Factor analysis of sports data via decomposition of matrices with grades*

Radim Belohlavek, Marketa Krmelova

Data Analysis and Modeling Lab (DAMOL)
Dept. Computer Science, Palacky University, Olomouc
radim.belohlavek@acm.org, marketa.krmelova@gmail.com

Abstract. The aim of this paper is to present experimental results on a recently developed method of factor analysis of data with graded, or fuzzy, attributes. The method utilizes formal concepts of data with graded attributes. In our previous papers, we described the factor model, the method, an algorithm to compute factors, and provided basic examples. In this paper, we perform a more extensive experimentation with this method. In particular, we apply the method to factor analysis of sports data. The aim of the paper is to demonstrate that the method yields reasonable factors, explain in detail how the factor model and the factors are to be understood, and to put forward new issues relevant to the method.

1 Introduction

1.1 Aim of This Paper

Recently, a considerable effort was devoted to the development of factor analysis and related methods for new types of data such as Boolean (binary) or ordinal. In our previous papers, we developed a method of factor analysis of Boolean data [5], i.e. data with Boolean attributes, and extended the problem and method to data with graded attributes [4, 6]. In the present paper, we use the method as well as the algorithm from [4, 6]. Due to the limited scope and the aim of this paper, we only provide a brief, mainly informal overview of the key notions involved, illustrate these notions by examples and refer the reader to [4, 6] for technical details. Our aim is to provide information sufficient to understand the experiments described in this paper. A full version of this paper will contain a detailed description of the method, a more comprehensive experimental section, formal treatment of some issues that we only discuss informally in this paper (cf. also Section 3), as well as a section putting the method being discussed into perspective of related methods of data analysis.

* We acknowledge support by the Grant No. P202/10/0262 of the Czech Science Foundation (R. Belohlavek); and by the IGA of Palacky University, No. PrF_20124029 (M. Krmelova).

ISBN 978–84–695–5252–0, Universidad de Málaga (Dept. Matemática Aplicada)

1.2 The Method, Factors, and Their Interpretation

In a broad sense, our method may be considered as implementing the general idea of factor analysis [1, 12]: Given an $n \times m$ object-attribute matrix I, one finds a decomposition

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

of I into a product of an $n \times k$ object-factor matrix A, a $k \times m$ matrix B, revealing thus k factors, i.e. new, possibly more fundamental attributes (or variables), which explain the original m attributes. We want $k < m$ and, in fact, k as small as possible to achieve parsimony: The n objects described by m attributes via I may then be described by k factors via A, with B representing a relationship between the original attributes and the factors. Contrary to classic factor analysis, which uses the calculus of real-valued matrices, we use the calculus of matrices over residuated lattices. That is, the entries of matrices involved are elements of a residuated lattice $\mathbf{L} = \langle L, \wedge, \vee, \otimes, \rightarrow, 0, 1 \rangle$, i.e. $I_{ij}, A_{il}, B_{lj} \in L$. The elements of L represent truth degrees, 0 and 1 are the smallest and largest one and correspond to "(fully) false" and "(fully) true"; $\wedge$ and $\vee$ denote the infimum and supremum, and $\otimes$ and $\rightarrow$ denote the truth functions on many-valued logic connectives of conjunction and implication. The product $\circ$ in (1) is defined by

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

Importantly, the entries of I, A, and B are interpreted the following way:

I_{ij} is the truth degree of the proposition "object i has attribute j",

A_{il} is the truth degree of the proposition "factor l applies to object i",

B_{lj} is the truth degree of "attribute j is one of the manifestations of factor l".

For the moment, think of i, j, and l as a particular athlete (object), good performance in long jump (attribute), and good speeding ability (factor). Using the principles of fuzzy logic [11], (2) and hence the whole factor model has the following meaning (this is even easy to see using intuition knowing that "exists" and "and" are modeled by $\bigvee$ and $\otimes$):

object i has attribute j if and only if
there exists factor l such that i has l (or, l applies to i) (3)
and j is one of the particular manifestations of l.

