Boolean Matrix Factorisation for Collaborative Filtering: An FCA-Based Approach

Dmitry I. Ignatov[1], Elena Nenova[2], Natalia Konstantinova[3], and Andrey V. Konstantinov[1]

[1] National Research University Higher School of Economics, Moscow
dignatov@hse.ru
[2] Imhonet, Moscow
[3] University of Wolverhampton, UK

Abstract. We propose a new approach for Collaborative filtering which is based on Boolean Matrix Factorisation (BMF) and Formal Concept Analysis. In a series of experiments on real data (MovieLens dataset) we compare the approach with an SVD-based one in terms of Mean Average Error (MAE). One of the experimental consequences is that it is enough to have a binary-scaled rating data to obtain almost the same quality in terms of MAE by BMF as for the SVD-based algorithm in case of non-scaled data.

Keywords: Boolean Matrix Factorisation, Formal Concept Analysis, Singular Value Decomposition, Recommender Algorithms

1 Introduction

Recommender Systems have recently become one of the most popular subareas of Machine Learning and Data Mining. In fact, the recommender algorithms based on matrix factorisation techniques (MF) are now considered industry standard [1].

Among the most frequently used types of Matrix Factorisation we should definitely mention Singular Value Decomposition (SVD) [2] and its various modifications like Probabilistic Latent Semantic Analysis (PLSA) [3] and SVD++ [4]. However, the existing similar techniques, for example, non-negative matrix factorisation (NMF) [5] and Boolean matrix factorisation (BMF) [6], seem to be less studied in the context of Recommender Systems. An approach similar to the matrix factorization is biclustering which was also successfully applied in recommender system domain [7,8]. For example, Formal Concept Analysis [9] can be also used as a biclustering technique and there are already several examples of its applications in recommenders' algorithms [10,11].

The aim of this paper is to compare the recommendation quality of the aforementioned techniques on the real datasets and try to investigate methods' interrelationship. It is especially interesting to conduct experiments and compare recommendation quality in case of an input matrix with numeric values and in case of a Boolean matrix in terms of Precision and Recall as well as MAE.

Moreover, one of the useful properties of matrix factorisation is its ability to keep reliable recommendation quality even in case of dropping some insufficient factors. For BMF this issue is experimentally investigated in section 4.

The novelty of the paper is defined by the fact that it is the first time when BMF based on Formal Concept Analysis [9] is investigated in the context of Recommender Systems.

The practical significance of the paper is determined by the demand of recommender systems' industry, that is focused on gaining reliable quality in terms of Mean Average Error (MAE), Precision and Recall as well as competitive time performance of the investigated method.

The rest of the paper consists of five sections. Section 2 is an introductory review of the existing MF-based recommender approaches. In section 3 we describe our recommender algorithm which is based on Boolean matrix factorisation using closed sets of users and items (that is FCA). Section 4 contains methodology of our experiments and results of experimental comparison of two MF-based recommender algorithms by means of cross-validation in terms of MAE and F-measure. The last section concludes the paper.

2 Introductory review of some matrix factorisation approaches

In this section we briefly describe two approaches to the decomposition of both real-valued and Boolean matrices.

2.1 Singular Value Decomposition (SVD)

Singular Value Decomposition (SVD) is a decomposition of a rectangular matrix $A \in \mathbb{R}^{m \times n} (m > n)$ into a product of three matrices

$$A = U \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V^T, \tag{1}$$

where $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are orthogonal matrices, and $\Sigma \in \mathbb{R}^{n \times n}$ is a diagonal matrix such that $\Sigma = diag(\sigma_1, \ldots, \sigma_n)$ and $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0$. The columns of the matrix U and V are called singular vectors, and the numbers σ_i are singular values [2].

In the context of recommendation systems rows of U and V can be interpreted as vectors of user's and items's loyalty (attitude) to a certain topic (factor), and the corresponding singular values as importance of the topic among the others. The main disadvantage lies in the fact that the matrix may contain both positive and negative numbers; the last ones are difficult to interpret.

The advantage of SVD for recommendation systems is that this method allows to obtain a vector of user's loyalty to certain topics for a new user without SVD decomposition of the whole matrix.

The evaluation of computational complexity of SVD according to [12] is $O(mn^2)$ floating-point operations if $m \geq n$ or more precisely $2mn^2 + 2n^3$.

