

Can We Measure the Interpretability of Factors?

Martin Trnecka and Marketa Trneckova

Department of Computer Science
Palacky University Olomouc, Czech Republic
martin.trnecka@gmail.com, marketa.trneckova@gmail.com

Abstract. Decomposition of matrices over some finite scale received a considerable attention in data mining research. The methods that perform such decomposition can be viewed as an implementation of factor analysis. Surprisingly, the main motivation that is behind the factor analysis, the interpretation of the factors, is given only a very small amount of attention, or is completely neglected, in current research. In this paper, we are arguing that the interpretation of factors is an important part of matrix decomposition and we propose a novel measure, based on simple structure from factor analysis, enabling the intererability measurement. Furthermore, we present an experimental evaluation of selected decomposition algorithms via our metric.

1 Motivation

Decomposition of matrices over some finite scale—especially a case where the scale contains only two elements, namely zero and one, called the Boolean matrix decomposition—has become one of the standard methods in data mining with applications to many fields. In a broad sense, these methods may be considered as implementing the general idea of classical factor analysis introduced by psychologist Charles Spearman [16].

The motivation for the factor analysis comes from the psychology and the social sciences. The general aim is to simplify complex data. More precisely to describe original data via new more fundamental variables called factors.

Boolean matrix decomposition (BMF) and in general the decomposition of matrix over some finite scale, always reflects the ideas of factor analysis. This is not surprising, because BMF also comes from the psychology, where Boolean data often occur (see e.g. [7]). On the other hand some aspects of factor analysis are neglected in contemporary literature including matrix decomposition methods, namely the quality of factors. In factor analysis, the quality of factors, more precisely the interpretability of factors, is on the first place.

The current direction of research focuses on creating new algorithms and evaluating their quality in relation to the number of factors and the size of the input data covered. If an analysis of factor interpretability is contained, it is done by hand on a small number of datasets. The main reason is that the interpretation of factors is subjective and very tedious. In addition, contemporary literature lacks a uniform methodology or a metric to measure the interpretability of factors.

The purpose of this paper is twofold. We argue that the interpretation of factors, which is often neglected, is important part of matrix decomposition method; and we propose a new measure, based on the simple structure from factor analysis, enabling an objective measurement of the interpretability of factors.

The main reason why we chose a simple structure as a criterion for the interpretability of factors is the historical interdependence of factor analysis and matrix decompositions. Simple structure is therefore the first choice, and for this reason this work should be seen as the first small step in the new field of research.

The rest of the paper is organized as follows. In the following Section 2 we provide a brief overview of current research involving matrix decomposition from the factor interpretation standpoint. Moreover, we describe in Section 2 basic approach to the factors interpretation in factor analysis. Then, in Section 3 basic notions, notation and formalization of the metric are presented. The metric is experimentally evaluated in Section 4. Section 5 draws a conclusion and future research direction.

2 Interpretation of Factors

2.1 Lost in the Flood of Algorithms

The current direction of matrix decomposition research focuses primarily on the production of new algorithms and improving existing ones. An overview of existing approaches and methods is beyond the scope of this paper (see e.g. [5,18] which provides comparison of the most commonly used algorithms). The factors interpretation has only a small amount of attention or it does not perform at all. Note that this is indeed a feature of contemporary research. Early works involving matrix decomposition usually contain a larger assessment—usually the analysis of the factors of one particular dataset—of factor interpretability.

One of the few exceptions is the work [4] that deals with the extensive detailed analysis of factors. However, this analysis is done manually. Works involving matrix decomposition do not contain any methodology or metrics to measure the interpretability of factors. This is very surprising, especially because these methods are inspired by classical factor analysis, where the interpretability of factors and its measurement is an elementary concept.

2.2 Good Factors Definition and Metric

In classical factor analysis, the question of whether the factor is good or bad is based on the law of parsimony, well known as Occam’s razor, i.e. we should pick the simplest explanation of facts. A solution, which is selected via the parsimony law, is called *simple structure*.

