

Leveraging Tagging for Neighborhood-aware Probabilistic Matrix Factorization

Le Wu¹, Enhong Chen², Qi Liu¹, Linli Xu², Tengfei Bao¹, Lei Zhang¹

^{1 2} School of Computer Science and Technology, University of Science and Technology of China
E-mail: ¹{wule,feiniaol,tfbao92,stone}@mail.ustc.edu.cn, ²{cheneh,linlixu}@ustc.edu.cn

ABSTRACT

Collaborative Filtering (CF) is a popular way to build recommender systems and has been successfully employed in many applications. Generally, two kinds of approaches to CF, the *local* neighborhood methods and the *global* matrix factorization models, have been widely studied. Though some previous researches target on combining the complementary advantages of both approaches, the performance is still limited due to the *extreme sparsity of the rating data*. Therefore, it is necessary to consider more information for better reflecting user preference and item content. To that end, in this paper, by leveraging the extra tagging data, we propose a novel unified two-stage recommendation framework, named **Neighborhood-aware Probabilistic Matrix Factorization (NHPMF)**. Specifically, we first use the tagging data to select neighbors of each user and each item, then add unique Gaussian distributions on each user's (item's) latent feature vector in the matrix factorization to ensure *similar users (items) will have similar latent features*. Since the proposed method can effectively explore the external data source (i.e., tagging data) in a unified probabilistic model, it leads to more accurate recommendations. Extensive experimental results on two real world datasets demonstrate that our NHPMF model outperforms the state-of-the-art methods.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval—*Information Filtering*

Keywords

Collaborative Filtering, Matrix factorization, Neighborhood method

1. INTRODUCTION

Collaborative filtering (CF) is a popular way to build recommender system [1] and has received significant success in various applications. It infers the interest of active users by collecting past behaviors (e.g. browsing history, click streams or products satisfaction expressed in ratings) from similar users or items without requiring the creation of explicit user and item profiles.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CIKM'12, October 29–November 2, 2012, Maui, HI, USA.
Copyright 2012 ACM 978-1-4503-1156-4/12/10 ...\$15.00.

Generally, there are two kinds of successful approaches to CF: neighborhood methods and matrix factorization models. Neighborhood methods (also known as K Nearest Neighbor, or KNN) can be further classified into two categories: user-oriented KNN method (UKNN) [6], and item-oriented KNN method (IKNN) [12]. They produce recommendations from like-minded users (UKNN) or similar items (IKNN). In contrast to neighborhood methods, Matrix Factorization (MF) approaches [9, 11] try to characterize both items and users in the same low latent space inferred from the sparse rating matrix.

In fact, these two kinds of CF approaches focus on exploiting different levels of structure information from the data. Neighborhood methods are most effective at detecting very localized relationships (the set of neighbors) to make predictions. Consequently, they are unable to capture the global weak signals encompassed in all of a user's ratings. On the other hand, matrix factorization models are generally effective at estimating the overall structure of the explicit rating matrix. However, these models are poor at detecting strong associations among a small set of closely related neighbor set [2, 8]. Obviously, there are complementary advantages for local neighborhood methods and the global matrix factorization models, and for making better recommendation it is necessary to combine the local information with the global structure of the data.

Though several unified frameworks have been proposed to combine KNN and MF models [3, 8, 13], the performance is still limited due to the extreme sparsity of the available data. As a matter of fact, in real world scenarios, the user rating data is very sparse and the density of observed ratings is often less than 1% [12]. Thus, it is hard for rating based KNNs to find the credible nearest neighbors, since they assume users have at least rated some items in common (UKNN) or two items have been co-rated by some users (IKNN). Meanwhile, relying only on the sparse rating data will lead to the over-specification problem for MFs [1]. Luckily, with the popularity of Web 2.0, more external information and metadata have emerged on the web, such as the tagging applications in the MovieLens website¹. These applications allow users to add free, personalized *tags* to the items. Recently, studies in [10] have made preliminary attempts to analyze the utility of these tags and found that user-generated tags are consistent with the web content and can capture the topics of user interest quite well. Therefore, it would be a very worthwhile endeavor to *use tagging data in finding local information for collaborative filtering, especially for those unified CFs*.