In principle, our method works as follows. We compute from I, using a greedy approximation algorithm from [6], a set

$$\mathcal{F} = \{\langle C_1, D_1 \rangle, \ldots, \langle C_k, D_k \rangle\} \subseteq \mathcal{B}(X, Y, I) \tag{4}$$

of formal fuzzy concepts of I, which gives us the decomposition as follows. Put

$$(A_{\mathcal{F}})_{il} = (C_l)(i) \quad \text{and} \quad (B_{\mathcal{F}})_{lj} = (D_l)(j), \tag{5}$$

i.e. $A_{\mathcal{F}}$ is an $n \times k$ matrix in which the lth column consists of grades assigned to objects by the lth concept extent C_l and $B_{\mathcal{F}}$ is a $k \times m$ matrix in which the lth row consists of grades assigned to attributes by the lth intent D_l. Then $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$, i.e. the matrices $A_{\mathcal{F}}$ and $B_{\mathcal{F}}$ induced by $\mathcal{F}$ provide us with

a decomposition of I. Moreover, the optimal decompositions (i.e. with least k possible) may in principle be obtained this way, in which sense using formal concepts as factors is an optimal strategy. Note however, that our algorithm computes suboptimal decompositions since the problem to compute an optimal decomposition is an NP-hard optimization problem.

We revisit these notions, particularly in Section 2.1, where we explain the notions involved using a particular example.

2 Examples and Experiments

Our aim in this section is to present the results of selected analyses, and thus to demonstrate a usefulness of the method, as well as to explain in detail the process of analysis, possibly even for a user who is not familiar with the technicalities of the method.

First, we point out some features common to the examples presented below. As the complete residuated lattice, we use as five-element Łukasiewicz chain. That is, the matrix degrees are taken from the set

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

and the operation $\otimes$ is given by

$$a \otimes b = \max(0, a + b - 1).$$

Many other choices are available, see e.g. [10]. 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):

Note that due to the well-known Miller's 7 ± 2 phenomenon [15], small scales with up to 7 ± 2 degrees are preferable to use because humans can understand and use such scales easily. For a reader not familiar with basics of many-valued logics let us note that the Łukasiewicz $\otimes$ (such as other many-valued conjunction) may be seen as a natural conjunction-like aggregation: the higher the truth values a and b of propositions A and B, the higher the truth value $a \otimes b$ of the conjunction $A\&B$.

2.1 2004 Olymphic 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 [6], 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 in the ten disciplines of decathlon, 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". That is, such an attribute applies to an athlete (object) to a degree

to which we consider the performance of the athlete 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 10.75sec in 100m gets 962 points, the result of 204cm 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 Tab. 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 from $L = \{0, 0.25, 0.5, 0.75, 1\}$ is accomplished using functions

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

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 $\text{round}(\frac{892-673}{968-673}) = \text{round}(0.74\ldots) = 0.75$. The matrix I allows us to interpret the athletes' results verbally. Namely, assigning to the degrees from L linguistic labels such as "not at all" to 0, "little bit" to 0.25, "half" to 0.5, "quite" to 0.75, and "fully" to 1, or the like, one may say that Sebrle's performance in 400m was quite good. 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 algorithm from [6] found a decomposition of I using six factors depicted in Fig. 2. The corresponding decomposition $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$ is depicted in Fig. 1. As explained in Section 1, cf. (5), the columns of $A_{\mathcal{F}}$ corresponding to $F_l = \langle C_l, D_l \rangle$ contain the degrees assigned to the athletes by C_l; likewise for the rows of $B_{\mathcal{F}}$, the attributes, and D_l.

Fig. 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 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.)

Table 1. 2004 Olympic Games Decathlon

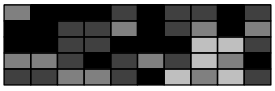
Scores of Top 5 Athletes

	10	*lj*	*sp*	*hj*	40	*hu*	*di*	*pv*	*ja*	15
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)

	10	*lj*	*sp*	*hj*	40	*hu*	*di*	*pv*	*ja*	15
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



Legend: 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.

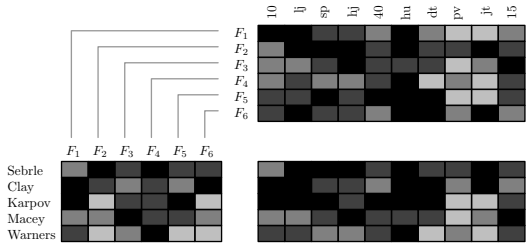


Fig. 1. Decomposition $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$. I, $A_{\mathcal{F}}$, and $B_{\mathcal{F}}$ are the bottom-right, bottom-left, and top matrix, respectively.

