear and intuitive explanation of this behavior is that if 0 and 1 are the only possible matrix entries, the decision of whether to round off the particular confidence value $c(p,q)$ to 0 or 1 by means of thresholding via τ is significant: if $c(p,q)$ is not rounded to 1, it is rounded to the *completely opposite* value, namely 0, and vice versa. In the setting with several degrees, as we observed, the actual value of τ becomes less significant as the number of degrees in scale L increases. Clearly, this is a good property for a user who, as a consequence, need not pay much attention to the choice of τ. Note that in the Boolean case, there are no principles available as to how to choose τ except for selecting τ *ad hoc*.

To examine the phenomenon under investigation, we employed synthetic data whose characteristics are provided by Table VI. The values of the coverage s as defined by (4) by the first factor and by all the factors obtained by ASSO$_L$ are displayed in Table VII. These values are observed for different values of τ (we use values around 0.9 which is recommended in the Boolean case and which also yields good results for fuzzy attributes). As is clear from the results, the coverage values for different values of τ tend to be the same as the number of truth values in L increases (i.e., as we go from Set 1 to Set 5). This kind of behavior is also apparent from Table VIII in which we observe the values of the stricter measure, $s_=$, instead of s. Let

TABLE VIII
COVERAGE $s_=$ BY THE FIRST FACTOR/BY ALL FACTORS OBTAINED BY ASSO$_L$ FOR DIFFERENT VALUES OF τ

dataset	$\tau = 0.85$	$\tau = 0.9$	$\tau = 0.95$
Set 1	0.74/0.76	0.72/0.77	0.66/0.80
Set 2	0.36/0.37	0.37/0.38	0.38/0.41
Set 3	0.11/0.12	0.11/0.12	0.10/0.11
Set 4	0.049/0.063	0.049/0.063	0.049/0.063
Set 5	0.005/0.008	0.005/0.008	0.005/0.008

us mention that in this table, the low values for scales L with a larger number of degrees, which correspond to low numbers of entries for which the input and the reconstructed matrices have equal values, correspond to the aim of ASSO$_L$ to generate approximate rather than exact decompositions (even with a low number of exactly equal values, the matrices's entries are very similar and thus approximately equal).

D. Discussion and Conclusion

The experiments demonstrate that ASSO$_L$ performs well in both approximately solving the *DBP*(L) problem as well as in computing from the data natural, well-interpretable factors. In comparison to the GRECOND$_L$ algorithm, the experiments reveal that the first couple of factors produced by ASSO$_L$ have a better coverage compared to the same number of factors produced by GRECOND$_L$. Beyond certain coverage, ASSO$_L$ stops producing factors and is not capable of computing exact decomposition of the input matrix, as opposed to GRECOND$_L$, which always achieves an exact decomposition and actually needs only a reasonably small number of factors to compute an almost exact decomposition. This is congruent with the aims of both these algorithms: While ASSO$_L$ is designed to solve the problem of computing a given small number of factors with a high coverage of data by these factors, GRECOND$_L$ is designed to compute exact or almost exact decompositions. Nevertheless, both ASSO$_L$ and GRECOND$_L$ produce natural and well-interpretable factors. As described earlier, a factor revealed by one of these algorithms need not be revealed by the other algorithm, which is a common phenomenon in factor analysis broadly conceived. Both ASSO$_L$ and GRECOND$_L$ may therefore be regarded as effective factorization algorithms which naturally complement each other in accordance with the purposes the are designed for.

Except for the Łukasiewicz operations $\otimes$ and $\rightarrow$, we also did experiments with other ones. We observed similar tendencies when comparing ASSO$_L$ to GRECOND$_L$. Let us also mention that we did not explore the interesting and challenging problem of how to select good operations $\otimes$ and $\rightarrow$ based on the data, which we will consider worth exploring in future.

Let us conclude by mentioning that employment of fuzzy logic methods to factor analysis of ordinal data is at an initial stage of exploration. Broadly conceived, further research on a possible combination of classical factor analysis methods and their variants developed for ordinal data on the one hand, and methods based on fuzzy logic is highly desirable to exploit the best of the two areas.

VI. Future Research

In particular, future research shall include the following topics:

1) Further theoretical progress in understanding factorizations and the corresponding matrix calculus of matrices over scales with truth degrees; particular attention needs to be paid to problems and issues degenerate in the Boolean case such as closeness of matrix entries or the choice of the aggregation operation $\otimes$.
2) Further development of factorization algorithms for ordinal data, which includes algorithms inspired by the existing classical factorization algorithms for Boolean matrices (BMF) algorithms, as well as exploitation of attribute dependencies in ordinal data; and further exploration of the present factorization scheme as regards its sensitivity to changes in the input data and in the choice of the parameter τ.
3) Exploration of noise in ordinal data which we have not considered here.
4) Development of real-case studies in factor analysis of ordinal data and development of applications in machine learning.
5) Let us also mention an interesting phenomenon in the setting of fuzzy attributes that we observed.

For Boolean data, the values I_{ij} in the input matrix I are approximated by 0 or 1 of $(A_{\mathcal{F}} \circ B_{\mathcal{F}})_{ij}$ only. Hence, in case of mismatch, the entry $\langle i, j\rangle$ contributes by $I_{ij} \leftrightarrow (A_{\mathcal{F}} \circ B_{\mathcal{F}})_{ij} = 0$ to the numerator in (4). As the number of truth degrees in L increases, the situation is different. For instance, if L contains $\frac{1}{2}$ and if $0 \leftrightarrow \frac{1}{2} = 1 \leftrightarrow \frac{1}{2} = \frac{1}{2}$, then the trivial matrix $A_{\mathcal{F}} \circ B_{\mathcal{F}}$ with all entries equal to $\frac{1}{2}$ always satisfies $s(I, A_{\mathcal{F}} \circ B_{\mathcal{F}}) \geq \frac{1}{2}$. One therefore has to be aware of this effect of presence in L of the "middle" degree ($\frac{1}{2}$) on the values of s. This issue needs to be carefully examined in future research.

