RSVD-based Dimensionality Reduction for Recommender Systems

Michal Ciesielczyk and Andrzej Szwabe

***Abstract*— We investigate dimensionality reduction methods from the perspective of their ability to produce a low-rank customer-product matrix representation. We analyze the results of using collaborative filtering based on SVD, RI, Reflective Random Indexing (RRI) and Randomized Singular Value Decomposition (RSVD) from the perspective of selected algebraic (i.e. application-independent) properties. We show that the Frobenius-norm optimality of SVD does not correspond to the optimal recommendation accuracy, when measured in terms of F1. On the other hand, a high collaborative filtering quality is achievable when a matrix decomposition – based on a combination of RRI and SVD referred to as RSVD-RRI – leads to increased diversity of low-dimensional eigenvectors. The diversity is observable from the perspective of cosine similarities analyzed in comparison to the analogical case of SVD. Such a feature is more desirable than the fidelity of the input matrix spectrum representation, despite the MSE-optimality of SVD.**

***Index Terms*— dimensionality reduction, random indexing, recommender system, singular value decomposition.**

I. Introduction

For more than a decade, SVD – the most popular matrix dimensionality reduction method – has been used as the key element of many Collaborative Filtering (CF) systems [1][2][3][4]. Recommendation systems based on collaborative filtering do not represent the only area of successful SVD applications: probably the best example from another domain is LSA – a method that is widely regarded as significantly enhancing the quality of text retrieval [5].

In the context of recommendation systems, one of the most challenging problems is to achieve feasibility of large-scale matrix data dimensionality reduction. This problem appears as especially challenging real-world e-commerce application scenarios, as large number of users usually rate or buy only a small percentage of available products [3][6].

There are two basic reasons for using dimensionality reduction in recommender systems. One of them is the ability to improve the precision of recommendations. Another one is the ability to reduce the complexity of online computations and storage requirements [6][3]. Recommendation algorithms usually divide their internal functions into two parts: and offline and an online component. Producing a low-dimensional representation of a customer-product matrix helps to decrease the complexity of online computations and reduces the volume of a recommender system database [3][4].

On the other hand, it is often stated that a large-scale SVD application requires computational resources beyond the reach of most researchers and that SVD is a technique applicable to data sets of small or medium sizes only [2][7]. This fact is the key motivation for work on scalable 'alternatives to SVD' [4][7][8]. In particular, RI is a method that has recently attracted an attention of the text retrieval research community [7][9][10][11].

It is worth noticing that researchers working on scalable alternatives to SVD have so far focused their efforts on the development of methods enabling application-specific performance improvement, rather than on analyzing how algebraic properties of the new methods influence the retrieval quality. In particular, although it is well-known that SVD provides dimensionality reduction that is optimal in the Frobenius norm sense [12] and that SVD effectively enhances Text Retrieval (TR) and collaborative filtering methods [1][5], the relation between the properties of SVD – especially singular values distribution – and the recommendation accuracy has not yet been thoroughly investigated [13][14].

This paper concentrates on using a sparse data set, as such a case better corresponds to many real-world application scenarios [4]. This complements the subject matter of [15], which discusses a case of comparatively dense data set that is investigated in [3]. Moreover, in this paper we analyze relations between algebraic properties of RI-based dimensionality reduction methods and the recommendation accuracy.

II. RI-based Dimensionality Reduction Methods for Collaborative Filtering

Although RI is intensively investigated by the text retrieval research community [7][9][10][11], it is still unexplored as a dimensionality reduction method for collaborative filtering. To our knowledge, the only research results in the field are those reported in [15]. For this reason the RI-based dimensionality reduction methods referred to in this paper have been implemented according to the descriptions provided by authors investigating TR applications [7][9].

Typically, the implementation of a recommender system involves using SVD as a 'black-box function' realized by means of one of the well-optimized libraries. Although RI is

Manuscript received March 30, 2011. This work was partly supported by the Polish Ministry of Science and Higher Education, grant N N516 196737, and by Poznan University of Technology under grant DS45-083/10.

Michal Ciesielczyk is with the Institute of Control and Information Engineering, Poznan University of Technology, Poznan, Poland (e-mail: Michal.Ciesielczyk@put.poznan.pl).

