STUDIA UNIV. BABEŞ–BOLYAI, INFORMATICA, Volume **LIX**, Special Issue 2, 2014
ICFCA 2014: 12th International Conference on Formal Concept Analysis, Cluj-Napoca, June 10-13, 2014

RANKING FORMAL CONCEPTS BY UTILIZING MATRIX FACTORIZATION

LENKA PISKOVÁ, TOMÁŠ HORVÁTH, AND STANISLAV KRAJČI

ABSTRACT. Formal Concept Analysis often produce huge number of formal concepts even for small input data. Such a large amount of formal concepts, which is intractable to analyze for humans, calls for a kind of a ranking of formal concepts according to their importance in the given application domain. In this paper, we propose a novel approach to rank formal concepts that utilizes matrix factorization, namely, a mapping of objects and attributes to a common latent space. The lower the distance between objects and/or attributes in the extent and/or intent of a formal concept in the latent space of factors, the more important the formal concept is considered to be. We provide an illustrative example of our approach and examine the impact of various matrix factorization techniques using real-world benchmark data.

1. INTRODUCTION

Formal Concept Analysis (FCA) [9] is a method for analyzing object-attribute data. In the basic setting, table entries are 1 or 0 indicating whether an object has a given attribute or not. FCA aims at finding so-called formal concepts (as well as the subconcept-superconcept relation among them) from this data. A formal concept is a formalization of the concept of a 'concept' which consists of two parts, a set of objects which forms its extension and a set of attributes which forms its intension [16]. The set of all concepts ordered by $\leq$ forms a complete lattice [9].

One of the obstacles in real-world application of FCA is that it often produces a huge number of formal concepts which can be exponential in the size of input data (see Table 3).

A kind of a ranking of resulting formal concepts would be beneficial for a human expert in the process of analyzing the formal concepts. We think

Received by the editors: 25 March 2014.
2010 *Mathematics Subject Classification.* 06-XX, 06Bxx.
1998 *CR Categories and Descriptors.* I.2.m [**Computing Methodologies**]: ARTIFICIAL INTELLIGENCE – *Miscellaneous* .
Key words and phrases. Formal Concept Analysis, formal concept, coherence, matrix factorization.

that such a ranking is domain and/or task specific and strongly depends on the actual needs of a user (i.e. a domain expert which is intended to use the outputs of FCA). Because of this, an approach to rank formal concepts should be intuitive, easily understandable and provide sufficient insight for a user.

In this paper, we propose a novel approach to rank formal concepts that utilizes matrix factorization, namely, a mapping of objects and attributes to a common latent space. The lower the distance between objects and/or attributes in the extent and/or intent of a formal concept in the latent space of factors, the more important the formal concept is considered to be. The presented approach is intuitive and easily explainable for users. It can be used together with other approaches to rank formal concepts.

2. Related Work

The reduction of the number of formal concepts and, thus, the size of concept lattices can be accomplished directly (removing formal concepts that do not satisfy a requirement) or in an indirect way (through handling formal contexts).

The aim of the approach in [5] is to find a decomposition of a Boolean (binary) matrix (formal context) with the smallest number of factors (that correspond to formal concepts) as possible. These factor concepts can be considered more important than other concepts of the formal context.

The usage of rank-k SVD in order to reduce the size of the corresponding concept lattice is proposed in [8]. However, SVD is not used to reduce the number of objects and/or attributes, but instead, to remove noise in an input table. Subsequently, the number of formal concepts is reduced. In [6], SVD is used to decompose a document-term matrix into a much smaller matrix where terms are related to a set of dimensions (factors) instead of documents. This term-dimension matrix is then converted into a binary matrix using a probabilistic approach.

The main idea of the JBOS (junction based on object similarity) method is that groups of similar objects are replaced by representative ones. The similarity of two objects is the sum of the weights of attributes in which the objects agree with each other (both objects have them or both do not have them) [7].

Another way is to reduce the number of formal concepts by means of attribute-dependency formulas (ADF) expressing the relative importance of attributes [3]. ADF depend on the purpose and have to be specified by an expert. Only formal concepts satisfying the set of ADF are relevant. The approach in [2] also utilizes background knowledge. After a user assigns weights to attributes, values of formal concepts are determined. Formal concepts with higher values are considered more important.

The idea of basic level of concepts appeared in [4]. Concepts in the basic level represent those concepts which are preferred by humans to use when describing the world around. The cohesion of a formal concept defined in [4], unlike the coherence proposed in this work, is a measure of whether the objects in its extent are pairwise similar.

The notion of the stability of a formal concept was introduced in [12]. The stability index indicates how much the intent of a concept depends on the set of objects in the extent (intentional stability). Extentional stability was defined analogously. Two other indices, probability and separation, are proposed in [11] and their performance on noisy data is discussed.

Another option to reduce the size of concept lattices is to consider only frequent formal concepts for a user given minimum support (Iceberg concept lattice) [15]. Note that a concept (A, B) is frequent if the fraction of objects that contain the attributes in B is above the minimum support threshold.

3. Formal Concept Analysis

A *formal context* is a triple (X, Y, R) consisting of a set $X = \{x_1, \ldots, x_n\}$ of objects, a set $Y = \{y_1, \ldots, y_m\}$ of attributes and a binary relation $R \subseteq X \times Y$ between them. We write $(x, y) \in R$ if the object x has the attribute y.

For a set $A \subseteq X$ of objects and a set $B \subseteq Y$ of attributes we define $A' = \{y \in Y : (\forall x \in A)(x, y) \in R\}$ and $B' = \{x \in X : (\forall y \in B)(x, y) \in R\}$. A' is the set of attributes common to the objects in A and B' is the set of objects which have all the attributes in B.