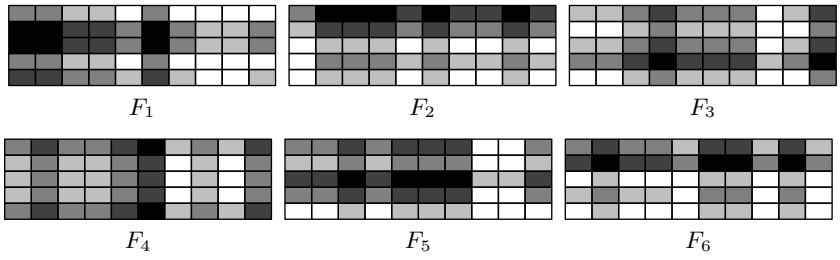


Fig. 2. Factor Concepts as Rectangular Patterns.

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}

Fig. 3 demonstrates what portion of matrix I is explained using the first l factors for $l = 1, \ldots, k$. In particular, the matrix labeled 56% just shows the rectangular pattern J_1 corresponding to F_1. The number indicates that 56% of the entries in I have the same value as in J_1, i.e. 56% of the data is explained by the first factor. 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, 82% of the data is explained by the first two factors. Since the first three factors explain 91% 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. Nevertheless, adding further the factors we see that the first four, five, and six factors explain 95%, 98%, and 100% of the data (the latter fact is obviously true because the factors completely decompose matrix I). Note also, that several of the $18\% = 100\% - 82\%$ of the entries not explained by the first two factors have values close to the corresponding entries of I, so a measure of closeness of J_l and I which takes into account also close entries, rather than exactly equal ones only, would yield a number larger than 82%. In any case, we can conclude from the visual inspection of the matrices that already the first two or three factors explain the data reasonably well. 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 Sebrle).

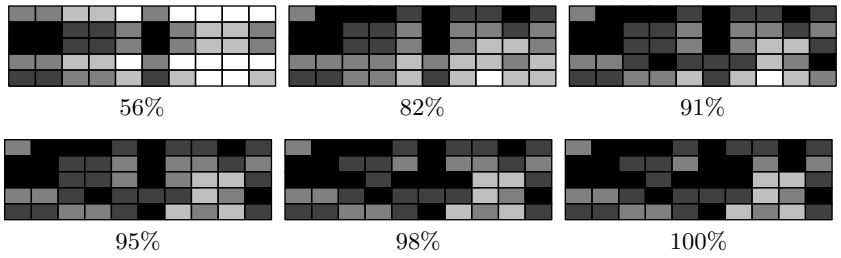


Fig. 3. $\bigvee$-superposition of Factor Concepts

Let us turn to the interpretation of the factors. For this purpose, Fig. 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. 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 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 the ability to apply very high force in a very short term (explosiveness). 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 that type among decathletes (196 cm and 98 kg) is the athlete to whom the factor applies to degree 1. These are the most important factors behind data matrix I.

2.2 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 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. Tab. 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 2.1. 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 reasonable similarity is apparent as well (presence of long

Table 3. 2004 Olympic Games Decathlon

Scores of Top 5 Athletes

	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

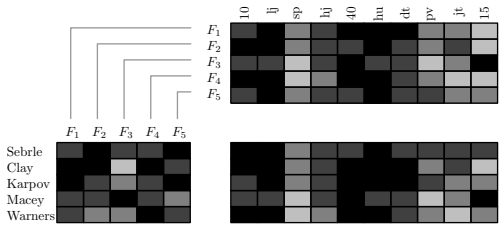


Fig. 4. Decomposition $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$.

jump and hurdles to degree 1, presence of 100 m and high jump to high degrees). A similar observation can be made on F_2 (connects Sebrle and Clay) and F_3 which is typical of Macey.

2.3 2004 Olympic Games Decathlon—Top 10

The results of the 5th–10th athletes in the 2004 Olympic Decathlon are depicted in Tab. 4. The matrix I corresponding to the top 10 athletes, along with a decomposition $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$ computed by the algorithm is depicted in Fig. 5. The same transformation from scores to degrees was used as in Section 2.1.