In 1947 Thurstone proposed five simple criteria of simple structure in his work [17]. These can be seen as informal, vague and verbally described definition of good factors. Thurstone’s criteria were as follows:

1. Each row of the rotated matrix should contain at least one zero.
2. In each factor the minimum number of zero loadings (see Section 3.2) should be the number of factors in the rotation.
3. For every pair of factors there should be variables with zero loadings on one and significant loadings on the other.
4. For every pair of factors a large portion of the loadings should be zero, at least in a matrix with a large number of factors.
5. For every pair of factors there should be only a few variables with significant loadings on both factors.

Two of these criteria, namely 3 and 5 are of overriding importance. Essentially, the criterion of simple structure is a factor matrix in which the factors each have a few high loadings.

Later in 1978, Cattell in [9], who continues in the Thurstone's work, argued that the simple structure factors are usually simple to interpret. There have been many attempts to formalize the simple structure (see e.g. [8]). The result of these attempts is an ad hoc formalization and a conclusion that there will never be a simple formula describing Thurstone's five criteria. Unfortunately, these approaches cannot be adopted in the case of decomposition of matrices over some finite scale, because they use a different calculus.

On the other hand, this kind of data can be handled using fuzzy logic. Moreover, in case of data over some finite scale, all Thurstone's criteria can be formalized via logical formulas. In the following section we will use the fuzzy logic to formalize Thurstone's criteria and create a metric allowing an objective analysis of factor interpretability.

3 Formalization of Metric

3.1 Basics from Fuzzy Logic

Fuzzy logic has been employed to handle the concept of partial truth, where the truth value may range between completely true and completely false. This approach has been proven to be useful in several areas and we refer to [1].

Let us consider a set L of truth values. We assume that this set is partially ordered (partial ordering is denoted by \leq), contains a least element 0 and a greatest element 1.

Let a and b are the truth degrees from L , then in L exists a truth value which is greater than both a and b . The least element that is greater or equal to both a and b is called *supremum* of a and b . Analogously, we can define *infimum* of a and b —the greatest element from L which is smaller or equal to both a and b . We define the *lower cone* of A by $\mathcal{L}(A) = \{a \in L | a \leq b \text{ for all } b \in A\}$ and the *upper cone* of A by $\mathcal{U}(A) = \{a \in L | b \leq a \text{ for all } b \in A\}$. If $\mathcal{L}(A)$ has a greatest element a , then a is called the *supremum* of A (denoted $\bigvee A$) and dually if $\mathcal{U}(A)$ has a least element a , then a is called the *infimum* of A (denoted $\bigwedge A$). In particular, we assume that the partial order \leq makes L a complete lattice [12] (i.e., arbitrary infima \bigwedge and suprema \bigvee exist in L). This assumption

is automatically satisfied if L is a finite chain (i.e. $a \leq b$ or $b \leq a$ for every $a, b \in L$), in which case $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$. We also need to define a logical conjunction operation (denoted by \otimes). We assume that \otimes is commutative, associative, has 1 as its neutral element ($a \otimes 1 = a = 1 \otimes a$), and distributes over arbitrary suprema, i.e. $a \otimes (\bigvee_{j \in J} b_j) = \bigvee_{j \in J} (a \otimes b_j)$. This leads to if a and b are truth degrees of propositions p_1 and p_2 , then $a \otimes b$ is the truth degree of proposition “ p_1 and p_2 ”.

Importantly, \otimes induces another operation, \rightarrow , called the *residuum of \otimes* , which plays the role of the truth function of implication and is defined by

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

Residuum, which may be looked at as a kind of division, satisfies an important technical condition called adjointness:

$$a \otimes b \leq c \text{ iff } a \leq b \rightarrow c,$$

which is also utilized below. This leads to algebraic structures called *residuated lattices*.