A *formal concept* of (X, Y, R) is a pair (A, B) where $A \subseteq X, B \subseteq Y, A' = B$ and $B' = A$. A and B are called the *extent* and the *intent* of the concept (A, B), respectively. The set of all concepts of (X, Y, R) is denoted by $\mathcal{B}(X, Y, R)$. $A \subseteq X$ $(B \subseteq Y)$ is an extent (intent) if and only if $A'' = A$ $(B'' = B)$.

We define a partial order $\leq$ on $\mathcal{B}(X, Y, R)$ by $(A_1, B_1) \leq (A_2, B_2)) \Leftrightarrow A_1 \subseteq A_2$ (equivalently, $B_1 \supseteq B_2$). The set of all concepts of (X, Y, R) ordered by $\leq$ constitutes the *concept lattice* $(\mathcal{B}(X, Y, R), \leq)$ of (X, Y, R) [16].

For more details on Formal Concept Analysis we refer to [9].

Example: The formal context in Table 1 induces 26 formal concepts:
$C_1 = (\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15\}, \emptyset)$,
$C_2 = (\{1, 2, 3, 5, 6, 7, 10, 12, 14\}, \{1\})$, $C_3 = (\{1, 9, 10, 11\}, \{4\})$,
$C_4 = (\{4, 8, 9, 10, 11, 13, 14, 15\}, \{6\})$,
$C_5 = (\{1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 15\}, \{8\})$,
$C_6 = (\{1, 10\}, \{1, 4\})$, $C_7 = (\{9, 10, 11\}, \{4, 6\})$,
$C_8 = (\{6, 8, 12\}, \{5, 8\})$, $C_9 = (\{4, 8, 9, 11, 13, 15\}, \{6, 8\})$,
$C_{10} = (\{8, 10, 13, 14, 15\}, \{6, 10\})$, $C_{11} = (\{1, 2, 3, 4, 5, 6, 7, 9, 11, 12\}, \{8, 9\})$,
$C_{12} = (\{10, 14\}, \{1, 6, 10\})$, $C_{13} = (\{1, 9, 11\}, \{4, 8, 9\})$,

Table 1. An illustrative example of a formal context of animals and their attributes (a cross in a row x and a column y indicates that the object x has the attribute y)

		1	2	3	4	5	6	7	8	9	10	11
		fur (hair)	feathers	scales	can fly	lives in water	lays eggs	produces milk	has a backbone	warm-blooded	cold-blooded	domestic
1	Bat	×			×			×	×	×		
2	Bear	×						×	×	×		
3	Cat	×						×	×	×		×
4	Chicken		×				×		×	×		×
5	Dog	×						×	×	×		×
6	Dolphin	×				×		×	×	×		
7	Elephant	×						×	×	×		
8	Frog					×	×		×		×	
9	Hawk		×		×		×		×	×		
10	Housefly	×			×		×				×	
11	Owl		×		×		×		×	×		
12	Sea lion	×				×		×	×	×		
13	Snake			×			×		×		×	
14	Spider	×					×				×	
15	Turtle			×			×		×		×	

$C_{14} = (\{8, 13, 15\}, \{6, 8, 10\})$, $C_{15} = (\{3, 4, 5\}, \{8, 9, 11\})$,
$C_{16} = (\{10\}, \{1, 4, 6, 10\})$, $C_{17} = (\{1, 2, 3, 5, 6, 7, 12\}, \{1, 7, 8, 9\})$,
$C_{18} = (\{4, 9, 11\}, \{2, 6, 8, 9\})$, $C_{19} = (\{13, 15\}, \{3, 6, 8, 10\})$,
$C_{20} = (\{8\}, \{5, 6, 8, 10\})$, $C_{21} = (\{1\}, \{1, 4, 7, 8, 9\})$,
$C_{22} = (\{6, 12\}, \{1, 5, 7, 8, 9\})$, $C_{23} = (\{3, 5\}, \{1, 7, 8, 9, 11\})$,
$C_{24} = (\{9, 11\}, \{2, 4, 6, 8, 9\})$, $C_{25} = (\{4\}, \{2, 6, 8, 9, 11\})$,
$C_{26} = (\emptyset, \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11\})$

4. Outline of Our Approach

Each formal context (input data for FCA) can be viewed as a matrix with n rows representing objects, m columns representing attributes and values 1 or 0 depending on whether an object has a given attribute or not. Thence, we can refer to a context as a matrix $R \in \{0, 1\}^{n \times m}$.

Consider a formal context R in Table 1, in which objects $x_1, \dots, x_n \in X$ are animals and $y_1, \dots, y_m \in Y$ are attributes which relate to animals, e.g. can fly, has a backbone, is warm-blooded, etc. Using a matrix factorization method we can create an approximation of a formal context R by a product of two or more matrices. Factorizing R means mapping the objects and attributes to a common k-dimensional latent space, the coordinates of which are called the factors. For attributes of animals, the discovered factors might measure obvious dimensions such as whether an animal is a mammal or whether an animal can fly; less well-defined dimensions such as whether an animal is dangerous or not; or, completely uninterpretable dimensions. For animals, each factor measures the extent to which the animal possesses the corresponding factor. Note that we are not concerned in the exact interpretation of the factors in this work since it belongs rather to areas of human sciences (psychology, sociology, etc.).

We use the idea of mapping of objects and attributes to a common latent factor space to define the coherence of a formal concept. The coherence is based on the distance between objects and/or attributes in the common latent factor space; objects which are close to each other share more common characteristics than objects which are remote from each other (similarly for attributes). For example, the distance between cat and dog should be small unlike the distance between cat and housefly. The attributes cold-blooded and warm-blooded should be remote from each other since these attributes exclude each other, i.e. if an animal is cold-blooded, it can not be warm-blooded and vice versa. Naturally, the location of objects and attributes in a latent factor space is dependent on an input (formal context) what will be seen later in Section 6.

5. Matrix Factorization

For the decomposition of formal contexts (Boolean matrices), Boolean Matrix Factorization is a natural choice. However, we also provided experiments with other factorization techniques, namely Singular Value Decomposition and Non-negative Matrix Factorization.

