Toward Factor Analysis of Educational Data

Eduard Bartl[1], Radim Belohlavek[1], and Alex Scharaschkin[2]

[1] Palacky University, Olomouc, Czech Republic,
eduard.bartl@upol.cz, radim.belohlavek@acm.org
[2] Oxford University, United Kingdom,
AScharaschkin@aqa.org.uk

Abstract. Factor analysis of Boolean and ordinal data became a significant research direction in data analysis. In this paper we present a case study involving a recently developed method of factor analysis of ordinal data which uses the apparatus of fuzzy logic and closure structures. In particular, the method uses formal concepts of the input data as factors and is utilized in our paper to analyze British educational data. The results of the analyses demonstrate that the method is capable of extracting natural and well-interpretable factors which provide insight into students' performances in tests. Our study represents an initial phase of a project of analyzing educational data by means of relational methods. Broader ramifications and further prospects regarding this project are also discussed.

1 Introduction and Paper Outline

Analysis of factors in various kinds of data represents an important topic in the domain of data analysis. The factors are thought of as hidden variables that are more fundamental than the directly observable variables using which the given data is described. Discovery of such factors enables one to better understand the data as well as to reduce its dimensionality. The best known factor-analytic methods are those designed for real-valued data and include the classical factor analysis, the singular value decomposition, principal component analysis, and non-negative matrix factorization; see e.g. [1,10,13,17,20]. As is well known, the application of such methods to Boolean and ordinal data is possible in principle but these classical methods suffer from poor interpretability when applied to such data. In the past years, a considerable effort has been devoted to the development of matrix methods for Boolean data; see e.g. [7,12,21,22,24] and the references therein. In our previous papers [4,5,6,8,9], we extended the factorization problem for Boolean data to ordinal data, which is of our main interest in this paper. Since in a development of new data analysis methods, explorations of real-case studies play a crucial role, our main aim in this paper is to add to our previous studies a further case study. The analyzed data comes from examination tests in the United Kingdom. Our efforts are part of a broader project whose goal is to explore methods of relational data analysis for analyzing educational data.

Our paper is organized as follows. In section 2, we present our method to the extent that both the principles as well as the user point of view are clarified.

© Dmitry I. Ignatov, Lhouari Nourine (Eds.): CLA 2018, pp. 191–206,
ISBN 978–80–244–5328–6, Department of Computer Science, Palacký University Olomouc, 2018.

In section 3, which is the main section of this paper, we describe the data, our selected analyses of the data, and provide discussion of the results obtained. Section 4 concludes the paper by summarizing our results, putting our work in context, and describing our future goals.

2 Our Method of Factor Analysis

2.1 The Basic Idea of Our Factor Model and Its Interpretation

The Factor Model We assume that the analyzed data is in the form of an $n \times m$ matrix I describing n objects (matrix rows) and m graded attributes (matrix columns). The matrix entries I_{ij} contain degrees (grades, levels) from a given scale L, such as $L = \{0, 1/2, 1\}$ or $L = \{0, 1/4, 1/2, 3/4, 1\}$. The entry I_{ij} represents the degree to which the object i has the attribute j. Thus, $I_{ij} = 0$ means that i does not have j at all, $I_{ij} = 1$ means that i has j to the full extent, and $I_{ij} = 3/4$ means that i has j to a large extent. For instance, the objects and attributes might be students and exam tests, respectively, and the entries I_{ij} might represent the extents to which student i succeeded in test j. The following is an example of a matrix I over the three-element scale:

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

In our model, one looks for a decomposition (or, factorization) of the input $n \times m$ object-attribute matrix I into an (exact or approximate) product

$$I = A \circ B, \tag{2}$$

of an $n \times k$ object-factor matrix A and a $k \times m$ factor-attribute matrix B (the entries of both A and B are again degrees from the scale L).

The matrix product $\circ$ is defined by

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

where $\otimes$ is an appropriate aggregation function generalizing the classical logical conjunction and $\bigvee$ is the supremum operation in the scale L ($\bigvee$ is max if L is a chain, i.e. linearly ordered); see below for details.

To understand the meaning of this factor model, consider first its particular case in which $L = \{0, 1\}$, i.e. the Boolean case. Then, (3) becomes

$$(A \circ B)_{ij} = \max_{l=1}^{k} \min(A_{il}, B_{lj}) \tag{4}$$

which is the well-known Boolean matrix product. Equivalently, (4) reads:

$(A \circ B)_{ij} = 1$ iff there exists $l \in \{1, \ldots, k\}$ such that $A_{il} = 1$ and $B_{lj} = 1$, from which it is immediate that the factor model has the following meaning:

$$\begin{array}{l}\text{object } i \text{ has attribute } j \text{ if and only if} \\ \quad \text{there exists factor } l \text{ such that } i \text{ has } l \text{ (or, } l \text{ applies to } i) \\ \quad \text{and } j \text{ is one of the particular manifestations of } l,\end{array} \tag{5}$$

which may be regarded as a verbal description of the model given by (2). Such description is certainly appealing and well understandable.

