Beyond Boolean Matrix Decompositions: Toward Factor Analysis and Dimensionality Reduction of Ordinal Data

Radim Belohlavek
Palacky University, Olomouc, Czech Republic
Email: radim.belohlavek@acm.org

Marketa Krmelova
Palacky University, Olomouc, Czech Republic
marketa.krmelova@gmail.com

***Abstract*—Boolean matrix factorization (BMF), or decomposition, received a considerable attention in data mining research. In this paper, we argue that research should extend beyond the Boolean case toward more general type of data such as ordinal data. Technically, such extension amounts to replacement of the two-element Boolean algebra utilized in BMF by more general structures, which brings non-trivial challenges. We first present the problem formulation, survey the existing literature, and provide an illustrative example. Second, we present new theorems regarding decompositions of matrices with ordinal data. Third, we propose a new algorithm based on these results along with an experimental evaluation.**

***Keywords*-ordinal data; matrix decomposition; Galois connection; concept lattice; aggregation; factor analysis**

I. Introduction

A. Motivation, Problem Description, Contributions

In this paper, we are concerned with extending the problems and methods of BMF toward a more general case. Instead of Boolean matrices, we consider matrices with entries in a partially ordered set L bounded by 0 and 1, such as the five-element scale $L = \{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\}$. In a Boolean matrix I, the entries represent presence/absence of attributes, i.e. $I_{ij} = 1/0$ if the object i has/does not have the attribute j. In the more general case, the entries represent degrees to which attributes are present, with 0 and 1 representing full absence and full presence.

Data represented by matrices over scales L appears in many areas. Examples are questionnaires in which the respondents (objects, matrix rows) provide answers to questions (attributes, matrix columns) in the form of grades (degrees, levels), such as the degree of satisfaction of the customer (respondent) with a particular product or service (question), or various kinds of performance data in which objects are evaluated in terms of attributes/criteria. In such data, the degrees in L are often small in number and have linguistic labels such as "bad", "weak", "medium", "good", "excellent", in congruence with the well-known Miller's 7 ± 2 phenomenon [11]. In addition, real-valued data may often conveniently be transformed to data with degrees by an established or *ad hoc* scaling. Clearly, Boolean matrices are a particular case for $L = \{0, 1\}$.

Supported by grant GACR No. P103/11/1456.

Central to this paper is the decomposition problem and its variants: Given an $n \times m$ (object-attribute) matrix I with entries in a scale L, find a decomposition of I into a product

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

of two matrices with entries in L, namely an $n \times k$ (object-factor) matrix A and a $k \times m$ (factor-attribute) matrix B with k reasonably small (more in Section II). The Boolean matrix product $\circ$, defined by

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

may naturally be generalized in the setting with grades. Namely, one may put

$$(A \circ B)_{ij} = \bigvee_{l=1}^{k} A_{il} \otimes B_{lj}, \tag{3}$$

where $\bigvee$ denotes supremum in L and $\otimes$ is an appropriate conjunctive aggregation function generalizing the truth function of classical conjunction. Such functions are studied in many-valued logics (more in Section II). Importantly, the model (1) retains its meaning from the Boolean case. Namely, according to the principles of many-valued logic [8], (1) says that the degree to which object i has attribute j equals to the truth degree of the proposition "there exists factor l such that l applies to i and j is one of the particular manifestations of l". Apart from this logical interpretation, (3) may simply be understood as a new type of matrix product.

We formulate two decomposition problems and present an illustrative example. Next, we present new results regarding the geometry of decompositions, which help answer the question of where to focus in computing decompositions. In view of these results, we propose a new decomposition algorithm for matrices over scales, inspired by [2], [3], [4], and provide an experimental evaluation.

B. Related Work

As the literature on matrix decompositions is too numerous to be included here, we survey only directly related work and refer the reader to [5] and [9] for further references. Regarding BMF, [12], in which NP-completeness of the basic problem is observed, is one of the first papers in this area. The interest in BMF in data mining is due to Miettinen: [9] with the ASSO algorithm whose extension for scales we use in experiments; further papers include [10] (selecting the number of factors using MDL). [7] is the first paper on "tiling" Boolean data, which is closely related to BMF. The utilization of formal concepts in BMF, their universality, and a a fast BMF algorithm and other issues are the content of [5]. Matrices over scales and other structures are examined in many papers, including those on matrices over semi-ring-like algebras and binary fuzzy relations [1], [8]. Direct predecessor to this paper are [2], [4].