2.2 Boolean Matrix Factorisation (BMF) based on Formal Concept Analysis (FCA)

Basic FCA definitions. Formal Concept Analysis (FCA) is a branch of applied mathematics and it studies (formal) concepts and their hierarchy [9]. The adjective "formal" indicates a strict mathematical definition of a pair of sets, called, the extent and intent. This formalisation is possible because of the use of the algebraic lattice theory.

DEFINITION 1. *Formal context* K is a triple (G, M, I), where G is a set of *objects*, M is a set of *attributes* , $I \subseteq G \times M$ is a binary relation.

The binary relation I is interpreted as follows: for $g \in G$, $m \in M$ we write gIm if the object g has the attribute m.

For a formal context $\mathbb{K} = (G, M, I)$ and any $A \subseteq G$ and $B \subseteq M$ a pair of mappings is defined:

$$A' = \{m \in M \mid gIm \text{ for all } g \in A\}, \quad B' = \{g \in G \mid gIm \text{ for all } m \in B\},$$

these mappings define Galois connection between partially ordered sets $(2^G, \subseteq)$ and $(2^M, \subseteq)$ on disjunctive union of G and M. The set A is called *closed set*, if $A'' = A$ [13].

DEFINITION 2. A *formal concept* of the formal context $\mathbb{K} = (G, M, I)$ is a pair (A, B), where $A \subseteq G$, $B \subseteq M$, $A' = B$ and $B' = A$. The set A is called the *extent*, and B is the *intent* of the formal concept (A, B).

It is evident that the extent and intent of any formal concept are closed sets.

The set of all formal concepts of a context $\mathbb{K}$ is denoted by $\mathfrak{B}(G, M, I)$.

The state-of-the-art surveys on advances in FCA theory and its applications can be found in [14,15].

Description of FCA-based BMF Boolean matrix factorization (BMF) is a decomposition of the original matrix $I \in \{0,1\}^{n \times m}$, where $I_{ij} \in \{0,1\}$, into a Boolean matrix product $P \circ Q$ of binary matrices $P \in \{0,1\}^{n \times k}$ and $Q \in \{0,1\}^{k \times m}$ for the smallest possible number of k. We define boolean matrix product as follows:

$$(P \circ Q)_{ij} = \bigvee_{l=1}^{k} P_{il} \cdot Q_{lj},$$

where $\bigvee$ denotes disjunction, and $\cdot$ conjunction.

Matrix I can be considered a matrix of binary relations between set X of objects (users), and a set Y of attributes (items that users have evaluated). We assume that xIy iff the user x evaluated object y. The triple (X, Y, I) clearly forms a formal context.

Consider a set $\mathcal{F} \subseteq \mathcal{B}(X, Y, I)$, a subset of all formal concepts of context (X, Y, I), and introduce matrices $P_{\mathcal{F}}$ and $Q_{\mathcal{F}}$:

$$(P_{\mathcal{F}})_{il} = \begin{cases} 1, i \in A_l, \\ 0, i \notin A_l, \end{cases} \quad (Q_{\mathcal{F}})_{lj} = \begin{cases} 1, j \in B_l, \\ 0, j \notin B_l. \end{cases},$$

where (A_l, B_l) is a formal concept from F. We can consider decomposition of the matrix I into binary matrix product $P_{\mathcal{F}}$ and $Q_{\mathcal{F}}$ as described above. The following theorems are proved in [6]:

Theorem 1. (Universality of formal concepts as factors). For every I there is $\mathcal{F} \subseteq \mathcal{B}(X, Y, I)$, such that $I = P_{\mathcal{F}} \circ Q_{\mathcal{F}}$.

Theorem 2. (Optimality of formal concepts as factors). Let $I = P \circ Q$ for $n \times k$ and $k \times m$ binary matrices P and Q. Then there exists a $\mathcal{F} \subseteq \mathcal{B}(X, Y, I)$ of formal concepts of I such that $|\mathcal{F}| \leq k$ and for the $n \times |\mathcal{F}|$ and $|\mathcal{F}| \times m$ binary matrices $P_{\mathcal{F}}$ and $Q_{\mathcal{F}}$ we have $I = P_{\mathcal{F}} \circ Q_{\mathcal{F}}$.