3.2 Basic Notions of Matrix Decomposition

In general, matrix decomposition aims at whether data involving objects and their directly observable attributes may be explained by a smaller number of different, more fundamental attributes called factors. For example, whether performances of students (directly observable attributes) may be described by some traits 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 —called a score matrix in the factor analysis terminology—and a $k \times m$ factor–attribute matrix B —called a loading matrix in the factor analysis terminology. What kind of matrices (real, Boolean, or other) and what kind of product \circ are involved determines the semantics of the factor model.

In this paper, we are mainly focused on the decomposition of matrices containing grades of certain scales L with the sup- \otimes product. In particular, the matrix entry I_{ij} is a degree to which attribute j applies to object i . Similarly, A_{il} and B_{lj} are the degrees to which factor l applies to object i and the degree to which attribute j is a (one particular) manifestation of factor l . The case where the scale L contains only two degrees (0 and 1), is called the Boolean matrix decomposition.

Equation (1) has the following meaning. Object i has attribute j if and only if there exists factor l such that i has l (or, l applies to i) and j is one of the particular manifestations of l . The meaning can be described by the formula

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

Let us note, in the Boolean case ($L = \{0, 1\}$), the meaning of equation (1) may be described via formula

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

There exist two concrete variants of the decomposition problem. These two problems reflect two important views on matrix decomposition. The first one—the *discrete basis problem* (DBP) [14]—emphasizes the importance of the first k (presumably the most important) factors. The second one—the *approximate factorization problem* (AFP) [5]—emphasizes the need to account for (and thus to explain) a prescribed portion of data, which is specified by error ε .

The DBP and AFP problems are generally known in BMF, but both problems can be generalized to problems over some scale L . For this purpose we need to define closeness of matrices over L .

Let $s_L : L \times L \rightarrow [0, 1]$ be an appropriate function measuring closeness of degrees in L . For matrices $I, J \in L^{n \times m}$, put

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

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

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

where $a \leftrightarrow b = \min(a \rightarrow b, b \rightarrow a)$ is the so-called *biresiduum* (many-valued equivalence from a logical point of view) of a and b . Let us note that the closeness coincides with the notion coverage in several papers.

The generalization of the AFP and DBP to the general decomposition over scale L follows:

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

3.3 Interpretability Metric

We approach the formalization of Thurstone’s five criteria according to the principles of mathematical fuzzy logic [1,12,13] as follows. We consider the factor

model 1 and the Royce [15] definition of factor, namely “factor is a construct operationally defined by its factor loadings”. In other words, factors are represented via attributes which are manifestation of them, i.e. factors are represented via rows of matrix B . This is very important aspect of our metric, because we can evaluate factors regardless of whether they do or not contain noise—noise is a big issue in matrix decompositions, see e.g. [14].

The following formalization of the five criteria described in Section 2.2 utilizes operations over scale L . As far as the choice of the operations on L is concerned, we use the Łukasiewicz t-norm in the formalization, due to some of its intuitive properties. We describe each criterion via single logical formula with the truth degrees from L . The degree of fulfillment of each criterion is determined by the degree of fulfillment of a particular formula.

3.4 Formalization of Thurstone’s Criteria

The first criterion. Each row contains at least one zero, i.e. for each factor there exist at least one attribute which is not particular manifestation of the factor. Formally, the first criterion can be described via formula $\forall i \exists j \neg B_{ij}$, i.e.

$$(\exists j \neg B_{1j}) \wedge (\exists j \neg B_{2j}) \wedge \dots \wedge (\exists j \neg B_{kj}),$$

where $(\exists j \neg B_{ij}) = (\neg B_{i1}) \vee (\neg B_{i2}) \vee \dots \vee (\neg B_{im})$.

