공학석사 학위논문

Review Rating Applied Social Recommender System

사용자 평론을 이용한 소셜 추천 기법

2013 년 8 월

서울대학교 대학원

전기·컴퓨터 공학부

장 해

Review Rating Applied Social Recommender System

사용자 평론을 이용한 소셜 추천 기법

지도교수 김 종 권

이 논문을 공학석사 학위논문으로 제출함

2013 년 4 월

서울대학교 대학원

전기·컴퓨터 공학부

장 해

장해의 석사학위논문을 인준함

2013 년 6 월

위 원 장 ______ 전 화 숙 ______ (인)

부위원장 ______ 김 종 권 ______ (인)

위 원 ______ 권 태 경 ______ (인)

Abstract

Review Rating Applied Social Recommender System

Hai Zhang
School of Computer Science and Engineering
The Graduate School
Seoul National University

Collaborative filtering (CF) has been widely used in recommender systems, which uses historical user ratings or purchase records as input to predict items that users may be interested. But this method does not reflect the real world situation. People in real world tend to ask their friends or expertise in that field before making a decision. By considering the social relation, some social recommend approaches had been proposed and achieve better performance comparing with traditional recommend methods. In these researches, the strength of social relation is usually measured by similarity between users. And only item rating data is used to calculate the similarity. However, if two friends have no items rated in common, the similarity between them will be zeros. In our research, we use another data source, review rating data, to find more similarity between users. By doing experiment on crawled Eipinions dataset, we confirm the improved prediction accuracy.

Keywords: Recommender System, Social Network, Review Rating, Matrix Factorization, Collaborative Filtering

Student Number: 2011-24086

Contents

Chapter 1 Introduction 1
1.1 Background 1
1.2 Main Idea and Contributions 3
1.3 Thesis Organization 4
Chapter 2 Related Work 5
2.1 Recommender System 5
2.2 Social Recommender System 7
Chapter 3 Problem Definition 9
Chapter 4 Social Recommender System 11
4.1. Matrix factorization 11
4.2. Social recommendation 13
4.3. Similarity clculation 15
Chapter 5 Experiments 18
5.1 Dataset 18
5.2 Metrics 19
5.3 Experimental results 20
Chapter 6 Conclusion 24

List of Figures

Figure 3.1: Social Network ………..9

Figure 3.2: User-Item Rating Matrix 9

Figure 3.3: User-Review Matrix ……………….......................................9

Figure 4.1: A Matrix Factorization Example 12

Figure 4.2: A Similarity Example ………..15

Figure 4.3: The Difference of Similarity Measured by Two Dataset …...17

Figure 5.1: MAE Comparison ……..21

Figure 5.2: RMSE Comparison ……..22

List of Tables

Table 4.1: Statistic of Similarity Number ……………….........................16

Table 5.1: Statistic of Crawled Dataset …………………........................18

Chapter 1 Introduction

1.1 Background

Recommender systems utilize existed user preferences or purchase history and other available information from users and items to recommend items that might interest users. With the development of Internet and the spread of Web 2.0, recommender system has been widely used in various website [1]. Amazon, Google News, Eipinions, and Netflix utilize recommender system to improve their competiveness. The most popular technology adopted by these websites is Collaborative filtering [2, 3, 4], which use the historical data as inputs and predict the proper items to user.

However, in real life, people tend not to just follow their own opinion when they make a decision. Instead, they usually ask some close friend or other people with the knowledge in similar field. It is a well-known phenomenon that explained by a social theory called Social Influence [5]. In addition, we could find lots of social relations on the Internet. For example, the social network service site Facebook let people make friends with other people using the same service. And consumer review and rating site Epinions allow users to make trust relationship with other experienced product reviewer in this site. Because of the available data of these social relationship, several algorithms [6, 7, 8] considering social relationship has been proposed. These algorithms try to apply the social influence effect to enhance the prediction accuracy of recommender system.