With a general scale L, we approach the situation according to the principles of formal fuzzy logic [2,15,16] 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 regard A_{il} as the truth degree $||\varphi(i,l)||$ of $\varphi(i,l)$, and B_{lj} as the truth degree $||\psi(l,j)||$ of $\psi(l,j)$, i.e.

$$||\varphi(i,l)|| = A_{il} \text{ and } ||\psi(l,j)|| = B_{lj}. \tag{6}$$

Now, according to fuzzy logic, the truth degree of the 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). 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 (5), is computed by

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

where $\bigvee$ denotes the supremum. Taking (6) into account, we see that a generalization of (4) to the case of multiple degrees in L is just given by the above formula (3). Therefore, even in presence of multiple degrees, the factor model (2) retains its simple meaning described by (5).

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 [14,25], used in fuzzy logic; see [15,16] for details. Grades of ordinal scales [19] are conveniently represented by numbers, such as the Likert scale $\{1,\ldots,5\}$, which naturally appears in our experiments (see below). We assume that these numbers are normalized and taken from the unit interval $[0,1]$, i.e. they form the five-element scale $L = \{0, 1/4, 1/2, 3/4, 1\}$ commonly used in fuzzy logic. In our analyses, we use the Łukasiewicz operations on this scale, i.e. we use

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

but many other examples are available; see e.g. [15].

2.2 Factors Utilized by Our Method

It follows from the above description that for any decomposition (2), 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, [14,26]) $C : \{1,\ldots,n\} \to L$ of objects and $D : \{1,\ldots,m\} \to L$ of attributes, see [3], that satisfies $C^{\uparrow} = D$ and $D^{\downarrow} = C$ where ${}^{\uparrow} : L^X \to L^Y$ and ${}^{\downarrow} : L^Y \to L^X$ are the concept-forming operators defined by

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

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 factor l. Using formal concepts as factors is optimal in the following sense [4]: Let for a set (we fix the numbering of its elements)

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

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

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

Then whenever $I = A \circ B$ for some $n\times k$ 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.

In our experiments, we use the basic greedy algorithm proposed in [8] for computing a set $\mathcal{F}$ of concepts for which $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$; see also [9] for computational complexity of the problem and the algorithm.

2.3 Explanation of Data by Factors

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 the product $A_{\mathcal{F}} \circ B_{\mathcal{F}}$, i.e. to the data reconstructed from the factors in $\mathcal{F}$. To measure closeness of I and $A_{\mathcal{F}} \circ B_{\mathcal{F}}$, we use the function $s(I, A_{\mathcal{F}} \circ B_{\mathcal{F}})$ defined by

$$s(I, A_{\mathcal{F}} \circ B_{\mathcal{F}}) = \frac{\sum_{i,j=1}^{n,m}(I_{ij} \leftrightarrow (A_{\mathcal{F}} \circ B_{\mathcal{F}})_{ij})}{n \cdot m}, \tag{7}$$

where $\leftrightarrow$ is the biresiduum (i.e. many-valued logical equivalence). This function is proposed in the above-mentioned papers. In fact, it turned out during our experiments that we need a slight generalization of this function, which we describe below.

3 Educational Data and Its Factor Analysis

3.1 A Broader Context

Analyzing students' performance is a task which constantly occupies educators. On a small scale, teachers are naturally interested in performances of their individual students as well as of their classes to help their students improve, with respect to educational aims and objectives. On a large scale, understanding students' performance is of great interest at the national level: Education experts attempt to understand the effects of current curricula and approaches to education to possibly improve educational policies. Our project fits into this picture. We attempt to analyze students' performances as assessed by the tasks they attempts in examinations. Unlike the common approaches, which are mostly based on classical statistical methods, we propose to utilize the recently developed method of factor analysis of ordinal data described in the previous section. The limited extent of this paper prevents us from describing the results we obtained to any larger extent as well as from describing broader ramifications of the findings and comparison to analyses obtained by alternative methods. We therefore present a fraction of our results only, which nevertheless meets our primary purpose in this paper, namely to demonstrate that our method is capable of revealing natural and well-interpretable factors hidden in the outputs of educational assessments.

3.2 The Data

We analyze anonymized data coming from the official school-leaving (so-called A-level) examination tests that are used in the UK by universities to select students. In brief, our overall aim in this project is to see what factors may explain the students' performances. In addition, we are interested in the question of the so-called *construct validity* [23] of the examinations, namely the extent to which students' responses, assessed as being at a particular level, match the intentions of the assessment designers in terms of the qualitative performance standard intended to broadly characterise responses at that level. This is the kind of question that is difficult to study using traditional quantitative methods.