II. DECOMPOSITION PROBLEM IN DETAIL

A. Scales and Aggregation

We assume that the partial order $\leq$ makes L a complete lattice [1], [8] (arbitrary infima $\bigwedge$ and suprema $\bigvee$ exist). This is automatically satisfied if L is a chain (i.e. $a \leq b$ or $b \leq a$ for every $a, b \in L$), in which case $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$. We also assume that $\otimes$ is commutative, associative, has 1 as its neutral element ($a \otimes 1 = a = 1 \otimes a$), and distributes over arbitrary suprema, i.e. $a \otimes (\bigvee_{j \in J} b_j) = \bigvee_{j \in J}(a \otimes b_j)$. Such $\otimes$ is considered a truth function of (many-valued) conjunction [8] and the classical conjunction is a particular case of $\otimes$ for $L = \{0, 1\}$. That is, if a and b are truth degrees of propositions p_1 and p_2, then $a \otimes b$ is the truth degree of $p_1 \& p_2$. Such L conforms to the structure of a complete residuated lattice and satisfies several properties, utilized in Section IV, for which we refer to [8]. Among the properties are natural properties of conjunction such as $a \otimes 0 = 0$ and monotony w.r.t. $\leq$. 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

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

Many examples of such scales are known in many-valued logic [8], among them those where L is the real unit interval $[0, 1]$ or its finite equidistant subinterval, i.e. $L = \{0, \frac{1}{n}, \ldots, \frac{n-1}{n}, 1\}$. In particular, we use scales with the Łukasiewicz operations, i.e.

$$a \otimes b = \max(0, a + b - 1) \text{ and } a \rightarrow b = \min(1, 1 - a + b),$$

which have convenient properties. (Many other $\otimes$ are available, such as $a \otimes b = \min(a, b)$, and all of them coincide with classical conjunction for $a, b \in \{0, 1\}$ [8].) In what follows, we assume that the scale L is equipped with the above operation $\otimes$ and, for simplicity, that L is a finite chain in the rest of this paper (the general results appear in an extended version of this paper).

B. Decomposition Problem and Its Two Variants

We now present two concrete variants, the discrete basis problem over L, DBP(L), and the approximate factorization problem over L, AFP(L), see [9], [3], [5] for the Boolean case. Let $s_L : L \times L \rightarrow [0, 1]$ be an appropriate function measuring closeness of degrees in L. For matrices $I, J \in L^{n \times m}$, put

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

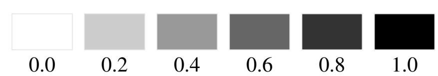
In general, we require $s_L(a, b) = 1$ if and only if $a = b$, and $s_L(0, 1) = s_L(1, 0) = 0$, in which case $s(I, J) = 1$ if and only if $I = J$. s_L may be defined by $s_L(a, b) = a \leftrightarrow b$, where $a \leftrightarrow b = \min(a \rightarrow b, b \rightarrow a)$ is the so-called biresiduum (degree of equivalence from a logical point of view) of a and b ($\rightarrow$ is the residuum of $\otimes$). For the Łukasiewicz operations we obtain $s_L(a, b) = 1 - |a - b|$ this way. One easily observes that the normalized Hamming distance of Boolean matrices I and J, often used in BMF, is just $1 - s(I, J)$ provided $s_L(a, b) = a \leftrightarrow b$. Problems:

– *DBP(L)*: Given $I \in L^{n \times m}$ and a positive integer k, find $A \in L^{n \times k}$ and $B \in L^{k \times m}$ that maximize $s(I, A \circ B)$.
– *AFP(L)*: Given I and prescribed error $\varepsilon \in [0, 1]$, find $A \in L^{n \times k}$ and $B \in L^{k \times m}$ with k as small as possible such that $s(I, A \circ B) \geq \varepsilon$.

In view of [9], [5] and the remarks above, DBP(L) and AFP(L) are clearly NP-hard optimization problems. Hence, we need to resort to algorithms providing approximate solutions.

III. ILLUSTRATIVE EXAMPLE

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 VI). 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 $\otimes$ (Section II-A). We represent the grades in L by shades of gray as follows:



.

The 5×11 object-attribute matrix I and its decomposition $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$ into the object-factor and factor-attribute matrices $A_{\mathcal{F}}$ and $B_{\mathcal{F}}$ are shown in Fig. 1. The decomposition is obtained using the algorithm described in Section V and utilizes a set $\mathcal{F}$ of formal concepts as factors (these notions are fully described below; we proceed on intuitive ground).

