Chapter 4
Experimental Evaluation on Matrix Decomposition Methods

Abstract In this chapter, we study the performance of described SVD and UV decomposition algorithms, against an improved version of the original item based CF algorithm combined with SVD. For the UV decomposition method, we will present the appropriate tuning of parameters of its objective function to have an idea of how we can get optimized values of its parameters. We will also answer the question if these values are generally accepted or they should be different for each data set. The metrics we will use are root-mean-square error (RMSE), precision, and recall. The size of a training set is fixed at 75 %, and we perform a fourfold cross-validation.

Keywords Experiments · SVD decomposition · UV decomposition

4.1 Data Sets

We perform experiments with two real data sets that have been used as benchmarks in prior work.

The first data set has been extracted from the www.epinions.com website (http://www.epinions.com). Epinions is a website that contains reviews of users on items such as electronics, movies, books, music, etc. Users also build their web of trust within the Epinions community. This web of trust is a list of trusted and distrusted users. Notice that trusted users' reviews are promoted, whereas distrusted users' reviews are less likely encountered. A review contains a rating between 1 and 5 and a free text message. Reviews can be commented and/or rated. This data set contains 131.828 users who have rated 841.372 items.

P. Symeonidis and A. Zioupos, *Matrix and Tensor Factorization Techniques for Recommender Systems*, SpringerBriefs in Computer Science,
DOI 10.1007/978-3-319-41357-0_4

The second data set has been extracted from the following GeoSocialRec website (http://delab.csd.auth.gr/geosocialrec), which is a location-based social network. We have chosen a small data set to test our algorithm in a case of severe sparseness and lack of user's data. The data set consists of 212 users who rated 649 locations. GeoSocialRec offers to its users the ability to connect to each other, to declare their position with a check-in, and to rate the places they have visited.

4.2 Sensitivity Analysis of the UV Decomposition Algorithm

4.2.1 Tuning of the k Latent Feature Space

In this section, we will examine how parameter k affects the effectiveness and performance of our algorithm. Recall that parameter k controls the size of U and V matrices which store user-latent features and item-latent features, respectively. Thus, a small number of k latent features mean that we keep both matrices thin, whereas the required storage space is small. To tune parameter k, we fix the number of algorithm's iteration at 5.000, which is the number of times that our algorithm is applied on data to predict ratings. The procedure goes like this: On the first algorithm's iteration, we predict missing ratings and store them in the training set. On the second algorithm's iteration, we repredict ratings using the information stored in the previous step and so on.

For the Epinions data set, as shown in Fig. 4.1, RMSE decreases with increasing values of parameter k. The best RMSE is attained when parameter k is equal to 16 latent features.

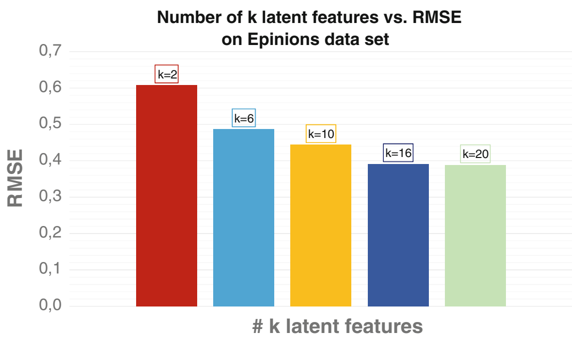

Fig. 4.1 RMSE versus different number of k latent features on the Epinions data set

In contrast, for the GeoSocialRec data set, as shown in Fig. 4.2, the best RMSE is attained for $k = 2$. That is, for different data sets, there can be no rule for selecting the optimal value of parameter k, as it depends on the size of the data set, the data sparseness, and probably other factors, which may be particular for each data set.