The data contains results of 2774 individual students' performances on a given examination in the subject "Government and Politics." The whole examination consists of four modules (i.e. four papers). Our data concerns the second module, which covers the current British governance.[3] Students choose two topics out of four possible. Our data contains the results for topic 2 (parliament) and topic 3 (the executive), which is the most popular combination. For each topic (2 and 3), the students answer three questions, i.e. six questions in total. The answer to each question, which is a piece of prose, is assessed by examiners with regard to three so-called assessment objectives, namely "knowledge and understanding," "analysis and evaluation," and "communication," with the exception

[3] The exam paper is available at `http://filestore.aqa.org.uk/sample-papers-and-mark-schemes/2016/june/AQA-GOVP2-QP-JUN16.PDF`.

of the first question in each topic for which only the first assessment objective is considered. As a result, 14 evaluations for each student examination (one for each topic, question, and permissible assessment objective) are obtained. A student obtains a mark in each of the 14 evaluations (the largest possible marks for the evaluations are mutually different in general). The sum of all marks gives the total mark of maximum value 80 assigned to this student, from which the grade for the student is obtained by a simple thresholding. The possible grades are A (the best grade, represented numerically by 5), B (represented by 4), C (3), D (2), E (1), and N (0). In addition, a simple scaling is defined for each of the 14 evaluations which assigns each possible mark for this evaluation a level on a five-element scale $0, \ldots, 4$ with 4 indicating the best performance. This scaling brings the evaluation data on a common scale.[4] Two exceptions are attributes 5 and 8 which are mapped on a three-element scale (level 0 represented by 0, level 1-2 represented by 1.5, and level 3-4 represented by 3.5). We nevertheless embed this three-element scale to the five-element one by the assignment $0 \mapsto 0$, 1-2 $\mapsto 1$, and 3-4 $\mapsto 3$. Each student examination is thus described by 14 fuzzy (graded) attributes over a five-element scale $L = \{0, {}^1/_4, {}^1/_2, {}^3/_4, 1\}$ whose degrees represent the levels $0, \ldots, 4$. A sample of the data sorted by the total marks is shown in Table 1: The first column represents the total marks, the second one represents the grades, and the remaining columns represent the 14 graded attributes, which are explained in Table 2.

3.3 Selected Analyses

The data described in the previous section may thus be represented by a 2774×14 matrix I with degrees in the scale $L = \{0, {}^1/_4, {}^1/_2, {}^3/_4, 1\}$, i.e. a matrix $I \in L^{2774 \times 14}$. A part of this matrix corresponding to the data from Table 1 is shown in Table 3.

We performed factor analyses of this data using the method described in section 2. In accordance to the intentions to understand the factors behind the various overall performance grades, we split the matrix I into 6 submatrices according to the grades. Thus, since there are 607 students who obtained grade A, we analyzed the corresponding 607×14 submatrix I_{A} of the whole matrix I and we performed this for the submatrices I_{G} for every grade G = A, ..., E, N.

Since our algorithm computes the factors one by one, from the most significant ones in terms of data coverage to the least significant until an exact factorization of the input matrix I_{G} is obtained, we observed the coverage of the data I_{G} by the first factor, by the first two factors, by the first three factors, and in general by the first $l = 1, \ldots, k$ factors where k is the total number of factors computed from the data. To measure coverage, which serves as an indicator of how well the data is explained by the factors, we first used the function (7) [5,6,8,9], which is a direct generalization of the coverage function from the Boolean case. We, however, observed a phenomenon not encountered in

[4] The marking scheme is described in `http://filestore.aqa.org.uk/sample-papers-and-mark-schemes/2016/june/AQA-GOVP2-W-MS-JUN16.PDF`.

Table 1: Sample of the examination data.

total	grade	k__cor__par	k__cor__exe	k__par__mac	e__par__mac	c__par__mac	k__cab__gov	e__cab__gov	c__cab__gov	k__par__mod	e__par__mod	c__par__mod	k__pow	e__pow	c__pow
78	5	4	4	4	4	3,5	4	4	3,5	4	4	4	4	4	4
77	5	4	4	4	4	3,5	3	2	3,5	4	4	4	4	4	4
76	5	3	3	4	4	3,5	4	4	3,5	4	4	4	4	4	4
75	5	2	4	4	4	3,5	3	3	3,5	4	4	4	4	4	4
75	5	3	3	4	3	3,5	4	4	3,5	4	4	4	4	4	4
74	5	4	3	4	3	3,5	4	4	3,5	4	4	4	4	4	4
73	5	4	3	4	3	3,5	3	3	3,5	4	4	4	4	4	4
73	5	3	4	4	3	3,5	4	3	3,5	4	4	4	4	4	4
73	5	3	3	4	4	3,5	4	3	3,5	4	4	4	4	3	4
73	5	4	3	4	4	3,5	4	4	3,5	3	4	4	4	4	4
73	5	4	4	4	3	3,5	4	3	1,5	4	4	4	4	4	4
73	5	3	4	4	3	3,5	3	4	3,5	4	4	4	4	4	4
73	5	4	4	4	4	3,5	3	3	3,5	4	4	4	4	4	4
73	5	3	3	4	4	3,5	3	4	3,5	4	4	4	4	4	4
73	5	3	2	3	3	3,5	4	4	3,5	4	3	4	4	4	4
72	5	4	4	4	4	3,5	4	4	3,5	4	4	4	3	3	3
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 2: Meaning of the 14 graded attributes.