Every factor F_l 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 the three kinds of friendliness and affection (attributes with high degrees in the first row of $B_{\mathcal{F}}$) and applies in particular to Labradors, Golden Retrievers and Beagels (breeds with high degrees in the first column of

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

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

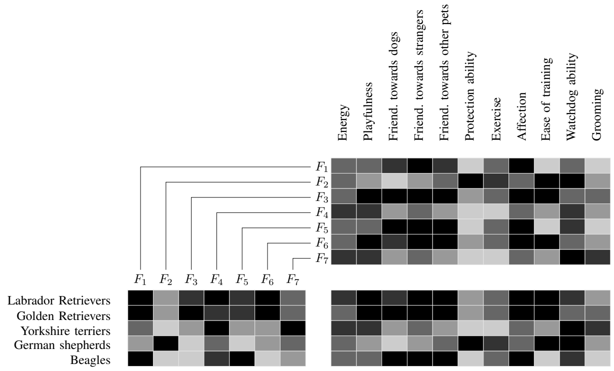


Figure 1: Decomposition $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$. I, $A_{\mathcal{F}}$, and $B_{\mathcal{F}}$ are the bottom-right, bottom-left, and top matrix, respectively.

$A_{\mathcal{F}}$), and to some extent to Yorkshires. The factor may hence be termed *friendliness*. In a similar manner, F_2 and F_3 may be interpreted as *guardian dog* and *dogs suitable for kids*. Interestingly, F_1, F_2, and F_3 explain, by and large, the whole data and hence, the other factors may be neglected. Namely, denoting by $A_{\mathcal{F}_3}$ and $B_{\mathcal{F}_3}$ the 5×3 and 3×11 matrices (parts of $A_{\mathcal{F}}$ and $B_{\mathcal{F}}$), the degree $s(I, A_{\mathcal{F}_3} \circ B_{\mathcal{F}_3})$ of similarity of I to $A_{\mathcal{F}_3} \circ B_{\mathcal{F}_3}$, i.e. reconstructability of the original data I from the first three factors, equals 0.92.

IV. Geometry of Decomposition and Essential Parts of Matrices over Scales

A. Basic Observations, Coverage, Formal Concepts

For two matrices $J_1, J_2 \in L^{n\times m}$ we put

$$J_1 \leq J_2 \text{ iff } (J_1)_{ij} \leq (J_1)_{ij} \text{ for every } i, j \qquad (5)$$

in which case we say that J_1 is *contained* in J_2. $J \in L^{n\times m}$ is called a *rectangle* if $J = C \circ D$ for some $C \in L^{n\times 1}$ and $D \in L^{1\times m}$. (In the Boolean case, rectangles are just tiles in terms of [7].) Unlike the Boolean case, the C and D for which $J = C \circ D$ are not unique. We say that a rectangle J covers $\langle i, j\rangle$ in I if $J_{ij} = I_{ij}$.

Lemma 1: The following conditions are equivalent:

(a) $I = A \circ B$ for some $A \in L^{n\times k}$ and $B \in L^{k\times m}$.

(b) There exist rectangles $J_1, \dots, J_k \in L^{n\times m}$ such that $I = J_1 \vee \cdots \vee J_k$, i.e. $I_{ij} = \max_{l=1}^{k}(J_l)_{ij}$.

(c) There exist rectangles $J_1, \dots, J_k \in L^{n\times m}$ contained in I such that every $\langle i, j\rangle$ in I is covered by some J_l.

Importantly, Lemma 1 allows us to consider the problem of decomposition of I as the problem of covering the entries in I by rectangles contained in I. Next we show that optimal in such coverage are rectangles that correspond to so-called formal concepts of I. A *formal concept* in $I \in L^{n\times m}$ [1] may be looked at as a pair $\langle C, D\rangle$ of vectors $C \in L^{1\times n}$ and $D \in L^{1\times m}$ satisfying $C^{\uparrow} = D$ and $D^{\downarrow} = C$, where

$$(C^{\uparrow})_j = \textstyle\bigwedge_{i=1}^{n}(C_i \rightarrow I_{ij}) \text{ and } (D^{\downarrow})_i = \bigwedge_{j=1}^{m}(D_j \rightarrow I_{ij}).$$

The set $\mathcal{B}(I) = \{\langle C, D\rangle \mid C^{\uparrow} = D, D^{\downarrow} = C\}$ of all formal concepts equipped with a partial order $\leq$, defined by $\langle C_1, D_1\rangle \leq \langle C_2, D_2\rangle$ iff $C_1 \leq C_2$ (iff $D_2 \leq D_1$), forms a complete lattice, called the *concept lattice* of I.

