Ann Math Artif Intell
DOI 10.1007/s10472-014-9398-6

Factor analysis of ordinal data via decomposition of matrices with grades

Radim Belohlavek · Marketa Krmelova

Abstract We present new results on a recently developed method of factor analysis of data with ordinal attributes. The method is based on the apparatus of logic, the theory of relations and ordered sets, and provides an alternative to traditional methods of factor analysis. It utilizes formal concepts as factors and we demonstrate on several examples using sports data that the factors produced by the method are reasonable and easy-to-interpret. In addition, we propose ways to address various natural questions regarding the method and its use and put forward new research issues.

Keywords Factor analysis · Relational data · Formal concept analysis · Graded attributes · Ordinal attributes

Mathematics Subject Classification (2010) 15A23 · 03B52 · 62H25 · 06A15

1 Introduction and paper outline

Factor analysis and related techniques based on matrix decompositions are important methods of data analysis. Perhaps the best known methods designed for real-valued matrices are singular value decomposition (SVD) and principal component analysis (PCA) [11, 14]. In presence of certain semantic issues, such as a difficulty to interpret negative coefficients, new methods are needed. A well-known example of these is the non-negative matrix factorization (NMF) [18]. In the past years, a considerable effort has been devoted to the development of matrix methods for Boolean (binary) data, particularly Boolean matrix factorization (BMF), see e.g. [8, 13, 19, 20, 23] and the references therein.

We acknowledge support by Grant No. P202/10/0262 of the Czech Science Foundation. A preliminary version of this paper was presented at CLA 2012.

R. Belohlavek · M. Krmelova (✉)
Department of Computer Science, Palacky University, Olomouc, Czech Republic
e-mail: marketa.krmelova@gmail.com

Published online: 25 January 2014

In our previous papers [1, 4, 9], we extended the matrix factorization problem for Boolean data and the method developed in [8] to ordinal data (see Section 2). We provided a framework for this problem, established basic results regarding optimal decompositions, transformations between the attribute and factor spaces, computational complexity of the problem, developed a greedy approximation algorithm for computing the decompositions, and demonstrated that the method is capable of revealing interesting and easily interpretable factors in ordinal data.

The aim of this paper is threefold. First, to provide a detailed, clean description of the factor model and its meaning. Second, we provide extensive examples of factor analysis of sports data. These are both important issues, somewhat neglected in our previous papers which focused on the formal and computational aspects. Third, we develop new theoretical concepts and results that address new conceptual issues arising in our factor analysis model.

2 The method: factor model, factors, and interpretation

2.1 The factor model: basic idea, technical description, and interpretation

Basic idea and the factor model In general, i.e. abstracting from the particular mathematical model involved, factor analysis aims at whether data involving objects and their directly observable attributes may be explained by a smaller number of different, hopefully more fundamental attributes which are called factors. For example, whether performances of students (directly observable attributes) may be described by some treats of their intelligence (factors). Formally, the input data is represented by an $n \times m$ object–attribute matrix I and the "explanation" means a decomposition

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

(exact or approximate) of I into a product $A \circ B$ of an $n \times k$ object–factor matrix A and a $k \times m$ factor–attribute matrix B. What kind of matrices (real, Boolean, or other) and what kind of product $\circ$ are involved determines the semantics of the factor model.

Our model involves matrices containing degrees (or grades) of certain scales L and the product is the sup-$\otimes$ product, as described below. In particular, the matrix entry I_{ij} is a degree to which attribute j applies to object i. Such degree may be but need not be numerical, e.g. $I_{ij} = 0.5$. Similarly, A_{il} and B_{lj} are the degree to which factor l applies to object i and the degree to which attribute j is a manifestation (one out of many possible manifestations) of factor l. The case in which the scale L contains only two degrees, 0 and 1, called the Boolean case in what follows, corresponds to Boolean matrices and Boolean factor analysis which is a special case of ours. Equation (1) in our model has the following meaning:

$$\begin{aligned}&\text{object } i \text{ has attribute } j \text{ if and only if}\\ &\quad\text{there exists factor } l \text{ such that } i \text{ has } l \text{ (or, } l \text{ applies to } i\text{)}\\ &\quad\text{and } j \text{ is one of the particular manifestations of } l\end{aligned} \tag{2}$$

which may be regarded as a verbal description of the model given by (1). Such description is certainly appealing and well understandable. In the Boolean case, in which $L = \{0, 1\}$, the verbal description leads to

$$(A \circ B)_{ij} = 1 \text{ iff there exists } l \in \{1, \ldots, k\} \text{ such that } A_{il} = 1 \text{ and } B_{lj} = 1,$$

which may equivalently be described by the well-known formula

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

for Boolean matrix composition. With a general scale L, we approach the situation according to the principles of (mathematical) fuzzy logic [2, 16, 17] as follows. We consider the formulas $\varphi(i,l)$ saying "object i has factor l" and $\psi(l, j)$ saying "attribute j is a manifestation of factor l", and consider A_{il} the truth degree $||\varphi(i,l)||$ of $\varphi(i,l)$ and B_{lj} the truth degree $||\psi(l, j)||$ of $\psi(l, j)$, i.e.

$$||\varphi(i,l)|| = A_{il} \text{ and } ||\psi(l, j)|| = B_{lj}. \tag{4}$$

Now, according to fuzzy logic, the truth degree of formula $\varphi(i,l)\&\psi(l, j)$ which says "object i has factor l and attribute j is a manifestation of factor l" is computed by

$$||\varphi(i,l)\&\psi(l, j)|| = ||\varphi(i,l)|| \otimes ||\psi(l, j)||$$

where $\otimes : L \times L \to L$ is a truth function of many-valued conjunction & (several reasonable functions exist), and hence the truth degree of $(\exists l)(\varphi(i,l)\&\psi(l, j))$ which says "there exists factor l such that object i has l and attribute j is a manifestation of l", i.e. the proposition involved in (2), is computed by

$$||(\exists l)(\varphi(i,l)\&\psi(l, j))|| = \bigvee_{l=1}^{k} ||\varphi(i,l)|| \otimes ||\psi(l, j)||, \tag{5}$$

where $\bigvee$ denotes the supremum. Given into account (4), we see that a generalization of (3) to the case of possibly intermediate degrees is given by

$$(A \circ B)_{ij} = \bigvee_{l=1}^{k} A_{il} \otimes B_{lj}. \tag{6}$$

Therefore, with $\circ$ given by (6), the factor model (1) retains its meaning (2) even in the case where intermediate degrees, such as 0.5, are allowed, as opposed to 0 and 1, which are the only truth values allowed in the Boolean case.

Scales of degrees and truth functions $\otimes$ and $\to$ Technically, we assume that the grades are taken from a partially ordered bounded scale L of certain type. In particular, we assume that L conforms to the structure of a complete residuated lattice, used in fuzzy logic, and refer the reader to [16, 17] for details. We only recall that being a complete residuated lattice means that the infima and suprema of arbitrary subsets of L exist w.r.t. to the partial order $\leq$ on L, and that L is equipped with a binary operation $\otimes$ which is commutative, associative, has 1 as its neutral element, and distributes over arbitrary suprema, i.e. $a \otimes \left(\bigvee_{j\in J} b_j\right) = \bigvee_{j\in J}(a \otimes b_j)$. From a logical point of view, the elements of L represent truth degrees, 1 and 0 corresponding to full truth and falsity, and $\otimes$ is a truth function of conjunction. In a sense, $\otimes$ may also be looked at as a monotone aggregation of truth degrees.

Another function, namely $\to$, which is called the residuum and which plays the role of the truth function of implication, is then uniquely associated to $\otimes$ by $a \to b = \max\{c \in L \mid a \otimes c \leq b\}$.

Residuum satisfies an important technical condition called adjointness, namely,

$$a \otimes b \leq c \text{ iff } a \leq b \to c.$$

Grades of ordinal scales are conveniently represented by numbers, such as the Likert scale $\{1, \ldots, 5\}$. In such a case we assume these numbers are normalized and taken from the unit interval $[0, 1]$. As an example, the Likert scale is represented by $L = \{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\}$.

In our experiments, we mainly use finite scales with the Lukasiewicz structure, i.e. we use equidistant subchains of [0, 1] with

$$a \otimes b = \max(0, a + b - 1) \text{ and } a \rightarrow b = \min(1, 1 - a + b), \tag{7}$$

but many other examples are available, see e.g. [16].

2.2 Factors, their interpretation and computation

From the above description, it is clear that for any decomposition (1), the lth factor ($l \in \{1, \ldots, k\}$) is represented by two parts: the lth column $A_{\_l}$ of A and the lth row $B_{l\_}$ of B. As shown in [4], optimal factors for a decomposition of I (see below) are provided by formal concepts associated to I. In detail, let $X = \{1, \ldots, n\}$ (rows/objects) and $Y = \{1, \ldots, m\}$ (columns/attributes). A formal concept of I is any pair $\langle C, D\rangle$ of L-sets (fuzzy sets [24]) $C : \{1, \ldots, n\} \rightarrow L$ of objects and $D : \{1, \ldots, m\} \rightarrow L$ of attributes, see [3], that satisfies $C^{\uparrow} = D$ and $D^{\downarrow} = C$ where ${}^{\uparrow} : L^X \rightarrow L^Y$ and ${}^{\downarrow} : L^Y \rightarrow L^X$ are the concept-forming operators defined by

$$C^{\uparrow}(j) = \textstyle\bigwedge_{i \in X}(C(i) \rightarrow I_{ij}) \text{ and } D^{\downarrow}(i) = \textstyle\bigwedge_{j \in Y}(D(j) \rightarrow I_{ij}).$$

The set of all formal concepts of I is denoted by $\mathcal{B}(X, Y, I)$ or just $\mathcal{B}(I)$. $C(i) \in L$ and $D(j) \in L$ are interpreted as the degree to which factor l applies to object i and the degree to which attribute j is a manifestation of l. Optimality of using formal concepts as factors means the following. Let for a set (we fix the numbering of its elements)

$$\mathcal{F} = \{\langle C_1, D_1\rangle, \ldots, \langle C_k, D_k\rangle\} \subseteq \mathcal{B}(X, Y, I) \tag{8}$$

of formal concepts denote by $A_{\mathcal{F}}$ and $B_{\mathcal{F}}$ the matrices defined by

$$(A_{\mathcal{F}})_{il} = (C_l)(i) \quad \text{and} \quad (B_{\mathcal{F}})_{lj} = (D_l)(j). \tag{9}$$

Then whenever $I = A \circ B$ for $n \times k$ and $k \times m$ matrices A and B, there exists a set $\mathcal{F} \subseteq \mathcal{B}(X, Y, I)$ $|\mathcal{F}| \leq k$ such that $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$, i.e. optimal decompositions are attained by formal concepts as factors.