There are several algorithms for finding $P_{\mathcal{F}}$ and $Q_{\mathcal{F}}$ by calculating formal concepts based on these theorems [6].

The algorithm we use (Algoritm 2 from [6]) avoids computation of all possible formal concepts and therefore works much faster [6]. Time estimation of the calculations in the worst case yields $O(k|G||M|^3)$, where k is the number of found factors, $|G|$ is the number of objects, $|M|$ is the number of attributes.

2.3 General scheme of user-based recommendations

Once a matrix of rates is factorized we need to learn how to compute recommendations for users and to evaluate whether a particular method handles this task well.

For the factorized matrices already well-known algorithm based on the similarity of users can be applied, where for finding K nearest neighbors we use not the original matrix of ratings $A \in \mathbb{R}^{m \times n}$, but the matrix $U \in \mathbb{R}^{m \times f}$, where m is a number of users, and f is a number of factors. After the selection of K users, which are the most similar to a given user, based on the factors that are peculiar to them, it is possible, based on collaborative filtering formulas to calculate the projected rates for a given user.

After generation of recommendations the performance of the recommender system can be estimated by measures such as Mean Absolute Error (MAE), Precision and Recall.

3 A recommender algorithm using FCA-based BMF

3.1 kNN-based algorithm

Collaborative recommender systems try to predict the utility (in our case rates) of items for a particular user based on the items previously rated by other users.

Memory-based algorithms make rating predictions based on the entire collection of previously rated items by the users. That is, the value of the unknown rating $r_{c,s}$ for a user c and item s is usually computed as an aggregate of the ratings of some other (usually, the K most similar) users for the same item s:

$$r_{c,s} = aggr_{c' \in \widehat{C}} r_{c',s},$$

where $\widehat{C}$ denotes a set of K users that are the most similar to user c, who have rated item s. For example, the function *aggr* may have the following form [16]

$$r_{c,s} = k \sum_{c' \in \widehat{C}} sim(c', c) \times r_{c',s},$$

where k serves as a normalizing factor and selected as $k = 1/ \sum_{c' \in \widehat{C}} sim(c, c')$.

Similarity measure between users c and c', $sim(c, c')$, is essentially an inverse distance measure and is used as a weight, i.e., the more similar users c and c' are, the more weight rating $r_{c',s}$ will carry in the prediction of $r_{c,s}$.

Similarity between two users is based on their ratings of items that both users have rated. The two most popular approaches are *correlation* and *cosine-based*.

To apply this approach in case of FCA-based BMF recommender algorithm we simply consider the user-factor matrices obtained after factorisation of the initial data as an input.

3.2 Scaling

In order to move from a matrix of ratings R where $R_{ij} \in (0, 1, 2, 3, 4, 5)$ to a Boolean matrix, and use the results of Boolean matrix factorisation, scaling is required. It is well known that scaling is a matter of expert interpretation of the original data. In this paper, we use binary scaling with different thresholds and compare the results in terms of MAE.

1. $I_{ij} = 1$ if $R_{ij} > 0$, else $I_{ij} = 0$ (user i rates item j).
2. $I_{ij} = 1$ if $R_{ij} > 1$, else $I_{ij} = 0$.
3. $I_{ij} = 1$ if $R_{ij} > 2$, else $I_{ij} = 0$.
4. $I_{ij} = 1$ if $R_{ij} > 3$, else $I_{ij} = 0$.

4 Experiments

To test our hypotheses and study the behavior of recommendations based on the factorisation of a ratings matrix by different methods we used MovieLens data. We used a part of the data, containing 100,000 ratings, and considered only users who have given more than 20 ratings.

The user ratings are split into two sets, a training set consisting of 80 000 ratings, and a test set consisting of 20 000 ratings. The original data matrix has the size of 943×1682, where the number of rows is the number of users and the number of columns is the number of rated movies (each movie has at least one vote).

4.1 The number of factors that cover $p\%$ of evaluations in an input data for SVD and BMF

The main purpose of matrix factorisation is a reduction of matrices dimensionality. Therefore we examine how the number of factors varies depending on the method of factorization, and depending on p % of the data that is covered by factorization. For BMF the coverage of a matrix is calculated as the ratio of the number of ratings covered by Boolean factorization to the total number of ratings.

$$\frac{|covered\_ratings|}{|all\_ratings|} \cdot 100\% \approx p_{BMF}\%, \tag{2}$$

For SVD we use the following formula:

$$\frac{\sum_{i=1}^{K} \sigma_i^2}{\sum \sigma_i^2} \cdot 100\% \approx p_{SVD}\%, \tag{3}$$

where K is the number of factors selected.