attribute	label in table 3	description
1	k__cor__par	demonstrates knowledge of core parliamentary principles
2	k__cor__exe	demonstrates knowledge of core principles of the executive
3	k__par__mac	demonstrates knowledge of parliamentary machinery
4	e__par__mac	explains and analyses aspects of parliamentary machinery
5	c__par__mac	communicates effectively about aspects of parliamentary machinery
6	k__cab__gov	demonstrates knowledge of cabinet government
7	e__cab__gov	explains and analyses aspects of cabinet government
8	c__cab__gov	communicates effectively about aspects of cabinet government
9	k__par__mod	demonstrates knowledge of pros and cons of parliamentary models
10	e__par__mod	explains and analyses pros and cons of parliamentary models
11	c__par__mod	communicates effectively about pros and cons of parliamentary models
12	k__pow	demonstrates a knowledge of power structures in government
13	e__pow	explains and analyses power structures in government
14	c__pow	communicates effectively about power structures in government

Table 3: Matrix with grades, or a formal fuzzy context, representing the data from Table 1.

1	1	1	1	3/4	1	1	3/4	1	1	1	1	1	1
1	1	1	1	3/4	3/4	1/2	3/4	1	1	1	1	1	1
3/4	3/4	1	1	3/4	1	1	3/4	1	1	1	1	1	1
1/2	1	1	1	3/4	3/4	3/4	3/4	1	1	1	1	1	1
3/4	3/4	1	3/4	3/4	1	1	3/4	1	1	1	1	1	1
1	3/4	1	3/4	3/4	1	1	3/4	1	1	1	1	1	1
1	3/4	1	3/4	3/4	3/4	3/4	3/4	1	1	1	1	1	1
3/4	1	1	3/4	3/4	1	3/4	3/4	1	1	1	1	1	1
3/4	3/4	1	1	3/4	1	3/4	3/4	1	1	1	1	3/4	1
1	3/4	1	1	3/4	1	1	3/4	3/4	1	1	1	1	1
1	1	1	3/4	3/4	1	3/4	1/4	1	1	1	1	1	1
3/4	1	1	3/4	3/4	3/4	1	3/4	1	1	1	1	1	1
1	1	1	1	3/4	3/4	3/4	3/4	1	1	1	1	1	1
3/4	3/4	1	1	3/4	3/4	1	3/4	1	1	1	1	1	1
3/4	1/2	3/4	3/4	3/4	1	1	3/4	1	3/4	1	1	1	1
1	1	1	1	3/4	1	1	3/4	1	1	1	3/4	3/4	3/4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

the previous analyzes reported in the literature which is due to the considerable size of the data (a 2774×14 matrix over a 5-element scale is comparable to a $2774 \times (14 \cdot 5) = 2774 \times 70$ Boolean matrix [6]). Namely, the accumulation of the biresidua $I_{ij} \leftrightarrow (A_{\mathcal{F}} \circ B_{\mathcal{F}})_{ij}$ by the summation in (7) makes the algorithm select also flat factors. By "flat" we mean that the entries $C_l(i) \otimes D_l(j)$, by which the factor $\langle C_l, D_l \rangle$ contributes to the explanation of the input data, are close to $1/2$. In many such cases, we would naturally prefer factors that are less flat even though their coverage as measured by (7) is slightly smaller, because such factors are more discriminative and thus more informative. To solve this problem, we adjusted the function (7) as follows: The new function, $s_c(I, A_{\mathcal{F}} \circ B_{\mathcal{F}})$, is defined by

$$s_c(I, A_{\mathcal{F}} \circ B_{\mathcal{F}}) = \frac{\sum_{i,j=1}^{n,m} (c(I_{ij} \leftrightarrow (A_{\mathcal{F}} \circ B_{\mathcal{F}})_{ij}))}{n \cdot m}, \tag{8}$$

where $c : L \to L$ is an appropriate increasing function satisfying $c(0) = 0$ and $c(1) = 1$. In our analyses, we used $c(a) = a^{q\sqrt{mn}}$ and obtained satisfactory results for $q = 0.1$, which we report below. The effect of using $c(I_{ij} \leftrightarrow (A_{\mathcal{F}} \circ B_{\mathcal{F}})_{ij})$ is the following. The value $I_{ij} \leftrightarrow (A_{\mathcal{F}} \circ B_{\mathcal{F}})_{ij}$ measures closeness of the values at the $\langle i, j \rangle$ entry of the original matrix I and the matrix $A_{\mathcal{F}} \circ B_{\mathcal{F}}$ reconstructed from the computed set $\mathcal{F}$ of factors. Transforming this value by the monotone c emphasizes entries that are very close while inhibiting those that are not so close. The rate of inhibition is parameterized by the geometric mean $\sqrt{mn}$ of the number m of attributes and the number n of rows of the matrix, and a parameter q.