5.1. **Boolean Matrix Factorization (BMF).** The aim of Boolean Matrix Factorization (BMF) is to find a decomposition of a given matrix $X \in \{0,1\}^{n\times m}$ into a Boolean matrix product

$$X = A \circ B \quad (\text{or } X \approx A \circ B)$$

of matrices $A \in \{0,1\}^{n\times k}$ and $B \in \{0,1\}^{k\times m}$. [5]

A Boolean matrix product $A \circ B$ is defined by

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

where max denotes the maximum and $\cdot$ the ordinary product.

A decomposition of $X = A \circ B$ corresponds to a discovery of k factors that exactly or approximately explain the data. The least k for which an exact decomposition $X = A \circ B$ exists is called the *Boolean rank* (Schein rank) of X.

There are two different problems to solve in BMF:

- Discrete Basis Problem (DBP): Given $X \in \{0,1\}^{n\times m}$ and an integer $k > 0$, find $A \in \{0,1\}^{n\times k}$ and $B \in \{0,1\}^{k\times m}$ that minimize $||X - A \circ B|| = \sum_{ij} |X_{ij} - (A \circ B)_{ij}|$. [14]
- Approximate Factorization Problem (AFP): Given X and an error $\varepsilon \geq 0$, find $A \in \{0,1\}^{n\times k}$ and $B \in \{0,1\}^{k\times m}$ with k as small as possible such that $||X - A \circ B|| \leq \varepsilon$. [5]

In this paper, we have used the greedy approximation algorithm for BMF described in [5] (where it is called Algorithm 2).

5.2. **Singular Value Decomposition (SVD).** Singular Value Decomposition (SVD) is a factorization of a matrix $X \in \mathbb{R}^{n\times m}$ by the product of three matrices $X = U\Sigma V^T$ where $U \in \mathbb{R}^{n\times n}$, $\Sigma \in \mathbb{R}^{n\times m}$ and $V \in \mathbb{R}^{m\times m}$ such that $U^TU = I, V^TV = I$ (where I is an identity matrix), column vectors of U (left-singular vectors) are orthonormal eigenvectors of XX^T, column vectors of V (right-singular vectors) are orthonormal eigenvectors of X^TX and Σ contains singular values of X at the diagonal in descending order.

We can create an approximation $\hat{X}$ of a matrix X as

$$X \approx \hat{X} = U\Sigma V^T,$$

where $U \in \mathbb{R}^{n\times k}$, $\Sigma \in \mathbb{R}^{k\times k}$ and $V \in \mathbb{R}^{m\times k}$.

5.3. **Non-negative Matrix Factorization (NMF).** Let X be an $n \times m$ non-negative matrix and $k > 0$ an integer. The goal of Non-negative Matrix Factorization (NMF) [13] is to find an approximation

$$X \approx WH,$$

where W and H are $n \times k$ and $k \times m$ non-negative matrices, respectively.

The matrices W and H are estimated by minimizing the function

$$D(X, WH) + Reg(W, H),$$

where D measures the divergence and Reg is an optional regularization function. The different types of NMF arise from using different cost functions for measuring the divergence between X and WH, and by regularization of W and/or H.

The quality of the approximation is quantified by a cost function D. The common cost function between two non-negative matrices A and B is the squared error (Frobenius norm)

$$D(A, B) = \sum_{ij} (a_{ij} - b_{ij})^2.$$

6. The Proposed Approach

The notation we use in this and subsequent sections is the following:

- $distance(x_1, x_2)$ denotes a distance between objects x_1 and x_2 in the latent space,
- $coherence(C)$ denotes the degree of coherence of a formal concept C

Let R be a formal context, $X = \{x_1, \ldots, x_n\}$ and $Y = \{y_1, \ldots, y_m\}$ be the sets of objects and attributes of R, respectively. After the decomposition of R each object $x_i \in X$ is represented by a k-dimensional vector of latent factors $(x_{i_1}, \ldots, x_{i_k})$ describing the object and each attribute $y_j \in Y$ is represented by a k-dimensional vector of factors $(y_{j_1}, \ldots, y_{j_k})$ describing the attribute. Obviously, some objects are close to each other, while other objects are far away from each other (depending on the distance between objects) in the space of factors.

In our experiments, we have used the Manhattan distance (L_1 distance) and the Euclidean distance (Euclidean metric, L_2 distance). The distance between two objects $x_1, x_2 \in X$ (i.e. k-dimensional vectors $(x_{1_1}, \ldots, x_{1_k})$ and $(x_{2_1}, \ldots, x_{2_k})$ of latent factors) is given by

(Manhattan distance) $$distance(x_1, x_2) = \frac{1}{k} \sum_{l=1}^{k} |x_{1_l} - x_{2_l}|$$

(Euclidean distance) $$distance(x_1, x_2) = \sqrt{\frac{1}{k} \sum_{l=1}^{k} (x_{1_l} - x_{2_l})^2}$$

To have $distance \in [0, 1]$ we put $\frac{1}{k}$ to the equations Manhattan distance and Euclidean distance. The distance between attributes or between an object and an attribute in the latent factor space can be computed similarly.

One natural way to measure the coherence of a formal concept is by using the distance (Manhattan, Euclidean) between the objects in the extent of the formal concept as

$$coherence_X^{\max}(A, B) = 1 - \max_{x_1, x_2 \in A} distance(x_1, x_2) \tag{1}$$

Alternatively, we might put

$$coherence_X^{\text{avg}}(A, B) = 1 - \frac{\sum_{\{x_1, x_2\} \subseteq A, x_1 \neq x_2} distance(x_1, x_2)}{|A|(|A| - 1)/2} \tag{2}$$

Simply, $coherence_X^{\max}(A, B)$ is computed by the maximum distance between any two objects in the extent of (A, B) and $coherence_X^{\text{avg}}(A, B)$ is computed using the average distance between two objects in the extent of (A, B). Thus, $coherence_X^{\max}(A, B) \leq coherence_X^{\text{avg}}(A, B)$.

