Matrix Factorization meets Social Network Embedding for Rating Prediction

Menghao Zhang, Binbin Hu, Chuan Shi(✉),
Bin Wu and Bai Wang

Beijing Key Lab of Intelligent Telecommunications Software and Multimedia,
Beijing University of Posts and Telecommunications,Beijing, China 100876
{Jack,hubinbin,shichuan,wubin,wangbai}@bupt.edu.cn

Abstract. Social recommendation becomes a current research focus, which leverages social relations among users to alleviate data sparsity and cold-start problems in recommender systems. The social recommendation methods usually employ simple similarity information as social regularization on users. Unfortunately, the widely used social regularization cannot make a good analysis of the users' social relation characteristics. In order to overcome the shortcomings of social recommendations, we propose a new framework for which combines network embedding and probabilistic matrix factorization. We make use of social relation features extracted from social networks, on top of which we learn an additional layer that uncovers the social dimensions that explain the variation in people's feedback. Furthermore, the influence of different social network embedding strategies on our framework are compared. Experiments on three real datasets validate the effectiveness of the proposed solution.

Keywords: Network Embedding, Matrix Factorization, Social Recommendation

1 Introduction

With the continuous development of the e-commerce, and the rapid growth in the number and variety of goods, customers need to spend much time to find what they want. The process of visiting a large number of irrelevant information will be certainly drowned in information overload problem and it will undoubtedly continue to lose customers. To confront these challenges, personalized recommender systems emerged. Obviously, in recommender systems, the recommendation methods are the most critical component. Collaborative filtering is one of the most important technology of recommender systems [1]. It recommends items, which have been evaluated positively by another similar user or by a set of such users. Probabilistic Matrix Factorization (PMF) [2] is one of the most successful collaborative filtering methods. And the goal of the PMF is to decompose user - item rating matrix into user factor matrix and item factor matrix. This matrix factorization model performs well on the sparse, large, and imbalanced datasets and scales linearly with the number of observations.

Traditional matrix factorization methods always ignore social relationships among users. However, in our real life, when we seek advice from our friends on restaurants or books, we are actually requesting verbal social recommendations. Hence, in order to provide more personalized recommendations to improve recommender systems, we need to incorporate users' social network information. A few social recommendation methods have been proposed [3, 4] on the basis of intuition that users' social relations can be employed to strengthen traditional recommender systems. And these social recommendation mehtods usually use user similarity and regularization constraints.

Recently, the deep learning model and social network embedding (SNE) methods have been widely studied [5–7]. Some researches apply the deep learning methods to the recommender systems [8, 9]. These methods provide novel approaches for recommendation systems. However, they all have a high time complexity and space complexity and they do not make full use of social relations to strengthen the rating prediction.

In this paper, we focus on the impact of SNE on social recommender systems. We will study how to combine social network embedding methods with probabilistic matrix factorization, and propose a framework named as matrix factorization meets social network embedding for rating prediction (called MERP). The framework takes care both the accuracy and efficiency of social recommender systems. The basic idea of the mehthod is to generate user social network features by SNE and user latent feature with PMF, then combine these two features to predict ratings. First of all, we utilize SNE in social information network to generate user social network features. Next, we put these features in optimized matrix factorization model which can learn user-item latent feature and social feature simultaneously. Finally, we combine latent feature and social feature to predict user-item rating. Our experiments show that the combination of SNE and PMF is more accurate than other rating prediction method in the social recommendation model.

2 The MERP Methods

In this section, we build our framework (MERP) which combines user social network embedding and matrix factorization method. Firstly, we describe our social network embedding model. And then we systematically interpret how to integrate social network information with rating information.

2.1 Social Network Embedding

Recent progress in representational learning for network embedding opened new ways for feature learning of discrete objects. Particularly, the Skip-gram model [10] aims to learn continuous words' feature representations by optimizing a neighborhood preserving likelihood objective. And some network embedding (also called network representation learning (NRL)) methods use Skip-gram model to learn the nodes latent factors. DeepWalk [5], LINE[7] and Node2vec [6] are three representative social network embedding methods (SNE).

One of the orginal methods in network embedding methods is DeepWalk. The basic steps of the algorithm are as follows: it scans over the nodes of a network, and for every node it aims to embed it such that the node's latent features can predict the nearby nodes (i.e., nodes inside some slide window). The node feature representations are learned by optimizing the likelihood objective using Skip-gram and Hierarchical Softmax. The Skip-gram objective is based on the distributional hypothesis. It states that nodes in similar places tend to have similar meanings. In other words, similar nodes tend to appear in similar node neighborhoods.