We now briefly describe the results for two grades, namely grade A and grade E. Grade A was attained by 607 students. Our algorithm obtained 36 factors from the 607×14 matrix I_{A}. The cumulative coverage of these factors is depicted in Fig. 1 and Table 4. The depicted coverage values corresponding to the sets $\mathcal{F}_l = \{F_1, \ldots, F_l\}$ consisting of the first l factors $F_i = \langle C_i, D_i \rangle$ computed by the algorithm are the values $s_c(I, A_{\mathcal{F}_l} \circ B_{\mathcal{F}_l})$ defined by (8). Thus, we can observe that the coverage by the first, the first two, and the first three factors is 0.417, 0.539, and 0.618, respectively. As one can see, a reasonable coverage (around 0.75 and more) is obtained by the first five factors already.

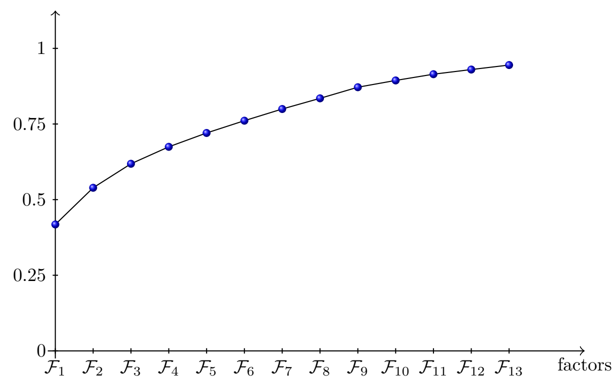

Fig. 1: Cumulative coverage by factors (grade A).

Table 4: Cumulative coverage by factors (grade A).

$\mathcal{F}_1$	$\mathcal{F}_2$	$\mathcal{F}_3$	$\mathcal{F}_4$	$\mathcal{F}_5$	$\mathcal{F}_6$	$\mathcal{F}_7$	$\mathcal{F}_8$	$\mathcal{F}_9$	$\mathcal{F}_{10}$	$\mathcal{F}_{11}$	$\mathcal{F}_{12}$	$\mathcal{F}_{13}$	...	$\mathcal{F}_{36}$
0.417	0.539	0.618	0.674	0.720	0.760	0.799	0.834	0.871	0.893	0.914	0.930	0.944	...	1

Let us now describe in detail the first three factors, i.e. the three most important ones according to the algorithm. The extent and the intent of the first factor, $F_1 = \langle C_1, D_1 \rangle$, is depicted in Fig. 2 and Fig. 3, respectively.

The intent, which conveys the meaning of each factor, is a fuzzy set assigning to every attribute y_i ($i = 1, \ldots, 14$) a value in the scale $L = \{0, {}^1\!/_4, {}^1\!/_2, {}^3\!/_4, 1\}$. This value is interpreted as the degree to which the particular attribute y_i is a manifestation of the given factor. That is, the degree to which good performance on the attribute y_i accompanies the presence of the factor. Fig. 3 displays such a fuzzy set for the first factor obtained by the algorithm, i.e. the fuzzy set D_1.

The extent is a fuzzy set assigning to every student (with grade A) a degree in the scale L to which the student possesses the given factor. Fig. 2 presents such a fuzzy set, i.e. C_1, for the first factor obtained. Since there are 607 students with grade A, the graph of C_1 is somewhat condensed (there are 607 points on the horizontal axis, hence 607 vertical bars indicating the assigned grades).

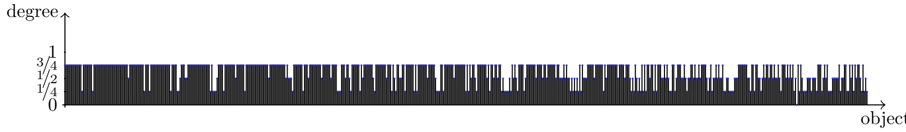

Fig. 2: Extent of F_1 (grade A).

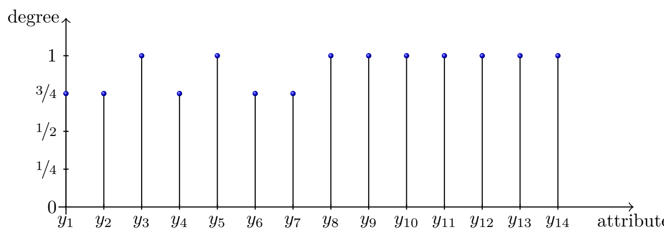

Fig. 3: Intent of F_1 (grade A).

In particular, one may observe from the intent D_1 and the description of the attributes $y_1, \ldots, y_{14}$ in Table 2 that the first factor may verbally be described as "excellent overall knowledge, and excellent analytical and communication skills," because the factor displays almost all attributes to the highest possible degree. From Fig. 2 one may see that this factor is possessed by most of the students who obtained grade A to the second-highest degree, $3/4$. Since the students are ordered on the horizontal axis by the their total marks, the graph also tells us that the factor appears in particular on the students with the highest total marks. Such a factor is a natural and expected one and from this viewpoint, our algorithm confirms the intuitive expectations.

The second factor, F_2, is depicted in Fig. 4 (extent) and Fig. 5 (intent). This factor may be interpreted as displaying very good overall knowledge with slightly limited communication skills and slightly limited knowledge of government power structures. Most of the students, particularly those with high total marks, possess this factor to a high degree, the best students even to the highest possible degree. Only one student does not possess this factor at all (i.e. the corresponding degree for this student in the extent is 0).