Formal concepts with similar objects (i.e. objects that share many common attributes) in their extents have a high degree of $coherence_X^{\max}$ and $coherence_X^{\text{avg}}$ (provided that similar objects are close to each other while distinct objects are remote from each other in the space of factors what will be seen later).

Similarly, the coherence of a formal concept can be measured using the distance between the attributes in its intent (denoted by $coherence_Y^{\max}$ and $coherence_Y^{\text{avg}}$). Alternatively, we can use the distance between both, the objects and attributes in the extent and intent of a formal concept, respectively (denoted by $coherence^{\max}$ and $coherence^{\text{avg}}$).

It is easy to see that if $(A_1, B_1) \leq (A_2, B_2)$, then $coherence_X^{\max}(A_1, B_1) \geq coherence_X^{\max}(A_2, B_2)$ and $coherence_Y^{\max}(A_1, B_1) \leq coherence_Y^{\max}(A_2, B_2)$.

Remark: From the above mentioned assumptions it follows that the decision of whether to use $coherence_Y^?$ or $coherence_X^?$, where ? = max or ? = avg, depends on user/expert preferences. $coherence_Y^?$ prefers specific concepts (a concept is specific if it consists of a few objects that share many attributes, see Fig. 2) while $coherence_X^?$ tends to prefer general concepts (a concept is general if it consists of many objects that have only a few attributes in common, see Fig. 3).

Now, we are able to assign a degree of coherence to each formal concept of a formal context. We consider formal concepts with higher degrees of coherence more important.

6.1. **Illustrative Example.** In this section we demonstrate our approach on a small example. It depends on the outcome of a matrix factorization method whether the results provided by our approach will be good or not. Therefore, we first address the problem of matrix factorization, and then we analyze the results themselves.

For the decomposition of the formal context in Table 1 we utilize Boolean Matrix Factorization (BMF) due to the good interpretability of factors. Using BMF the animals and their attributes are mapped to 8-dimensional space of latent factors which is shown in Fig. 1.

The interpretation of the factors might be the following: The first factor can be interpreted as the property of being a mammal (manifestations of the factor are: *fur (hair)*, *produces milk*, *has a backbone*, *warm-blooded*) and the second factor can be interpreted as the property of being a bird (manifestations of this factor are *feathers*, *lays eggs*, *has a backbone*, *warm-blooded*). The other factors relates to the attributes *cold-blooded*, *scales*, *can fly*, *lives in water*, *domestic*, *fur (hair)*, respectively.

The animals *bear* and *elephant* are mapped to the same point in the space of latent factors. The same is true for *cat* and *dog*, *dolphin* and *sea lion*, *hawk*

FIGURE 1. The decomposition of the formal context in Table 1 using BMF

		factor 1	factor 2	factor 3	factor 4	factor 5	factor 6	factor 7	factor 8
1	Bat	×				×			×
2	Bear	×							×
3	Cat	×						×	×
4	Chicken		×					×	
5	Dog	×						×	×
6	Dolphin	×					×		×
7	Elephant	×							×
8	Frog			×			×		
9	Hawk		×			×			
10	Housefly			×		×			×
11	Owl		×			×			
12	Sea lion	×					×		×
13	Snake			×	×				
14	Spider			×					×
15	Turtle			×	×				

	fur (hair)	feathers	scales	can fly	lives in water	lays eggs	produces milk	has a backbone	warm-blooded	cold-blooded	domestic
factor 1	×						×	×	×		
factor 2		×				×		×	×		
factor 3						×				×	
factor 4			×			×		×		×	
factor 5				×							
factor 6					×			×			
factor 7								×	×		×
factor 8	×										

and *owl* as well as *snake* and *turtle*. According to Table 1 these animals agree with each other, i.e. either all of the animals have some attribute or none of them. In the contrary, for the animals *owl* and *sea lion*, if one of the animals has a factor, then the second one does not have the factor. Correspondingly, if *owl* has an attribute in the formal context in Table 1, then *sea lion* does not have the attribute and vice versa (except for the attributes *has a backbone* and *warm-blooded* contained in the manifestations in both of the first two factors).

In the space of factors, the attributes *lays eggs* and *cold-blooded* differ only in the second factor. All the animals in the formal context in Table 1 except for *chicken*, *hawk* and *owl* (e.g. except for birds) agree on these attributes.

After the decomposition of the formal context in Table 1, we can measure the distance between animals (objects) and/or their properties (attributes) in the space of factors. Using Manhattan distance, $distance(bear, elephant) = 0$, $distance(owl, sea\ lion) = \frac{5}{8}$, $distance(lays\ eggs, cold\text{-}blooded) = \frac{1}{8}$.

It is important to notice that the location of objects and attributes in the common factor space depends on the formal context, mainly on the selection of appropriate attributes. A formal context that does not contain "good" attributes may cause that different animals will have many factors in common.

For each formal concept of the formal context in Table 1 we can compute the degree of coherence. For example,
$coherence_X^{\max}(C_{18}) = 1 - \frac{2}{8} = 0.75$,
$coherence_X^{\mathrm{avg}}(C_{18}) = 1 - \frac{\frac{2}{8}+\frac{2}{8}+0}{3} = 1 - \frac{1}{6} = \frac{5}{6}$,
$coherence_Y^{\max}(C_{18}) = 1 - \frac{4}{8} = 0.5$,
$coherence_Y^{\mathrm{avg}}(C_{18}) = 1 - \frac{\frac{2}{8}+\frac{4}{8}+\frac{2}{8}+\frac{4}{8}+\frac{4}{8}+\frac{2}{8}}{6} = 1 - \frac{3}{8} = \frac{5}{8}$.