By $\text{rank}_{\mathbf{L}}(I)$ we denote the smallest k for which the above decomposition of I exists and call it the (**L**-) *rank* of I.

Remark 1 There is a useful way to visualize factor concepts from (8). Namely, a concept $\langle C_l, D_l\rangle$ may be visualized by the $n \times m$ matrix J_l defined by $(J_l)_{ij} = A_{il} \otimes B_{lj}$. One may then also observe the matrix $I_p = \bigvee_{l=1}^{p} J_l$ for $p = 1, \ldots, k$, which results by adding together the first p factors (note that $I_k = I$).

In our experiments, we use the greedy algorithm proposed in [9] for computing a set $\mathcal{F}$ of concepts for which $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$. Note also that, due to [22], the problem of computing a smallest $\mathcal{F}$ with $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$ is an NP-hard optimization problem [9] and that, as a result, the above-mentioned algorithm is an approximation algorithm. The algorithm proceeds in a greedy manner and picks as factors the formal concepts that explain the most of the yet unexplained data. In doing so, the candidate formal concepts are constructed from attribute concepts of the input matrix I, i.e. concepts corresponding to single attributes. The construction consists in a particular extension of the candidate concept obtained so far by adding consecutively the most promising attribute and degree to its intent. The algorithm always computes an exact decomposition of I and may be stopped earlier after a decomposition sufficiently close to I is obtained.

2.3 Illustrative example

In this section, we present a small example which illustrates the notions introduced above. Examples on real data are presented in Section 4. Consider a three-element Łukasiewicz chain, i.e. the set $L = \{0, 1/2, 1\}$ equipped with the operations defined by (7), and the following matrix I:

$$\begin{pmatrix} 1/2 & 1 & 1 & 1/2 & 1/2 \\ 1/2 & 1 & 1 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 & 1 & 1 \\ 0 & 0 & 1 & 1/2 & 1 \end{pmatrix} \tag{10}$$

The pair $\langle C_1, D_1\rangle$, where

$$\langle C_1(1), C_1(2), C_1(3), C_1(4)\rangle = \langle 1, 1, 1/2, 0\rangle,$$
$$\langle D_1(1), D_1(2), D_1(3), D_1(4), D_1(5)\rangle = \langle 1/2, 1, 1, 1/2, 1/2\rangle$$

is a formal concept of I. That is, the concept fully applies (to degree 1) to the first and the second object, partly applies (to degree 1/2) to the third object, and does not apply at all (to degree 0) to the fourth object; on the side of attributes, the second and third ones are characteristic (to degree 1) of this concept.

In order to see it, we need to check $C_1^{\uparrow} = D_1$ and $D_1^{\downarrow} = C_1$, i.e. to check $C_1^{\uparrow}(j) = D_1(j)$ for each $j = 1, \ldots, 5$, and $D_1^{\downarrow}(i) = C_1(i)$ for each $i = 1, \ldots, 4$. For $j = 1$ we obtain

$$\begin{aligned} C_1^{\uparrow}(1) &= \textstyle\bigwedge_{i \in X}(C_1(i) \to I_{i1}) \\ &= (C_1(1) \to I_{11}) \wedge (C_1(2) \to I_{21}) \wedge (C_1(3) \to I_{31}) \wedge (C_1(4) \to I_{41}) \\ &= (1 \to 1/2) \wedge (1 \to 1/2) \wedge (1/2 \to 0) \wedge (0 \to 0) = 1/2 = D_1(1), \end{aligned}$$

and similarly for $j = 2, 3, 4$. Conversely, we obtain

$$\begin{aligned} D_1^{\uparrow}(1) &= \textstyle\bigwedge_{j \in Y}(D_1(j) \to I_{1j}) = (D_1(1) \to I_{11}) \wedge (D_1(2) \to I_{12}) \wedge \\ &\quad \wedge (D_1(3) \to I_{13}) \wedge (D_1(4) \to I_{14}) \wedge (D_1(5) \to I_{15}) \\ &= (1/2 \to 1/2) \wedge (1 \to 1) \wedge (1 \to 1) \wedge (1/2 \to 1/2) \wedge (1/2 \to 1/2) = 1 = C_1(1), \end{aligned}$$

and similarly for $i = 2, 3, 4, 5$.

The concept lattice $\mathcal{B}(I)$ contains in total 11 formal concepts (and may be computed by existing algorithms designed for this purpose). In addition to $\langle C_1, D_1\rangle$, two other concepts in $\mathcal{B}(I)$ are

$$\langle C_2, D_2\rangle \text{ with } C_2 \text{ and } D_2 \text{ represented by } \langle 1/2, 1/2, 1, 1/2\rangle \text{ and } \langle 0, 1/2, 1/2, 1, 1\rangle,$$

and

$$\langle C_3, D_3\rangle \text{ with } C_3 \text{ and } D_3 \text{ represented by } \langle 1/2, 1/2, 1/2, 1\rangle \text{ and } \langle 0, 0, 1, 1/2, 1\rangle.$$

The set $\mathcal{F} = \{\langle C_1, D_1\rangle, \langle C_2, D_2\rangle, \langle C_3, D_3\rangle\}$ is a set of factor concepts of I. Namely, the corresponding matrices $A_{\mathcal{F}}$ and $B_{\mathcal{F}}$ are

$$A_{\mathcal{F}} = \begin{pmatrix} 1 & 1/2 & 1/2 \\ 1 & 1/2 & 1/2 \\ 1/2 & 1 & 1/2 \\ 0 & 1/2 & 1 \end{pmatrix} \text{ and } B_{\mathcal{F}} = \begin{pmatrix} 1/2 & 1 & 1 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 & 1 & 1 \\ 0 & 0 & 1 & 1/2 & 1 \end{pmatrix},$$

and one easily verifies that $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$. For instance,

$$\begin{aligned}(A_{\mathcal{F}} \circ B_{\mathcal{F}})_{11} &= \bigvee_{l=1}^{3} (A_{\mathcal{F}})_{1l} \otimes (B_{\mathcal{F}})_{l1} \\ &= (A_{\mathcal{F}})_{11} \otimes (B_{\mathcal{F}})_{11} \vee (A_{\mathcal{F}})_{12} \otimes (B_{\mathcal{F}})_{21} \vee (A_{\mathcal{F}})_{13} \otimes (B_{\mathcal{F}})_{31} \\ &= 1/2 \vee 0 \vee 0 = 1/2 = I_{11}.\end{aligned}$$

3 Explanation of data by factors

In this section we propose a way to address the following related problems. First, what does it mean that a set of formal concepts explains well (or to a certain extent) a given dataset? Second, what does it mean that good factors of a given dataset explains well another dataset? Third, how is the present method related to the possibility of scaling the ordinal attributes to Boolean ones and performing Boolean factor analysis? We use the notions proposed in this section in the experiments in Section 4.

3.1 Explanation of I by formal concepts of I

If a set $\mathcal{F} \subseteq \mathcal{B}(X, Y, I)$ of formal concepts of I satisfies $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$, we intuitively regard $\mathcal{F}$ as fully explaining the data represented by I and call $\mathcal{F}$ a set of *factor concepts*. In general, however, we are interested in small $\mathcal{F}$ for which I is close enough to $A_{\mathcal{F}} \circ B_{\mathcal{F}}$. If I is an $n \times m$ matrix, we put

$$s_{=}(I, A_{\mathcal{F}} \circ B_{\mathcal{F}}) = \frac{|\{\langle i, j\rangle; I_{ij} = (A_{\mathcal{F}} \circ B_{\mathcal{F}})_{ij}\}|}{n \cdot m},$$

and say that $\mathcal{F}$ *explains* $100 \cdot s_{=}(I, A_{\mathcal{F}} \circ B_{\mathcal{F}})$% *of data* represented by I. Clearly, this means that $100 \cdot s_{=}(I, A_{\mathcal{F}} \circ B_{\mathcal{F}})$% of all the $n \times m$ entries have the same values in I and $A_{\mathcal{F}} \circ B_{\mathcal{F}}$. In a sense, this is a pessimistic approach because it ignores the case where I_{ij} is close to but different from $(A_{\mathcal{F}} \circ B_{\mathcal{F}})_{ij}$. Closeness of degrees I_{ij} and $(A_{\mathcal{F}} \circ B_{\mathcal{F}})_{ij}$ can naturally be represented by the biresiduum $I_{ij} \leftrightarrow (A_{\mathcal{F}} \circ B_{\mathcal{F}})_{ij}$ defined by $a \leftrightarrow b = (a \to b) \wedge (b \to a)$ [15], because $\leftrightarrow$ represents the truth function of logical equivalence. If $L = [0, 1]$ (the most common choice), closeness of degrees may naturally be taken into account by replacing $s_{=}$ by

$$s_{\approx}(I, A_{\mathcal{F}} \circ B_{\mathcal{F}}) = \frac{\sum_{i,j=1}^{n,m} (I_{ij} \leftrightarrow (A_{\mathcal{F}} \circ B_{\mathcal{F}})_{ij})}{n \cdot m}.$$

Since $I_{ij} \leftrightarrow (A_{\mathcal{F}} \circ B_{\mathcal{F}})_{ij} = 1$ if $I_{ij} = (A_{\mathcal{F}} \circ B_{\mathcal{F}})_{ij}$, we obtain $s_{=}(I, A_{\mathcal{F}} \circ B_{\mathcal{F}}) \leq s_{\approx}(I, A_{\mathcal{F}} \circ B_{\mathcal{F}})$ for every $\mathcal{F}$.