For an (indexed) set $\mathcal{F} = \{\langle C_1, D_1\rangle, \dots, \langle C_k, D_k\rangle\} \subseteq \mathcal{B}(I)$ of formal concepts of I, define the $n \times k$ and $k \times m$ matrices $A_{\mathcal{F}}$ and $B_{\mathcal{F}}$ by

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

By Lemma 1, $A_{\mathcal{F}} \circ B_{\mathcal{F}}$ is the $\vee$-superposition of the rectangles $C_l^{\mathrm{T}} \circ D_l$. The following theorem shows that formal concepts of I are optimal factors for the approximate decompositions of I providing a *from-below approximation* of I, i.e. $A \circ B \leq I$ (generalizing [2] which concerns exact decompositions $I = A \circ B$).

Theorem 1: Let for $I \in L^{n\times m}$ there exist $A \in L^{n\times k}$ and $B \in L^{k\times m}$ such that $A \circ B \leq I$. Then there exists a set $\mathcal{F} \subseteq \mathcal{B}(I)$ of formal concepts of I with $|\mathcal{F}| \leq k$ such that for the $n \times |\mathcal{F}|$ and $|\mathcal{F}| \times m$ matrices $A_{\mathcal{F}}$ and $B_{\mathcal{F}}$ over L we have

$$s(I, A_{\mathcal{F}} \circ B_{\mathcal{F}}) \geq s(I, A \circ B).$$

B. Essential Parts of Matrices over Scales

Interestingly, an inspection of the concept lattice $\mathcal{B}(I)$ reveals an important fact—the possibility to differentiate the role of matrix entries for decompositions. In, particular, we identify so-called essential part of I, the part to focus on when computing decompositions.

Definition 1: $J \leq I$ is called an *essential part* of I if J is minimal w.r.t. $\leq$ having the property that for every $\mathcal{F} \subseteq \mathcal{B}(I)$ we have: if $J \leq A_{\mathcal{F}} \circ B_{\mathcal{F}}$ then $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$.

In other words, the coverage of J by formal concepts of I guarantees the coverage of all entries in I. It turns out that intervals in $\mathcal{B}(I)$ play a crucial role. For $C \in L^{1\times n}$, $D \in L^{1\times m}$, put $\gamma(C) = \langle C^{\uparrow\downarrow}, C^{\uparrow}\rangle$ and $\mu(D) = \langle D^{\downarrow}, D^{\downarrow\uparrow}\rangle$, and denote by $\mathcal{I}_{C,D}$ the interval

$$\mathcal{I}_{C,D} = [\gamma(C), \mu(D)].$$

Let $\{^{a}/i\}$ denote the "singleton" vector C with zero components except $C_i = a$. Let

$$\mathbf{I}_{ij} = \{\mathcal{I}_{\{^{a}/i\},\{^{b}/j\}} \mid a \otimes b = I_{ij}\}$$

and put

$$\mathcal{I}_{ij} = \bigcup \mathbf{I}_{ij}.$$

We now present an important theorem that identifies which formal concepts of I cover a given entry of I.

Theorem 2: A rectangle corresponding to $\langle E, F\rangle \in \mathcal{B}(X, Y, I)$ covers $\langle i, j\rangle$ in I iff $\langle E, F\rangle \in \mathcal{I}_{ij}$.

Denote by $\mathcal{E}(I) \in L^{n\times m}$ the matrix over L defined by

$$(\mathcal{E}(I))_{ij} = \begin{cases} I_{ij} & \text{if } \mathcal{I}_{ij} \text{ is non-empty and minimal w.r.t. } \subseteq, \\ 0 & \text{otherwise,} \end{cases}$$

where $\subseteq$ denotes inclusion of intervals in $\mathcal{B}(I)$. The following two theorems provide are the main result in this section and are utilized in the new algorithm.

Theorem 3: $\mathcal{E}(I)$ is the unique essential part of I.

Theorem 4: Let $\mathcal{G} \subseteq \mathcal{B}(\mathcal{E}(I))$ be a set of factor concepts of $\mathcal{E}(I)$, i.e. $\mathcal{E}(I) = A_{\mathcal{G}} \circ B_{\mathcal{G}}$. Then every set $\mathcal{F} \subseteq \mathcal{B}(I)$ containing for each $\langle C, D\rangle \in \mathcal{G}$ at least one concept from $\mathcal{I}_{C,D}$ is a set of factor concepts of I, i.e. $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$.