Formal concepts $C_1, \ldots, C_{26}$ and their degrees of coherence are shown in Table 2. Remember that the higher the coherence, we consider the formal concept more important.

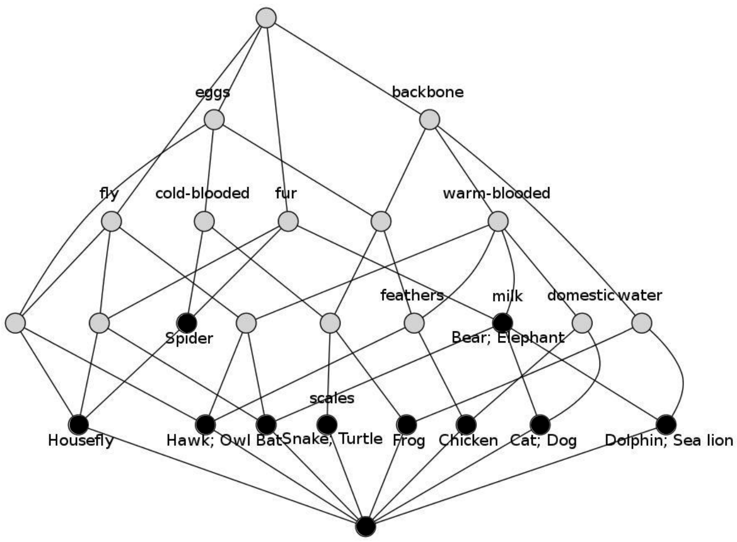


FIGURE 2. The concept lattice of animals. The formal concepts with $coherence_X^{\mathrm{avg}}$ greater or equal to 0.85 using BMF are highlighted in black

Table 2. The coherence (computed using Manhattan distance) of the formal concepts of Table 1 rounded to two decimal places

	intent of (A, B)											coherence of (A, B)					
	fur (hair)	feathers	scales	can fly	lives in water	lays eggs	produces milk	has a backbone	warm-blooded	cold-blooded	domestic	$coherence^{\max}$	$coherence^{\text{avg}}$	$coherence_X^{\max}$	$coherence_X^{\text{avg}}$	$coherence_Y^{\max}$	$coherence_Y^{\text{avg}}$
1.												0.38	0.60	0.38	0.60	1.00	1.00
2.	×											0.50	0.78	0.50	0.76	1.00	1.00
3.				×								0.63	0.75	0.63	0.71	1.00	1.00
4.						×						0.38	0.64	0.38	0.63	1.00	1.00
5.								×				0.25	0.57	0.38	0.59	1.00	1.00
6.	×			×								0.63	0.73	0.75	0.75	0.63	0.63
7.				×		×						0.50	0.70	0.63	0.75	0.50	0.50
8.					×			×				0.38	0.65	0.63	0.75	0.50	0.50
9.						×		×				0.38	0.60	0.50	0.63	0.50	0.50
10.						×				×		0.50	0.76	0.63	0.75	0.88	0.88
11.								×	×			0.25	0.63	0.38	0.66	0.75	0.75
12.	×					×				×		0.38	0.65	0.88	0.88	0.38	0.58
13.				×				×	×			0.25	0.60	0.63	0.75	0.25	0.50
14.						×		×		×		0.38	0.70	0.75	0.83	0.38	0.58
15.								×	×		×	0.50	0.71	0.63	0.75	0.50	0.67
16.	×			×		×				×		0.38	0.60	1.00	1.00	0.38	0.58
17.	×						×	×	×			0.25	0.74	0.75	0.85	0.38	0.65
18.		×				×		×	×			0.38	0.68	0.75	0.83	0.50	0.63
19.			×			×		×		×		0.38	0.74	1.00	1.00	0.38	0.65
20.					×	×		×		×		0.38	0.60	1.00	1.00	0.38	0.56
21.	×			×			×	×	×			0.25	0.61	1.00	1.00	0.25	0.60
22.	×				×		×	×	×			0.38	0.67	1.00	1.00	0.38	0.63
23.	×						×	×	×		×	0.38	0.70	1.00	1.00	0.38	0.65
24.		×		×		×		×	×			0.25	0.64	1.00	1.00	0.25	0.58
25.		×				×		×	×		×	0.50	0.68	1.00	1.00	0.50	0.63
26.	×	×	×	×	×	×	×	×	×	×	×	0.25	0.63	1.00	1.00	0.25	0.63

The most coherent formal concepts using $coherence_X^{\text{avg}}$ (the 4th column in Table 2) sorted in descending degree are C_{16}, $C_{19} - C_{25}$, C_{12} and C_{17} (Fig. 2). The formal concepts C_{19}, C_{22}, C_{23}, C_{12} and C_{17} can be named as "reptiles" (*snake*, *turtle*), "sea mammals" (*dolphin*, *sea lion*), "pets" (*cat*, *dog*), "invertebrate animals" (*housefly*, *spider*) and "mammals" (*bat*, *bear*, *cat*, *dog*, *dolphin*, *elephant*, *sea lion*), respectively.