Table 1. Number of factors for SVD and BMF at different coverage level

p%	100%	80%	60%
SVD	943	175	67
BMF	1302	402	223

4.2 MAE-based recommender quality comparison of SVD and BMF for various levels of evaluations coverage

The main purpose of matrix factorisation is the reduction of matrices dimensionality. As a result some part of the original data remains not covered, so it was interesting to explore how the quality of recommendations changes based on different factorisations, depending on the proportion of the data covered by factors.

Two methods of matrix factorisation were considered: BMF and SVD. The fraction of data covered by factors is defined in subsections 2 and 3.

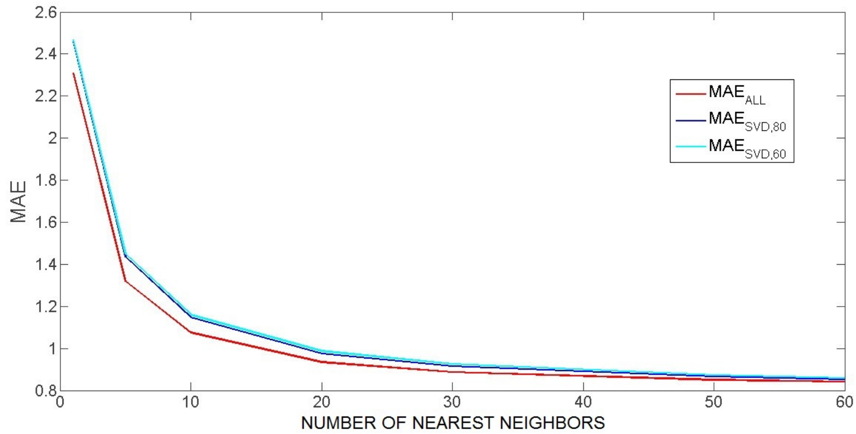

Fig. 1. MAE dependence on the percentage of the data covered by SVD-decomposition, and the number of nearest neighbors.

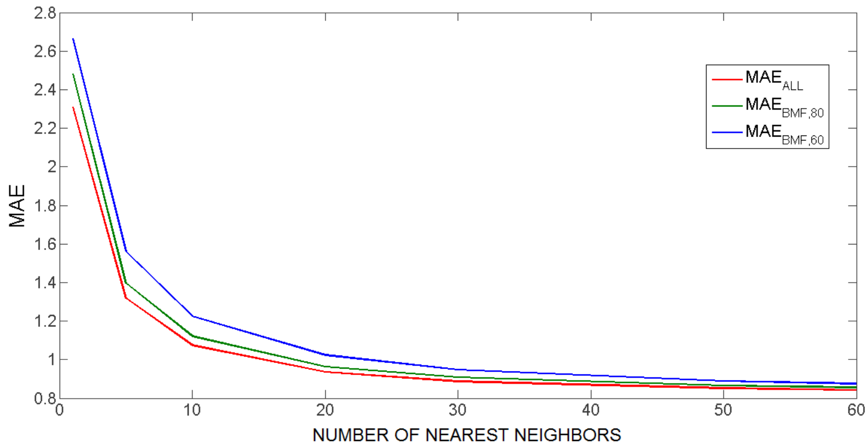

Fig. 2. MAE dependence on the percentage of the data covered by BMF-decomposition, and the number of nearest neighbors.

Fig. 1 shows that MAE_{SVD60}, calculated for the model based on 60% of factors, is not very different from MAE_{SVD80}, calculated for the model built for 80% factors. At the same time, for the recommendations based on a Boolean factorization covering 60% and 80% of the data respectively, it is clear that increasing the number of factors improves MAE, as shown in Fig. 2.