2.2 MERP

In recommender systems, an efficient and effective approach is to factorize the user-item rating matrix. Its basic formulation assumes the following model to predict the preference of a user u toward an item i:

$$\hat{x}_{u,i} = \mu + b_u + b_i + \gamma_u^T \gamma_i \tag{1}$$

where μ is global offset, b_u and b_i are user and item bias terms, and γ_u and γ_i are vectors describing latent factors of user u and item i.

We firstly introduce DeepWalk network embedding method to implement our model. In recommendation system, social relations provide an independent source for recommendation. We think of our user social relations as a network, then we utilize DeepWalk to learn a latent space representation of social interactions θ_u. Our extended predictor takes the form

$$\hat{x}_{u,i} = \mu + b_u + b_i + \gamma_u^T \gamma_i + \theta_u^T \theta_i \tag{2}$$

where μ, b_u, b_i, γ_u, γ_i are as in Eq. 1. θ_u, θ_i are social factors whose inner product models the social interaction between u and i.

One naive way to implement the above model would be to directly use social network embedding features f_u of user u as θ_u in the above equation. However, this would present issues due to the high dimensionality of the features in question.Therefore, we propose to learn an embedding kernel which linearly transforms such high-dimensional features into a much lower-dimensional (say 16 or so) 'social rating' space:

$$\theta_u^T = f_u^T E^T \tag{3}$$

Here, E is a matrix embedding DeepWalk feature space into social space and f_i is the original social feature vector for user u. The numerical values of the projected dimensions can then be interpreted as the extent to which a user exhibits a particular social rating factor. This embedding is efficient in the sense

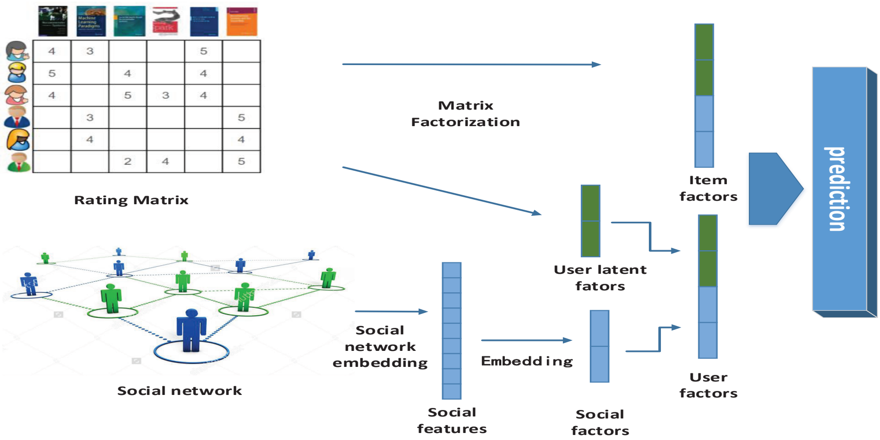


Fig. 1. Diagram of the preference predictor of MERP.

that all users share the same embedding matrix which significantly reduces the number of parameters to learn.

Next, we introduce a visual bias term b' whose inner product with f_u models items' overall property toward the social appearance of a given user. In summary, our final model is shown in Figure 1 and our final prediction formula is:

$$\hat{x}_{u,i} = \mu + b_u + b_i + \gamma_u^T \gamma_i + (f_u^T E^T)\theta_i + f_u^T b' \tag{4}$$

2.3 Model Learning and Discussion

We blend the users' social network embeddings in to matrix factorization framework for learning the parameters of our model. And we adopt SGD to optimize the following objective:

$$\begin{aligned} \mathscr{L} = min \sum_{u,i \in U,I} (\hat{x}_{ui} - x_{ui})^2 + \lambda(\|\gamma_u\|_F^2 + \|\gamma_i\|_F^2 \\ + b_u^2 + b_i^2) + \beta_\theta \|\theta_i\|_F^2 + \beta_E \|E\|_F^2 + \beta_b {b'}^2 \end{aligned} \tag{5}$$

As mentioned above, we select three different methods in social network embedding: DeepWalk, LINE, Node2vec. These three methods have their own advantages. LINE applies to large-scale information network. DeepWalk adopt random walk strategy in the network node search, while Node2vec provides two strategies: depth-first search (DFS) and breadth-first search (BFS). These two strategies make Node2vec learn node representations obeying two principles: the ability to learn representations that embed nodes from the same network community closely together, as well as to learn representations where nodes with similar structural roles get embedded together. In social information network, we find that BFS outperforms DFS which illustrates that embed nodes from the same network community closely is more suitable in social recommendation.