Next, consider the most coherent concepts using the coherence measured on the attributes only $coherence_Y^{\max}$ (the 5th column in Table 2). The concepts in descending order of the degree of coherence are $C_2 - C_5$, C_{10}, C_{11}, C_6, C_7, C_8, C_9, C_{15}, C_{18} and C_{25} (Fig. 3). The concepts C_5, C_{10}, C_{11}, C_8, C_{15} and

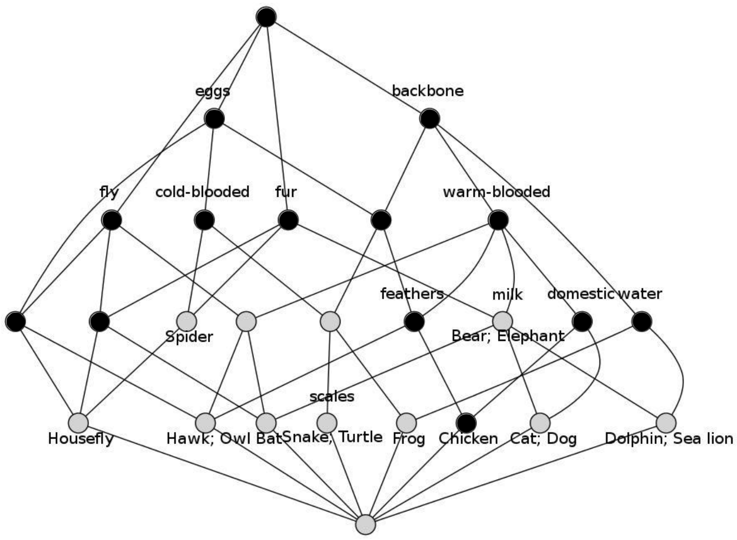


FIGURE 3. The concept lattice of animals. The formal concepts having $coherence_Y^{\max}$ greater or equal to 0.75 using BMF are highlighted in black

C_{18} represent "vertebrate animals", "cold-blooded animals", "warm-blooded animals", "aquatic animals", "domestic animals" and "birds".

The interpretation of other results (i.e. these provided by $coherence^{\max}$, $coherence^{\text{avg}}$, $coherence_X^{\max}$ and $coherence_Y^{\text{avg}}$) is left to the reader.

The decision of whether to compute the coherence on objects or attributes as well as the use of $coherence^{\max}$ or $coherence^{\text{avg}}$ depends on the purpose of the concrete application of FCA on the data and also on other user-related factors.

7. EXPERIMENTS

In this section, we present some experiments we have performed to give a deeper insight into the behaviour of the proposed method for ranking formal concepts. Benchmark data sets used for these experiments are taken from the UCI Machine Learning Repository [1] characteristics of which are shown in Table 3.

7.1. **Experiment 1.** In the first experiment, we have explored the degrees of coherence that are assigned to formal concepts.

Using Boolean Matrix Factorization (BMF) for the decomposition of formal contexts we have found out that (see Table 4):

TABLE 3. Some characteristics of the used data sets in our experiments

Dataset	# Objects	# Attributes	# Factors (BMF)	# Formal Concepts
Car	1728	25	25	12640
Spect Heart	267	46	46	2135549
Tic-tac-toe	958	29	29	59505
Wine	178	68	57	24423

- Many concepts of the formal contexts (data sets) have the same degree of coherence if we measure the coherence using maximum distance between objects and/or attributes in the factor space. For example, for wine data set the total number of formal concepts is 24423, each of which is assigned 1 out of 12 degrees of coherence (using $coherence^{\max}$).
- The number of distinct coherence values computed by the maximum distance is identical using either of the two distance measures (Manhattan, Euclidean).
- $coherence_X^{\max}$ and $coherence_X^{\mathrm{avg}}$ provide more distinct coherence values than $coherence_Y^{\max}$ and $coherence_Y^{\mathrm{avg}}$, respectively. It follows from the fact that the number of objects is greater than the number of attributes for each data set.
- It is appropriate to utilize the Euclidean distance instead of the Manhattan distance for measuring the coherence, because the number of distinct degrees of coherence that are assigned to formal concepts is greater if we use the Euclidean distance (what is not surprising, since the factor matrices are binary, i.e. contain only 0s and 1s).
- For tic-tac-toe data the number of distinct degrees of coherence assigned to formal concepts using $coherence_Y^{\max}$ and $coherence_Y^{\mathrm{avg}}$ is very small, because each attribute possesses a unique factor no other attribute has.

We can conclude that, using BMF, the same coherence degree is assigned to many formal concepts (see Table 4). These concepts are then ranked at the same position which is not useful for a user.

We have also carried out similar experiment where SVD and NMF were used for the decomposition of formal contexts. Remember that factor matrices generated by SVD and NMF are real-valued matrices. Therefore, using the average distance between objects and/or attributes in a factor space, almost all formal concepts take on different degrees of coherence. Further, using the maximum distance between objects and/or attributes in a space of factors, the number of distinct degrees of coherence is greater (in comparison to the case of using BMF) when we use SVD or NMF for the decomposition of formal contexts.

TABLE 4. The numbers of distinct values of coherence that the formal concepts of the formal contexts (data sets) take on using the respective ways of measuring the coherence (BMF was used for decomposition of data sets)

	Manhattan distance						Euclidean distance					
Dataset	$coherence^{\max}$	$coherence^{\text{avg}}$	$coherence_X^{\max}$	$coherence_X^{\text{avg}}$	$coherence_Y^{\max}$	$coherence_Y^{\text{avg}}$	$coherence^{\max}$	$coherence^{\text{avg}}$	$coherence_X^{\max}$	$coherence_X^{\text{avg}}$	$coherence_Y^{\max}$	$coherence_Y^{\text{avg}}$
Car	7	826	8	609	6	28	7	2326	8	1381	6	57
Spect Heart	15	231041	25	153976	5	127	15	2072196	25	1969571	5	372
Tic-tac-toe	8	1156	10	1048	2	2	8	6557	10	5402	2	6
Wine	12	4906	16	3127	17	651	12	24360	16	16868	17	8132

The use of SVD and NMF in our approach allow us to differentiate better between formal concepts with respect to degrees of coherence, and thus are better for ranking formal concepts according to their coherence. From this point of view, it is also appropriate to measure the coherence utilizing the average distance (not the maximum distance) between objects and/or attributes in the extents and/or intents of formal concepts, respectively.

7.2. **Experiment 2.** The aim of the second experiment is to investigate the impact of matrix factorization methods on the selection of important (coherent) formal concepts. Hence, we have compared the top-k most coherent formal concepts resulting from our approach by using various methods of matrix decomposition.