Table 2. MAE for SVD and BMF at 80% coverage level

Number of neighbors	1	5	10	20	30	50	60
MAE_{SVD80}	**2,4604**	1.4355	1.1479	0.9750	0.9148	0.8652	**0.8534**
MAE_{BMF80}	2.4813	**1.3960**	**1.1215**	**0.9624**	**0.9093**	**0.8650**	0.8552
MAE_{all}	2.3091	1.3185	1.0744	0.9350	0.8864	0.8509	0.8410

Table 2 shows that MAE for recommendations built on a Boolean factorisation covering 80 % of the data, for the number of neighbors less than 50, is better than the MAE for recommendations built on SVD factorization. It is also easy to see that difference of MAE_{SVD80} and MAE_{BMF80} from MAE_{all} is no more than $1 - 7\%$.

4.3 Comparison of kNN-based approach and BMF by Precision and Recall

Besides comparison of the algorithms using MAE, other evaluation metrics can also be exploited, for example, $Recall = \frac{|objects\_in\_recommendation \cap objects\_in\_test|}{|objects\_in\_test|}$, $Precision = \frac{|objects\_in\_recommendation \cap objects\_in\_test|}{|objects\_in\_recommendation|}$ and $F1 = \frac{2 \cdot Recall \cdot Precision}{Recall + Precision}$.

Usually the larger $F1$ (F-measure) is, the better is recommendation algorithm.

Figure 3 shows the dependence of the evaluation metric on the percentage of data covered by BMF-decomposition, and the number of nearest neighbors. The number of objects to recommend was chosen to be 20. The figure shows that the recommendation based on the Boolean decomposition, is worse than recommendations generated on the full matrix of ratings.

4.4 Influence of scaling on the recommendations quality for BMF in terms of MAE

Another aspect that was interesting to examine was the impact of scaling described in subsection 3.2 on the quality of recommendations. All four options from 3.2 of scaling were considered. The distribution of ratings in the data is shown in Table 3

Table 3. The rating distribution of Movie Lens data

Rating	1	2	3	4	5
Part of all rates %	6.1	11.4	27.2	34.1	21.2

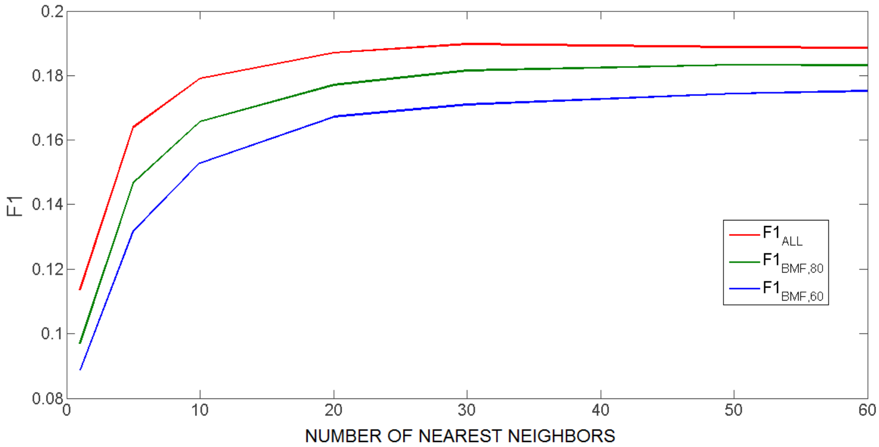

Fig. 3. *$F1$ dependence on the percentage of data covered by BMF-decomposition, and the number of nearest neighbors.*

For each of the Boolean matrices we have calculated its Boolean factorisation, covering 80 % of the data. Then recommendations are calculated just like in subsection 4.2. It can be seen on Figure 4.4 that MAE_1 is almost the same as MAE_0, and $MAE_{2,3}$ is better than MAE_1, when the number of nearest neighbors is more than 30.