The second criterion. In each factor, the minimum number of zero loadings should be the number of factors, i.e. in each factor, there is at least k attributes that are not manifestation of this factor. Formally,

$$\forall i \exists j_1 \exists j_2 \dots \exists j_k (\neg B_{ij_1} \wedge \neg B_{ij_2} \wedge \dots \wedge \neg B_{ij_k} \wedge j_1 \neq j_2 \neq \dots \neq j_k).$$

The third criterion. For every pair of factors there should be variables with zero loadings on one and significant loadings on the other. Formally,

$$\forall i_1 \forall i_2 \exists j (B_{i_1 j} \wedge \neg B_{i_2 j}) \wedge (\neg B_{i_1 j} \wedge B_{i_2 j}).$$

The fourth criterion. For every pair of factors a large portion of loadings should be zero (at least in a matrix with large number of factors. We need to define what the “large portion” means, i.e. how many attributes do not manifest one or second (or both) factors. Let us denote “large portion” by lp and $B^{ij} = B_{i_} \cup B_{_j}$ ($B_{i_}$ denotes row i of matrix B), then formally

$$\forall i_1 \forall i_2 \exists j_1 \exists j_2 \dots \exists j_{lp} (\neg B_{j_1}^{i_1 i_2}) \wedge (\neg B_{j_2}^{i_1 i_2}) \wedge \dots \wedge (\neg B_{j_{lp}}^{i_1 i_2}) \wedge j_1 \neq j_2 \neq \dots \neq j_{lp}.$$

The fifth criterion. For every pair of factors there should be only a few attributes that manifest both factors. Similarly as in the previous case, we need to define a “few”.

$$\forall i_1 \forall i_2 \exists j_1 \exists j_2 \dots \exists j_{few} (B_{i_1 j_1} \wedge B_{i_2 j_1}) \wedge (B_{i_1 j_2} \wedge B_{i_2 j_2}) \wedge \dots$$

$$\cdots \wedge (B_{i_1 j_{few}} \wedge B_{i_2 j_{few}}) \wedge j_1 \neq j_2 \neq \cdots \neq j_{few}.$$

The formalization via above presented logical formulas strictly says how a set of factors satisfies each criteria, but it does not take into account how many factors (pairs of factors) do not meet the criterion.

We can analogously define less strict measure which takes into account for how many factors (pairs of factors) each criterion holds, i.e instead of minimum value for each factor (pair of factors), we take mean of this values.

We denote the first variant of metric as the strict metric and the second variant as the partially strict metric in Section 4—which provides experimental evaluation of our metric.

4 Experimental Evaluation

The following section is devoted to the experimental evaluation of metrics described in Section 3. We compare three algorithms for the matrix decomposition problem, namely GRECOND_L [6], GRESS_L [3] and ASSO_L [3]. The first two are based on formal concept analysis [11]. Let us note, these algorithms provide the decomposition of matrices over a finite scale L . All of them are inspired by the existing BMF algorithms.

4.1 Real-World Data

Since we are interested in the interpretability of factors, we perform experiments only on the real-world datasets—which, unlike synthetic data, are influenced by real factors. We used the following datasets.

Dog breeds dataset represents 151 dog breeds and their 13 attributes such as for example Playfulness, Protection ability, Affection or Ease of training. For detailed analysis see [3].

Decathlon extends the dataset from [6] to 28 athletes and their performance in 10 disciplines of decathlon. A detail analysis of this data can be found in [2].

IPAQ consists of international questionnaire data involving 4510 respondents answering 16 questions using a three-element scale regarding physical activity. The questions include those regarding their age, sex, body-mass-index (BMI), health, to what extent the person bicycles, walks, etc. For more detail see [3].

Music [3] consists of results of a study inquiring how people perceive speed of song depending on various song characteristics. The data consists of a 900×26 matrix over a six-element scale L , representing a questionnaire involving 30 participants who were presented 30 music samples.

Rio dataset [18] represents 87×31 matrix I obtained from <https://www.rio2016.com/en/medal-count> and consists of 87 countries that obtained any medal in one of 31 sport (such as Archery, Athletics, Badminton, Basketball, Boxing, ...) on Olympics games in Rio de Janeiro 2016. L contains four grades—1 means that country won at least one gold medal, $\frac{2}{3}$ at least silver medal, $\frac{1}{3}$ at least one bronze medal and 0 no medal in this sport. This dataset is very sparse in comparison with other presented datasets.