Along this line, in this paper, we propose a novel two-stage CF framework named **Neighborhood-aware Probabilistic Matrix Factorization (NHPMF)** to improve recommendation accuracy. In the first stage, we utilize the tagging data to obtain the neighbors

¹<http://www.movielens.umn.edu/>

of each user and each item. In the second stage, those obtained neighborhood information is incorporated into the factorization of the rating matrix to ensure *similar users(items) will have similar latent features*. To be specific, the combination is achieved by using local neighborhood information in the priors of the matrix factorization model, where the latent feature vectors of each user(item) are dependent on those of her(its) neighbor set. To the best of our knowledge, NHPMF is one of the few attempts to explore external tagging data for better incorporating both user neighborhood and item neighborhood into matrix factorization by a unified probabilistic framework. Extensive experimental results on two real world datasets show that our model has better performance than the state-of-the-art methods, especially when the active user has only a few rating records.

2. RELATED WORK

In this section, we review the literatures related to our work from the following two aspects.

Combining KNN and MF Although KNN and MF methods have been widely studied in CF respectively, only several researches have focused on combining these two methods to improve the recommendation accuracy [2, 3, 5, 8, 13]. During the Netflix competition, researchers have found that there are complementary advantages for both local neighborhood methods and global matrix factorization models and combining these two methods' results will improve recommendation accuracy [2]. Researches in [3, 13] have integrated those two approaches, but they suggested using neighborhood methods to post-process MF results rather than a unified model. Koren in [8] proposed to combine IKNN and MF into the regularized least squares problem, where item similarities and the latent factors are learned from the data through gradient descent. The experimental results on Netflix data set showed the prediction accuracy is improved by their model. Gu et. al [5] proposed a unified model for CF based on graph regularized weighted nonnegative matrix factorization. They used user demographic and item genre information to constructed neighborhood graphs and incorporated user and item graphs in weighted nonnegative matrix factorization. Nevertheless, in the real world, it's hard to collect the content information of users and items due to the expensive expert labeling efforts and privacy concerns. Our model is more intuitive and general since users take initiative to annotate web resources, which makes it easier to collect tagging data.

Besides those mentioned above, with the advent of online social networks in recent years, the social network based recommendation has emerged. Some work in this area, such as [7], has combined UKNN and MF for recommendation, where user neighborhood information is learned from the social structure in the social network. Our work has essential difference with their methods since they only focused on combing social friends information into MF. Our work emphasizes to incorporate the UKNN and IKNN together into MF, where the neighborhood information can be learned from rich external data sources (e.g., tagging history, social relationships, user demographic information, item contents).

Utilizing Tagging Data for CF With the development of Web 2.0 applications, tagging techniques have become a popular tool for users to semantically describe web resources. Researchers in [10] have found that tags are good at representing users' interests and thus reflect the web content quite well. Recently, people have started to consider utilizing tagging data to enhance CF based recommendation accuracies [14, 15]. TagRec is introduced in [15], where a factor analysis model is proposed based on a unified probabilistic matrix factorization by utilizing both users' tagging and rat-

ing information. Further, in [14], a regularization term is added to the objective function of probabilistic matrix factorization to make two user latent feature vectors as similar as possible if they have similar tagging history.

In summary, to the best of our knowledge, none of the previous work has unified the complementary advantages of both local user neighborhood, item neighborhood and global matrix factorization models together in the CF in a unified probabilistic model, where the local information is learned by exploiting other data sources.

3. NEIGHBORHOOD-AWARE PROBABILISTIC MATRIX FACTORIZATION

The main challenge for our model is the way to incorporate the local neighborhood information into the global matrix factorization of the rating matrix, where the neighborhood information is learned from external data sources. Since previous research work have found that tags can represent user interests and reflect the web content quite accurately [10], it is natural to select neighbors of users and items by the tagging data. In addition, in the real world, the interests and behavior of user i is similar to that of her neighbor set N_{U_i} , and the attributes of item j are similar to those of its neighbor set N_{V_j} . Based on this intuition, in the matrix factorization process we should make sure that *similar users(items) will have similar latent feature vectors*, and in this way the external neighborhood information is incorporated in the global MF of the rating matrix.