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

Compared to the factors from Section 2.1, the factors in this example are generally different although some similarities are apparent. For example, factor

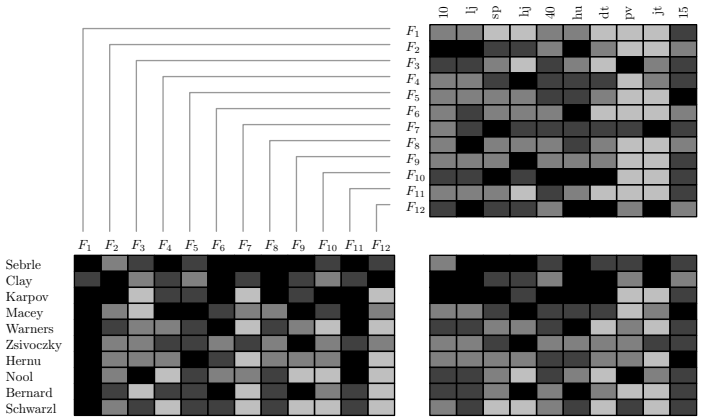


Fig. 5. Decomposition $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$.

F_2 here is exactly the same (has same intent) as F_1 in Section 2.1, F_{12} is the same as F_6 in Section 2.1, and F_4 is almost the same as F_3 in Section 2.1.

We might nevertheless be interested in the question of how well the factors from Section 2.1 explain the new dataset regarding the top 10 athletes. A reasonable way to proceed is the following. Consider the set of 6 concepts of the new, 10×10 matrix I, namely,

$$\mathcal{G} = \{G_1 = \langle P_1, Q_1 \rangle, \ldots, G_6 = \langle P_6, Q_6 \rangle\}$$

obtained from the factors $F_1 = \langle C_1, D_1 \rangle, \ldots, F_6 = \langle C_6, D_6 \rangle$ by

$$P_1 = D_1^{\downarrow}, Q_1 = P_1^{\uparrow}, \ldots, P_6 = D_6^{\downarrow}, Q_6 = P_6^{\uparrow},$$

i.e. every factor G_l is the concept of the 10×10 matrix I generated by the intent of F_l. This way, we do not have $I = A_{\mathcal{G}} \circ B_{\mathcal{G}}$ in general, as can be seen from this example. Nevertheless, the first factor G_1 explains 50% of the data, the first two factors 69%, the first three factors 80%, the first four factors 86%, the first five factors 89%, and all factors in $\mathcal{G}$ explain 91% of the data. Hence, one may conclude that the factors of the 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 $A_{\mathcal{G}} \circ B_{\mathcal{G}}$ (the matrix reconstructed from the factors in $\mathcal{G}$).

2.4 2004 Olympic Games Modern Pentathlon

Another 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

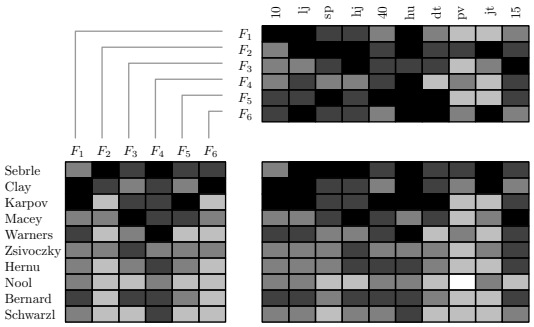


Fig. 6. Matrices $A_{\mathcal{G}} \circ B_{\mathcal{G}}$ (bottom-right), $A_{\mathcal{G}}$ (bottom-left), and $B_{\mathcal{G}}$ (top).

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.

Tab. 5 contains the results of the 2004 Olympic Games modern pentathlon of the top 10 athletes. The corresponding matrix I and its decomposition into $A_{\mathcal{F}} \circ B_{\mathcal{F}}$ is depicted in Fig. 7. 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 achieve in discipline j in the 2004 Olympic Games modern pentathlon. Note that $H_{sh} = 1168$, $L_{sh} = 892$; $H_{fe} = 1000$, $L_{fe} = 664$; $H_{sw} = 1376$, $L_{sw} = 1140$; $H_{ri} = 1172$, $L_{ri} = 584$; $H_{ru} = 1116$, $L_{ru} = 752$.

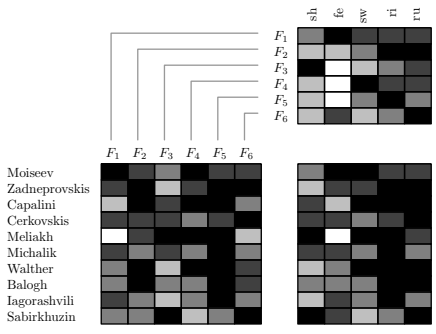


Fig. 7. Decomposition $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$.