Table 1. Effectiveness Experimental Results on three datasets (The improvement is based on PMF)

Dataset	Training	Metrics	PMF	UserMean	ItemMean	NMF	BPMF	SoMF	ERP	MERP
Douban	80%	MAE	0.6272	0.6922	0.6600	0.6265	0.5986	0.6047	0.6724	**0.5848**
		Improve		-10.36%	-5.23%	0.32%	4.56%	3.59%	-7.21%	6.76%
		RMSE	0.7870	0.8668	0.8283	0.8015	0.7651	0.7625	0.8478	**0.7396**
		Improve		-10.14%	-5.25%	-1.84%	2.78%	3.11%	-7.73%	6.02%
	60%	MAE	0.6251	0.6896	0.6619	0.6314	0.6066	0.6129	0.6790	**0.5861**
		Improve		-9.89%	-5.89%	-1.01%	2.96%	1.95%	-8.62%	6.24%
		RMSE	0.7825	0.8621	0.8298	0.8063	0.7712	0.7745	0.8611	**0.7416**
		Improve		-10.17%	-6.04%	-3.04%	1.44%	1.02%	-10.04%	5.23%
	40%	MAE	0.6342	0.6951	0.6712	0.6657	0.6303	0.6399	0.6873	**0.5888**
		Improve		-9.60%	-5.83%	-4.97%	0.61%	-0.90%	-8.37%	7.16%
		RMSE	0.7950	0.8683	0.8737	0.8596	0.8032	0.8029	0.8820	**0.7474**
		Improve		-9.22%	-9.90%	-8.13%	-1.03%	-1.00%	-10.94%	5.99%
	20%	MAE	0.6600	0.6979	0.6848	0.7243	0.6895	0.7132	0.7382	**0.5957**
		Improve		-5.74%	-3.76%	-9.74%	-4.47%	-8.06%	-11.85%	9.74%
		RMSE	0.8276	0.8744	0.8764	0.9396	0.8848	0.8891	0.9890	**0.7538**
		Improve		-5.65%	-5.90%	-13.53%	-6.91%	-7.43%	-19.50%	8.92%
Yelp	80%	MAE	0.8155	0.8500	0.8177	0.8208	0.8889	0.8390	0.8922	**0.7847**
		Improve		-4.33%	-0.27%	-0.65%	-9.00%	-2.88%	-9.41%	3.78%
		RMSE	1.0420	1.0894	1.0618	1.0474	1.1664	1.0733	1.2282	**1.0032**
		Improve		-4.55%	-1.90%	-0.52%	-11.94%	-3.00%	-17.87%	3.72%
	60%	MAE	0.8281	0.8550	0.8286	0.8307	0.9295	0.8549	0.9229	**0.7969**
		Improve		-3.25%	-0.06%	-0.31%	-12.24%	-3.24%	-11.45%	3.77%
		RMSE	1.0560	1.0942	1.0771	1.0619	1.2192	1.0884	1.2877	**1.0152**
		Improve		-3.62%	-2.00%	-0.56%	-15.45%	-3.07%	-21.94%	3.86%
	40%	MAE	0.8470	0.8691	0.8468	0.8535	1.0091	0.8666	0.9871	**0.8091**
		Improve		-2.61%	0.02%	-0.77%	-19.14%	-2.29%	-16.54%	4.47%
		RMSE	1.0832	1.1133	1.0990	1.0905	1.3240	1.0976	1.3973	**1.0258**
		Improve		-2.78%	-1.46%	-0.67%	-22.23%	-1.33%	-29.00%	5.30%
	20%	MAE	0.8894	0.8973	0.8828	0.8935	1.0810	0.8711	1.1419	**0.8359**
		Improve		-0.89%	0.74%	-0.46%	-21.54%	2.06%	-28.39%	6.02%
		RMSE	1.1332	1.1525	1.1434	1.1404	1.4067	1.1005	1.6341	**1.0519**
		Improve		-1.70%	-0.90%	-0.64%	-24.14%	2.89%	-44.20%	7.17%
Epinions	80%	MAE	0.8730	0.9385	0.8981	0.8520	0.9473	0.8615	0.9129	**0.8293**
		Improve		-7.50%	-2.88%	2.41%	-8.51%	1.32%	-4.57%	5.01%
		RMSE	1.1167	1.2115	1.1628	1.1061	1.2427	1.1083	1.1846	**1.0638**
		Improve		-8.49%	-4.13%	0.95%	-11.28%	0.75%	-6.08%	4.74%
	60%	MAE	0.8951	0.9467	0.9132	0.8665	0.9935	0.8768	0.9349	**0.8429**
		Improve		-5.79%	-2.02%	3.20%	-10.99%	2.04%	-4.45%	5.83%
		RMSE	1.1454	1.2217	1.1791	1.1258	1.3035	1.1264	1.2105	**1.0757**
		Improve		-6.66%	-2.31%	1.71%	-13.80%	1.66%	-5.68%	6.09%
	40%	MAE	0.9253	0.9626	0.9444	0.8887	1.0731	0.9016	0.9666	**0.8643**
		Improve		-4.03%	-2.06%	3.96%	-15.97%	2.56%	-4.46%	6.59%
		RMSE	1.1844	1.2477	1.2145	1.1345	1.4111	1.1636	1.2438	**1.0968**
		Improve		-5.34%	-2.54%	4.21%	-19.14%	1.76%	-5.02%	7.40%
	20%	MAE	0.9599	1.0033	0.9983	0.9290	1.2144	0.9198	1.0278	**0.9080**
		Improve		-4.52%	-4.00%	3.22%	-26.51%	4.18%	-7.07%	5.41%
		RMSE	1.2282	1.3002	1.2718	1.2068	1.5947	1.2029	1.3077	**1.1374**
		Improve		-5.86%	-3.55%	1.74%	-29.84%	2.06%	-6.47%	7.39%