Example 1 Consider the matrix I and the concepts $\langle C_1, D_1\rangle$, $\langle C_2, D_2\rangle$, and $\langle C_3, D_3\rangle$ of I from Section 2.3. For $\mathcal{F}_1 = \{\langle C_1, D_1\rangle\}$, $\mathcal{F}_2 = \{\langle C_1, D_1\rangle, \langle C_2, D_2\rangle\}$, and $\mathcal{F}_3 = \{\langle C_1, D_1\rangle, \langle C_2, D_2\rangle, \langle C_3, D_3\rangle\}$, we obtain

$$A_{\mathcal{F}_1} \circ B_{\mathcal{F}_1} = \begin{pmatrix} 1/2 & 1 & 1 & 1/2 & 1/2 \\ 1/2 & 1 & 1 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \qquad A_{\mathcal{F}_2} \circ B_{\mathcal{F}_2} = \begin{pmatrix} 1/2 & 1 & 1 & 1/2 & 1/2 \\ 1/2 & 1 & 1 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 & 1 & 1 \\ 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix},$$

and $A_{\mathcal{F}_3} \circ B_{\mathcal{F}_3} = I$. One then obtains

$$s_{=}(I, A_{\mathcal{F}_1} \circ B_{\mathcal{F}_1}) = \frac{15}{20} = 0.75, \quad s_{=}(I, A_{\mathcal{F}_2} \circ B_{\mathcal{F}_2}) = \frac{18}{20} = 0.9,$$
$$s_{=}(I, A_{\mathcal{F}_3} \circ B_{\mathcal{F}_3}) = 1, \text{ and}$$
$$s_{\approx}(I, A_{\mathcal{F}_1} \circ B_{\mathcal{F}_1}) = \frac{15.5}{20} = 0.775, \quad s_{\approx}(I, A_{\mathcal{F}_2} \circ B_{\mathcal{F}_2}) = \frac{18.5}{20} = 0.925,$$
$$\text{and } s_{\approx}(I, A_{\mathcal{F}_3} \circ B_{\mathcal{F}_3}) = 1.$$

This is easily seen from the matrices $I \leftrightarrow (A_{\mathcal{F}_p} \circ B_{\mathcal{F}_p})$ defined by $(I \leftrightarrow (A_{\mathcal{F}_p} \circ B_{\mathcal{F}_p}))_{ij} = (I_{ij} \leftrightarrow (A_{\mathcal{F}_p} \circ B_{\mathcal{F}_p})_{ij})$, for which we obtain

$$(I \leftrightarrow (A_{\mathcal{F}_1} \circ B_{\mathcal{F}_1})) = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1/2 & 0 \end{pmatrix}, \quad (I \leftrightarrow (A_{\mathcal{F}_2} \circ B_{\mathcal{F}_2})) = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1/2 \end{pmatrix},$$
$$(I \leftrightarrow (A_{\mathcal{F}_3} \circ B_{\mathcal{F}_3})) = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

These matrices describe closeness of the entries of I and $A_{\mathcal{F}_p} \circ B_{\mathcal{F}_p}$—notice that $\leftrightarrow$ is described by

$\leftrightarrow$	0	1/2	1
0	1	1/2	0
1/2	1/2	1	1/2
1	0	1/2	1,

whence, in paticular, the number of 1s in $I \leftrightarrow (A_{\mathcal{F}_p} \circ B_{\mathcal{F}_p})$ is just the number of entries in which I coincides with $(A_{\mathcal{F}_p} \circ B_{\mathcal{F}_p})$.

3.2 Explanation of I by factors of J: general case

Let I and J be two matrices describing the sets X_1 and X_2 of objects by a common set Y of attributes. How can we answer the question of whether a set $\mathcal{F} \subseteq \mathcal{B}(X, Y, I)$ of possibly good factors of I is a set of good factors of J? The concepts in $\mathcal{F}$ may not directly be used as concepts of J because for $\langle C, D\rangle \in \mathcal{F}$ we have $C \in L^{X_1}$ while we need $\in L^{X_2}$ for factors of J. A natural option is to consider instead of $\mathcal{F}$ the set of concepts of J that are generated by the intents of the factors in $\mathcal{F}$, i.e. the set

$$\mathcal{F}^J = \left\{\langle D^{\downarrow_J}, D^{\downarrow_J \uparrow_J}\rangle \mid \langle C, D\rangle \in \mathcal{F}\right\}, \tag{11}$$

because the intents represent the meanings of concepts ($^{\downarrow_J}$ and $^{\uparrow_J}$ denote the concept-forming operators induced by J). One may then use $s_{=}(I, A_{\mathcal{F}_J} \circ B_{\mathcal{F}_J})$ or $s_{\approx}(I, A_{\mathcal{F}_J} \circ B_{\mathcal{F}_J})$ to asses how well the factors $\mathcal{F}$ of I explain the data represented by J.

3.3 Explanation of I by factors of J: extending the set of objects

Of a particular importance is the particular case where J results by adding rows to I (i.e. adding objects to those represented by I). Let us thus assume that $X_1 \subseteq X_2$ and that $I_{ij} = J_{ij}$ for $i \in X_1$ and $j \in Y$. We may proceed as in Section 3.2 but the following observation presents a convenient simplification of the set $\mathcal{F}^J$.

Observation 1 *For the above notation,*

$$\mathcal{F}^J = \{\langle D^{\downarrow_J}, D\rangle \mid \langle C, D\rangle \in \mathcal{F}\}.$$

Proof We need to show that every intent of I is an intent of J. We have $D = C^{\uparrow_I}$. Consider the L-set $E \in L^{X_2}$ defined by $E(i) = C(i)$ for $i \in X_1$ and $E(i) = 0$ for $i \in X_2 - X_1$. Then

$$E^{\uparrow_J}(j) = \bigwedge_{i \in X_2}(E(i) \to J_{ij}) = \bigwedge_{i \in X_1}(E(i) \to J_{ij}) \wedge \bigwedge_{i \in X_2 - X_1}(E(i) \to J_{ij})$$
$$= \bigwedge_{i \in X_1}(C(i) \to I_{ij}) \wedge 1 = \bigwedge_{i \in X_1}(C(i) \to I_{ij}) = C^{\uparrow_I} = D,$$

proving that D is an intent of J. □

Therefore an intent of a factor of I is also an intent of a possible factor of a larger dataset J. An interesting problem for future reseach is how to select from a possibly large dataset J a smaller I such that the factors of I explain J reasonably well.

3.4 Relationship to Boolean factorization of ordinally scaled attributes

We now turn to the following question. In view of the existing methods of conceptual scaling of ordinal data [12] and of the existing methods of Boolean factor analysis [8, 13, 19], it seems natural to perform factor analysis of an input matrix I with grades as follows. First, one transforms I by ordinal scaling to a Boolean matrix $I^\times$. Second, one performs Boolean factor analysis to $I^\times$ and interprets the obtained set $\mathcal{F}^\times$ of factors of $I^\times$ in an appropriate way, taking the scaling procedure into account. We show in this section theoretically and in the next section experimentally that such approach has severe limitations and that factor analyzing I directly using the method examined in this paper has a significant advantage.

Given an input matrix $I \in L^{n\times m}$, consider the matrix $I^\times \in \{0, 1\}^{n\times(m\cdot|L|)}$ defined by

$$I^\times_{ij_a} = \begin{cases} 1 & \text{if } a \leq I_{ij}, \\ 0 & \text{otherwise}, \end{cases}$$

for $i = 1, \ldots, n$, $j = 1, \ldots, m$, $a \in L$ (we assume a fixed sorting of the elements in L so that the order of columns in $I^\times$ is fixed). That is, $I^\times$ is the Boolean matrix resulting from I by simple ordinal scaling. In a sense, each graded attribute j is replaced by a collection of Boolean attributes j_a ($a \in L$); j_a applies to object i if i has j to a degree at least a. The concept lattices and other structures associated to $I^\times$ and their relationships to those associated to I are studied in [7, 10] and are utilized in what follows.

Recalling that $\mathrm{rank}_{\mathbf{2}}(I^\times)$ and $\mathrm{rank}_{\mathbf{L}}(I)$ denote the Boolean rank of $I^\times$ and the $\mathbf{L}$-rank of I, respectively, i.e. the smallest numbers of factors using which $I^\times$ and I may be explained (factorized), we may formulate the following theorem.

Theorem 1 *For every I,* $\mathrm{rank}_{\mathbf{L}}(I) \leq \mathrm{rank}_{\mathbf{2}}(I^\times)$.

Proof Let $A_{\mathcal{F}^\times} \circ B_{\mathcal{F}^\times} = I^\times$ where $|\mathcal{F}^\times| = \mathrm{rank}_{\mathbf{2}}(I^\times)$. Due to [8, Theorem 2], we may safely assume that $\mathcal{F}^\times \subseteq \mathcal{B}(I^\times)$. Due to [7, 10], the ordinary concept lattice $\mathcal{B}(I^\times)$ is embedded in the fuzzy concept lattice $\mathcal{B}(I)$ via the mapping taking every $\langle C^\times, D^\times\rangle \in \mathcal{B}(I^\times)$ to $\langle C, D\rangle \in \mathcal{B}(I)$ with $D(j) = \bigvee_{j_a \in D} a$ and $C = c(C^\times)^{\uparrow_I\downarrow_I}$ where $c(C^\times)$ is the characteristic function of $C^\times$. Let $\mathcal{F} = \{\langle C, D\rangle \mid \langle C^\times, D^\times\rangle \in \mathcal{B}(I^\times)\}$ denote the counterparts of the factors in $\mathcal{F}^\times$. To prove the claim, it is sufficient to show that $A_{\mathcal{F}} \circ B_{\mathcal{F}} = I$.