Motivated by the above analysis, we propose a novel two-stage framework, **Neighborhood-aware Probabilistic Matrix Factorization (NHPMF)**, to improve recommendation accuracy. In the first stage, the user-tag and item-tag interaction matrixes are used to select the neighbor sets of users and items respectively. In the second stage, we add Gaussian priors on users' and items' latent feature vectors to ensure each user(item)'s latent feature vector is centered around the mean of her(its) neighborhood. The priors will lead to a corresponding regularization term that penalizes the latent vector divergence of each user(item) from her(its) neighbor set. Table 1 lists the mathematical notations used in this paper.

Table 1: Mathematical notations

Notation	Description
M, N, L	# of users, items and tags respectively
D	Dimension of latent feature representation
R_{M*N}	User-item rating matrix
P_{M*L}	User-tag interaction matrix
Q_{N*L}	Item-tag interaction matrix
U_{D*M}	User's latent feature matrix
V_{D*N}	Item's latent feature matrix
N_{U_i}	The set of neighbors of user i
N_{V_j}	The set of neighbors of item j
\hat{R}_{M*N}	Approximation of the rating matrix R_{M*N}

3.1 Neighborhood Selection

As stated earlier, it is natural to select neighbors of each user from the user-tag interaction matrix, where P_{il} is the *tf*idf* weight of user i to tag l : $P_{il} = tf(i, l) \times \ln(\frac{M}{df(l)})$. Then the similarity $s(i, j)$ between user i and j , can be measured by computing the cosine similarity of the angle between the two users in the L dimensional tag space:

$$s(i, j) = \cos(\vec{i}, \vec{j}) = \frac{\vec{i} \cdot \vec{j}}{\|\vec{i}\|_F * \|\vec{j}\|_F} \quad (1)$$

After calculating similarities between users, it is easy to select the K nearest neighbors of user i , N_{U_i} . Identical steps can be used to characterize the similarity value $t(j, k)$ between each pair

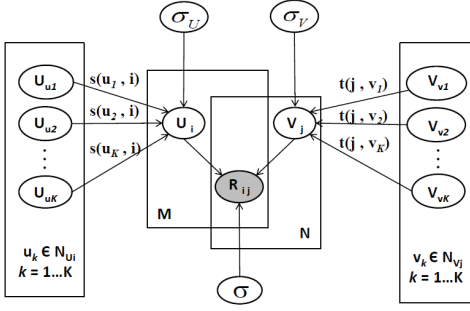


Figure 1: Graphical representation of NHPMF Framework. For each user i and item j in the recommender, her(its) latent features not only affect the rating matrix, but are also influenced by her(its) neighborhood set.

of item j and k , and the K nearest neighbors of item j are denoted N_{V_j} . For notational convenience in later chapters, we normalize the similarities between each user i and her neighborhood N_{U_i} , and those between each item j and its neighborhood N_{V_j} to ensure $\sum_{l \in N_{U_i}} s(i, l) = 1$ and $\sum_{l \in N_{V_j}} t(j, l) = 1$.

3.2 Neighborhood-aware Probabilistic Matrix Factorization

Naturally, the behavior of user i is similar to that of her neighbor set N_{U_i} and the attributes of item j are similar to those of its neighbor set N_{V_j} . Based on this intuition, we formulate the following equations:

$$U_i = \sum_{l \in N_{U_i}} s(i, l) * U_l + \theta_U, \quad \theta_U \sim N(0, \sigma_U^2 \mathbf{I}) \quad (2)$$

$$V_j = \sum_{l \in N_{V_j}} t(j, l) * V_l + \theta_V, \quad \theta_V \sim N(0, \sigma_V^2 \mathbf{I}) \quad (3)$$

In the above two equations, each user's and item's latent feature vector is composed of two terms. The first term characterizes the group feature of the user(item), which is the weighted average of her(its) neighborhood. The second term emphasizes the uniqueness of each user and item feature vector, which could diverge from her(its) neighborhood to an extent. The divergence is controlled by the variance parameter σ_U^2 and σ_V^2 in Eq. (2) and Eq. (3). The smaller the variance, the less possible that the feature vector diverges from that of her(its) neighbor set. When the variance approximates zeros, the second term vanishes. Using simple mathematical transformations, Eq. (2) and Eq. (3) will turn to the following two equations:

$$p(U|S, \sigma_U^2) = \prod_{i=1}^M N\left(\sum_{l \in N_{U_i}} s(i, l) * U_l, \sigma_U^2 \mathbf{I}\right) \quad (4)$$

$$p(V|T, \sigma_V^2) = \prod_{j=1}^N N\left(\sum_{l \in N_{V_j}} t(j, l) * V_l, \sigma_V^2 \mathbf{I}\right) \quad (5)$$