3 Experiments

In this section, we perform experiments on multiple real-world datasets and present the result analysis.

3.1 Datasets and Metrics

We use three popular datasets to validate the effectiveness of our model. The Douban dataset [11] includes 1000 users and 5000 movies with 176308 movie ratings ranging from 1 to 5. As for the Yelp dataset [11], it includes 9581 users and 14037 items with 171109 ratings ranging from 1 to 5. And the Epinions dataset is a larger dataset, consisting 40163 users and 139738 items with 664827 ratings ranging from 1 to 5. We use two common metrics to evaluate the performance of different methods, namely Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) [11].

3.2 Compared Method

We consider the following baselines to validate the effectiveness of MERP. (1) PMF [2] / NMF [12]: They are typical matrix factorization methods. (2) $UserMean/ItemMean$: It employs a user / item's mean rating to predict the missing ratings directly. (3) $BPMF$ [13]: It is a matrix factorization method with Bayesion framework. (4) $SoMF$ [3]: It is the MF based recommendation method with social regularization.

After obtaining the social network embeddings, we could calculates users' similarity with them directly and makes predictions with weighted average method, which is called **ERP**. Instead of the native idea, our model **MERP** is a more flexible model to leverage the social network embeddings.

3.3 Effectiveness Experiments

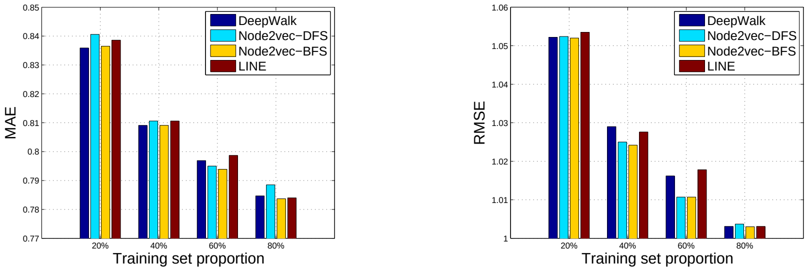
This section will validate the effectiveness of MERP through comparing its different variations to baselines. For a fair comparison of PMF, UserMean, ItemMean, NMF, BPMF, SoMF, we use the same parameters in both methods. For all the experiments in this experiments, the latent factor number is fixed to 10. In MERP, we set social latent factors $D = 64$, social embedding features $F = 10$, the regularization coefficient λ, β_θ, β_E, β_b are set to trivial values 0.05,0.001,0.001,0.1. In this experiment, we choose DeepWalk method as our network embedding method, we have fixed the window size(5) and the walk length(10) to emphasize local structure.

For these three datasets, we use different ratios(80%, 60%, 40%, 20%) of data as training data. For example, the training data 80% means that we select 80% of the ratings from user-item rating matrix as the training data to predict the remaining 20% of ratings. The random selection was carried out 10 times independently in all the experiments.We report the average results on three different datasets and also record the improvement of all methods compared to the baseline PMF.

The performance of all the methods are shown in Table 1. And we can get the following conclusions. MERP always perform better than the original methods on each dataset and all ratios. To some extent, MERP is considered to be a combination of two methods PMF and ERP. MERP always performs better than PMF and ERAP which illustrates the effectiveness of the combination of these two methods. By comparing the three datasets, we can find that on more sparse datasets the MERP can performs better.