Since for every $\langle C^{\times}, D^{\times}\rangle \in \mathcal{B}(I^{\times})$ we have $c(C^{\times}) \subseteq c(C^{\times})^{\uparrow_I \downarrow_I}$, the set $\mathcal{G} = \{\langle c(C^{\times}), D\rangle \mid \langle C^{\times}, D^{\times}\rangle \in \mathcal{F}^{\times}\}$ satisfies $A_{\mathcal{G}} \circ B_{\mathcal{G}} \leq A_{\mathcal{F}} \circ B_{\mathcal{F}}$. Because $A_{\mathcal{F}} \circ B_{\mathcal{F}} \leq I$ for any $\mathcal{F} \subseteq \mathcal{B}(I)$ [4], it is sufficient to prove $I \leq A_{\mathcal{G}} \circ B_{\mathcal{G}}$. Let thus $I_{ij} = b$ for an arbitrary $\langle i, j\rangle$. Then $I^{\times}_{ij_b} = 1$, hence $A_{\mathcal{F}^{\times}} \circ B_{\mathcal{F}^{\times}} = I^{\times}$ implies that there exists $\langle C^{\times}, D^{\times}\rangle \in \mathcal{B}(I^{\times})$ such that $i \in C^{\times}$ and $j_b \in D^{\times}$, i.e. $(c(C^{\times}))(i) = 1$ and $D(j) \geq b$. It thus follows

$$I_{ij} = b = 1 \otimes b \leq (c(C^{\times}))(i) \otimes D(j) \leq \bigvee_{\langle E,F\rangle \in \mathcal{G}} E(i) \otimes F(j) = (A_{\mathcal{G}} \circ B_{\mathcal{G}})_{ij},$$

proving $I \leq A_{\mathcal{G}} \circ B_{\mathcal{G}}$. □

Therefore, as far as the number of factors (usually considered as measuring goodness of explanation) is concerned, we are not worse off when directly factorizing I compared to the scale-and-Boolean-factorize method.

Remark 2 The reason, partly apparent from the proof, is that the spaces of factors to attain optimal factorizations of I and $I^{\times}$ are the concept lattices $\mathcal{B}(I)$ and $\mathcal{B}(I^{\times})$. Now, $\mathcal{B}(I^{\times})$ is in general smaller than $\mathcal{B}(I)$, in fact embedded in $\mathcal{B}(I)$ in a natural way [7, 10]. Thus, the space of possible factors of I is richer than that of $I^{\times}$, leaving Theorem 1 a natural consequence of this fact. In addition to Theorem 1, this fact is manifested in the experimental results provided in Section 4.4.

The following example shows that the estimation is not tight and that in fact, $\mathrm{rank}_{\mathbf{L}}(I)$ can be significantly smaller than $\mathrm{rank}_{\mathbf{2}}(I^{\times})$.

Example 2 Let $L = \{0 = a_1, a_2, \ldots, a_n = 1\}$, consider an $n \times 1$ matrix I with $I_{i1} = a_i$. Then $I^{\times}$ is an $n \times n$ matrix, the square staircase matrix, given by $I^{\times}_{pq} = 1$ if $p \geq q$ and $= 0$ if $p < q$. It is a matter of moment's reflection to see that $\mathrm{rank}_{\mathbf{L}}(I) = 1$ and $\mathrm{rank}_{\mathbf{2}}(I^{\times}) = n$.

4 Examples and experiments

In this section we present results of selected analyses of ordinal data. Our purpose is to explain in detail the process of factor analysis in technical terms as well as informally, and to demonstrate the usefulness of the presented method.

Throughout this section, we use small finite scales L such as

$$L = \{0, 0.25, 0.5, 0.75, 1\}.$$

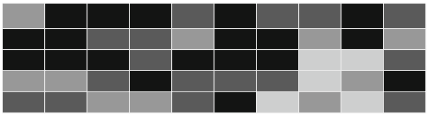
Due to the well-known Miller's 7 ± 2 phenomenon [21], small scales L with up to 7 ± 2 degrees are preferable because humans can understand and use such scales easily. In particular, people attach to the degrees linguistic labels such as "not at all" to 0, "somewhat" to 0.25, "middle" to 0.5, "rather" to 0.75, and "fully" to 1, and the like. We mainly use the five-element Łukasiewicz chain (a particular complete residuated lattice), i.e. the matrix degrees are elements in $L = \{0, 0.25, 0.5, 0.75, 1\}$ and the operations $\otimes$ and $\rightarrow$ are given by (7).

We represent the degrees by shades of gray as follows (this also emphasizes the fact that the truth degrees have a symbolic, rather than numerical, meaning):

Table 1 2004 Olympic Games Decathlon

	10	*lj*	*sp*	*hj*	40	*hu*	*di*	*pv*	*ja*	15
Scores of top 5 athletes										
Sebrle	894	1020	873	915	892	968	844	910	897	680
Clay	989	1050	804	859	852	958	873	880	885	668
Karpov	975	1012	847	887	968	978	905	790	671	692
Macey	885	927	835	944	863	903	836	731	715	775
Warners	947	995	758	776	911	973	741	880	669	693
Matrix I with graded attributes (input to the method)										
Sebrle	0.50	1.00	1.00	1.00	0.75	1.00	0.75	0.75	1.00	0.75
Clay	1.00	1.00	0.75	0.75	0.50	1.00	1.00	0.50	1.00	0.50
Karpov	1.00	1.00	1.00	0.75	1.00	1.00	1.00	0.25	0.25	0.75
Macey	0.50	0.50	0.75	1.00	0.75	0.75	0.75	0.25	0.50	1.00
Warners	0.75	0.75	0.50	0.50	0.75	1.00	0.25	0.50	0.25	0.75

Graphical representation of matrix I



10—100 meters sprint race; lj—long jump; sp—shot put; hj—high jump; 40—400 meters sprint race; hu—110 meters hurdles; di—discus throw; pv—pole vault; ja—javelin throw; 15—1500 meters run

For a reader not familiar with basics of many-valued logics let us note that the Łukasiewicz $\otimes$ (such as other many-valued conjunctions) may be seen as a natural conjunction-like aggregation: the higher the truth values a and b of propositions φ and ψ, respectively, the higher the truth value $a \otimes b$ of the conjunction $\varphi\&\psi$.

4.1 2004 Olympic Games Decathlon—top 5

We start with a detailed description of factor analysis of top 5 athletes in the 2004 Olympic Decathlon and use this example as a reference example in the subsequent sections (this data is also used in [9], but our analysis here is slightly different since we use a different transformation of the athletes' results to grades). Our method is particularly suitable for analyzing such data for the following reasons. The raw data, i.e. the actual results, can naturally be transformed to data with graded attributes, i.e. to a matrix I. Namely, for every discipline d, one may consider a graded attribute "good performance in d": such an attribute applies to an athlete (object) to a degree to which we consider the performance of the athlete to be a good performance. This is a natural, generally applicable idea. However, in our case, the IAAF (International Association of Athletics Federations) provides us with decathlon scoring tables (http://www.iaaf.org, IAAF Scoring Tables for Combined Events) using which one transforms the actual results to scores from an ordinal scale, namely the interval of integers $[0, 1, \ldots, 1400]$, which is common to all disciplines. For example, the result of 10.75 s

Table 2 Factor concepts

F_i	Extent	Intent
F_1	{$^{.5}$/Sebrle, Clay, Karpov, $^{.5}$/Macey, $^{.75}$/Warners}	{10, lj, $^{.75}$/sp, $^{.75}$/hj, $^{.5}$/40, *hu*, $^{.5}$/di, $^{.25}$/pv, $^{.25}$/ja, $^{.5}$/15}
F_2	{Sebrle, $^{.75}$/Clay, $^{.25}$/Karpov, $^{.5}$/Macey, $^{.25}$/Warners}	{$^{.5}$/10, lj, sp, hj, $^{.75}$/40, *hu*, $^{.75}$/di, $^{.75}$/pv, ja, $^{.75}$/15}
F_3	{$^{.75}$/Sebrle, $^{.5}$/Clay, $^{.75}$/Karpov, Macey, $^{.5}$/Warners}	{$^{.5}$/10, $^{.5}$/lj, $^{.75}$/sp, hj, $^{.75}$/40, $^{.75}$/ *hu*, $^{.75}$/di, $^{.25}$/pv, $^{.5}$/ja, 15}
F_4	{Sebrle, $^{.75}$/Clay, $^{.75}$/Karpov, $^{.75}$/Macey, Warners}	{$^{.5}$/10, $^{.75}$/lj, $^{.5}$/sp, $^{.5}$/hj, $^{.75}$/40, *hu*, $^{.25}$/di, $^{.5}$/pv, $^{.25}$/ja, $^{.75}$/15}
F_5	{$^{.75}$/Sebrle, $^{.5}$/Clay, Karpov, $^{.75}$/Macey, $^{.25}$/Warners}	{$^{.75}$/10, $^{.75}$/lj, sp, $^{.75}$/hj, 40, *hu*, di, $^{.25}$/pv, $^{.25}$/ja, $^{.75}$/15}
F_6	{$^{.75}$/Sebrle, Clay, $^{.25}$/Karpov, $^{.5}$/Macey, $^{.25}$/Warners}	{$^{.75}$/10, lj, $^{.75}$/sp, $^{.75}$/hj, $^{.5}$/40, hu, di, $^{.5}$/pv, ja, $^{.5}$/15}

in 100 m gets 962 points, the result of 204 cm in high jump gets 927 points, etc. A table with actual scores may then be transformed to a matrix I with graded attributes using an appropriate set L of truth degrees and an appropriate transformation function.