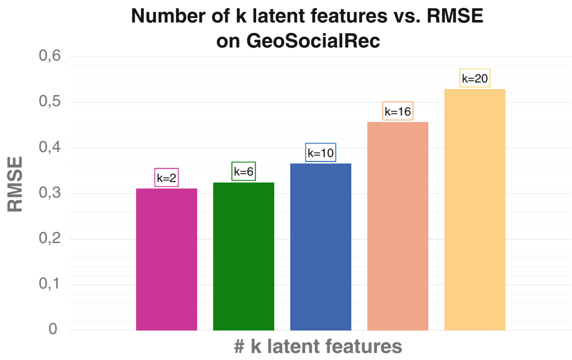

Fig. 4.2 RMSE versus different number of k latent features on the GeoSocialRec data set

4.2.2 *Tuning of Parameter* β

Parameter β is responsible to control the magnitudes of item-latent features and user-latent features of U and V matrices, respectively. Parameter β overcomes the problem of overfitting. For both data sets, as shown in Figs. 4.3 and 4.4, RMSE drops slightly with decreasing values of the parameter β. In particular, with smaller values of β, we get an improvement of 20 % in the Epinions data set and only 2 % in GeoSocialRec data set. As expected, data overfitting does not exist in the GeoSocialRec data set, since there is not enough training data to bias our algorithm. In contrast, parameter β is important in the case of the Epinions data set, where it improves drastically the performance of our algorithm.

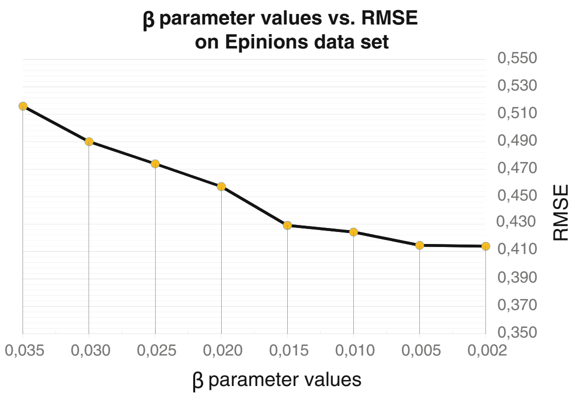

Fig. 4.3 RMSE versus different values of parameter β on the Epinions data set

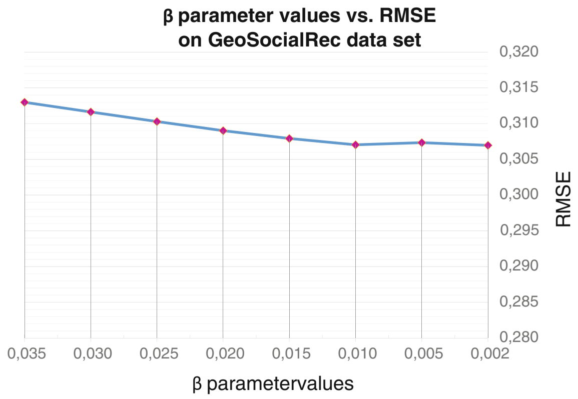

Fig. 4.4 RMSE versus different values of parameter β on the GeoSocialRec data set

4.2.3 *Optimizing Algorithm's Parameters*

In this section, we present the algorithm's performance with default values versus the performance after tuning the parameters of the objective function.

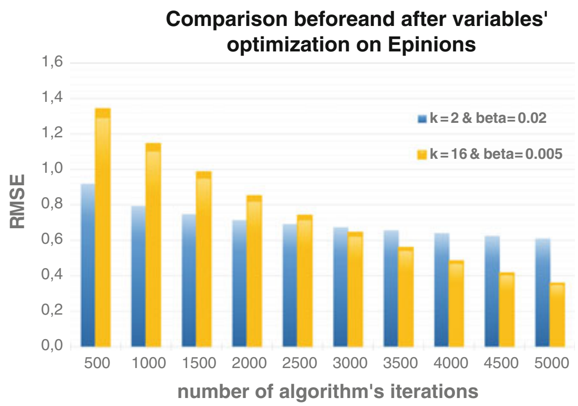

Fig. 4.5 RMSE results using default versus optimized parameters of the objective function on the GeoSocialRec