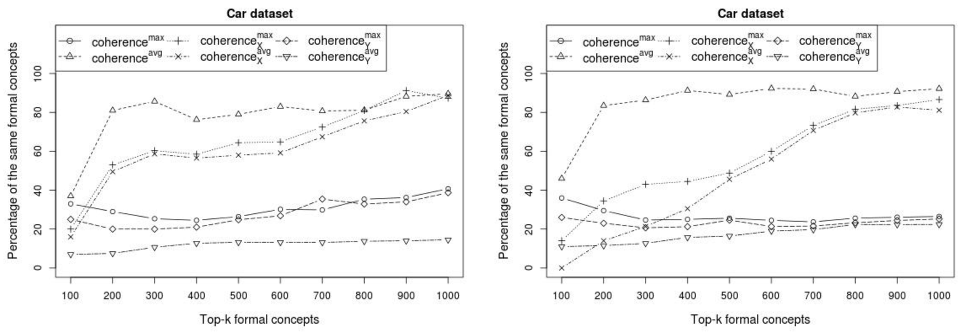
Based on the conclusions of the previous experiment, matrix factorization methods that decompose a matrix into a product of real-valued matrices are better if we want to rank formal concepts according to their degrees of coherence. Thus, in this experiment we have used SVD and NMF for matrix decomposition.

For a formal concept (A, B) it holds that if $|A| \leq 1$ ($|B| \leq 1$), then $coherence_X^? = 1$ ($coherence_Y^? = 1$), where $? = \max$ or $? = \text{avg}$. The number of formal concepts satisfying these conditions, and thus having the corresponding degrees of coherence equal to 1 are shown in Table 5. Since this assertion holds regardless of the selected method of matrix factorization, we did not consider such formal concepts in this experiment. Obviously, formal concepts that do not satisfy these conditions can also have degrees of coherence equal to 1.

The results of the comparison of the top-k most coherent formal concepts using SVD and NMF are shown in Fig. 4 – Fig. 7.

TABLE 5. The number of formal concepts (A, B) such that $|A| \leq 1$ or $|B| \leq 1$

Dataset	# Formal Concepts (A, B) having $\lvert A\rvert \leq 1$	# Formal Concepts (A, B) having $\lvert B\rvert \leq 1$
Car	1729	23
Spect Hear	215	44
Tic-tac-toe	959	30
Wine	169	37



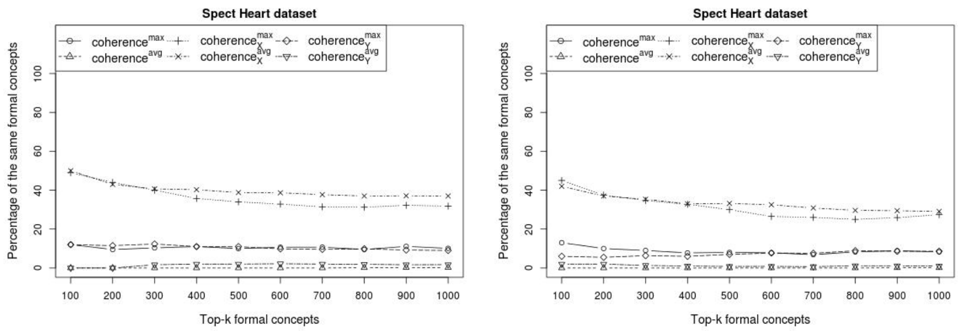
(a) Manhattan distance (b) Euclidean distance

FIGURE 4. The percentage of the same formal concepts from the top-k formal concepts using SVD and NMF for the decomposition of Car dataset using Manhattan distance (Fig. 4(a)) and Euclidean distance (Fig. 4(b)).

The used matrix factorization method has only a little influence on formal concepts provided by $coherence^{\text{avg}}$ (except for the Spect Heart dataset). Approximately 80% of formal concepts provided by $coherence^{\text{avg}}$ are the same if we decompose data sets using SVD or NMF (for the Car and Tic-tac-toe datasets). On the other hand, $coherence^{\text{max}}$, $coherence_Y^{\text{max}}$ and $coherence_Y^{\text{avg}}$ are quite sensitive on the selected method of matrix decomposition. The results are similar regardless of the computation of the distance in a space of factors (Manhattan distance, Euclidean distance).

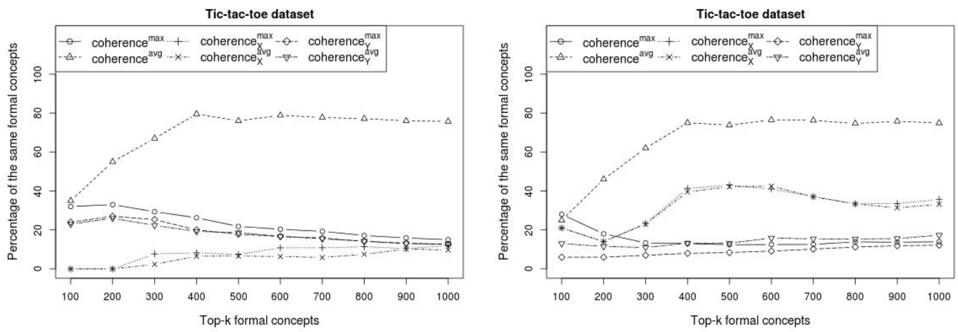
8. CONCLUSIONS

We have introduced a novel approach to rank (and thus to reduce the number of) formal concepts utilizing different types of matrix factorization methods. Besides the intuitive choice, the Boolean Matrix Factorization technique (BMF), we have utilized also Singular Value Decomposition (SVD) and Non-negative Matrix Factorization (NMF). As our experiments showed, using



(a) Manhattan distance (b) Euclidean distance

FIGURE 5. The percentage of the same formal concepts from the top-k formal concepts using SVD and NMF for the decomposition of Spect Heart dataset using Manhattan distance (Fig. 5(a)) and Euclidean distance (Fig. 5(b)).