One of the unsolved problem in recommender systems is data sparsity, comparing to the large number of users as well as items in the recommender system. User rating to the items is so sparse that almost 99% of user-item matrix has no value. With the social relationship available in recommender system. Researchers have a new data source that could be used in recommend system. For example, a user in Amazon.com may have no purchase record. And by traditional collaborative filtering method, recommender system could not provide a recommendation. However, this user may have some friend relationship, and if the friends of him or her have purchase records, system could make a recommendation to the user.

In addition, people may have lots of social relationship and friends, but the influence from each friend should be different. In order to reflect the strength of relationship, the author of [6] proposed Vector Space Similarity (VSS) [9] and Pearson Correlation Coefficient (PCC) [10] to measure the similarity between friends. These two similarity functions are calculated based on the same item sets that both two friend rated. But there is a problem, if two users with social relation have no rating item in common, then their similarity is 0. However, there exists the possibility that the two users have similar rating pattern and preference, and they just have not rated a common item. We call this problem as similarity loss problem in this paper. To find more accurate similarity between users, we try to use another data source-review rating.

In product review and rating site, such as Epinions and CIAO, users could not only rate items, they could also rate reviews written by other users. And this is a much bigger dataset than item rating, because users rate more reviews than items. By mapping the review rating onto values from 1 to 5, we use this dataset to calculate the similarity between users and we could get more precise similarity between users compare to using the item rating dataset.

1.2 Main Idea and Contributions

We propose the social recommender system using review rating data to find similarity between users. We extend the social recommendation model [6] which combines matrix factorization method and social regularization scheme. The previous scheme [6] is using item rating when calculate similarity between users in previous work. We additionally use review rating to find more similarity between users. We apply our idea to extend model [6] and do experiment with real world dataset crawled from Epinions to validate our proposed method.

Our contributions of this paper are as follows:

- We use the review rating dataset in product review and rating site, which is the first trial in the social recommender systems.
- We crawled the Epinions dataset and processed the dataset to use the dataset in our experiment.
- We extend the social recommendation model by applying review rating dataset to compute the similarity and get better recommendation performance.

1.3 Thesis Organization

The rest of this paper is organized as follow. Related work (recommender system, social recommender system) are discussed in section 2. In section 3, we provide the problem definition. Section 4 details of our method for social recommendation and similarity calculation. Our experiments are presented in Section 5. Finally, we conclude the paper in Section 6.

Chapter 2 Related Work

2.1 Recommender System

Recommender systems use historical system data on user purchase and item rating preference and other kinds of data on user (for example, profile and location) and items (for example, category and tags) to predict the items that the system users may like. One of the widely used methods is Collaborative Filtering (CF) [2, 3, 4], which is further defined by [9] as a broader set of techniques that use previous preferences to predict a new ones. Because CF only use the simple user-item matrix as input without other information, it could be easily adopted to real world. Amazon [1] and Netflix use CF in their recommender system to provide recommend service to their customers.

CF could be divided into two main categories, memory-based and model-based approach. And memory-based could be further divided into user based method [7, 11] and item-based method [1, 12]. The user based approach evaluates the preference of a user to an item based on the rating history of similar users. On the contrary, the item based approach finds a user's preference on an item from the ratings of similar items by the same user. Both user based method and item based method utilize similarity function such as PCC and VSS to calculate similarity between users or similarity between items. Since PCC considers rating pattern of every user, it is usually more accurate than VSS.

In contrast, the model-based CF methods use the historical data to build models which are then used for predicting new preferences. These method at first use training data as input to the model, and use test data to find parameter value in the model. Several well-known models are Latent factor model [13], Bayesian hierarchical model [14], Clustering model [15]. And matrix factorization model [16] makes an assumption that any user or item in the recommend system could be describe by a number of hidden factors, which could be considered as latent factors, and by multiplying user-factor matrix and item-factor matrix, the unknown rating values in user-item rating matrix could be calculated. Because matrix factorization model could get a good accuracy prediction even though the dataset is very sparse, it is regarded as the most popular CF recommendation model [6].