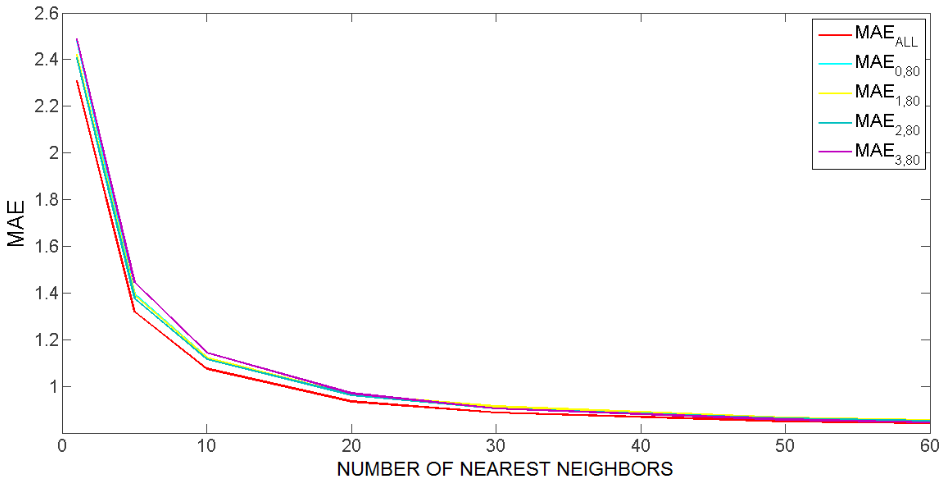

Fig. 4. MAE dependence on scaling and number of nearest neighbors for 80% coverage.

4.5 Influence of data filtering on MAE for BMF kNN-based approach

Besides the ability to search for K nearest neighbors not in the full matrix of ratings $A \in \mathbb{R}^{n \times m}$, but in the matrix $U \in \mathbb{R}^{m \times f}$, where m is a number of users, and f is a number of factors, Boolean matrix factorization can be used for data filtering. Since the algorithm as an output returns not only users-factors and factors-objects matrices, but also the ratings that were not used for factoring, we can try to search for users, similar to the target user, based on the matrix consisting only of ratings used for the factorisation.

Just as before to find the nearest neighbors cosine measure is used, and the predicted ratings are calculated as the weighted sum of the ratings of nearest users.

Figure 5 shows that the recommendations built on user-factor matrix, are better than recommendations, constructed on matrix of ratings filtered with boolean factorization.

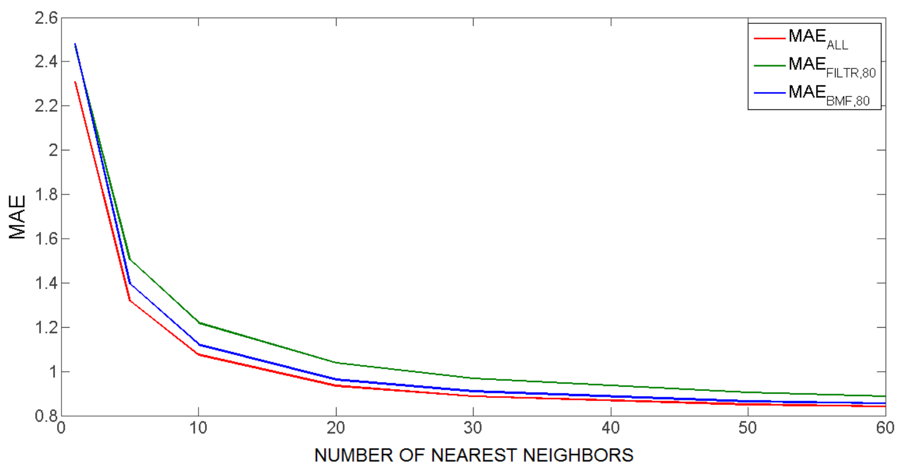

Fig. 5. MAE dependence on data filtration algorithm and the number of nearest neighbors.

5 Conclusion

In the paper we considered two methods of Matrix Factorisation which are suitable for Recommender Systems. They were compared on real datasets. We investigated BMF behaviour as part of recommender algorithm. We also conducted several experiments on recommender quality comparison with numeric matrices, user-factor and factor-item matrices in terms of F-measure and MAE. We showed that MAE of our BMF-based approach is not greater than MAE of

SVD-based approach for the same coverage percentage of BMF and p-level of SVD.

We have also investigated how data filtering, namely scaling, influences on recommendations' quality. In terms of MAE, the BMF-based collaborative filtering algorithm demonstrates almost the same level of quality before scaling (full information) and after (considerable information loss).

Even though the reported results were obtained on a freely abvailable datasets and therefore they can be easily reproduced, in case of another datasets (different type of items or data size and its density) additional tests seems to be necessary.