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

Legend: *sh*—shooting; *fe*—fencing; *sw*—swimming; *ri*—riding; *ru*—running.

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 corresponds to the intuitive idea that the disciplines are diverse and require diverse skills.

3 Concusions, Further Issues and Future Work

We presented several examples of factor analysis of sports data using a recently developed method that utilizes formal concepts as factors. Our main aim was to explain the method, to illustrate the key notions used in the method, and to demonstrate how one can understand the results of the method. It turns out from the examples that the method yields reasonable factors and that the results of the method are easy to understand.

Due to the limited scope of this paper, we presented only a limited number of examples and limited comments on the presented examples. In addition to the examples presented in this paper, we performed factor analyses of further decathlon data (namely, the World Championships), figure skating, and ice hockey players performance. We refrained from formalizing some of the issues involved, such as "explanation of data by factors", "similarity of factors", how well the factors of one dataset serve as good factors of another dataset, etc., and used these notions with their informal meaning only. We therefore also skipped theoretical results regarding these notions, as well as further notions and results that may help us answer further natural questions that arise in the context of the presented method, such as the influence of the choice of the scale of degrees, the operation $\otimes$, the influence of the transformation from scores to degrees, and the like.

More examples with detailed comments as well as a detailed treatment of some of issues mentioned above will appear in the full version of this paper. An interesting question that is to be a subject of our future research is a comparison, experimental and possibly also theoretical, of relationships of the presented method with related methods that involve matrix decomposition, notable the non-negative matrix factorization [8, 14].

References

1. Bartholomew D. J., Knott M.: *Latent Variable Models and Factor Analysis, 2nd Ed.*, London, Arnold, 1999.
2. Bartl E., Belohlavek R., Konecny J.: Optimal decompositions of matrices with grades into binary and graded matrices. Annals of Mathematics and Artificial Intelligence 59(2)(2010), 151–167.
3. Belohlavek R.: Concept lattices and order in fuzzy logic. *Annals of Pure and Applied Logic* **128**(1–3)(2004), 277–298.
4. Belohlavek R.: Optimal decompositions of matrices with entries from residuated lattices. Journal of Logic and Computation, doi: 10.1093/logcom/exr023, online: September 7, 2011.
5. Belohlavek R., Vychodil V.: Discovery of optimal factors in binary data via a novel method of matrix decomposition. Journal of Computer and System Sciences 76(1)(2010), 3–20.
6. Belohlavek R., Vychodil V.: Factor analysis of incidence data via novel decomposition of matrices. LNAI 5548(2009), 83-97.
7. Ganter B., Wille R.: *Formal Concept Analysis. Mathematical Foundations.* Springer, Berlin, 1999.
8. Golub G., Van Loan C.: *Matrix Computations.* Johns Hopkins University Press, 1996.
9. J. A. Goguen. The logic of inexact concepts. *Synthese* **18** (1968–9), 325–373.
10. Gottwald S.: *A Treatise on Many-Valued Logics.* Research Studies Press, Baldock, Hertfordshire, England, 2001.
11. Hájek P.: *Metamathematics of Fuzzy Logic.* Kluwer, Dordrecht, 1998.
12. Harman H. H. : *Modern Factor Analysis, 2nd Ed.* The Univ. Chicago Press, Chicago, 1970.
13. Krantz H. H., Luce R. D., Suppes P., Tversky A.: *Foundations of Measurement.* Vol. I (Additive and Polynomial Representations), Vol. II (Geometric, Threshold, and Probabilistic Represenations), Vol. III (Represenations, Axiomatization, and Invariance). Dover Edition, 2007.
14. Lee D., Seung H.: Learning the parts of objects by non-negative matrix factorization. *Nature* **401**(1999), 788–791.
15. Miller G. A.: The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.* **63**(1956), 81–97.
16. Stockmeyer L. J.: The set basis problem is NP-complete. IBM Research Report RC5431, Yorktown Heights, NY, 1975.
17. Ward M., Dilworth R. P.: Residuated lattices. *Trans. Amer. Math. Soc.* **45** (1939), 335–354.
18. Zadeh L. A.: Fuzzy sets. *Inf. Control* **8**(1965), 338–353.