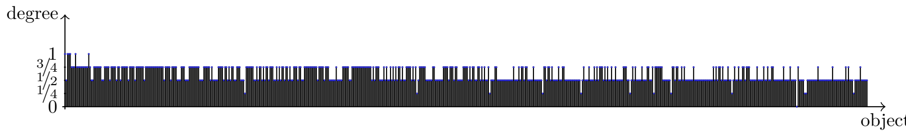

Fig. 4: Extent of F_2 (grade A).

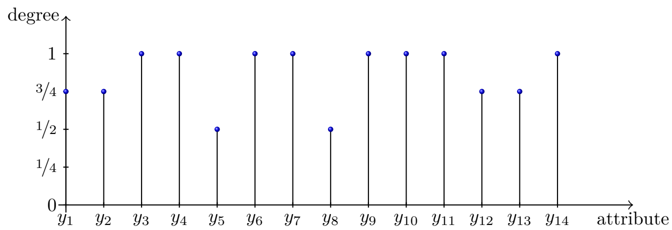

Fig. 5: Intent of F_2 (grade A).

The third factor, F_3, is depicted in Fig. 6 (extent) and Fig. 7 (intent). It may be interpreted as manifesting excellent knowledge in the first two questions (attributes y_1 and y_2), only a moderate knowledge of parliamentary machinery (y_3, y_4, and y_5), virtually no knowledge of cabinet government (y_6, y_7, and y_8), and reasonable knowledge of parliamentary models (y_9, y_{10}, and y_{11}) and government power structures (y_{12}, y_{13}, and y_{14}). This factor, which is possessed by many students to a high degree, is considerably discriminative and therefore interesting.

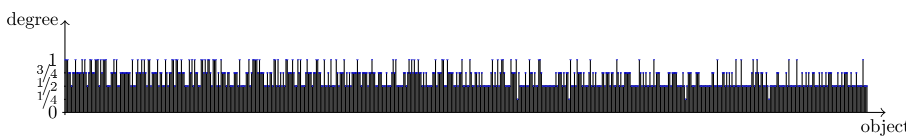

Fig. 6: Extent of F_3 (grade A).

Let us now turn to the analysis of performances of students who obtained grade E. Due to limited scope, our main purpose is to demonstrate that our method reveals different factors from the data for grade E compared to the data for grade A, which is in accordance with intuitive expectations. Grade E was attained by 322 students and our algorithm obtained 29 factors from the 322×14 matrix I_{E}. The cumulative coverage of these factors is depicted in Fig. 8 and Table 5. The depicted coverage values again correspond to the sets $\mathcal{F}_l = \{F_1, \ldots, F_l\}$ consisting of the first l factors $F_i = \langle C_i, D_i \rangle$ computed by the algorithm. As with grade A, we can observe that a reasonable coverage (around 0.75 and more) is obtained by the first five factors already.

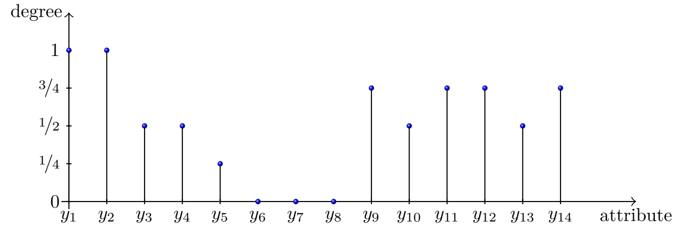

Fig. 7: Intent of F_3 (grade A).

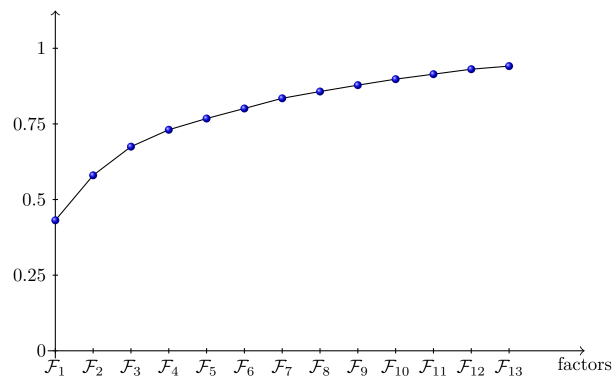

Fig. 8: Cumulative coverage by factors (grade E).

Table 5: Cumulative coverage by factors (grade E).

$\mathcal{F}_1$	$\mathcal{F}_2$	$\mathcal{F}_3$	$\mathcal{F}_4$	$\mathcal{F}_5$	$\mathcal{F}_6$	$\mathcal{F}_7$	$\mathcal{F}_8$	$\mathcal{F}_9$	$\mathcal{F}_{10}$	$\mathcal{F}_{11}$	$\mathcal{F}_{12}$	$\mathcal{F}_{13}$	...	$\mathcal{F}_{29}$
0.431	0.580	0.674	0.730	0.767	0.801	0.834	0.857	0.878	0.897	0.914	0.930	0.941	...	1

The most significant factor in the data for grade E is depicted in Fig. 9 (extent) and Fig. 10 (intent). We may observe that the factor is possessed by most students to the degree $3/4$ and by a considerably high number of students even to the highest possible degree. This factor may be described as manifesting no or very limited knowledge in all questions except for questions regarding parliamentary models and governmental power structures, for which the students who possess this factor exhibit moderate performance with respect to all three assessment objectives.