4.2 Assessment of the Interpretability Metric

Obtained results for each of five Thurstone’s criteria and a total value of the simple structure are presented in Table 1 (strict measure) and Table 2 (partially strict measure). We provide the results for sets of factors with the values of closeness (column s) 0.75, 0.85, 0.9, 0.95 and 1 (which corresponds with the values of coverage 75%, 85%, 90%, 95% and 100% of the input data). Value NA means, that the particular algorithm can not obtain a prescribed coverage.

One may observe that the best results provides (in case of strict measure) GRESS $_L$ algorithm which outperforms GRECOND $_L$ and ASSO $_L$ on **Breeds**, **Decathlon** and **IPAQ**. On the **Music** and **Rio** data, GRECOND $_L$ produces slightly better results than GRESS $_L$. ASSO $_L$ is not able to reach higher coverage and usually provides worse results, but on **Music** and **Rio** data it outperforms both GRECOND $_L$ and GRESS $_L$.

We obtain similar results for partially strict measure. For this metric GRECOND and GRESS produce higher values than in the case of strict metric. Additionally GRESS $_L$ outperforms GRECOND $_L$ on **IPAQ** data and produces almost identical results on **Rio** data. ASSO $_L$ produces very similar results as in the case of strict metric.

From Tables 1 and 2 it is obvious that the simple structure firstly fail on second criterion especially for high closeness (in both GRECOND $_L$ and GRESS $_L$) since usually they need more factors than the number of attributes to achieve a prescribed coverage. ASSO $_L$ is algorithm for solving DBP, usually it is not able to obtain full coverage of input data. On the other hand, first factors obtained by ASSO $_L$ cover larger portion of data, so for example in **Rio** dataset we need only one factor to obtain coverage slightly higher than 90%. This is the reason why the total value of simple structure is equal to 1.

In [3] authors discuss problem of factor with values “around the middle”. These factors are the reason why ASSO $_L$ produces results which returns lower values on the criteria, that depend on the number of zeros, namely criterion 1, criterion 2 and criterion 5. In these criteria GRESS $_L$ returns better factorization than GRECOND $_L$ on almost all of the datasets. The reason is probably the logic behind the factor selection which particular algorithm utilizes. Unlike GRECOND $_L$ algorithm GRESS $_L$ algorithm takes into account different role of entries, namely it utilizes the so-called essential entries [5].

Some observations that depend on data itself are for example, that **Decathlon** dataset does not contain any 0 as input value, so neither GRECOND $_L$ nor GRESS $_L$ fully satisfy criterion 1. **IPAQ** dataset has much more objects than attributes so we need more factors to obtain higher coverage. This is the reason why from closeness 0.9 it fails in criterion 2.

4.3 Application to Boolean Matrix Decomposition

BMF is probably the most popular class of matrix decomposition over finite scale—in this case the scale L contains only two elements, namely zero and one. Factors produced by BMF algorithms can be analyzed without any problems via

our metric. We perform several experiments and we observe how good is the set of factors from the simple structure perspective.

There exist several algorithms for BMF based on different ideas. We used the following algorithms: GRECOND, GREESS, ASSO, HYPER, PANDA, TILING (for more details see e.g. [5]) and 8M [10]. Like in graded case, some of them are usually not able to achieve 100% coverage, namely ASSO, PANDA and 8M. We evaluate all of them on well known real data such as for example *Americas-small*, *DBLP*, *Emea*, *Chess* and *Mushroom*. All of them are well known and widely used. Description and characteristics of these datasets can be found e.g. in [5].

We present only basic observation. A broader analysis of results delivered by BMF algorithms is left to an extended version of this paper.

In the Boolean case, the third criterion is always true. It can be understood as: for every pair of factors, there should be an attribute which is manifestation of one of them and is not manifestation of the other one.

The best factors in terms of above defined measures are obtained by algorithm HYPER—for almost all datasets it returns value 1 (for closeness ≤ 0.95). The main reason is that HYPER usually selects factors including only one attribute. Such factors are really easy to interpret. On the other hand, this shows a drawback of the simple structure, because such factors are not useful.