V. GREESS$_L$ALGORITHM

GREESS$_L$, inspired by the algorithm in [3], is based on the results in Section IV and some other facts mentioned below, is primarily designed for AFP(L), but can also be used for DBP(L). The pseudocode describes the algorithm computing an exact decomposition of I but an easy modification makes it an algorithm computing ε-approximate decompositions (in l. 3, stop when precision ε is reached).

In the pseudocodes, $\emptyset$ denotes the empty set or the vector full of zeroes, depending on the context, $F \vee \{^a/_j\}$ denotes F with the component j updated to $F_j \vee a$, and $C \otimes D$ denotes the crossproduct of C and D, i.e. the rectangle for which $(C \otimes D)_{ij} = C_i \otimes D_j$. Moreover, $cov(U, F, J)$ and $cov_I(U, D, \mathcal{E})$ denote the number of $\langle i, j\rangle \in U$ covered in I by the rectangle $F^{\downarrow_J} \otimes F^{\downarrow_J\uparrow_J}$ and $(D^{\downarrow_\mathcal{E}})^{\uparrow_I\downarrow_I} \otimes (D^{\downarrow_\mathcal{E}\uparrow_\mathcal{E}})^{\downarrow_I\uparrow_I}$, respectively. The fuzzy-set-like notation $\{^a/_j\} \in C^{\uparrow_I} \setminus F$ means $F_j < a \leq C_j^{\uparrow_I}$.

COMPUTEINTERVALS first computes $\mathcal{E}(I)$ and then a set $\mathcal{G}$ of factors of $\mathcal{E}(I)$, each $\langle C, D\rangle \in \mathcal{G}$ representing the interval $\mathcal{I}_{C,D}$ in $\mathcal{B}(I)$ from which it is possible to obtain a decomposition of I according to Theorem 4. In fact, we use an improvement of Theorem 4: for $\mathcal{G}$, it suffices that the crossproducts $C^{\uparrow_I\downarrow_I} \otimes D^{\downarrow_I\uparrow_I}$ corresponding to $\langle C, D\rangle \in \mathcal{G}$ cover all entries in I (l. 12). The formal concepts in $\mathcal{G}$ are computed in a greedy manner from $\mathcal{E}(I)$ in l. 5–10. The entries covered by $C^{\uparrow_I\downarrow_I} \otimes D^{\downarrow_I\uparrow_I}$ are removed from U. The selection is repeated until U is empty. With $\mathcal{G}$ obtained this way, GREESS$_L$ performs a greedy search for factors, in the intervals $\mathcal{I}_{C,D}$, $\langle C, D\rangle \in \mathcal{G}$, in l. 3–21. For every $\mathcal{I}_{C,D}$, we select the formal concept in $\mathcal{I}_{C,D}$ with best coverage in l. 6–11 in a manner similar to the one used in COMPUTEINTERVALS. The best found concept $\langle E', F'\rangle$ over all the intervals is then added to $\mathcal{F}$ in l. 18; $\mathcal{I}_{C',D'}$ in which $\langle E', F'\rangle$ was found is removed from $\mathcal{G}$ in l. 19. Correctness of GREESS$_L$ easily follows from Theorems 3 and 4 and the above considerations.

Algorithm 1: GREESS$_L$

Input: matrix I with entries in scale L
Output: set $\mathcal{F}$ of factors for which $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$
1 $\mathcal{G} \leftarrow$ COMPUTEINTERVALS(I)
2 $U \leftarrow \{\langle i, j\rangle | I_{ij} > 0\}$; $\mathcal{F} \leftarrow \emptyset$
3 **while** *U is non-empty* **do**
4 $\quad s \leftarrow 0$
5 $\quad$**foreach** $\langle C, D\rangle \in \mathcal{G}$ **do**
6 $\quad\quad J \leftarrow D^{\downarrow_I} \otimes C^{\uparrow_I}$; $F \leftarrow \emptyset$; $s_{\langle C,D\rangle} \leftarrow 0$
7 $\quad\quad$**while** *exists* $\{^a/_j\} \in C^{\uparrow_I} \setminus F$ *s.t.* $cov(U, F \vee \{^a/_j\}, J) > s_{\langle C,D\rangle}$ **do**
8 $\quad\quad\quad$select $\{^a/_j\}$ maximizing $cov(U, F \vee \{^a/_j\}, J)$
9 $\quad\quad\quad F \leftarrow (F \vee \{^a/_j\})^{\downarrow_J\uparrow_J}$; $E \leftarrow (F \vee \{^a/_j\})^{\downarrow_J}$
10 $\quad\quad\quad s_{\langle C,D\rangle} \leftarrow cov(U, F, J)$
11 $\quad\quad$**end**
12 $\quad\quad$**if** $s_{\langle C,D\rangle} > s$ **then**
13 $\quad\quad\quad \langle E', F'\rangle \leftarrow \langle E, F\rangle$
14 $\quad\quad\quad \langle C', D'\rangle \leftarrow \langle C, D\rangle$
15 $\quad\quad\quad s \leftarrow s_{\langle C,D\rangle}$
16 $\quad\quad$**end**
17 $\quad$**end**
18 $\quad$**add** $\langle E', F'\rangle$ **to** $\mathcal{F}$
19 $\quad$**remove** $\langle C', D'\rangle$ **from** $\mathcal{G}$
20 $\quad$**remove** from U entries $\langle i, j\rangle$ covered by $E' \otimes F'$ in I
21 **end**
22 **return** $\mathcal{F}$