The second most significant factor for grade E is depicted in Fig. 11 (extent) and Fig. 12 (intent). This factor is possessed by almost all students to degree $1/2$ and by several of them even to degree $3/4$. None of the students with grade E possesses this factor to the highest possible degree. The factor is manifested by a very limited knowledge of the first two questions (y_1 and y_2) and moderate knowledge of the remaining questions except for the question about governmental power structures where the manifested performance is severely limited.

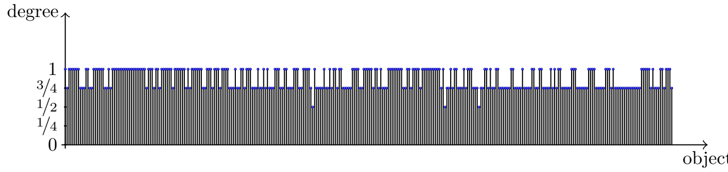

Fig. 9: Extent of F_1 (grade E).

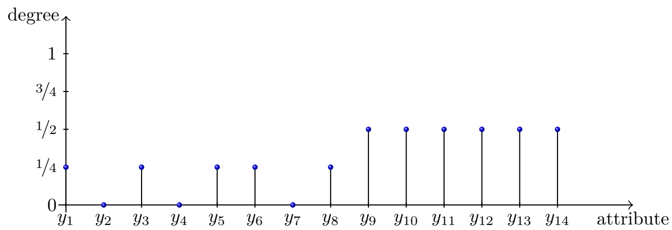

Fig. 10: Intent of F_1 (grade E).

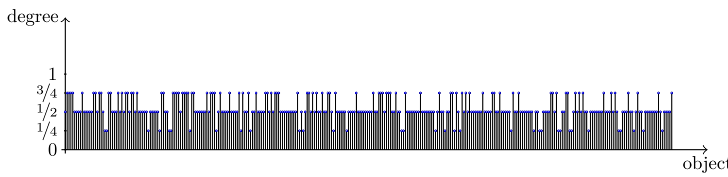

Fig. 11: Extent of F_2 (grade E).

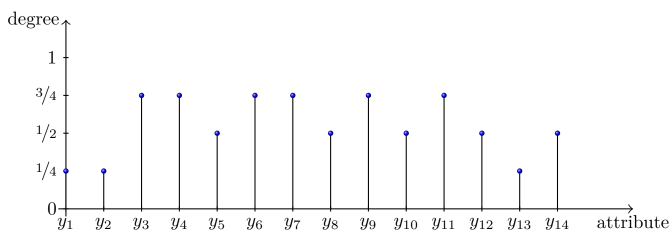

Fig. 12: Intent of F_2 (grade E).

The third most significant factor for grade E is depicted in Fig. 13 (extent) and Fig. 14 (intent). As is apparent from the intent of this factor, the factor is manifested by reasonably good knowledge in most of the questions. Nevertheless, the factor is possessed to very small degrees by the students with grade E and, therefore, is not as significant as the previous factors.

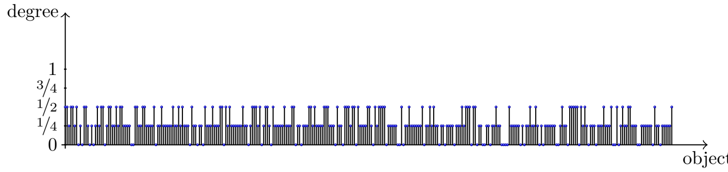

Fig. 13: Extent of F_3 (grade E).

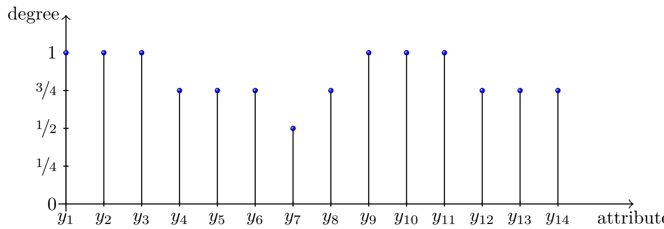

Fig. 14: Intent of F_3 (grade E).

4 Conclusions and Further Steps

The purpose of this paper is twofold. For one, we provide further analyses of real-world data using the recently developed method of factor analysis of ordinal data described in [4,5,6,8,9]. Secondly, we provide some first steps in our long-term project of utilizing relational methods of data analysis, in particular the methods related to formal concept analysis [11], in understanding students' performance data.

We demonstrated by our analyses that students' performance data, which consists of a collection of ordinal attributes, may naturally be subject to analysis by the methods we explore, in particular by the present method of factor analysis. We also demonstrated that the method yields naturally interpretable factors from data which are easy to understand, adding thus further evidence of a practical value of the method.