Andrzej Szwabe is with the Institute of Control and Information Engineering, Poznan University of Technology, Poznan, Poland (e-mail: Andrzej.Szwabe@put.poznan.pl).

much less popular than SVD, once implemented, it may be comparatively easily applied to either TR or CF systems. From such a technical perspective, the dimensionality reduction methods based on a single step of SVD, RI or RRI may be regarded as algorithmically simple. Unlike RSVD – these methods are quite precisely described in the relevant literature [7][9][11]. Therefore we have decided to focus on presenting the relatively complex algorithm of RSVD [15].

It is worth stressing that the ability to provide a complete set of three (approximated) SVD-like matrices is one of the key aspects in which RSVD (in both of its variants: RSVD-RI and RSVD-RRI) differs from the existing methods based on RI.

In contrast to the existing RI-based methods - which are aimed at providing dimensionality reduction that serves a ‘functional substitute for SVD’ [7][8][9] - the method proposed in this paper assumes performing RI-based dimensionality reduction that accompanies, rather than replaces SVD. The basic idea is to use RI for the preliminary dimensionality reduction (of a relatively low computational complexity), then apply SVD to improve orthogonality of the column vectors of the context vectors matrix.

Fig. 1 summarizes both variants of the RSVD algorithm (RSVD-RI and RSVD-RRI) that have been presented more in more detail in [15]. In this paper, we use the naming convention proposed in [7][9][12] and followed in [15]. In particular, n is the number of random vectors, d – the number of dimensions used for random vectors representation, and s – the number of non-zero dimensions of a random vector (so-called seed, s << d).

III. Quality of Collaborative Filtering with RI-based Dimensionality Reduction

A. RSVD evaluation methodology

The proposed methodology for evaluation of RSVD is based on the following preliminary assumptions:

- RSVD is aimed at substituting SVD, so results of RSVD should be compared to SVD and evaluated from the perspective of the ability of obtaining RSVD results as similar to given SVD results as possible.
- As we target application of RSVD in recommender systems, we should focus on applications involving extensive dimensionality reduction, e.g. in case of experiments on relatively dense (as seen from the typical density of e-commerce user-product matrices) ML100k data set we should focus on the numbers of dimensions not higher than around thirty [3].
- In any case of using RSVD (at least for d < r, where r represents the rank of X) one cannot expect to obtain results of RSVD exactly the same as these obtained by means of SVD; moreover, as no universal RSVD evaluation criterion exists, selected evaluation criteria have to correspond to particular application scenarios.
- In order to evaluate RSVD in a way that abstracts from any application-specific scenario, some objective measure has to be used – it should be well established in the literature on recommender systems and correspond to algebraic properties of SVD.