Algorithm 2: COMPUTEINTERVALS

Input: matrix I with entries in scale L
Output: set $\mathcal{G} \subseteq \mathcal{B}(\mathcal{E}(I))$
1 $\mathcal{E} \leftarrow \mathcal{E}(I)$
2 $U \leftarrow \{\langle i, j\rangle | \mathcal{E}_{ij} > 0\}$
3 **while** *U is non-empty* **do**
4 $\quad D \leftarrow \emptyset$; $s \leftarrow 0$
5 $\quad$**while** *exists* $\{^a/_j\} \in D$ *s.t.* $cov_I(U, D \vee \{^a/_j\}, \mathcal{E}) > s$ **do**
6 $\quad\quad$select $\{^a/_j\}$ maximizing $cov_I(U, D \vee \{^a/_j\}, \mathcal{E})$
7 $\quad\quad D \leftarrow (D \vee \{^a/_j\})^{\downarrow_\mathcal{E}\uparrow_\mathcal{E}}$; $C \leftarrow (D \vee \{^a/_j\})^{\downarrow_\mathcal{E}}$
8 $\quad\quad s \leftarrow cov_I(U, D, \mathcal{E})$
9 $\quad$**end**
10 $\quad$**add** $\langle C, D\rangle$ **to** $\mathcal{G}$
11 $\quad$**remove** from U all $\langle i, j\rangle$ covered by $C^{\uparrow_I\downarrow_I} \otimes D^{\downarrow_I\uparrow_I}$ in I
12 **end**
13 **return** $\mathcal{G}$

VI. EXPERIMENTAL EVALUATION

In the evaluation on real and synthetic data, we use GREESS$_L$ and another algorithm, ASSO$_L$, which is inspired by the well-known ASSO algorithm [9] and presented in detail in an extended version of this paper. We observe the ability of the extracted factors to explain (reconstruct) the input data by means of $s(I, A_{\mathcal{F}} \circ B_{\mathcal{F}})$, where $\mathcal{F}$ is the examined set of factors (usually the first k factors obtained) as well as the interpretation of factors.

dataset	size	$\|L\|$	$\|\|I\|\|$	$\|\|\mathcal{E}(I)\|\|$	$\|\|\mathcal{E}(I)\|\|/\|\|I\|\|$
Breeds	151×11	6	1963	362	0.184
Decathlon	28×10	5	266	59	0.221
IPAQ	4510×16	3	41624	1281	0.031
Music	900×26	7	20377	5952	0.292

Table II: Real data

A. *Real Data*

The basic characteristics of data are shown in Table II. ($||A||$ denotes the number of non-zero entries in matrix A). Notice that the reduction in the number of non-zero entries in $\mathcal{E}(I)$ compared to I is significant.

Dog Breeds[2] extends the dataset from Section III to 151 breeds. GREESS$_L$ found 20 factors providing an exact decomposition of the 151×11 matrix I. Already the set $\mathcal{F}_3$ consisting of the first three factors explains a large portion of the data, namely $s(I, A_{\mathcal{F}_3} \circ B_{\mathcal{F}_3}) = 0.795$. Among these is a factors which represents dogs that excel in sports (such as agility, flyball, frisbee) or guide, service, and therapy dogs, such as Golden Retriever, Labrador Retriever, or Papillon. Another is naturally interpreted as guardian dog. Interestingly, these two factors are similar to factors F_3 and F_2 described in Section III. In fact, the factors F_1, F_2, and F_3 from Section III, when extended to the 151×11 matrix, cover 0.85 of the matrix according to s, illustrating an interesting natural property that we observed in several examples. ASSO$_L$, on the other hand, computed 7 factors, which are not so easy to interpret compared to those from GREESS$_L$—a feature that we observed in most examples we examined.