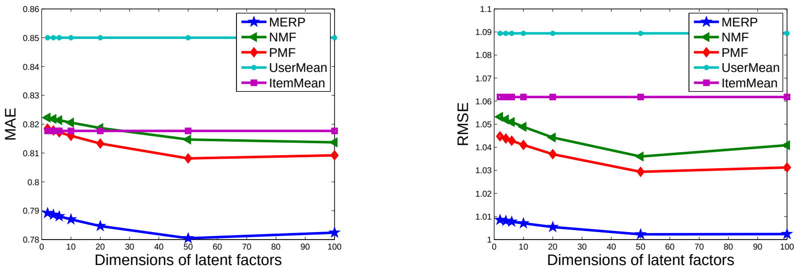
3.4 Impact of Different Strategies

Experiments in this section will validate the sensitivity of different network embedding strategies in MERP. Here we compare two metods in MERP: DeepWalk and Node2vec. In Node2vec, there are two different strategies: breadth first search (BFS) and depth first search (DFS). So we compare a total of three



(a) MAE (b) RMSE

Fig. 2. The comparison of different strategies in MERP.



(a) MAE (b) RMSE

Fig. 3. The comparison on different latent dimensions.

strategies: DeepWalk, Node2vec-BFS, Node2vec-DFS with the same experiment setting as Section 3.3.

As shown in Figure 2(a), all the methods improve as training set proportion increases. However, when we try different strategies, we find that the Node2vec-BFS performs better then other network embedding strategies. The result illustrates that breadth first search strategy is more suitable in social recommendation.

3.5 Parameter Study

For matrix factorization based methods, the latent dimension is an important parameter to tune. And our model also involves such a parameter. We vary it from 0 to 100 with a step of 10, and examine how the performance changes with regard to the latent dimension. As shown in Figure 3(a), using 50 latent dimension yields the best performance and MERP performs better in all latent dimesions.

4 Conclusion

In this paper, we propose a framework which combines social network embedding and matrix factorization to predict the unknown user-item ratings in the rating matrix. In order to fully utilize the social information network to solve data sparseness, scalability, and predictive quality in social recommendation problem, we introduce social network embedding method into matrix factorization. MERP makes a better performance on rating prediction accuracy. We analyze the performance of MERP in terms of its dependency on training set size, different network search strategies and latent dimension number, our method performs well in all cases.

5 Acknowledgements

This work is supported by the National Natural Science Foundation of China (No. 61772082, 61375058), the National Key Research and Development Program of China (2017YFB0803304), and the Beijing Municipal Natural Science Foundation (4182043).

References

1. Zhang, M., Hu, B., Shi, C., Wang, B.: Local low-rank matrix approximation with preference selection of anchor points. In: International Conference on World Wide Web Companion. (2017) 1395–1403
2. Mnih, A., Salakhutdinov, R.R.: Probabilistic matrix factorization. In: NIPS. (2008) 1257–1264
3. Ma, H., Zhou, D., Liu, C., Lyu, M.R., King, I.: Recommender systems with social regularization. In: WSDM. (2011) 287–296
4. Tang, J., Hu, X., Liu, H.: Social recommendation: a review. Social Network Analysis and Mining **3**(4) (2013) 1113–1133
5. Perozzi, B., Al-Rfou, R., Skiena, S.: Deepwalk: Online learning of social representations. In: SIGKDD. (2014) 701–710
6. Grover, A., Leskovec, J.: node2vec: Scalable feature learning for networks. In: SIGKDD. (2016) 855–864
7. Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., Mei, Q.: Line: Large-scale information network embedding. In: WWW. (2015) 1067–1077
8. Wang, H., Wang, N., Yeung, D.Y.: Collaborative deep learning for recommender systems. In: SIGKDD. (2015) 1235–1244
9. Liang, D., Altosaar, J., Charlin, L., Blei, D.M.: Factorization meets the item embedding: Regularizing matrix factorization with item co-occurrence. In: RecSys. (2016) 59–66
10. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. In: ICLR Workshop. (2013)
11. Shi, C., Hu, B., Zhao, W.X., Yu, P.S.: Heterogeneous information network embedding for recommendation. arXiv preprint arXiv:1711.10730 (2017)
12. Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. Nature **401**(6755) (1999) 788–791
13. Salakhutdinov, R., Mnih, A.: Bayesian probabilistic matrix factorization using markov chain monte carlo. In: ICML. (2008) 880–887