Although CF is easy to implement and could also make a proper performance, the data sparsity problem in recommender system is still an unsolved problem. Researchers try to find other available data and theory that could improve the predict accuracy. Social relation and social influence theory provide new research topic to recommender system.

2.2 Social Recommender System

In real life, when people make a decision, such as buy a product or see a movie, they not only decide by their own opinions, but also consider other people's opinions. Maybe a close friend or an expert in that area. Previous research [17] analyzes the impact of trust networks in product review site and the author finds users with trust relation have more similar pattern than users without trust relation. Another research [18] also prove the existence of social influence by t-test. And there are some methods proposed to utilize both user-item matrix and social relation data. Authors in [7] propose a factor analysis model based on the probabilistic model which consider both user-item matrix and trust relation matrix under an assumption that users are affected by the same latent factors in user-item matrix and trust relation matrix. Even though the model get a better performance than previous collaborative filtering approach, the model does not reflect the interpersonal relation, and thus far from a real social recommendation. [8] proposes a method based on probabilistic matrix factorization with graphical model and consider the social trust network. This approach reflects both the user's own tastes and user friends' taste. However, the taste of user's friends does not affect user's feature vector. SocialMF model is proposed in [19], which considers social information in user's feature vector. However, it does not measure the strength of social relation and treat all social relation as same. In real world situation, however, the strength of each social relation is different, to measuring the difference between social relation, [6] uses similarity function to measure how similar the two users' preference is by Pearson Correlation Coefficient (PCC) and Vector Space Similarity (VSS). Two users with social relation may have rated items in common, according to the rating value of common items, two users' similarity could be calculated by the above two similarity functions. If two users with social relation have not rated the same item, even though their preference are very similar, proposed model in [6] cannot measure the similarity value. To address this problem, we try to use review rating data additionally to reflect the similarity between users and improve the recommendation accuracy.

Chapter 3 Problem Definition and Preliminaries

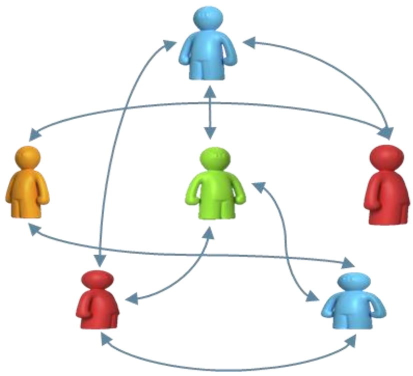

Figure 3.1: Social Network

	v_1	v_2	v_3	v_4	v_5
u_1	?	3	4	2	?
u_2	4	?	?	5	?
u_3	?	5	?	?	3
u_4	2	?	?	?	?
u_5	?	4	4	?	?
u_6	?	?	?	4	5

Figure 3.2: User-Item Rating Matrix

	r_1	r_2	r_3	r_4	r_5
u_1		4		5	
u_2		3	2	4	
u_3	3				4
u_4	4		5		
u_5				2	4
u_6		4			

Figure 3.3: User-Review Matrix

Different from traditional recommender system only considers the user-item rating matrix, our approach considers social relations between users. In real world social recommendation, we consider friends opinions when make a decision. To reflect friends' influence on our decision, we should know the friends network and the preferences of these friends, which can be modeled by the example of social network graph in Figure 3.1 and user-item matrix in Figure 3.2. While in our work we additionally consider user-review matrix in Figure 3.3, which are ratings to reviews written by reviewers.

In this example, there are 6 users (from u_1 to u_6), 5 items (from v_1 to v_5) and 5 reviews (from r_1 to r_5) with user social relations (edges). Each edge in Figure 3.1 represent the social connection between two users. The social relation could be friendship liked bidirectional relationship or social trust liked one directional relationship, which depends on the dataset we work on. In our research, each user rates items on a 5-point integer scale to express the extent of the preference of each item.

The problem we want to solve in this paper is how to quickly and accurately predict the "?" entities in user-item rating matrix by utilizing the three data source.