Decathlon[3] We analyzed the results of 2004 Olympic decathlon and represented them by a 28×10 matrix I (28 athletes, 10 disciplines of decathlon) using a five-element scale L. Using GREESS$_L$, we obtained 10 factors that we consulted with an experienced decathlon coach. Among the most important factors are the factor *speed*, containing to high degrees the attributes 100m, long jump, and hurdles; *explosiveness*, containing long jump, shot put, high jump, and javelin; and a factor containing high jump and 1500m, typical of light-weight athletes. All these factors were found natural by the decathlon coach.

IPAQ Data[4] consists of international questionnaire data involving 4510 respondents answering 16 questions using a three-element scale, regarding physical activity. The questions include those regarding their age, sex, body-mass-index (BMI), health, to what extent the person bicycles, walks, etc. GREESS$_L$ produced 17 factor concepts. As with the other examples, the first three factors may be considered sufficient to explain the data. These may be interpreted as as healthy people with good education who cycle on a regular basis; people with normal BMI who walk on a daily basis; and people who are employed, own a car, and cycle on a regular basis. All these groups are considered important according to an expert kinanthropologist.

Music Data[5] consists of results of a study inquiring how people perceive speed of song depending of various song characteristics. The data consists of a 900×26 matrix over a six-element scale L, representing a questionnaire involving 30 participants who were presented 30 music samples (29 complex complex ones and one simple tone), and 26 attributes regarding emotional experience of participants, using a 6-element scale L. Using GREESS$_L$, we obtained 29 factors. The authors of this study examined the factors and concluded that the groups of music samples corresponding to the factors are meaningful and that the factors can be interpreted in terms of emotional experience. For example, an interesting factor with a good coverage contained songs No. 5, 7, 16, and 26, all of which are melancholic. Among other interesting factors are the one manifested to a high degree by attributes Ugly and Violent; the one manifested by Restful, Safe, Stable, and Inert; and the factor manifested by Successful, Valuable, Meaningful, and Significant.

Coverage of Data by Factors Table III displays the numbers of factors produced by the two algorithms that are needed to achieve a prescribed coverage represented by s, see (4). For example, the first row corresponding to Breeds says that we need the set $\mathcal{F}_2$ consisting of the first two factors produced by ASSO$_L$ in order for the coverage to be at least 0.75, i.e. $s(I, A_{\mathcal{F}_2} \circ A_{\mathcal{F}_2}) \geq 0.75$, while in case of GREESS$_L$, we need the first three factors. "NA" indicates that the prescribed coverage is not achievable by the factors produced by ASSO$_L$.

B. *Synthetic Data*

We used Set 1–4, each consisting of 500 $n \times m$ matrices I obtained as products of $n \times k$ and $k \times m$ randomly generated matrices A and B with prescribed probability distributions over L, see Table IV. Table VI displays selected results of coverage s by the first k factors for the datasets and the two algorithms. We also include the percentage $s_=$ of entries $\langle i, j \rangle$ for which $I_{ij} = (A_{\mathcal{F}_k} \circ B_{\mathcal{F}_k})_{ij}$, where $\mathcal{F}_k$ is the set of the first k factors, i.e. $s_=$ is stronger than s. "–" for ASSO$_L$ means that no new factors were produced.

C. *Discussion*

The first couple of factors produced by ASSO$_L$ tend to have a better coverage compared to GREESS$_L$. On the other hand, beyond certain coverage, ASSO$_L$ stops producing factors and is not able to compute an (exact) decomposition

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

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

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

[5]Flaska K., Cakirpaloglu P., Identification of the multidimensional model of subjective time experience. Int. Conf. Time Perspective, Converging Paths in Psychology, Portugal, September 4–8, 2012 (extended version submitted).

dataset	s	number of factors needed Asso$_L$	GreEss$_L$
Breeds	0.75	2	3
	0.85	3	7
	0.95	NA	11
	1	NA	15
Decathlon	0.75	2	3
	0.85	4	5
	0.95	NA	8
	1	NA	10
IPAQ	0.75	1	10
	0.85	1	12
	0.95	NA	15
	1	NA	17
Music	0.75	2	7
	0.85	NA	14
	0.95	NA	25
	1	NA	29