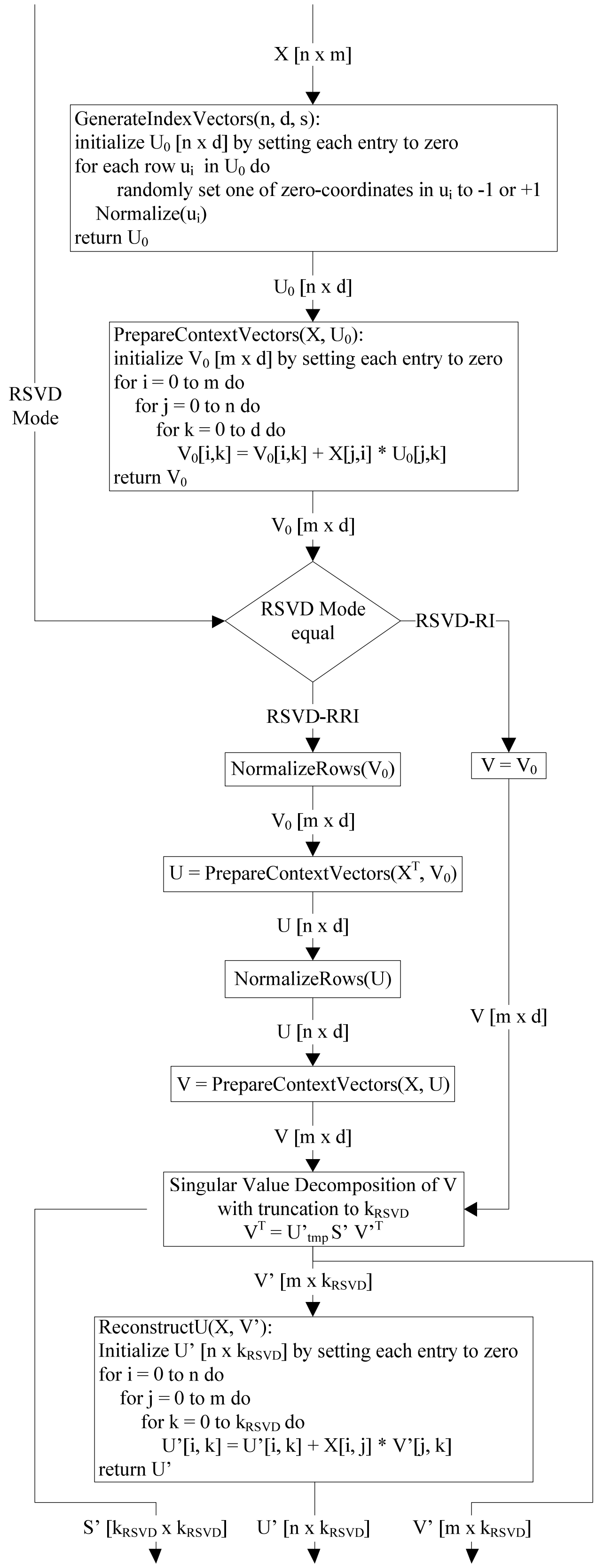

Figure 1. The algorithm of RSVD in both of its variants: RSVD-RI and RSVD-RRI.

- In order to evaluate quality of recommendations provided by a collaborative filtering system based on RSVD, the widely used F1 measure should be used: similarly to the authors of widely referenced papers [3][16] we follow the IR-based approach to recommendation accuracy evaluation by assuming that the aim of a recommender system is to recommend a small set of items that are the most likely to be attractive to a given user, rather than to provide accurate estimates of all items not rated by the user [17].

From the preliminary assumptions stated above, we have derived the following final assumptions for the RSVD evaluation methodology:

- RSVD should be evaluated from the perspective of the ability of obtaining decomposition results that are as close to SVD results as possible. The 'reference SVD results' that RSVD is to be compared with, should be specified as a small set of predefined configuration cases. Each of the reference SVD configurations should be specified by selecting the number of dimensions used for low-dimensional data representation. In all experiments RSVD is used as a substitute for SVD in the most straightforward way: in the case of RSVD all application configuration options or other parameters (other than the number of dimensions) are set exactly as in the case of SVD.
- In order to perform additional 'visual' evaluation of RSVD, the results of RSVD-based dimensionality reduction should be compared to analogical results of SVD application; the comparison should include RSVD statistics (in the form of histograms) presenting cosine similarities of column or row vectors – the vectors forming the low-dimensional user or item matrices (i.e. URSVD and VRSVD matrices) - should be compared to analogical statistics of SVD results
- As we basically target the collaborative filtering scenario, neither the measurement of RMSE nor the statistics showing vector similarities of U and V matrices can be used as the fully reliable and complete RSVD evaluation method. The mentioned methods should be regarded as supporting rather than constituting the main, recommendation-oriented RSVD accuracy evaluation procedure. An application-specific comparison of RSVD and SVD, accompanied by a computation times comparison should be provided as the most important results of the experimental RSVD evaluation.

In order to meet the assumptions above, we have followed the approach to experimental evaluation of a recommender system proposed in a widely cited paper by Sarwar et al [3]. In particular, we have chosen the F1 measure which is widely known in the field of IR [5] [17] and applicable in real recommender system application scenarios aimed at providing the best possible Top-N (in the case of our experiments Top-10) recommendations.

B. Recommendation Accuracy Evaluation

RI may be used to reconstruct an input matrix by calculating each of its entries as a dot product between a given index vector and a given context vector [Santosh]. On the other hand, RRI is much less suitable for the input matrix reconstruction task [7].

We have compared RSVD-based and RI-based systems to the frequently referenced SVD-based collaborative filtering algorithms presented in [3][18]. As rate estimation methods are usually able to achieve better results than kNN methods [2][15], we have compared RSVD-RI and RSVD-RRI to SVD from the perspective of their applicability to a rate-estimating system. All the methods were evaluated by using one of the most widely referenced CF data sets – the MovieLens ML100k set [3].