Taking the neighborhood information into account does not change the conditional distributions of the observed ratings given the user and item latent feature vectors. Thus, the conditional distribution over the observed rating is:

$$p(R|U, V, \sigma^2) = \prod_{i=1}^M \prod_{j=1}^N [\mathcal{N}(R_{ij}|U_i^T V_j, \sigma^2)]^{Y_{ij}} \quad (6)$$

According to Eq. (4), Eq. (5) and Eq. (6), the corresponding graphical representation of our model is shown in Figure 1. Then through Bayesian inference, the log of the posterior distribution over user and item latent factors is given by Eq. (7):

$$p(U, V|R, \sigma^2, \sigma_U^2, \sigma_V^2) \propto p(U|S, \sigma_U^2) * p(V|T, \sigma_V^2) * p(R|U, V, \sigma^2) \quad (7)$$

Keeping the hyperparameters (σ^2, σ_U^2 and σ_V^2) fixed, maximizing the log posterior in Eq. (7) is equivalent to minimizing the following sum-of-squared cost function.

$$\begin{aligned} E &= \frac{1}{2} \sum_{i=1}^M \sum_{j=1}^N Y_{ij} (R_{ij} - U_i^T V_j)^2 \\ &+ \frac{1}{2} \lambda_U \sum_{i=1}^M \|(U_i - \sum_{l \in N_{U_i}} s(i, l) * U_l)\|_F^2 \\ &+ \frac{1}{2} \lambda_V \sum_{j=1}^N \|(V_j - \sum_{l \in N_{V_j}} t(j, l) * V_l)\|_F^2 \end{aligned} \quad (8)$$

In the above equation, $\lambda_U = \sigma^2 / \sigma_U^2$, $\lambda_V = \sigma^2 / \sigma_V^2$. The objective function has three terms and it is smoothed by the parameter λ_U and λ_V , which naturally fuses local neighborhood information with global matrix factorization in recommender systems. The parameter λ_U controls how much the user neighborhood influences while λ_V controls how much the item neighborhood influences on the error function. With smaller values of λ_U and λ_V , we rely less on neighborhood information.

A local minimum of the objective function Eq. (8) can be found by performing gradient descent on U_i and V_j for each user i and each item j given the derivatives below.

$$\begin{aligned} \frac{\partial E}{\partial U_i} &= \sum_{j=1}^N (R_{ij} - U_i^T V_j)(-V_j) + \lambda_U (U_i - \sum_{l \in N_{U_i}} s(i, l) * U_l) \\ &- \lambda_U \sum_{i \in N_{U_i}} s(i, i)(U_i - \sum_{j \in N_{U_i}} s(j, l) * U_j) \end{aligned} \quad (9)$$

$$\begin{aligned} \frac{\partial E}{\partial V_j} &= \sum_{i=1}^M (R_{ij} - U_i^T V_j)(-U_i) + \lambda_V (V_j - \sum_{l \in N_{V_j}} t(j, l) * V_l) \\ &- \lambda_V \sum_{j \in N_{V_i}} t(l, j)(V_i - \sum_{i \in N_{V_i}} t(i, l) * V_i) \end{aligned} \quad (10)$$

3.3 Discussion on NHPMF

The NHPMF model incorporates the tag-based neighborhood information into matrix factorization for recommendation. The intuition behind this method is that the latent feature vector of each user(item) is similar to those of her(its) neighbors'. Eq. (4) and Eq. (5) ensure that the latent factor of each user(item) is composed of two parts, one is a weighted sum of her(its) neighbors' characters, and the other is her(its) own characters. Thus in NHPMF, each user's and each item's latent feature vector has a unique Gaussian distribution, where its mean is learned from the rating data and dependent on her(its) neighbors. Therefore, NHPMF is different from related PMF model [11], where it assumes all users and items obey the same zero mean Gaussian distribution, which is too general and rough to capture user's and item's personality when applied to real world applications.