GREESS, GRECOND and TILING algorithm returns comparable results (GREESS is slightly better than GRECOND and both are slightly better than TILING). All of them work in the similar way (they use formal concepts [11] as factors) and all of them do not meet the fifth criterion for high coverage.

8M, PANDA and ASSO return a set of factors which cover small amount data (in many cases they explain less than 80% of input data). Surprisingly, factors delivered via these algorithms produce the best results from the simple structure standpoint.

5 Conclusion and Future Research

We proposed a novel metric, based on a simple structure from factor analysis, for the measurement of the interpretability of factors delivered by matrix decomposition algorithms—more precisely algorithms that provide decomposition of matrices over some finite scale. Simple structure is defined via five criterion which we formalized via mathematics of Fuzzy logic. We proposed two variants of the metric, strict and partially strict and we experimentally evaluated the results produced by GRECOND_L , GREESS_L and ASSO_L algorithm. Additionally we provide a brief overview of experimental evaluation of selected BMF algorithms.

The observed results encourage us to the following future research directions. First, to explore different ways of the interpretability measuring. Second, to provide extensive evaluation of results produced by BMF algorithms.

Table 1: Strict metric

dataset	s	GRECOND _L					ASSOL					GRESSL							
		c1	c2	c3	c4	c5	total	c1	c2	c3	c4	c5	total	c1	c2	c3	c4	c5	total
Breeds	0.75	0.75	0.5	0.75	0.5	0.5	0.5	0.25	0.5	0.25	0.25	0.25	0.25	1	1	0.75	1	0.75	0.75
	0.85	0.5	0.5	0.75	0.5	0.5	0.5	0.5	0.25	0.5	0.25	0	0	1	0.75	0.75	0.75	0.5	0.5
	0.9	0.5	0.5	0.5	0.5	0.25	0.25	NA	NA	NA	NA	NA	NA	1	0.75	0.75	0.75	0.5	0.5
	0.95	0.5	0	0.5	0.5	0.25	0	NA	NA	NA	NA	NA	NA	1	0.25	0.75	0.75	0.5	0.25
Decathlon	1	0.5	0	0.5	0.5	0.25	0	NA	NA	NA	NA	NA	NA	1	0	0.75	0.75	0.5	0
	0.75	0.75	0.5	0.75	0.5	0.5	0.5	0.5	0.5	1	1	1	0.5	0.75	0.5	0.5	0.5	0.5	0.5
	0.85	0.75	0.5	0.75	0.5	0.25	0.25	NA	NA	NA	NA	NA	NA	0.75	0.5	0.5	0.5	0.5	0.5
	0.9	0.5	0.25	0.5	0.5	0.25	0.25	NA	NA	NA	NA	NA	NA	0.75	0.25	0.5	0.5	0.5	0.25
IPAQ	0.95	0.5	0.25	0.5	0.5	0.25	0.25	NA	NA	NA	NA	NA	NA	0.75	0.25	0.5	0.5	0.5	0.25
	1	0.5	0	0.5	0.5	0.25	0	NA	NA	NA	NA	NA	NA	0.75	0	0.5	0.5	0.25	0
	0.75	1	1	1	1	1	1	1	0.5	0.5	0.5	0	0	1	1	1	1	1	1
	0.85	1	1	1	1	1	1	1	0	0.5	0.5	0	0	1	1	1	1	1	1
Music	0.9	1	0	1	1	1	0	NA	NA	NA	NA	NA	NA	1	1	1	1	1	1
	0.95	1	0	1	1	1	0	NA	NA	NA	NA	NA	NA	1	0	1	1	1	0
	1	1	0	0.5	1	1	0	NA	NA	NA	NA	NA	NA	1	0	1	1	1	0
	0.75	0.5	0.33	1	0.33	0.17	0.17	0.5	0.33	0.67	0.33	0.33	0.33	0.33	0.33	0.83	0.33	0.33	0
Rio	0.85	0.5	0.17	0.83	0.33	0.17	0.17	NA	NA	NA	NA	NA	NA	0.33	0.17	0.83	0.33	0	0
	0.9	0.5	0.17	0.83	0.33	0.17	0.17	NA	NA	NA	NA	NA	NA	0.33	0	0.83	0.33	0	0
	0.95	0.5	0	0.67	0.33	0.17	0	NA	NA	NA	NA	NA	NA	0.33	0	0.83	0.33	0	0
	1	0.5	0	0.17	0.33	0.17	0	NA	NA	NA	NA	NA	NA	0.33	0	0.17	0.33	0	0
Rio	0.75	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.67
	0.85	1	0.67	1	1	0.67	0.67	1	1	1	1	1	1	1	1	1	1	1	0.67
	0.9	1	0.33	1	1	0.67	0.33	1	1	1	1	1	1	1	0.67	0.67	0.67	0.33	0.33
	0.95	1	0.33	1	0.67	0	0	NA	NA	NA	NA	NA	NA	1	0.67	0.67	0.67	0.33	0.33
Rio	1	1	0	0.33	0.67	0	0	NA	NA	NA	NA	NA	NA	1	0	0.67	0.67	0.33	0