To evaluate the impact that the randomness of generating index vectors and choosing the appropriate test set has on the results of the presented methods, we accompanied each black curve representing a series of averaged values by a set of 8 non-averaged grey curves. Each of the non-averaged plots represents an individual case of an experiment result. The results of the comparison are presented in Fig. 2.

In general, the F1@10 results that we have obtained indicate that methods based on random vectors that have been proposed in the literature, i.e. RI and RRI, are not suitable for dimensionality reduction supporting collaborative filtering. In particular, the F1 values obtained for the RI-based rate estimation are lower than 0.1 and we have decided to exclude them from Fig. 2.

As we have confirmed experimentally, RSVD may be successfully used to obtain results that are arbitrarily similar, to or even better than, the SVD results. They are also much more accurate (in terms of F1) than the results obtained by RI-based methods proposed in [10] and [7]. Both RSVD variants feature high scalability, as the preliminary data dimensionality reduction [15] may be flexibly adapted to the amount of available computational resources (i.e., memory and processing power, which are directly dependent on parameter d).

Surprisingly, the trading recommendation accuracy for decreased memory consumption – a unique feature of both RSVD variants, which enable us to effectively cope with the well-known SVD bottleneck problem – does not lead to a degradation recommendation quality, when compared to the SVD-based method.

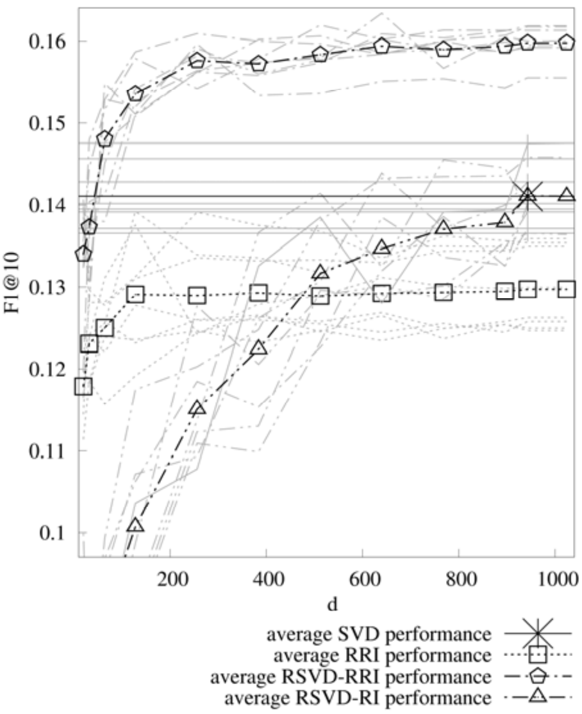

Figure 2. The trade-off between the recommendation quality and data dimensionality reduction: each series of F1 as a function of the dimensions number, for the case of k=6 and s=1.

It should be taken into account that in many real-world scenarios involving the decomposition of large matrices, the requirement of the full representation of input matrix in the memory (throughout the whole process of matrix decomposition) frequently makes large-scale SVD practically impossible [2][7].

The online computational complexity of a recommendation system depends on the number of dimensions used to represent items and users. From such perspective, RSVD may be regarded as a method that enables large-scale matrix decomposition, but which, in contrast to RI and RRI, still makes it possible to obtain low-dimensionality results that are relatively close to those provided by means of the regular SVD.

C. Relation Between Recommendation Accuracy and Selected Statistical Measures

We have investigated the way each of the selected algebraic properties of the methods under comparison is correlated with the recommendation accuracy. As we focus on the rate estimation approach to recommendation generation (more promising than kNN), we analyzed the results obtained with the use of the RSVD-based algorithm – F1@10 results [15] – from the perspective of two leading input matrix reconstruction error measures – Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) [17]. These measures have allowed us to statistically compare the real user-item ratings in the test set against the predicted values from the reconstructed matrix.

The results that we obtained are presented in Fig. 3. It is easy to see that RSVD-RI 'tries' to achieve a better result in terms of MAE, while RSVD-RRI minimizes the RMSE more effectively than MAE.