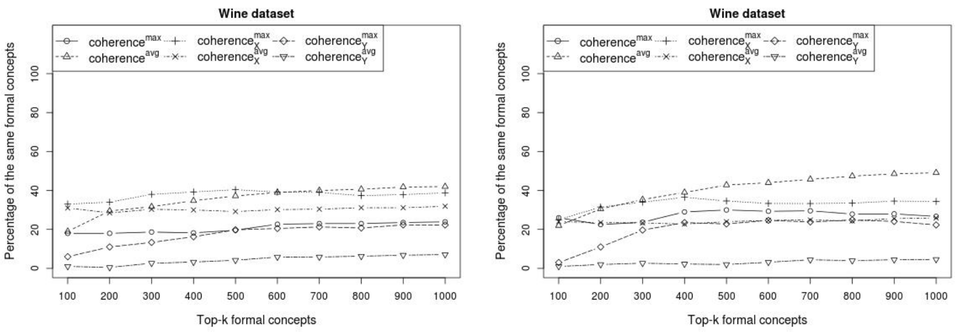


(a) Manhattan distance (b) Euclidean distance

FIGURE 6. The percentage of the same formal concepts from the top-k formal concepts using SVD and NMF for the decomposition of Tic-tac-toe dataset using Manhattan distance (Fig. 6(a)) and Euclidean distance (Fig. 6(b)).

BMF in our approach results in a case when only a small number of distinct values are assigned to formal concepts and thus many formal concepts have the same degree of coherence which is not helpful in ranking. However, having just a small number of different ranking degrees could be interesting in some cases of application of FCA to data.

The main research issue we would like to focus on the following issues:



(a) Manhattan distance (b) Euclidean distance

FIGURE 7. The percentage of the same formal concepts from the top-k formal concepts using SVD and NMF for the decomposition of Wine dataset using Manhattan distance (Fig. 7(a)) and Euclidean distance (Fig. 7(b)).

- Experimental evaluation of the proposed approach on several real-world data sets including qualitative evaluation of the results by domain experts.
- Comparison with other techniques to select important formal concepts, in particular with the one for selecting basic level concepts [4].

Acknowledgements: This publication is the result of the Project implementation: University Science Park TECHNICOM for Innovation Applications Supported by Knowledge Technology, ITMS: 26220220182, supported by the Research & Development Operational Programme funded by the ERDF and VEGA 1/0832/12.

References

[1] K. Bache, M. Lichman, *UCI Machine Learning Repository.* 2013. URL http://archive.ics.uci.edu/ml.

[2] R. Bělohlávek, J. Macko, *Selecting Important Concepts Using Weights.* International Conference on Formal Concept Analysis, LNAI, vol. 6628, Springer, 2011, pp. 65-80.

[3] R. Bělohlávek, V. Sklenář, *Formal Concept Analysis Constrained by Attribute-Dependency Formulas*, International Conference on Formal Concept Analysis, LNAI, vol. 3403, Springer, 2005, pp. 176-191.

[4] R. Bělohlávek, M. Trnečka, *Basic Level of Concepts in Formal Concept Analysis.* International Conference on Formal Concept Analysis, LNAI, vol. 7278, Springer, 2012, pp. 28-44.

[5] R. Bělohlávek, V. Vychodil, *Discovery of optimal factors in binary data via a novel method of matrix decomposition.* Journal of Computer and System Sciences, vol. 76, 2010, pp. 3-20.

[6] V. Codocedo, C. Taramasco, H. Astudillo, *Cheating to achieve Formal Concept Analysis over a large formal context.* Concept Lattices and Their Application, 2011, pp. 349-362.
[7] S. M. Dias, N. J. Vieira, *Reducing the Size of Concept Lattices: The JBOS approach.* Concept Lattices and Their Application, CEUR WS, vol. 672, 2010, pp. 80-91.
[8] P. Gajdoš, P. Moravec, V. Snášel, *Concept Lattice Generation by Singular Value Decomposition.* Concept Lattices and Their Application, 2004, pp. 102-110.
[9] B. Ganter, R. Wille, *Formal concept analysis: Mathematical foundations.* Springer, 1999.
[10] J. Han, H. Cheng, D. Xin, X. Yan, *Frequent pattern mining: current status and future directions.* Data Mining and Knowledge Discovery, vol. 15, 2007, pp. 55-86.
[11] M. Klimushkin, S. Obiedkov, C. Roth, *Approaches to the Selection of Relevant Concepts in the Case of Noisy Data.* International Conference on Formal Concept Analysis, LNAI, vol. 5986, Springer, 2010, pp. 255-266.
[12] S. O. Kuznetsov, *On stability of a formal concept.* Annals of Mathematics and Artificial Intelligence, vol. 49(1-4), 2007, pp. 101-115.
[13] D. D. Lee, H. S. Seung, *Algorithms for non-negative matrix factorization.* Advances in neural information processing systems, vol. 13, 2001.
[14] P. Miettinen, T. Mielikäinen, A. Gionis, G. Das, H. Mannila, *The Discrete Basis Problem.* IEEE Transactions on Knowledge and Data Engineering, vol. 20(10), 2008, pp. 1348-1362.
[15] G. Stumme, *Efficient Data Mining Based on Formal Concept Analysis.* DEXA, Springer, LNCS, vol. 2453, 2002, pp. 534-546.
[16] R. Wille, *Restructuring lattice theory: An approach based on hierarchies of concepts.* Ordered Sets, vol. 83, 1982, pp. 445-470.

Institute of Computer Science, Faculty of Science, Pavol Jozef Šafárik University, Jesenná 5, Košice, Slovakia
E-mail address: {lenka.piskova, tomas.horvath, stanislav.krajci}@upjs.sk