The top table in Table 1 contains the results of top 5 athletes according to the IAAF scoring tables. The second table from the top contains the corresponding matrix I, i.e. the matrix with degrees from the five-element scale L, and the bottom table contains its graphical representation. The transformation from the table with scores to the matrix with degrees in $L = \{0, 0.25, 0.5, 0.75, 1\}$ is accomplished using the functions

$$s_j : [0, \dots, 1400] \to L \text{ defined by } s_j(p) = \mathit{round}\left(\frac{p - L_j}{H_j - L_j}\right) \tag{12}$$

where j is an attribute (discipline), and L_j and H_j are the lowest and the highest scores achieved by all the athletes (i.e. not only the top 5) who participated in the competition, and *round* is the function rounding the numbers in [0, 1] to their closest values in L. Note that in this competition, we have $L_{10} = 746$, $L_{lj} = 723$, $L_{sp} = 657$, $L_{hj} = 644$, $L_{40} = 673$, $L_{hu} = 755$, $L_{dt} = 622$, $L_{pv} = 673$, $L_{jt} = 598$, $L_{15} = 466$, and $H_{10} = 989$, $H_{lj} = 1050$, $H_{sp} = 873$, $H_{hj} = 944$, $H_{40} = 968$, $H_{hu} = 978$, $H_{dt} = 905$, $H_{pv} = 1035$, $H_{jt} = 897$, $H_{15} = 791$. Therefore, the degree assigned to Sebrle in 400m is $\mathit{round}\left(\frac{892-673}{968-673}\right) = \mathit{round}(0.74\dots) = 0.75$. Even though we lose some information using such rounding to five degrees, the information preserved still allows us to perform a reasonable analysis, which is shown next. The degrees in I allow us to interpret the athletes' results verbally, e.g. by assigning "not at all" to 0, "little bit" to 0.25, "half" to 0.5, "quite" to 0.75, and "fully" to 1.

The algorithm from [9] found a decomposition of I using six factors depicted in Table 2. The corresponding decomposition $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$ is depicted in Fig. 1. As explained in Section 2.2, cf. (9), the columns of $A_{\mathcal{F}}$ corresponding to $\langle C_l, D_l\rangle \in \mathcal{F}$ contain the degrees assigned to the athletes by C_l; likewise for the rows of $B_{\mathcal{F}}$, the attributes, and D_l.

Figure 2 shows rectangular patterns using which the factors may be visualized. Each rectangular pattern labeled F_l is actually the matrix J_l resulting as the Cartesian product of the extent C_l and the intent D_l of F_l, i.e. we have $(J_l)_{ij} = C_l(i) \otimes D_l(j)$. (For readers

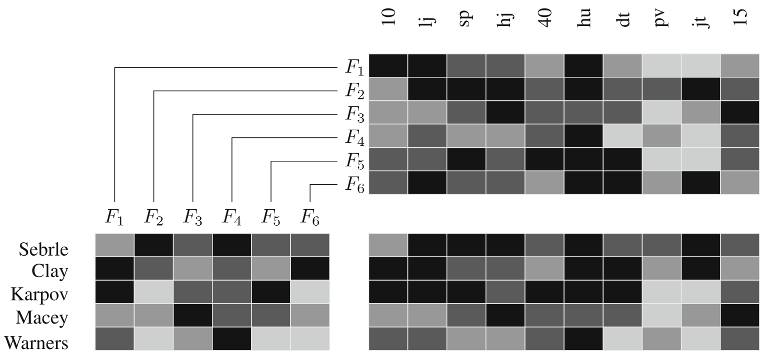


Fig. 1 Decomposition $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$. I, $A_{\mathcal{F}}$, and $B_{\mathcal{F}}$ are the *bottom-right*, *bottom-left*, and *top* matrix, respectively

familiar with the ordinary FCA, let us note that these patterns are the rectangles corresponding to formal concepts and that in the general situation with degrees, the concepts cannot be uniquely restored from these patterns.)

Figure 3 demonstrates what percentage of matrix I is explained using the first l factors for $l = 1, \ldots, k$, i.e. by the set $\mathcal{F}_l = \{\langle C_1, D_1\rangle, \ldots, \langle C_l, D_l\rangle\}$. In particular, the matrix labeled 67 % (42 %) just shows the rectangular pattern J_1 corresponding to F_1. The number indicates that 42 % of the entries in I have the same value as in J_1, i.e. $s_=(I, A_{\mathcal{F}_1} \circ B_{\mathcal{F}_1}) = 0.42$, and that if we also take closeness of truth degrees into account, 67 % of the data is explained by the first factor, i.e. $s_\approx(I, A_{\mathcal{F}_1} \circ B_{\mathcal{F}_1}) = 0.67$,. The second matrix contains $J_1 \vee J_2$, i.e. it illustrates what happens when we add the second factor. As we can see, 86 % of the data is explained by the first two factors. Since the first three factors explain 93 % of the data, one might say that the first three factors account for most of the data, are most important, and the rest of the factors may be omitted. This fact may also be observed by visual inspection of the superposition of the first three factors and its comparison to I. Nevertheless, adding further the factors we see that the first four, five, and six factors explain 96 %, 98.5 %, and 100 % of the data (the latter fact simply means that the six factors exactly decompose the matrix I).

Let us turn to the interpretation of the factors. For this purpose, Table 2 is crucial since it contains all the information about the factors. Note however that Fig. 2 is also helpful as it shows the clusters corresponding to the factor concepts which draw together the athletes and their performances in the events. Factor F_1: F_1 applies to Sebrle to degree 0.5, to both Clay and Karpov to degree 1, to Macey to degree 0.5, and to Warners to degree 0.75.

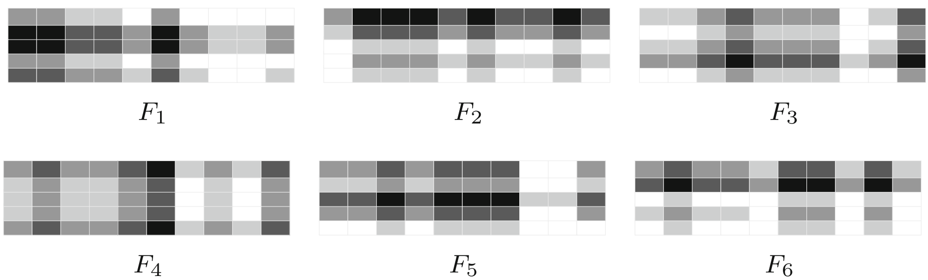


Fig. 2 Factor concepts as rectangular patterns

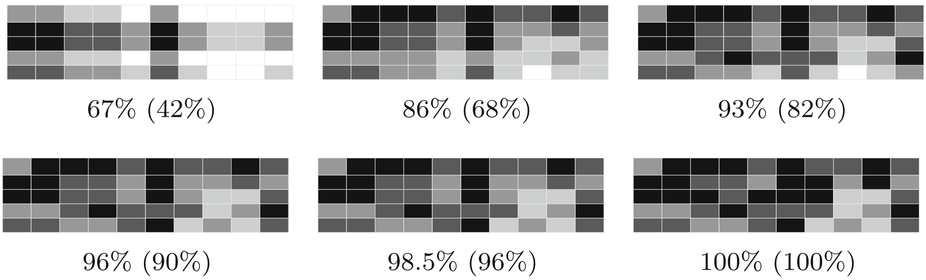


Fig. 3 Matrices $A_{\mathcal{F}_l} \circ B_{\mathcal{F}_l}$, representing the $\bigvee$-superposition of the set $\mathcal{F}_l = \{\langle C_1, D_1\rangle, \ldots, \langle C_l, D_l\rangle\}$ of the first l factors, $l = 1, \ldots, 6$. The corresponding percentage $100 \cdot s_{\approx}\%$ $(100 \cdot s_{=}\%)$ of I explained by the first l factors

Furthermore, this factor applies to attribute 10 (100 m) to degree 1, to attribute *lj* (long jump) to degree 1, to attribute *sp* (shot put) to degree 0.75, etc. This means that an excellent performance (degree 1) in 100 m, an excellent performance in long jump, a very good performance (degree 0.75) in shot put, etc. are particular manifestations of this factor. On the other hand, only a relatively weak performance (degree 0.25) in javelin throw and pole vault are manifestations of this factor. All the manifestations of this factor with degree 1 are 100 m, long jump, and 110 m hurdles. This factor can be interpreted as speed, i.e. the ability to run fast for short distances. Note that this factor applies particularly to Clay and Karpov which is well known in the world of decathlon. Factor F_2: Similarly, since the manifestations of this factor with degree 1 are long jump, shot put, high jump, and javelin, F_2 can be interpreted as explosiveness, i.e. the ability to apply very high force in a very short term. F_2 applies particularly to Sebrle, and then to Clay, who are known for this ability. Factor F_3: Manifestations with grade 1 are high jump and 1500 m. This factor is typical for lighter, not very muscular athletes. Macey, who is evidently this type of decathlete (196 cm and 98 kg), is the athlete to whom the factor applies to degree 1. These are the most important factors behind the data matrix I. Note that the fact that the revealed factors are reasonable was confirmed to us by an experienced decathlon coach who also pointed out to us that F_2 (explosiveness) is known to be well-developed by the Czech school of decathlon (hence it applies to Sebrle).

4.2 2004 Olympic Games Decathlon—top 5 by their best results

In this example, we take the top 5 athletes of the 2004 Olympic Decathlon but we take their best performances during their decathlon competitions, instead of their actual performances

Table 3 2004 Olympic Games Decathlon

Scores of top 5 athletes—overall best scores in competitions										
	10	*lj*	*sp*	*hj*	40	*hu*	*di*	*pv*	*ja*	15
Sebrle	942	1089	880	944	921	1002	859	972	907	798
Clay	1010	1050	868	887	944	1022	993	941	920	670
Karpov	931	1073	910	915	968	984	929	1004	743	729
Macey	940	1002	841	944	998	931	836	849	799	990
Warners	947	1022	800	831	978	973	824	886	692	693

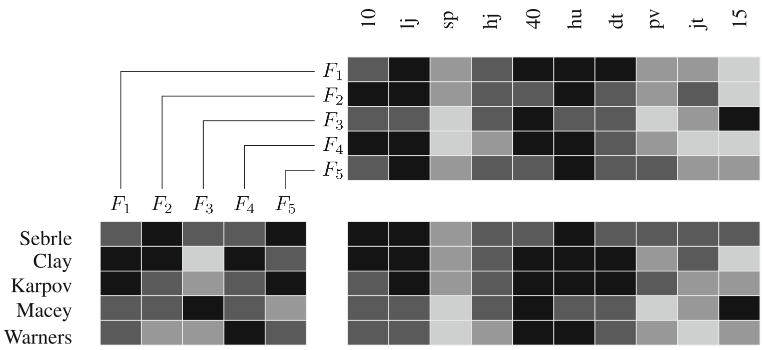


Fig. 4 Decomposition $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$

in a single event such as the 2004 Olympics. Taking best performances may be reasonable if we want to avoid a possible bad luck in a particular discipline such as a bad start in 100 m. Table 3 contains the scores. The corresponding matrix I and its decomposition into $A_{\mathcal{F}} \circ B_{\mathcal{F}}$ is depicted in Fig. 4. Here, the transformation from points to degrees is defined as follows. For discipline j, we put

$$s_j(p) = \begin{cases} 1 & \text{for } p \in [H_j, H_j - 100), \\ 0.75 & \text{for } p \in [H_j - 100, H_j - 200), \\ 0.5 & \text{for } p \in [H_j - 200, H_j - 300), \\ 0.25 & \text{for } p \in [H_j - 300, H_j - 400), \\ 0 & \text{for } p \leq H_j - 400, \end{cases}$$

where H_j is the highest score ever achieved during a decathlon competition for discipline j. Note that $H_{10} = 1042$; $H_{lj} = 1117$; $H_{sp} = 1048$; $H_{hj} = 1061$; $H_{40} = 1025$; $H_{hu} = 1064$; $H_{di} = 993$; $H_{pv} = 1152$; $H_{ja} = 1040$; $H_{15} = 963$.