Moreover, in contrast to RSVD-RI (less effective in the recommendation task than RSVD-RRI), RSVD-RRI assures a nearly linear dependence not only between F1 and MAE, but also between F1 and MSE. Based on such an observation, one may conclude that, in general, the ability to minimize both MAE and RMSE at the same time appears to be an appropriate means for achieving high accuracy of RSVD-based recommendation.

IV. Algebraic Properties of RI-based Dimensionality Reduction Methods

Both RI and RRI may be used to reconstruct an input matrix by calculating each of its entries as a dot product between an appropriate index vector and an appropriate context vector. However, as indicated in our experiments, both RI and RRI are by far less suitable as means for the input matrix reconstruction than RSVD is [15].

We have compared RSVD-RI and RSVD-RRI with SVD from the perspective of algebraic properties that have a direct impact on the ability of successful input matrix reconstruction. For both RSVD variants the same configuration case has been presented (s=1, d=512).

From the algebraic point of view, three differences between SVD results and the corresponding results of RSVD that have a direct impact on the accuracy of RSVD are:

1) U'RSVD columns lack the full orthogonality.

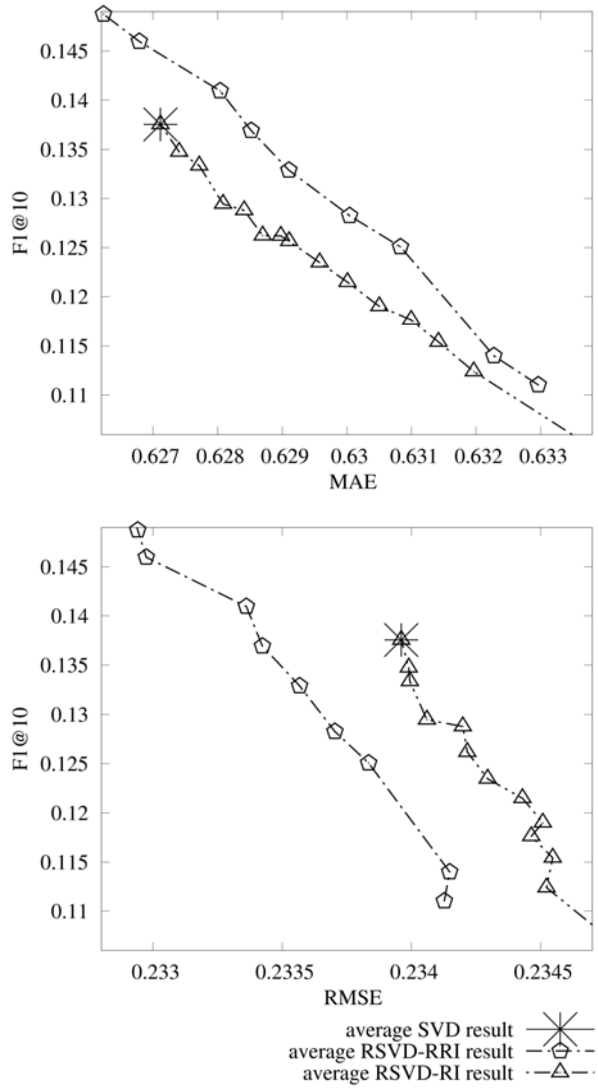

Figure 3. The correspondence between F1 and different input matrix reconstruction measures (MAE/RMSE) for RSVD-RI and RSVD-RRI, for the case of x=0.2 and a configuration optimal for x=0.2: k=6 and s=1.

2) 1. Orthogonality of the V' rows occurs at the number of dimensions reduced to d, instead of the original number of dimensions, i.e. at n. (i.e. instead of the full-dimensional orthogonality).
3) Concentration of RSVD-based matrix spectral components ('pseudo-singular values') on the most principal components is less effective than in the case of using SVD.

In order to analyze the mentioned impact of algebraic differences between SVD and RSVD on the accuracy of RSVD, we have gathered statistical data on cosine similarities of low-dimensional row vectors of V'k and analogical data on row vectors of U'k and U'k, where k=6. We measured how frequently the vector similarities belong to specified regular intervals. All pairs of non-zero row vectors of U'k and all pairs of non-zero row vectors of V'k were analyzed this way. The results are presented in Fig. 4 in the form of normalized histograms of Vk and Uk vector similarities.