Finally, to get deeper insights, we observe that our NHPMF model is actually a unified framework of KNN and PMF. Under certain circumstances, it degenerates to the following three methods:

- when $\forall i, j \in \text{userset} : s(i, j) = 0$,
 $\forall i, j \in \text{itemset} : t(i, j) = 0$;

That is to say, we have no other external data source to extract the neighbors of users and items. According to Eq. (4) and Eq. (5), users and items' feature vectors turn to be the zero-mean Gaussian prior. The model degenerates to the PMF model [11].

- when $\forall i, j \in \text{userset} : s(i, j) = 0$;

In this case, we have no external data source to extract the neighbors of users, but we can calculate the similarities between items. This scenario is common in online shopping websites. The attributes of items are carefully described by

human efforts for better presenting to consumers, but it is hard to get the user profiles due to privacy concerns. Under this situation, user feature vectors obey a common zero-mean Gaussian distribution. While the feature vector of each item still has a unique Gaussian distribution with its mean centered around its neighbors. The model turns out to be the fusion of IKNN and PMF.

- when $\forall i, j \in \text{itemset} : t(i, j) = 0$;
This scenario is similar to the above one, but this time we have no external data source to extract the neighbors of items. For example, in some social recommendation websites, such as Epinions², we have not only the rating information but also the social interactions or connections between users. This social network structure can be used to form neighborhood information of users. Under this situation, item feature vectors obey a common zero-mean Gaussian distribution. While the feature vector of each user still has a unique Gaussian distribution with its mean centered around her neighbors. The model turns out to be the fusion of UKNN and PMF.

4. EXPERIMENTAL EVALUATION

In this section, we conduct several experiments on two real world datasets and demonstrate: (1) the effectiveness comparison between NHPMF and other state-of-the-art methods; (2) the influence of the parameter settings in the NHPMF; (3) the performance of the NHPMF when users have few or even no observed ratings.

Dataset All experiments are performed on two real world datasets containing both rating and tagging data: the public MovieLens dataset³, and the Douban dataset⁴ that we crawled from the web. MovieLens has 10 million ratings and 100,000 tag applications to 10,000 movies by 72,000 users. The initial crawled Douban dataset contains 10,000 users' movie ratings and their tagging history. In both datasets, we remove those tags which are annotated by less than five distinct users and five distinct items for the tagging applications. Then we select those users and items that have at least five distinct related tags. Finally we filter out those users and items that have no tagging history at all. More statistics of the pre-processed dataset can be seen in Table 2.

Table 2: General statistics of the MovieLens and Douban

Statistics	MovieLens	Douban
# of users	467	4,213
# of items	2,437	17,671
# of tags	1,408	4,343
# of rating records	157,798	1,107,620
Average tags per user	24.38	71.87
Average tags per item	11.44	13.38

Baseline Methods and Evaluation Metrics We compare our model with the following baseline methods: UKNN [6], IKNN [12], PMF [11], SVD++ [8], TagiCoFi [14] and TagRec [15]. We choose these methods as baselines because they can be categorized into neighborhood models (UKNN, IKNN), matrix factorization models (PMF, TagiCoFi and TagRec) and the hybrid of the above two (SVD++). Also, these baseline methods can be classified according to whether they use tagging information or not. TagiCoFi and TagRec use both the tagging information and rating data for rating

²<http://www.epinions.com/>

³<http://www.grouplens.org/node/73>

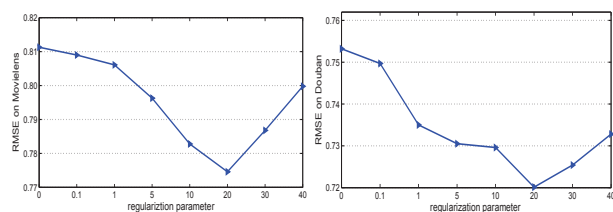
⁴<http://www.douban.com/>

prediction, the other methods use only the rating matrix. The evaluation metric we use in our experiments is root mean squared error (RMSE) [8], which is widely used to measure the performance of rating prediction accuracy in CF.