Table 2: Partially strict metric

dataset	s	GRECOND _L					ASSOL					GRESS _L								
		c1	c2	c3	c4	c5	total	c1	c2	c3	c4	c5	total	c1	c2	c3	c4	c5	total	
Breeds	0.75	0.87	0.75	0.87	0.62	0.5	0.5	0.6	0.35	0.55	0.25	0.25	0.25	1	1	0.98	1	0.88	0.88	
	0.85	0.83	0.75	0.91	0.61	0.51	0.51	0.6	0.25	0.52	0.25	0.24	0.24	1	0.93	0.94	0.99	0.84	0.84	
	0.90	0.86	0.64	0.88	0.66	0.53	0.53			NA				1	0.81	0.94	0.95	0.83	0.81	
	0.95	0.9	0.02	0.91	0.75	0.63	0.02			NA				1	0.6	0.95	0.96	0.84	0.6	
	1	0.94	0	0.92	0.83	0.72	0			NA				1	0.02	0.96	0.96	0.85	0.02	
Decathlon	0.75	0.75	0.5	0.75	0.5	0.5	0.5	0.5	0.5	1	1	1	0.5	0.75	0.67	0.67	0.5	0.5	0.5	
	0.85	0.88	0.63	0.79	0.54	0.46	0.46			NA				0.8	0.55	0.75	0.55	0.5	0.5	
	0.9	0.79	0.5	0.73	0.53	0.45	0.45			NA				0.83	0.54	0.78	0.58	0.5	0.5	
	0.95	0.78	0.38	0.72	0.53	0.46	0.38			NA				0.84	0.5	0.78	0.6	0.51	0.5	
	1	0.83	0	0.74	0.58	0.5	0			NA				0.82	0	0.73	0.58	0.49	0	
IPAQ	0.75	1	1	1	1	1	1	1	0.5	0.67	0.5	0	0	1	1	1	1	1	1	
	0.85	1	1	1	1	1	1	1	0.04	0.68	0.5	0.07	0.04	1	1	1	1	1	1	
	0.9	1	0.86	1	1	1	0.86			NA				1	1	1	1	1	1	
	0.95	1	0	1	1	1	0			NA				1	0.93	1	1	1	0.93	
	1	1	0	1	1	1	0			NA				1	0	1	1	1	0	
Music	0.75	0.92	0.89	1	0.79	0.62	0.62	0.58	0.5	0.67	0.33	0.33	0.33	0.89	0.86	0.99	0.73	0.57	0.57	
	0.85	0.96	0.9	0.99	0.9	0.71	0.71			NA				0.95	0.9	1	0.88	0.8	0.8	
	0.9	0.97	0.87	0.99	0.93	0.81	0.81			NA				0.96	0.85	1	0.92	0.89	0.85	
	0.95	0.98	0.35	0.99	0.95	0.87	0.35			NA				0.97	0.67	1	0.94	0.92	0.67	
	1	0.99	0	0.99	0.97	0.93	0			NA				0.98	0	0.99	0.96	0.94	0	
Rio	0.75	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.99	0.99
	0.85	1	0.98	1	1	0.99	0.98	1	1	1	1	1	1	1	1	1	1	1	0.99	0.99
	0.9	1	0.96	1	1	0.98	0.96	1	1	1	1	1	1	1	0.95	1	1	1	0.97	0.95
	0.95	1	0.93	1	1	0.95	0.93			NA				1	0.92	1	1	1	0.96	0.92
	1	1	0	0.99	0.99	0.9	0			NA				1	0	1	1	1	0.99	0.91