Table III: Quality of decompositions (real data).

dataset	size	$\|L\|$	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	$[\frac{1}{8}\ \frac{1}{8}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}]$

Table IV: Synthetic data.

of I, while GreEss$_L$ always is, with a reasonably small number of factor needed for coverage very close to 1. This is congruent with the fact that Asso$_L$ and GreEss$_L$ are primarily designed for DBP(L) and AFP(L), respectively, as well as with the available evidence from the Boolean case. We found that GreEss$_L$ produces easier interpretable factors compared to Asso$_L$. This is partly because if $|L| > 2$ (non-Boolean case), rectangles with values "around the middle" in L, such as 0.5, which may be produced as factors by Asso$_L$ but not in general by GreEss$_L$, have good coverage and are thus sometimes selected by Asso$_L$. This issue needs to be carefully examined in future research. We also performed experiments comparing GreEss$_L$ with the algorithm in [4]. On average, GreEss$_L$ requires 30% less factors to achieve a prescribed coverage.

VII. Conclusions

We presented theoretical results regarding the geometry of decompositions of data over scales. We presented a new decomposition algorithm and provided its experimental evaluation. We demonstrated that decompositions, generalizing BMF, help reveal interesting factors in data over scales. Future research shall include: (a) further theoretical advancement matrices over scales; (b) further design of algorithms including those inspired by the existing BMF algorithms; (c) development of case studies in factor analysis of ordinal data and applications of the decomposition methods in machine learning in presence of ordinal data; (d) comparison to other relevant decomposition algorithms such as the non-negative matrix factorization.

dataset	avg $\|\|I\|\|$	avg $\|\|\mathcal{E}(I)\|\|$	avg $\|\|\mathcal{E}(I)\|\|/\|\|I\|\|$
Set 1	2452	193	0.079
Set 2	2499	358	0.143
Set 3	4998	614	0.123
Set 4	10000	2130	0.213

Table V: Characteristics of synthetic data.

dataset	k	coverage $s/s_=$ by the first k factors Asso$_L$	GreEss$_L$
Set 1	1	0.849/0.699	0.638/0.398
	4	0.884/0.759	0.874/0.760
	8	–	0.961/0.927
	14	–	1/1
Set 2	1	0.841/0.416	0.688/0.287
	4	0.878/0.555	0.900/0.679
	8	–	0.974/0.905
	12	–	0.999/0.994
	16	–	1/1
Set 3	1	0.900/0.604	0.729/0.345
	5	0.947/0.794	0.899/0.614
	10	–	0.948/0.885
	20	–	0.983/0.959
	47	–	1/1
Set 4	1	0.872/0.585	0.674/0.416
	5	0.930/0.725	0.872/0.606
	10	–	0.937/0.752
	20	–	0.986/0.945
	38	–	1/1

Table VI: Coverage s and $s_=$ by the first k factors.

References

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

[2] R. Belohlavek, *Optimal decompositions of matrices over residuated lattices*, J. Logic Comput. 22(2012), 1405–1425.

[3] R. Belohlavek, M. Trnecka, *From-below approximations in Boolean matrix factorization: Geometry and new algorithm.* (submitted, available at arXiv).

[4] R. Belohlavek, V. Vychodil, *Factor analysis of incidence data via novel decomposition of matrices.* LNAI 5548(2009), 83–97.

[5] R. Belohlavek, V. Vychodil, *Discovery of optimal factors in binary data via a novel method of matrix decomposition.* J. Computer and System Sciences 76(1)(2010), 3–20.

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

[7] F. Geerts, B. Goethals, T. Mielikäinen, *Tiling databases*, Proc. Discovery Science 2004, pp. 278–289.

[8] S. Gottwald, *A Treatise on Many-Valued Logics*. Research Studies Press, Baldock, Hertfordshire, England, 2001.

[9] P. Miettinen, T. Mielikäinen, A. Gionis, G. Das, H. Mannila, *The discrete basis problem*, IEEE TKDE 20(2008), 1348–62.

[10] Miettinen P., Vreeken J., *Model order selection for Boolean matrix factorization*, ACM SIGKDD 2011, pp. 51–59.

[11] G. T. Miller, *The magical number seven, plus or minus two*, Psychol. Rev. 63(1956), 81–97.

[12] D. S. Nau et al., *A mathematical analysis of human leukocyte antigen serology*, Math. Biosci. 40(1978), 243–270.