Effectiveness Comparison Table 3 and Table 4 report the RMSE values of all the algorithms under different settings of neighbor set size K and latent feature dimension D respectively. The neighbor set size K in neighborhood methods is set to 30, 50, 70 and 100 respectively while the dimensionality of the latent feature vectors D in matrix factorization methods is set to 5, 10, 20 and 30 respectively. As to the other parameters, we perform parameter tuning in advance for each method and use the best settings found in all the experiments for fairness.

From the comparison, we have the following observations: (1) Our method NHPMF, which incorporates the neighborhood information into matrix factorization, performs the best in all situations. Generally, NHPMF improves the RMSE of the best results of UKNN, IKNN, PMF, SVD++, TagiCoFi and TagRec by about 7%, 3%, 4.5%, 2.5%, 6% and 4%. (2) The results for Douban are better than the results for MovieLens for all methods, possibly because Douban is a much richer dataset since there are more tagging history and ratings per person and per item (see Table 2). (3) In pure KNN methods, the neighbor set size K plays an important role in model accuracy. Increasing K in a certain range does intend to improve the overall recommendation accuracy [4]. However, there's no benefit to increase K in NHPMF, because adding more neighbors will cover more global efforts, matrix factorization model can do well. This phenomenon is also noticed in [8]. (4) In matrix factorization models, the parameter of latent dimension D will affect the model accuracy. In general, larger values of D will give us more flexibility to represent both users and items in the latent space, leading to better performance. However, due to the time complexity of matrix factorization, which is linear with respect to D , larger values of D will require more time in training the MF model.

Impact of Regularization Parameter Figure 3 compares the RMSE of our model for the different ranges of regularization parameter on MovieLens and Douban dataset. For convenience, we set $\lambda_U = \lambda_V = \lambda$, and λ is set to be 0, 0.1, 1, 5, 10, 20, 30, 40 respectively. The other parameters are set as $K = 50$ and $D = 20$. We observe that the value of λ impacts the recommendation results significantly, which demonstrates that incorporating neighborhood information actually improves the recommendation accuracy. As λ increases in the range $[0, 20]$, the prediction accuracy improves quickly at first, but when λ surpasses 20, the prediction accuracy decreases. Clearly in this figure, the best regularization parameter setting for both datasets is $\lambda_U = \lambda_V = 20$.



(a) MovieLens Dataset (b) Douban Dataset

Figure 2: Impact of regularization parameter.

Performance with Different Sparsity One critical challenge in CF is to provide accurate recommendations when users supply only a few ratings or even have no rating history at all. In order to compare our model with the other methods under this situation, we split the 10% of the Douban dataset for training and the remaining for testing. Then we group all of the users based on the number of

Table 3: RMSE comparisons for different neighborhood size $K(D=20)$.

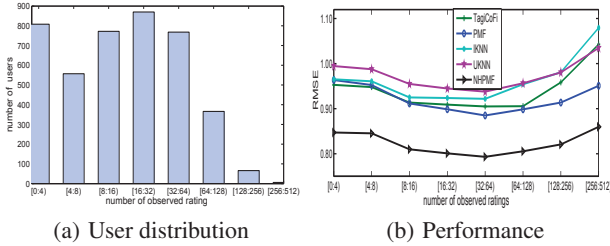
Model	K = 30		K = 50		K = 70		K = 100	
	MovieLens	Douban	MovieLens	Douban	MovieLens	Douban	MovieLens	Douban
UKNN	0.8494	0.8081	0.8400	0.7912	0.8358	0.7810	0.8326	0.7713
IKNN	0.8113	0.7551	0.8034	0.7449	0.8032	0.7398	0.8031	0.7351
SVD++	0.7970	0.7339	0.7962	0.7335	0.7957	0.7332	0.7951	0.7332
NHMPF	0.7764	0.7204	0.7746	0.7201	0.7749	0.7213	0.7746	0.7212

Table 4: RMSE comparisons for different latent feature dimension $D(K=50)$.

Model	D = 5		D = 10		D = 20		D = 30	
	MovieLens	Douban	MovieLens	Douban	MovieLens	Douban	MovieLens	Douban
PMF	0.8241	0.7512	0.8181	0.7502	0.8150	0.7479	0.8140	0.7483
SVD++	0.8006	0.7347	0.7974	0.7341	0.7962	0.7335	0.7949	0.7333
TagiCoFi	0.8212	0.7737	0.8142	0.7740	0.8132	0.7738	0.8164	0.7730
TagRec	0.8184	0.7421	0.8157	0.7417	0.8130	0.7405	0.8102	0.7401
NHMPF	0.7937	0.7257	0.7818	0.7236	0.7746	0.7201	0.7730	0.7188

the observed ratings in the training data and make predictions on different user groups.