As a future research direction we would like to investigate the proposed approaches in case of graded and triadic data [17,18] and reveal whether there are some benefits for the algorithm's quality in usage of least-squares data imputation techniques [19]. In the context of matrix factorisation we would also like to test our approach for the quality assessment of the recommender algorithms that we performed on some basic algorithms (see bimodal cross-validation in [20]).

Acknowledgments. We would like to thank Radim Belohlavek, Vilem Vychodil, Sergei Kuznetsov, Vladimir Bobrikov and anonymous reviewers for their comments, remarks and explicit and implicit help during the paper preparations. This work was supported by the Basic Research Program at the National Research University Higher School of Economics in 2012-2014 and performed in the Laboratory of Intelligent Systems and Structural Analysis. First author was also supported by Russian Foundation for Basic Research (grant #13-07-00504).

References

1. Koren, Y., Bell, R., Volinsky, C.: Matrix factorization techniques for recommender systems. Computer **42**(8) (2009) 30–37
2. Elden, L.: Matrix Methods in Data Mining and Pattern Recognition. Society for Industrial and Applied Mathematics (2007)
3. Hofmann, T.: Unsupervised learning by probabilistic latent semantic analysis. Machine Learning **42**(1-2) (2001) 177–196
4. Koren, Y.: Factorization meets the neighborhood: A multifaceted collaborative filtering model. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '08, New York, NY, USA, ACM (2008) 426–434
5. Lin, C.J.: Projected gradient methods for nonnegative matrix factorization. Neural Comput. **19**(10) (October 2007) 2756–2779
6. Belohlavek, R., Vychodil, V.: Discovery of optimal factors in binary data via a novel method of matrix decomposition. Journal of Computer and System Sciences **76**(1) (2010) 3 – 20 Special Issue on Intelligent Data Analysis.
7. Symeonidis, P., Nanopoulos, A., Papadopoulos, A., Manolopoulos, Y.: Nearest-biclusters collaborative filtering based on constant and coherent values. Information Retrieval **11**(1) (2008) 51–75
8. Ignatov, D.I., Kuznetsov, S.O., Poelmans, J.: Concept-based biclustering for internet advertisement. In: Data Mining Workshops (ICDMW), 2012 IEEE 12th International Conference on. (Dec 2012) 123–130

9. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations. Springer, Berlin/Heidelberg (1999)
10. du Boucher-Ryan, P., Bridge, D.: Collaborative recommending using formal concept analysis. Knowledge-Based Systems **19**(5) (2006) 309 – 315 {AI} 2005 {SI}.
11. Ignatov, D.I., Kuznetsov, S.O.: Concept-based recommendations for internet advertisement. In Belohlavek, R., Kuznetsov, S.O., eds.: Proc. of The Sixth International Conference Concept Lattices and Their Applications (CLA'08), Palacky University, Olomouc (2008) 157–166
12. Trefethen, L.N., Bau, D.: Numerical Linear Algebra. 3rd edition edn. SIAM (1997)
13. Birkhoff, G.: Lattice Theory. eleventh printing edn. Harvard University, Cambridge, MA (2011)
14. Poelmans, J., Ignatov, D.I., Kuznetsov, S.O., Dedene, G.: Formal concept analysis in knowledge processing: A survey on applications. Expert Syst. Appl. **40**(16) (2013) 6538–6560
15. Poelmans, J., Kuznetsov, S.O., Ignatov, D.I., Dedene, G.: Formal concept analysis in knowledge processing: A survey on models and techniques. Expert Syst. Appl. **40**(16) (2013) 6601–6623
16. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Trans. on Knowl. and Data Eng. **17**(6) (June 2005) 734–749
17. Belohlavek, R., Osicka, P.: Triadic concept lattices of data with graded attributes. International Journal of General Systems **41**(2) (2012) 93–108
18. Belohlavek, R.: Optimal decompositions of matrices with entries from residuated lattices. Journal of Logic and Computation (2011)
19. Wasito, I., Mirkin, B.: Nearest neighbours in least-squares data imputation algorithms with different missing patterns. Comput. Stat. Data Anal. **50**(4) (February 2006) 926–949
20. Ignatov, D.I., Poelmans, J., Dedene, G., Viaene, S.: A new cross-validation technique to evaluate quality of recommender systems. In Kundu, M., Mitra, S., Mazumdar, D., Pal, S., eds.: Perception and Machine Intelligence. Volume 7143 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2012) 195–202