It seems natural that the factors in this case are different from those in the example in Section 4.1. Nevertheless, we can see that F_1 applies to degree 1 to Clay and Karpov in both examples and applies to the other athletes to similar degrees in both examples as well. Nevertheless, the intents of the first factor are different although a considerable similarity of the intents of these factors is apparent, namely the presence of long jump and hurdles to degree 1, presence of 100 m and high jump to high degrees. Therefore, F_1 may reasonably be interpreted as speed in this example, as in case of the first factor in Section 4.1. A similar observation can be made on F_2 and F_3. Namely, F_2 connects Sebrle and Clay and and may, as the second factor in Section 4.1, be interpreted as explosiveness because it contains

Table 4 2004 Olympic Games Decathlon

Scores of the 5th–10th athletes	10	*lj*	*sp*	*hj*	40	*hu*	*di*	*pv*	*ja*	15
Zsivoczky	881	847	809	915	842	856	780	819	790	748
Hernu	867	859	768	831	874	942	761	849	704	782
Nool	906	942	744	698	870	874	706	1035	758	704
Bernard	931	930	777	915	855	953	762	731	667	704
Schwarzl	865	932	729	749	826	942	714	941	683	721

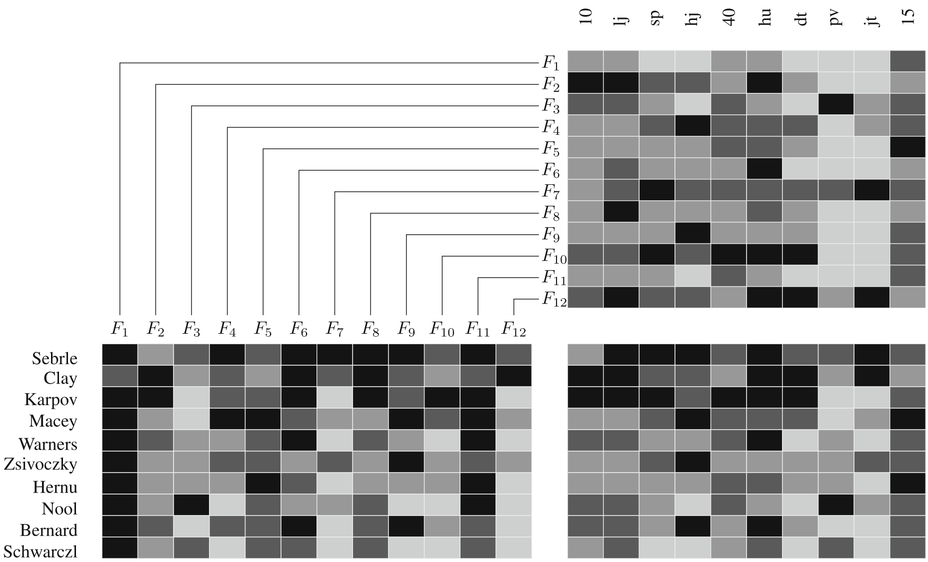


Fig. 5 Decomposition $J = A_{\mathcal{F}} \circ B_{\mathcal{F}}$ for J representing the top 10 athletes

100 m, long jump, hurdles, and javelin to high degrees. F_3 which is typical of Macey and a good performance in 400 m and 1500 m is typical of it, as in the case of the third factor in Section 4.1.

4.3 2004 Olympic Games Decathlon—top 10

The results of the 6th–10th athletes in the 2004 Olympic Decathlon are depicted in Table 4. The matrix J corresponding to the top 10 athletes, along with a decomposition $J = A_{\mathcal{F}} \circ B_{\mathcal{F}}$ computed by the algorithm from [9] is depicted in Fig. 5. The same transformation from scores to degrees was used as in Section 4.1. The corresponding percentage $100 \cdot s_{\approx}\%$ ($100 \cdot s_{=}\%$) of I explained by the first $l = 1, \ldots, 12$ factors is 71 %, 81.3 %, 85.3 %, 92 %, 93.5 %, 94.8 %, 96.8 %, 97.3 %, 97.8 %, 99 %, 99.3 %, 100 % for $s_{\approx}$ and 35 %, 52 %, 64 %, 76 %, 82 %, 86 %, 90 %, 92 %, 94 %, 97 %, 98 %, 100 % for $s_{=}$, hence again, already a small number of factors explains a reasonable portion of data.

Compared to the factors from Section 4.1, the factors in this example are generally different although some similarities are apparent. For example, factor F_2 here is exactly the same (has same intent) as F_1 in Section 4.1, F_{12} is the same as F_6 in Section 4.1, and F_4 is almost the same as F_3 in Section 4.1.

With this example, we get to the question from Section 3.3. That is, we ask how well do the six factors in $\mathcal{F}$ of I from Section 4.1 of top 5 athletes explain the new dataset regarding the top 10 athletes. For this purpose, we consider the set $\mathcal{G} = \mathcal{F}^J$ of formal concepts of J, see Observation 1, and consider the similarities $s_{\approx}(J, A_{\mathcal{G}_l} \circ B_{\mathcal{G}_l})$ and $s_{=}(J, A_{\mathcal{G}_l} \circ B_{\mathcal{G}_l})$ of J to the superposition of the set $\mathcal{G}_l$ consisting of the first l factors in $\mathcal{G}$, $l = 1, \ldots, 6$.

As one may check, $\mathcal{G}_1$ (first factor) explains 66 % (33 % for $s_{=}$) of J, $\mathcal{G}_2$ (first two factors) 78 % (49 %), $\mathcal{G}_3$ (first three factors) 88 % (66 %), $\mathcal{G}_4$ (first four factors) 91 % (74 %), $\mathcal{G}_5$ (first five factors) 92 % (78 %), and the set $\mathcal{G} = \mathcal{F}^J$ of all six factors explains 93 % (80 %) of the data represented by J. Hence, one may conclude that the factors of the

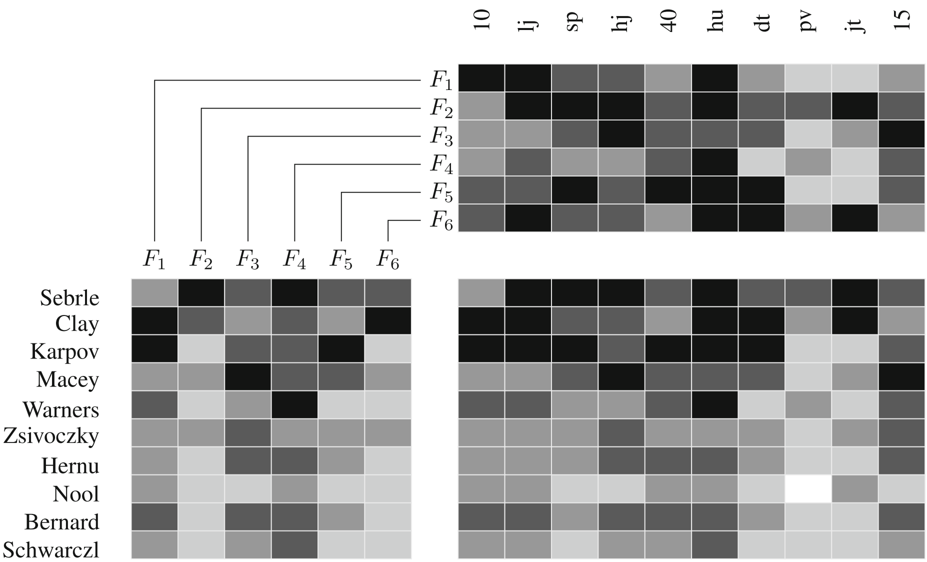


Fig. 6 Matrices $A_{\mathcal{F}^J} \circ B_{\mathcal{F}^J}$ (bottom-right), $A_{\mathcal{F}^J}$ (*bottom-left*), and $B_{\mathcal{F}^J}$ (*top*)

top 5 athletes explain reasonably well also the results of all the top 10 athletes. The matrices involved are depicted in Fig. 6. Note that one may clearly observe the similarity between I (the original matrix) and $A_{\mathcal{F}^J} \circ B_{\mathcal{F}^J}$ (the matrix reconstructed from the factors in $\mathcal{F}^J$).

We obtain analogous results even if we take for J the first top 20 athletes in the 2004 Olympic Decathlon. Namely, the first $l = 1, \ldots, 6$ factors in $\mathcal{F}^J$ in this case explain 63 % (26 %), 72 % (39 %), 81 % (52 %), 85 % (58 %), 87 % (63 %), and 88 % (65 %) of the 20×10 matrix J, respectively.