References

1. Belohlavek, R.: *Fuzzy Relational Systems: Foundations and Principles*. Kluwer Academic/Plenum Press New York (2002)
2. Belohlavek, R., Krmelova, M.: Factor analysis of sports data via decomposition of matrices with grades. In: Szathmary, L., Priss, U. (eds.) *Proceedings of The Ninth International Conference on Concept Lattices and Their Applications*. pp. 293–304 (2012)
3. Belohlavek, R., Krmelova, M.: Beyond boolean matrix decompositions: Toward factor analysis and dimensionality reduction of ordinal data. In: Xiong, H., Karypis, G., Thuraisingham, B.M., Cook, D.J., Wu, X. (eds.) *2013 IEEE 13th International Conference on Data Mining*. pp. 961–966. IEEE Computer Society (2013)
4. Belohlavek, R., Krmelova, M.: Factor analysis of ordinal data via decomposition of matrices with grades. *Ann. Math. Artif. Intell.* **72**(1-2), 23–44 (2014)
5. Belohlavek, R., Trnecka, M.: From-below approximations in boolean matrix factorization: Geometry and new algorithm. *J. Comput. Syst. Sci.* **81**(8), 1678–1697 (2015)
6. Belohlavek, R., Vychodil, V.: Factor analysis of incidence data via novel decomposition of matrices. In: Ferré, S., Rudolph, S. (eds.) *Proceedings of 7th International Conference, ICFCA 2009. Lecture Notes in Computer Science*, vol. 5548, pp. 83–97. Springer (2009)
7. Boeck, P.D., Rosenberg, S.: Hierarchical classes: Model and data analysis. *Psychometrika* **53**, 361–381 (1988)
8. Carroll, J.B.: An analytical solution for approximating simple structure in factor analysis. *Psychometrika* **18**(1), 23–38 (1953)
9. Cattell, R.B.: *The scientific use of factor analysis in behavioral and life sciences*. Springer US (1978)
10. Dixon, W.: *Bmdp statistical software manual to accompany the 7.0 software release*, vols 1–3. (1992)
11. Ganter, B., Wille, R.: *Formal concept analysis - mathematical foundations*. Springer (1999)
12. Gottwald, S.: *A Treatise on Many-Valued Logics*, vol. 3. research studies press Baldock (2001)
13. Hajek, P.: *Metamathematics of fuzzy logic*, vol. 4. Springer Science & Business Media (1998)
14. Miettinen, P., Mielikäinen, T., Gionis, A., Das, G., Mannila, H.: The discrete basis problem. *IEEE Trans. Knowl. Data Eng.* **20**(10), 1348–1362 (2008)
15. Royce, J.R.: Factors as theoretical constructs. *American Psychologist* **18**(8), 522 (1963)
16. Spearman, C.: "General intelligence," objectively determined and measured. *The American Journal of Psychology* **15**(2), 201–292 (1904)
17. Thurstone, L.L.: *Multiple factor analysis*. University of Chicago Press: Chicago (1947)
18. Trneckova, M.: *Formal concept analysis with ordinal attributes*. Ph.D. thesis, Palacky University Olomouc, Czech Republic (2017)