The limited scope of this paper does not allow us to go into the ramifications of our results obtained so far for educational policy makers. Formulating such ramifications is the ultimate goal for our research. Nevertheless, a proper methodology and experimental basis has first to be developed. Our present method and the reported experiments are to be considered as the first steps in this regard. The natural next steps seem to be the following. Firstly, we plan to further develop the present method of factor analysis. One direction is to adjust the method to be capable of extracting factors with a pattern preferred by the users of the method. An example is the flatness of the factors mentioned above. Another direction, which emerged during our experiments, is to allow a reasonable interaction with the user of the method. As the factors are generated one-by-one, we plan to provide the user with the option to accept or reject a candidate factor, hence the option to control the very process of factorization. In face of the extent of the data, we also plan to explore a possible statistical enhancement of our method. Secondly, we plan to compare the results of our factor analyses to the results obtained by alternative factor-analytic methods, as well as put our work in further works on analyzing educational data by relational methods, e.g. [18]. Thirdly, we plan to explore further methods related to formal concept analysis in analyzing students' performance data.

Acknowledgment

R. Belohlavek and E. Bartl acknolwedge support by the ECOP (Education for Competitiveness Operational Programme) project No. CZ.1.07/2.3.00/20.0059, which was co-financed by the European Social Fund and the state budget of the Czech Republic. The present research has been conducted in the follow-up period of this project.

References

1. Bartholomew, D.J., Knott M.: Latent Variable Models and Factor Analysis. 2nd Ed., Arnold, London (1999)
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. Annals of Pure and Applied Logic **128** (1–3), 277–298 (2004)
4. Belohlavek, R.: Optimal decompositions of matrices with entries from residuated lattices. J. Logic and Computation **22** (6), 1405–1425 (2012)
5. Belohlavek, R., Krmelova, M.: Beyond Boolean matrix decompositions: Toward factor analysis and dimensionality reduction of ordinal data. In: 13th International Conference on Data Mining, pp. 961–966. IEEE (2013)
6. Belohlavek, R., Krmelova, M.: Factor analysis of ordinal data via decomposition of matrices with grades. Ann. Mathematics and Artificial Intelligence **72** (1), 23–44 (2014)
7. Belohlavek, R., Vychodil, V.: Discovery of optimal factors in binary data via a novel method of matrix decomposition. J. Computer and System Sciences **76** (1), 3–20 (2010)
8. Belohlavek, R., Vychodil, V.: Factor analysis of incidence data via novel decomposition of matrices. LNAI **5548**, 83—97 (2009)
9. Belohlavek, R., Vychodil, V.: Factorization of matrices with grades. Fuzzy Sets and Systems **292**, 85–97 (2016)
10. Eldén, L.: Matrix Methods in Data Mining and Pattern Recognition. SIAM, Philadelphia, PA (2007)
11. Ganter, B., Wille, R.: Formal Concept Analysis. Mathematical Foundations. Springer, Berlin (1999)
12. Geerts, F., Goethals, B., Mielikäinen, T.: Tiling databases. LNCS **3245**, 278–289 (2004)
13. Golub, G., Van Loan, C.: Matrix Computations. Johns Hopkins University Press (1996)
14. Goguen, J.A.: The logic of inexact concepts. Synthese **18**, 325–373 (1968–9)

15. Gottwald, S.: A Treatise on Many-Valued Logics. Research Studies Press, Baldock, Hertfordshire, England (2001)
16. Hájek, P.: Metamathematics of Fuzzy Logic. Kluwer, Dordrecht (1998)
17. Harman, H.H.: Modern Factor Analysis. 2nd Ed. The Univ. Chicago Press, Chicago (1970)
18. Ignatov, D.I., Mamedova, S., Romashkin, N., Shamshurin, I.: What Can Closed Sets of Students and Their Marks Say? EDM, 223–228 (2011)
19. 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)
20. Lee, D., Seung, H.: Learning the parts of objects by non-negative matrix factorization. Nature **401**, 788–791 (1999)
21. Miettinen, P., Mielikäinen, T., Gionis, A., Das, G., Mannila, H.: The discrete basis problem. IEEE Trans Knowledge and Data Eng **20** (10), 1348–1362 (2008)
22. Miettinen, P., Vreeken, J.: Model order selection for Boolean matrix factorization. Proc. ACM SIGKDD, 51–59 (2011)
23. Newton, P.E., Shaw, S.D.: Validity in Educational and Psychological Assessment. SAGE, London (2014)
24. Tatti, N., Mielikäinen, T., Gionis, A., Mannila, H.: What is the dimension of your binary data? In: The 2006 IEEE Conference on Data Mining, IEEE Computer Society, 603–612 (2006)
25. Ward, M., Dilworth, R.P.: Residuated lattices. Trans. Amer. Math. Soc. **45**, 335–354 (1939)
26. Zadeh, L.A.: Fuzzy sets. Inf. Control **8**, 338–353 (1965)