For the Epinions data set, as shown in Fig. 4.5, blue bars present RMSE levels calculated with the initial algorithm's parameter values, whereas yellow bars present RMSE levels after optimizing the parameters of the algorithm's objective function. As expected, as algorithm's iterations for rating predictions are increased, RMSE of optimized parameters outperforms the performance of initial values. Please notice that the improvement of optimized parameters over initial variables is quite significant for this data set. The main reason is that there is enough information in the training data so that the UV decomposition algorithm can make accurate "generalizations" for the test data.

For the GeoSocialRec data set, as shown in Fig. 4.6, improvement of RMSE on GeoSocialRec data set is not significant. The main reason is that this data set is very small and there is not enough information in the training set.

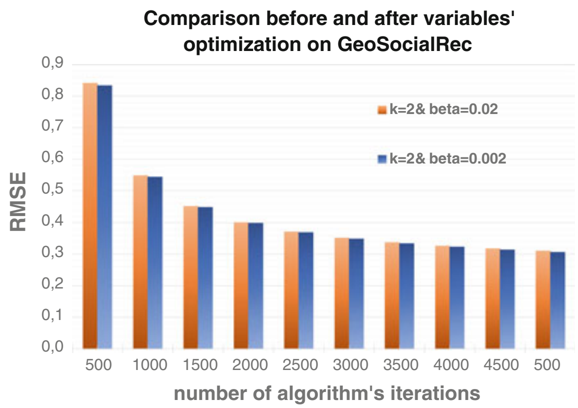

Fig. 4.6 RMSE results using default versus optimized parameters of the objective function on the GeoSocialRec

4.3 Comparison to Other Decomposition Methods

In this section, we compare the UV decomposition algorithm (with parameters that attained the best performance in previous experiments) against the following methods:

- CUR-decomposition algorithm, which confronts the problem of high density in the factorized matrices (a problem that is faced mainly when using the SVD decomposition method) [1, 3, 4]. This algorithm is denoted as CUR.
- item-based CF combined with SVD [6]. Item-based CF is an improved version [2] of the well-known item-based CF algorithm that weights similarities by the number of common ratings among items. This variation of item-based CF weights the similarity sim between two items with a parameter γ, as follows: $\frac{\max(c,\gamma)}{\gamma} \cdot sim$, where c is the number of co-rated users. The best value of parameter γ is fixed at 4 and 2 for the Epinions and the GeoSocialRec data set, respectively. This algorithm is denoted as item-based SVD.

We will compare all three algorithms in terms of precision, recall, and RMSE measures. Please notice that RMSE works well for measuring how accurately an algorithm predicts the rating of a randomly selected item, but may fail to evaluate whether an algorithm will provide accurate item recommendations [5]. Therefore, precision–recall evaluation measures, in addition to RMSE, are able to better measure the quality of recommendations. Moreover, we will use precision–recall diagrams because they can reveal the robustness of each algorithm in attaining high recall with minimal losses in terms of precision. We examine the top-N ranked item list, which is recommended to a target user, starting from the top item. In this case, recall and precision vary as we proceed with the examination of the top-N list of recommended items.

For the Epinions data set, in Fig. 4.7a, we plot a precision versus a recall curve for all three algorithms. As expected, all algorithms' precision falls as N increases. In contrast, as N increases, recall for all algorithms increases as well. The UV decomposition algorithm attains the best results.