References

[1] S. Arora and B. Barak, *Computational Complexity: A Modern Approach*. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[2] E. Bartl, "Minimal solutions of generalized fuzzy relational equations: Probabilistic algorithm based on greedy approach," *Fuzzy Sets Syst.*, vol. 260, pp. 25–42, 2015.

[3] E. Bartl and G. J. Klir, "Fuzzy relational equations in general framework," *Int. J. General Syst.*, vol. 43, pp. 1–18, 2014.

[4] R. Belohlavek, *Fuzzy Relational Systems: Foundations and Principles*. New York, NY, USA: Kluwer, 2002.

[5] R. Belohlavek, "Optimal decompositions of matrices with entries from residuated lattices," *J. Logic Comput.*, vol. 22, pp. 1405–1425, 2012.

[6] R. Belohlavek and M. Krmelova, "Factor analysis of ordinal data via decomposition of matrices with grades," *Ann. Math. Artif. Intell.*, vol. 72, pp. 23–44, 2014.

[7] R. Belohlavek and V. Vychodil, "Discovery of optimal factors in binary data via a novel method of matrix decomposition," *J. Comput. Syst. Sci.*, vol. 76, no. 1, pp. 3–20, 2010.

[8] R. Belohlavek and V. Vychodil, "Factorization of matrices with grades," *Fuzzy Sets Syst.*, vol. 292, no. 1, pp. 85–97, 2016.

[9] K. Flaska and P. Cakirpaloglu, "Identification of the multidimensional model of subjective time experience," in *International Studies in Time Perspective*. Coimbra, Portugal: Imprensa da Universidade de Coimbra, 2013, pp. 259–273.

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

[11] F. Geerts, B. Goethals, and T. Mielikäinen, "Tiling databases," in *Proc. Discovery Sci.*, 2004, pp. 278–289.

[12] J. S. Golan, *Semirings and Their Applications*. New York, NY, USA: Springer, 1999.

[13] S. Gottwald, *A Treatise on Many-Valued Logics*. Baldock, U.K.: Research Studies Press, 2001.

[14] P. Hájek, *Metamathematics of Fuzzy Logic*. New York, NY, USA: Kluwer, 1998.

[15] D. Lee and H. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.

[16] P. Miettinen, "The Boolean column and column-row matrix decompositions," *Data Mining Knowl. Discovery*, vol. 17, pp. 39–56, 2008.

[17] P. Miettinen, "Sparse Boolean matrix factorizations," in *Proc. IEEE Int. Conf. Data Mining*, 2010, pp. 935–940.

[18] P. Miettinen, T. Mielikäinen, A. Gionis, G. Das, and H. Mannila, "The discrete basis problem," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 10, pp. 1348–62, Oct. 2008.

[19] P. Miettinen and J. Vreeken, "Model order selection for Boolean matrix factorization," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2011, pp. 51–59.

[20] G. T. Miller, "The magical number seven, plus or minus two," *Psychol. Rev.*, vol. 63, pp. 81–97, 1956.

[21] D. S. Nau *et al.*, "A mathematical analysis of human leukocyte antigen serology," *Math. Biosci.*, vol. 40, pp. 243–270, 1978.

[22] D. A. Simovici and C. Djeraba, *Mathematical Tools for Data Mining*. New York, NY, USA: Springer, 2008.

[23] L. Stockmeyer, "The set basis problem is NP-complete," IBM, Yorktown, NY, USA, Tech. Rep. RC5431, 1975.

[24] N. Tatti, T. Mielikäinen, A. Gionis, and H. Mannila, "What is the dimension of your binary data?" in *Proc. IEEE 6th Int. Conf. Data Mining*, 2006, pp. 603–612.

[25] L. A. Zadeh, "Probability measures of fuzzy events," *J. Math. Anal. Appl.*, vol. 23, pp. 421–427, 1968.

Radim Belohlavek (SM'07) received the Ph.D. degree in computer science from Technical University of Ostrava, in 1998 and Ph.D degree in mathematics from Palacky University Olomouc, in 2001. He is currently a Professor in computer science with Palacky University, Olomouc, Czech Republic. In the past, he was a tenured Professor with the State University of New York, USA. His academic interests include the areas of uncertainty and information, data analysis, fuzzy logic and fuzzy sets, applied algebra, and logic. He has authored or coauthored four books (Kluwer, Springer, MIT Press, Oxford University Press) and more than 150 journal and conference papers in these areas.

Dr. Belohlavek is a member of Editorial Boards of several journals including *J. Computer and System Sciences*, *Fundamenta Informaticae*, and *Fuzzy Sets and Systems*, and a member of ACM.

Marketa Trneckova received the Ph.D. degree in computer science from Palacky University, Olomouc, Czech Republic, in 2017.

She has authored or coauthored several papers in journals and conferences, including the *Annals of Mathematics, Artificial Intelligence,* and IEEE International Conference on Data Mining. Her research interests include data analysis and fuzzy logic.

Dr. Trneckova is a member of ACM.