The histograms are coupled by a diagram of singular values aimed at showing the similarity (in terms of distribution) of RSVD 'pseudo-singular values' (i.e. the diagonal values of SRSVD) to the 'true' singular values, i.e. the diagonal values of SRSVD for the case of x=1 and their squares. SVD results for the case x=1 (referred to as 'SVD, x=1') are presented on the same plot in order to visualize the full data set (that is the 'hidden' subject of the recommendation systems task).

As shown in Fig. 4, there are a few important differences between SVD properties and the corresponding properties of RSVD that have an impact on the magnitude of input matrix reconstruction errors. Rows of U'RSVD are seldom orthogonal to each other. Orthogonality of the V'RSVD rows occurs at the number of dimensions reduced to d, instead of the original number of dimensions, i.e. at n. (i.e. instead of the full-dimensional orthogonality).

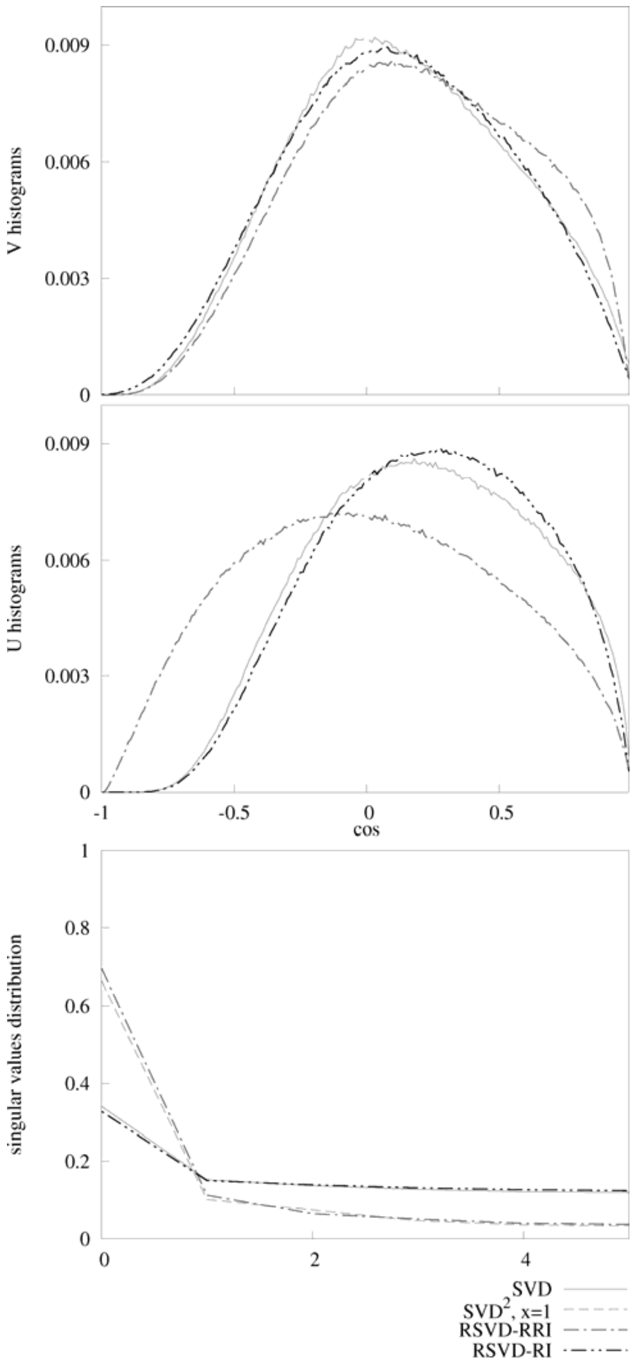

Figure 4. Selected algebraic properties of RSVD (RSVD-RI and RSVD-RRI) in comparison to SVD: V and U vectors' average similarity histograms and the corresponding distributions of singular values (diagonal values of S_SVD and S_RSVD matrices).

The concentration of RSVD-RI-decomposed matrix spectral components ('RSVD-RI pseudo-singular values') on the most principal components is slightly weaker than in the case of using SVD. On the other hand, the concentration of RSVD-RRI-decomposed matrix spectral components ('RSVD-RRI pseudo-singular values') on the most principal components is significantly stronger than in the case of using SVD – the singular values of a matrix pre-processed by means of RRI are very close to squared singular values of the original matrix.