4.4 2004 Olympic Games Decathlon—Boolean factorization of ordinal scaling

We performed experiments with the possible approach examined in Section 4.4. That is, instead of the analysis in Section 4.1, we transformed the input matrix $I \in L^{5\times 10}$ with grades in $L = \{0, 0.25, 0.5, 0.75, 1\}$ to a Boolean matrix $I^\times \in \{0,1\}^{5\times(10\cdot 5)} = \{0,1\}^{5\times 50}$ and computed a set $\mathcal{G}^\times \subseteq \mathcal{B}(I^\times)$ of factors of $I^\times$ using the algorithm in [8]. Even though the decomposition algorithms are only approximation algorithms, the experimental results confirm the theoretical ones from Section 4.4 that the number of factors of I is in general smaller than that of $I^\times$. In this case, we obtained 8 factors of $I^\times$, compared to the 6 factors obtained for I. The 8 factors, $F_1, \ldots, F_8$, are depicted in a concise way in Fig. 7. As before, $A_{\mathcal{G}^\times}$ is the bottom-left matrix and its columns represent the factor extents, which are now ordinary sets of objects (athletes). To save space, the 8×50 Boolean matrix $B_{\mathcal{G}^\times}$ is represented by the top 8×10 matrix with grades as follows. For every attribute y (10, lj,…, 15), instead of the 5 columns $y_0, y_{0.25}, \ldots, y_1$ of $B_{\mathcal{G}^\times}$, the 8×10 matrix contains just one column which contains in row F_l the largest degree a for which y_a belongs to the intent of F_l. This way, the intent of F_l, an ordinary set of the scaled Boolean attributes y_a, is uniquely described because if y_b is in the intent and $c \leq b$, then y_c in in the intent as well. The corresponding percentage $100 \cdot s_{\approx}\%$ (which is the same as $100 \cdot s_{=}\%$ in the Boolean case) of

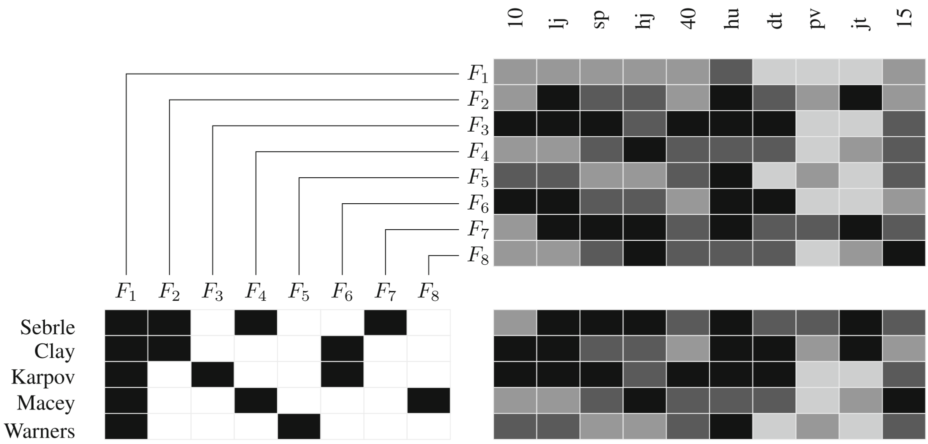


Fig. 7 Decomposition of $I^{\times} = A_{\mathcal{G}^{\times}} \circ B_{\mathcal{G}^{\times}}$

$I^{\times}$ explained by the first $l = 1, \ldots, 8$ factors is 76 %, 85 %, 91 %, 95 %, 98 %, 98.8 %, 99.6 %, and 100 %, respectively.

The factors may naturally be compared to those from Section 4.1 and the concise representation of the intents used in Fig. 7 facilitates this comparison. We may notice that F_6 here is very similar to F_1 from Section 4.1, and that F_7 and F_8 are vary similar (have even the same intents) to F_2 and F_3 from Section 4.1. These factors are clearly interpretable (see Section 4.1). The rest of the factors in Section 4.1, F_4, F_5, and F_6, are very similar to F_5, F_3, and F_2 here, but their interpretation is not so intuitively clear as for the above three. Moreover, the remaining factors here, F_1 and F_4, have no counterparts among those in Section 4.1 and seem not very interesting, particularly F_1, which can be described as "at least good performance in 10, lj, sp, hj, 40, 15, and at least rather good performance in hurdles" applies to all athletes.

To conclude, our experiments confirm that when using the alternative approach examined in Section 4.4, the number of factors to explain the data is larger. Moreover, the first, i.e. the most important, factors are not so clearly interpretable and perhaps also not so interesting compared to those obtained by the method examined in this paper, which directly works with degrees.

4.5 2004 Olympic Games modern pentathlon

Another kind of sport that contains several disciplines and may be interesting for factor analysis is the modern pentathlon. The five disciplines are, however, rather diverse and it is therefore challenging to think of natural factors in this sport. Recall that modern pentathlon consists of pistol shooting, fencing, 200 m freestyle swimming, show jumping, and a 3 km cross-country run. Except for the fencing competition, athletes do not directly compete against one another in the five events. Instead, a better absolute performance results in a higher score and the sum of all the scores for the disciplines gives the overall total score of a given athlete.

Table 5 2004 Olympic Games modern pentathlon

Scores of top 10 athletes	*sh*	*fe*	*sw*	*ri*	*ru*
Moiseev	1036	1000	1376	1032	1036
Zadneprovskis	1000	916	1308	1088	1116
Capalini	1084	776	1336	1116	1080
Cerkovskis	1096	916	1252	1004	1088
Meliakh	1168	692	1332	1144	1004
Michalik	1108	888	1260	1144	932
Walther	952	832	1336	1116	1084
Balogh	1036	804	1240	1172	1044
Iagorashvili	988	916	1252	1172	948
Sabirkhuzin	1156	888	1216	908	1034

sh—shooting; fe—fencing; sw—swimming; ri—riding; ru—running

Table 5 contains the results of the 2004 Olympic Games modern pentathlon of the top 10 athletes. To transform the scores of discipline j to degrees, we used the function

$$s_j(p) = \begin{cases} 1 & \text{for } p \in [H_j, H_j - \frac{1}{5}(H_j - L_j)), \\ 0.75 & \text{for } p \in [H_j - \frac{1}{5}(H_j - L_j), H_j - \frac{2}{5}(H_j - L_j)), \\ 0.5 & \text{for } p \in [H_j - \frac{2}{5}(H_j - L_j), H_j - \frac{3}{5}(H_j - L_j)), \\ 0.25 & \text{for } p \in [H_j - \frac{3}{5}(H_j - L_j), H_j - \frac{4}{5}(H_j - L_j)), \\ 0 & \text{for } p \leq H_j - \frac{4}{5}(H_j - L_j), \end{cases}$$

where H_j and L_j are the highest and the lowest score achieved in discipline j in the 2004 Olympic Games modern pentathlon. Note that $H_{sh} = 1168$, $L_{sh} = 892$; $H_{fe} = 1000$, $L_{fe} = 664$; $H_{sw} = 1376$, $L_{sw} = 1140$; $H_{ri} = 1172$, $L_{ri} = 584$; $H_{ru} = 1116$, $L_{ru} = 752$.

The corresponding matrix I and its decomposition into $A_{\mathcal{F}} \circ B_{\mathcal{F}}$ is depicted in Fig. 8. Note that of all the factors computed, F_2 is probably most interesting because it is actually known in the world of modern pentathlon. Namely, F_2's manifestations are riding and cross-country run which is typical for athletes who are in a good physical shape and have good endurance. Each of the other factors more or less corresponds to a single discipline which reflects the intuitive idea that the disciplines are diverse and require diverse skills. Not also that having disciplines that require diverse skills was the purpose of modern pentathlon.

4.6 Toward the choice of the scale of degrees

In the above examples, we used a five-element scale L equipped with Łukasiewicz operations. Our reason for using five degrees is motivated by the 7 ± 2 phenomenon described in the beginning of Section 4. We indeed experienced with a larger number of degrees, say 10, the undesirable effect which is in accordance with the 7 ± 2 phenomenon. Namely, among the factors extracted appear factors which, although mutually distinct, appear similar to a human expert because they involve close degrees, which are intuitively not distinct enough, such as 0.7 and 0.8. This impairs the interpretation of the factors in that even though each of the extracted factors alone makes a relatively good sense, the extracted factors do not satisfy the intuitive requirement of clear distinctiveness from the other factors. Moreover,

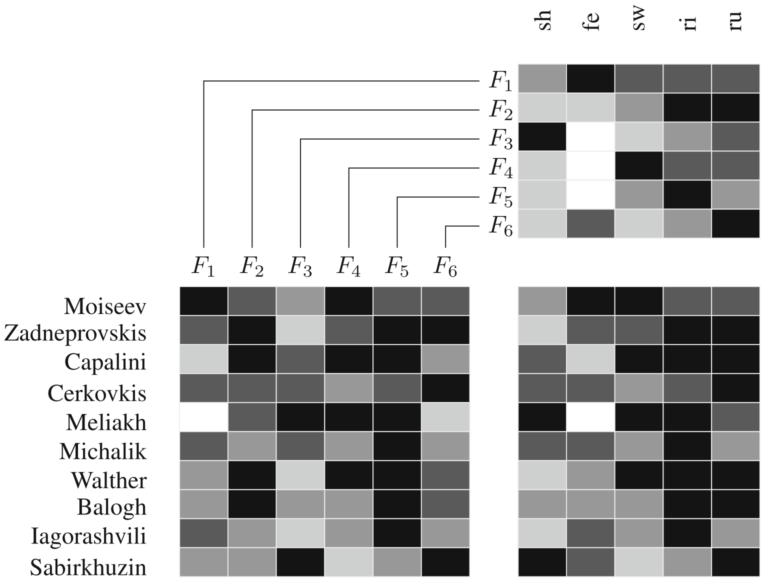


Fig. 8 Decomposition $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$

for the purpose of identifying and linguistically labeling the most important factors, a relatively smaller number of degrees seems sufficient. In addition to that, with more degrees of L in the input matrix I, resulting in the process of scaling via (12), I becomes more complex and thus the number of factors to decompose I gets larger. Therefore, "too much precision" introduced by having a larger L, impairs the practical aspects of the analysis several ways. This is illustrated in Fig. 9, which shows the result of decomposition of matrix I representing the same decathlon data as in Section 4.1 but using a ten-element scale $L = \{0, 0.1, \ldots, 0.9, 1\}$. We can see that, for instance, F_1, F_3, and F_6 are rather similar to F_4, F_1, and F_3 from Section 4.1. However, several factors here are mutually similar and