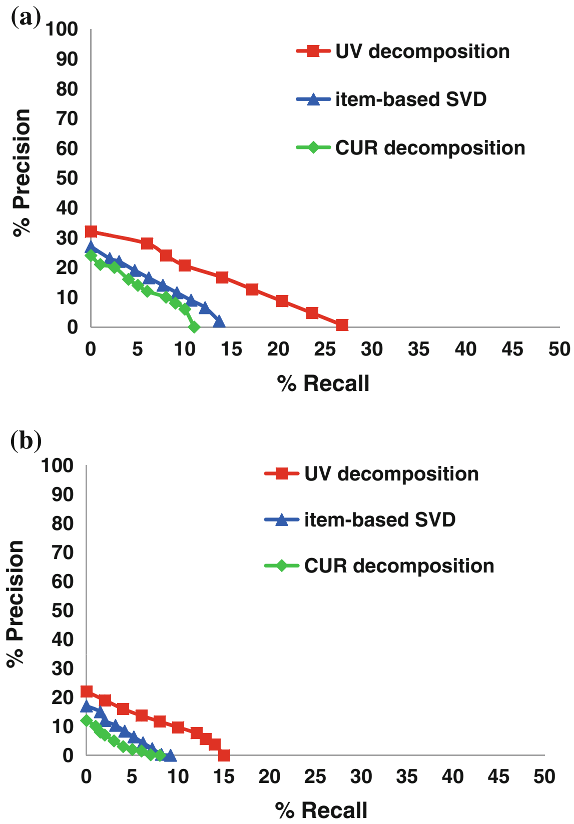

Fig. 4.7 Accuracy performance of algorithms in terms of precision–recall for the **a** Epinions and **b** GeoSocialRec data sets

For the GeoSocialRec data set, in Fig. 4.7b we also plot a precision versus recall diagram. The UV decomposition algorithm again outperforms the item-based SVD and CUR algorithms. Notice that the accuracy performance of all algorithms for the GeoSocialRec data set is lower than those for the Epinions data set. The reason is possibly because the latter has more ratings per user and can be considered a more dense data set.

Next, we measured RMSE for all three examined algorithms on the Epinions and GeoSocialRec data sets. The results are summarized in Table 4.1. Again, UV decomposition clearly outperforms the item-based SVD and CUR algorithms in terms of RMSE. As shown, UV decomposition attains the lowest RMSE values and item-based SVD is the second best algorithm in both data sets.

Table 4.1 RMSE values for all three algorithms on two real data sets

Algorithm	Epinions data set	GeoSocialRec data set
UV decomposition	0.38	0.32
Item-based SVD	0.76	0.98
CUR decomposition	0.79	1.08

A smaller value means a better performance

In conclusion, we have to notice that good results of RMSE may not fully characterize users' experience in web-based recommender systems (Amazon, eBay, etc.), which propose a top-N ranked list of items to a user. The basic reason is that an error of size ϵ has the same impact on RMSE regardless of where that error places the item in a top-N ranking. In contrast, the results of precision–recall can better characterize the users' experience in the aforementioned systems.

References

1. Drineas, P., Kannan, R., Mahoney, M.W.: Fast monte carlo algorithms for matrices III: computing a compressed approximate matrix decomposition. SIAM J. Comput. **36**(1), 184–206 (2006)
2. Herlocker, J., Konstan, J., Riedl, J.: An empirical analysis of design choices in neighborhood-based collaborative filtering algorithms. Inf. Retrieval **5**(4), 287–310 (2002)
3. Mahoney, M.W., Drineas, P.: Cur matrix decompositions for improved data analysis. Proc. Natl. Acad. Sci. **106**(3), 697–702 (2009)
4. Mahoney, M.W., Maggioni, M., Drineas, P.: Tensor-cur decompositions for tensor-based data. SIAM J. Matrix Anal. Appl. **30**(3), 957–987 (2008)
5. McLauglin, R., Herlocher, J.: A collaborative filtering algorithm and evaluation metric that accurately model the user experience. In: Proceedings of ACM SIGIR Conference, pp. 329–336 (2004)
6. Symeonidis, P., Nanopoulos, A., Papadopoulos, A., Manolopoulos, Y.: Collaborative recommender systems: combining effectiveness and efficiency. Expert Syst. Appl. **34**(4), 2995–3013 (2008)