Using each normalized histogram as a representation of a probability distribution function enables us to show how the diversification of user vector directions leads to a higher information entropy of cosine similarities (within the low-dimensional user vector set) than that observed when SVD or RSVD-RI are used. On the other hand, the application of RSVD-RRI makes a distribution of singular values 'less informative', i.e. less influenced by dimensionality reduction.

V. CONCLUSIONS

The results of the experiments presented in this paper allow us to conclude that, at least as far as collaborative filtering accuracy is concerned, the basic RI-based dimensionality reduction methods (RI and RRI) presented in the literature (as components of TR systems) are significantly less effective than SVD. A simple substitution of SVD by RI or RRI leads to a severe recommendation quality degradation. On the other, hand – when appropriately combined with SVD – both RI and RRI feature several interesting properties. Such a combination – referred to as RSVD – is potentially much more suitable for collaborative filtering than the most widely used methods based on standard SVD.

As we have shown, appropriately post-processed results of RI (index vectors and context vectors) may be made very similar to low-dimensional SVD results. After being normalized and orthogonalized (by means of low-dimensional SVD), the vectors forming the basis of the context vectors space (e.g. a space of row vectors of the input matrix) may be used together with the input matrix in order to produce useful approximations of both the corresponding 'new index vectors' (e.g. representing column vectors of the input matrix) and the 'pseudo-singular values'. Matrices that are obtained with the use of this method (referred to as the RSVD-RI method) may serve as an arbitrarily precise approximation of SVD-produced matrices (the precision directly depends on the random vectors dimensionality).

The execution times (single-threaded, on a 'low-end' PC) for the compared methods on the ML1M dataset have been presented in Fig. 5. As it is has been shown, the RI-based methods allow to significantly reduce the computational complexity.

Apart from the ability of SVD approximation scalability described above, the results of our experiments show that the application of RI-based input matrix pre-processing (more precisely: based on RRI) may – rather counter intuitively – lead to an increase in the recommendation quality, despite the lack of the Frobenius norm optimality. RSVD-RRI makes it possible to outperform the widely referenced SVD-based collaborative filtering algorithms not only in terms of computational efficiency (a feature typical of any RI-based dimensionality reduction), but also in terms of recommendation accuracy [15].

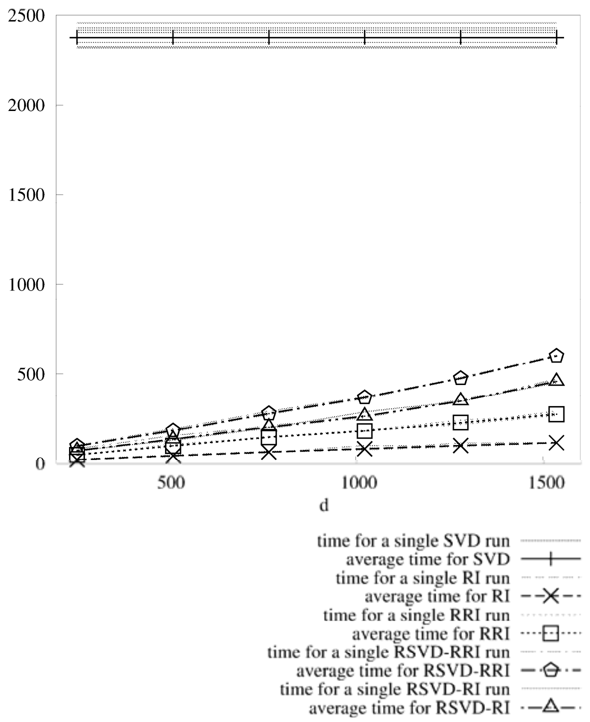

Figure 5. Execution times(in seconds) of the compared matrix decomposition methods (SVD and the three methods based on RI – RSVD, RI and RRI) used on the ML1M dataset.

In order to put some light on this phenomenon, we have compared RSVD-RRI to SVD (as well as to RSVD-RI, serving as an approximated substitute for SVD) from the perspective of several algebraic properties. We have shown that, thanks to using RSVD, cosine similarities of low-rank user vectors are statistically more ‘meaningful’ (less concentrated around zero) than the analogical similarities obtained when SVD is used. When modeled probabilistically (with each normalized histogram representing a probability distribution function), such diversification of user vector directions may be understood as leading to an increase in information entropy.