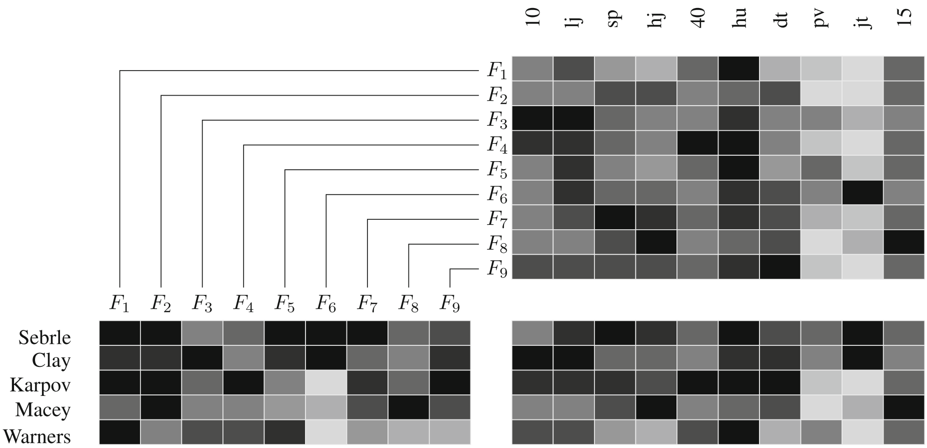


Fig. 9 Decomposition $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$

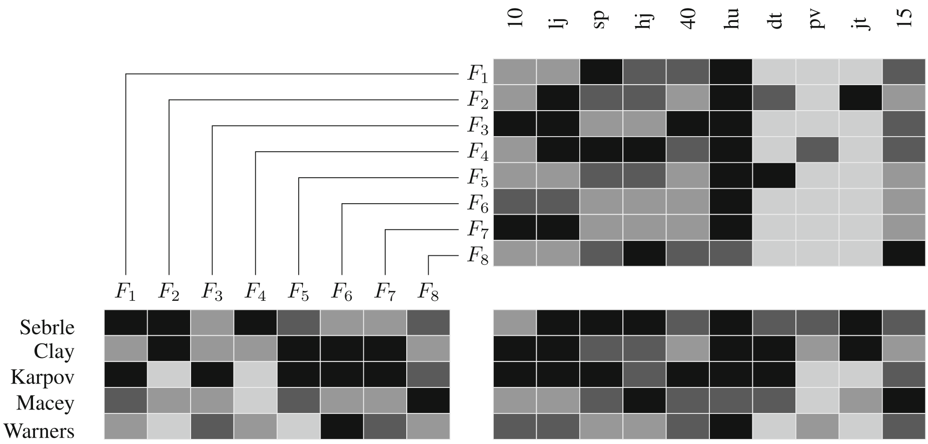


Fig. 10 Decomposition using Gödel operations

difficult to distinguish. These observations point to a possibly interesting topic for future research, namely working out a suitable concept of distinguishability (analogous to independence or orthogonality) which would help solve these issues and would possibly be useful even for small scales.

As far as the choice of the operations on L is concerned, we mainly use Łukasiewicz, as in the above examples, because of some of its intuitive properties. For example, the implication $\rightarrow$ naturally corresponds to the natural distance in [0, 1], resulting e.g. in the property that for the logical equivalence $a \leftrightarrow b = (a \rightarrow b) \wedge (b \rightarrow a)$ we have $a \leftrightarrow b = 1 - |a - b|$. Recall that the Łukasiewicz conjunction, along with the Gödel ($a \otimes b = \min(a, b)$, whose residuum is $a \rightarrow b = 1$ if $a \leq b$ and $a > b = b$ if $a > b$) and Goguen conjunction ($a \otimes b = a \cdot b$, whose residuum is $a \rightarrow b = 1$ if $a \leq b$ and $a > b = b/a$ if $a > b$) are the basic conjunctions (namely, any other continuous conjunction is an ordinal sum of the three [16]). Our experience is that with the other operations, we obtain different factors but the corresponding sets of factors contain important factors in common. Figure 10 shows the result of decomposition of the matrix I from Section 4.1 using Gödel operations. Notice that F_1, F_2, and F_3 from Section 4.1 are similar to F_7, F_2, and F_8 here and let us note that F_1, F_2, and F_3 from Section 4.1 also have their counterparts among the factors obtained using the Goguen operations. Apparently, a more profound treatment of the problem of the choice of the operations on the scales is needed. This concerns not only factor analysis and FCA in fuzzy setting but fuzzy set theory and its applications in general. In this regard, the theory of measurement from mathematical psychology seems an appropriate framework for such considerations (see [5] for some first steps in FCA in a fuzzy setting).

5 Conclusions and future work

We presented several examples of factor analysis of sports data using a recently developed method that utilizes formal concepts as factors. In addition, we proposed ways to answer natural questions regarding how well a set of factors explains a given dataset. Our main aim was to explain the method, to illustrate the key notions used in the method, and to demonstrate how one can interpret the results of the method. It turns out from the examples that the method yields reasonable and, in a sense, robust factors and that the results of the method are easy to understand. We performed analyses of other decathlon competitions as well as other sports data such as figure skating competitions and ice hockey players' performance and obtained results that confirm the aforementioned assessment of the method.

Future research shall include several topics, including those mentioned in Section 4.6. One is to exploit better the notion of similarity which naturally appears in data with grades. One particular topic we plan to explore is to modify the existing algorithm from [9] to take similarity of degrees into account when computing factors. Another topic consists in a suitable simplification of the input data utilizing similarity of matrix entries with the purpose of making it easier to compute factors, yet retaining good reconstructability of the original input data. Another issue related to similarity is a possible sampling procedure for the purpose of obtaining from the possibly large input data a smaller but representative subcollection whose factors may be used as factors of the original dataset. We would also like to mention [6], which appeared during the revision of the present paper, where further theoretical results leading to a new algorithm are presented—these results exploit the structure of the concept lattice associated to I which is worth furher examination. Some of the notions and results, in particular of Sections 3.2 and 3.3, are to be considered the first steps in addressing in a more comprehensive way the respective problems. Another appealing question that shall be a subject of future research is a comparison, experimental and possibly theoretical, of relationships of the presented method with related methods that involve matrix decomposition, notably the non-negative matrix factorization [14, 18].

References

1. Bartl, E., Belohlavek, R., Konecny, J.: Optimal decompositions of matrices with grades into binary and graded matrices. Ann. Math. Artif. Intell. **59**(2), 151–167 (2010)
2. Belohlavek, R.: Fuzzy Relational Systems: Foundations and Principles. Kluwer Academic/Plenum Publishers, New York (2002)
3. Belohlavek, R.: Concept lattices and order in fuzzy logic. Ann. Pure Appl. Logic **128**(1–3), 277–298 (2004)
4. Belohlavek, R.: Optimal decompositions of matrices with entries from residuated lattices. J. Logic Comput. **22**(6), 1405–1425 (2012)
5. Belohlavek, R.: Ordinally equivalent data: a measurement-theoretic look at formal concept analysis of fuzzy attributes. Int. J. Approx. Reason. **54**(9), 1496–1506 (2013)
6. Belohlavek, R., Krmelova, M.: Beyond Boolean matrix decompositions: toward factor analysis and dimensionality reduction of ordinal data, pp. 961–966. ICDM (2013)
7. Belohlavek, R., Sklenar, V., Zacpal, J.: Crisply generated fuzzy concepts. LNAI **3403**, 268–283 (2005)
8. Belohlavek, R., Vychodil, V.: Discovery of optimal factors in binary data via a novel method of matrix decomposition. J. Comput. Syst. Sci. **76**(1), 3–20 (2010)
9. Belohlavek, R., Vychodil, V.: Factor analysis of incidence data via novel decomposition of matrices. LNAI **5548**, 83–97 (2009)
10. Belohlavek, R., Vychodil, V.: Formal concept analysis and linguistic hedges. Int. J. Gen. Syst. **41**(5), 503–532 (2012)
11. Eldén, L.: Matrix Methods in Data Mining and Pattern Recognition. SIAM, Philadelphia (2007)
12. Ganter, B., Wille, R.: Formal Concept Analysis. Mathematical Foundations. Springer, Berlin (1999)
13. Geerts, F., Goethals, B., Mielikäinen, T.: Tiling databases. DS. LNCS **3245**, 278–289 (2004)
14. Golub, G., Van Loan, C.: Matrix Computations. Johns Hopkins University Press, Baltimore (1996)
15. Goguen, J.A.: The logic of inexact concepts. Synthese **18**(1968–1969), 325–373
16. Gottwald, S.: A Treatise on Many-Valued Logics. Research Studies Press, Baldock (2001)
17. Hájek, P.: Metamathematics of Fuzzy Logic. Kluwer, Dordrecht (1998)
18. Lee, D., Seung, H.: Learning the parts of objects by non-negative matrix factorization. Nature **401**, 788–791 (1999)

19. Miettinen, P., Mielikäinen, T., Gionis, A., Das, G., Mannila, H.: The discrete basis problem. IEEE Trans. Knowl. Data Eng. **20**(10), 1348–1362 (2008)
20. Miettinen, P., Vreeken, J.: Model order selection for Boolean matrix factorization. Proc. ACM SIGKDD, 51–59 (2011)
21. Miller, G.A.: The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol. Rev. **63**, 81–97 (1956)
22. Stockmeyer, L.J.: The set basis problem is NP-complete. IBM Research Report RC5431, Yorktown Heights (1975)
23. Tatti, N., Mielikäinen, T., Gionis, A., Mannila, H.: What is the dimension of your binary data? In: The 2006 IEEE conference on data mining (ICDM 2006), pp. 603–612. IEEE Computer Society (2006)
24. Zadeh, L.A.: Fuzzy sets. Inf. Control. **8**, 338–353 (1965)