Figure 3(a) summarizes the user group distribution of the training data. Obviously, this is a very sparse dataset since 19.18% of the users have at most three rating scores. The total number of rating records in train data is 110,410. On average, each user rates 26.2 items, where there are totally 17,671 items in the dataset. And the density of the training rating data is $110,410 / (4213 * 17671) = 0.148\%$.

**Figure 3: Performance comparison on different user rating scales.**

Pure CF methods, which only rely on the rating matrix, can not work well under this situation, since little preference information is available to provide any basis for recommendation. However, NHMPF can solve this problem to some extent. When user i 's rating information is limited, NHMPF can still rely on her tagging record and neighborhood information to infer her latent feature vector. As shown in Figure 3(b), NHMPF outperforms the other methods under all user groups, especially when few user ratings are given.

5. CONCLUSION

In this paper, we propose a unified framework to incorporate the complementary advantages of two popular approaches in CF: local neighbor method and global matrix factorization model. Meanwhile, we explore the possibility of using external tagging data other than the rating data to capture local relationships for users or items. Specifically, tagging data is first used to find neighbors of each user(item) and the neighborhood information is then incorporated into the factorization of rating matrix to make similar users(items) have similar latent features. Thus our method can better capture the local and global relationships of users and items, leading to more accurate recommendations. Experimental results on two real world datasets show that our model outperforms state-of-the-art collaborative filtering methods.

Acknowledgment The work was supported by grants from Natural Science Foundation of China (No.61073110), the Key Program of National Natural Science Foundation of China (No.60933013), the Research Fund for the Doctoral Program of Higher Education

of China (No.20093402110017) and the National Major Special Science and Technology Projects (No.2011ZX04016 – 071).

6. 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 TKDE*, pages 734–749, 2005.
- [2] R. M. Bell and Y. Koren. Lessons from the netflix prize challenge. *ACM SIGKDD Explorations Newsletter*, pages 75–79, 2007.
- [3] R. M. Bell and Y. Koren. Scalable collaborative filtering with jointly derived neighborhood interpolation weights. In *Proceedings of IEEE ICDM*, pages 43–52, 2007.
- [4] M. Deshpande and G. Karypis. Item-based top-n recommendation algorithms. *ACM TOIS*, pages 143–177, 2004.
- [5] Q. Gu, J. Zhou, and D. Chris. Collaborative filtering: Weighted nonnegative matrix factorization incorporating user and item graphs. In *Proceedings of SIAM SDM*, pages 199–210, 2010.
- [6] J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl. An algorithmic framework for performing collaborative filtering. In *Proceedings of ACM SIGIR*, pages 230–237, 1999.
- [7] M. Jamali and M. Ester. A matrix factorization technique with trust propagation for recommendation in social networks. In *Proceedings of ACM RecSys*, pages 135–142, 2010.
- [8] Y. Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proceeding of ACM SIGKDD*, pages 426–434, 2008.
- [9] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *IEEE Computer*, pages 30–37, 2009.
- [10] X. Li, L. Guo, and Y. Zhao. Tag-based social interest discovery. In *Proceeding of ACM WWW*, pages 675–684, 2008.
- [11] R. Salakhutdinov and A. Mnih. Probabilistic matrix factorization. *Advances in neural information processing systems*, pages 1257–1264, 2008.
- [12] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of ACM WWW*, pages 285–295, 2001.
- [13] G. Takács, I. Pilászy, B. Németh, and D. Tikk. Matrix factorization and neighbor based algorithms for the netflix prize problem. In *Proceedings of ACM RecSys*, pages 267–274, 2008.
- [14] Y. Zhen, J. Wu, and D.-Y. Yeung. Tagicofi: tag informed collaborative filtering. In *Proceedings of ACM RecSys*, pages 69–76, 2009.
- [15] T. C. Zhou, H. Ma, I. King, and M. R. Lyu. Tagrec: Leveraging tagging wisdom for recommendation. In *Proceedings of IEEE CSE*, pages 194–199, 2009.