Moreover, our experiments show that the ability to minimize both MAE and RMSE at the same time appears to be an appropriate means for achieving a high F1-measured accuracy of the RSVD-based recommendation.

References

[1] G. Adomavicius and A. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions", IEEE Transactions on Knowledge and Data Engineering 17 (6) 2005, 734–749.

[2] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems", IEEE Computer, 42(8):30–37, 2009.

[3] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Riedl. "Application of dimensionality reduction in recommender system – a case study", In ACM WebKDD 2000 Web Mining for E-Commerce Workshop, Boston, MA, USA, 2000.

[4] J.B. Schafer, J. A. Konstan, and J. Riedl, “E-Commerce Recommendation Applications”, Data Mining and Knowledge Discovery, vol. 5, nos. 1/2, pp. 115–153, 2001.

[5] C. D. Manning, P. Raghavan, and H. Schütze. An "Introduction to Information Retrieval", Cambridge University Press, 2008.

[6] Yehuda Koren, Robert Bell Chris Volinsky, Matrix factorization techniques for recommender systems. IEEE Computer, 42(8):30–37, 2009.

[7] T. Cohen, R. Schaneveldt, and D. Widdows, “Reflective Random Indexing and Indirect Inference: A Scalable Method for Discovery of Implicit Connections”, Journal of Biomedical Informatics, 2009.

[8] V. Vasuki and T. Cohen, "Reflective random indexing for semi-automatic indexing of the biomedical literature". Journal of biomedical informatics, 2010.

[9] P. Kanerva, J. Kristoferson, and A. Holst. "Random Indexing of text samples for latent semantic analysis", In L.R. Gleitman and A.K. Josh (eds.), Proc. 22nd Annual Conference of the Cognitive Science Society (U Pennsylvania), p. 1036. Mahwah, New Jersey: Erlbaum, 2000.

[10] M. Sahlgren. "An introduction to random indexing. In: Proceedings of the methods and applications of semantic indexing workshop at the 7th international conference on terminology and knowledge engineering (TKE)", Copenhagen, Denmark, 2005.

[11] T. Cohen and D. Widdows. "Empirical distributional semantics: Methods and biomedical applications", JBI, vol. 42, no. 2, pp. 390–405, 2009.

[12] D. I. Martin and M. W. Berry, "Mathematical foundations behind latent semantic analysis". In: L.T. McNamara and S.W. Sennis, Editors, Handbook of latent semantic analysis, Lawrence Erlbaum Associates, 2007.

[13] Z. Huang, D. Zeng, and H. Chen, "A Comparison of Collaborative-Filtering Recommendation Algorithms for E-commerce", IEEE Intelligent Systems, vol. 22, Issue 5, pp. 68–78, New Jersey, 2007.

[14] H. Yan, W. I. Grosky, and F. Fotouhi. "Augmenting the power of LSI in text retrieval: Singular value rescaling", Data Knowl. Eng., vol. 64, pp. 108–125, Amsterdam, 2008.

[15] M. Ciesielczyk, A. Szwabe, and B. Prus-Zajaczkowski. "Interactive Collaborative Filtering with RI-based Approximation of SVD", in Procs. of 3rd International Conference on Computational Intelligence and Industrial Application, in press, IEEE Press, 2010.

[16] P. Symeonidis, A. Nanopoulos, A. N. Papadopoulos, and Y. Manolopoulos. Collaborative recommender systems: Combining efectiveness and eciency. Expert Syst. Appl., 34(4):2995-3013, 2008.

[17] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, “Evaluating Collaborative Filtering Recommender Systems,” ACM Trans. Information Systems, vol. 22, no. 1, pp. 5–53, 2004.

[18] Santosh S. Vempala, The Random Projection Method, American Mathematical Society, Series in Discrete Mathematics and Theoretical Computer Science, Vol. 65, 2004.

[19] M. Rashid, S. K. Lam, G. Karypis, and J. Riedl, "ClustKNN: A Highly Scalable Hybrid Model&MemoryBased CF Algorithm", Computer Science and Engineering, University of Minnesota, Minneapolis.