Chapter 4 Social Recommender System

In this section, we analyse the social recommendation problem based on matrix factorization model and social regularization. Then we utilize the review rating dataset to similarity calculation.

4.1. Matrix Factorization

Matrix factorization is one of the most effective method in collaborative filtering, which considers both user preferences and item characteristics could be explained by some numbers of latent factors. According to this method, a user's preference could be represent as a vector $\boldsymbol{U}_i$, and an item's characteristic could also be represent as a vector V_j.

Considering a user-item rating matrix $R = (R_{ij})_{M\times N}$, where there are M users and N items. Matrix factorization approach try to approximate the rating matrix R by a multiplication of user transpose matrix U^T and item matrix V,

$$R \approx U^T \times V \tag{1}$$

where U= $(U_{ij})_{M\times f}$, V = $(V_{ij})_{N\times f}$, and f is the numbers of latent factor. Matrix U and V could be regarded as the user latent matrix and item latent matrix. Since user usually rate a small portion of items, the matrix R is usually very sparse.

We define I= $(I_{ij})_{M\times N}$, if $R_{ij} \neq 0$, $I_{ij} = 1$; else $I_{ij} = 0$. Then we can rewrite the objective loss function as

$$\mathrm{L(U,V)} = \sum_{ij} I_{ij} \left(R_{ij} - U_i^T V_j\right)^2 \tag{2}$$

In order to avoid over fitting, two regularization terms are added into equation 2.

$$L(U, V) = \sum_{ij} I_{ij} (R_{ij} - U_i^T V_j)^2 + \lambda_1 ||U||_F^2 + \lambda_2 ||V||_F^2 \qquad (3)$$

$||.||_F^2$ represents the Fresenius norm. A simple example is provided as Figure 4.1.

$R_{m \times n}$	Item 1	Item 2	...	Item n
User 1	3	5	...	5
User 2	1	?	...	2
:	:	:	:	:
User m	5	1	...	0

=

$U_{f \times m}$	d_1	d_2
User 1	1.3	-1.1
User 2	-0.5	1.5
:	:	:
User m	-0.6	1.0

×

$V_{f \times n}$	Item 1	Item 2	...	Item n
d_1	0.8	-1.6	...	0.9
d_2	0.7	1.3	...	-0.4

Figure 4.1: Matrix Factorization Example

4.2. Social Matrix Factorization

When make a decision, people tend to take account of friends' opinion because we trust friends' advice and their preference. And we try to represent this social influence by a mathematical model. If a user is affected by his or her friends, user's taste with the friends should be similar. Thus, we consider the difference between user and his or her friends individually, we name this a social regularization part.

$$\beta \sum_{i=1}^{m} \sum_{f \in F^{+}(i)} Sim(i, f) \left\| U_i - U_f \right\|_F^2 \tag{4}$$

where β>0, which is the impact of social network in this model. $F^{+}(i)$ means the group that user i trusts other people, namely u_i's outer link friends. U_i denotes the feature of user u_i and U_f denotes the feature of user u_i's friends feature. The value of Sim(i, f) means the difference of feature vector U_i and U_f. If the value of Sim(i, f) is near to 1, then user i and i's friend j have similar taste and reflect friend j's taste more. On the other hand, if the value is near to 0, then user i accepts friend j's opinion less. In section 4.3, the similarity calculation will be discussed in more detail. Hence, the social recommendation model could be noted as follow:

$$\begin{aligned} \mathrm{L(U, V)} = \sum_{ij} I_{ij} \left(R_{ij} - U_i^T V_j\right)^2 + \beta \sum_{i=1}^{m} \sum_{f \in F^{+}(i)} Sim(i, f) \left\| U_i - U_f \right\|_F^2 \\ + \lambda_1 ||U||_F^2 + \lambda_2 ||V||_F^2 \end{aligned} \tag{5}$$

Since trust relation is asymmetric, which means the relation is one directional, we consider not only outer link friends but also inner link friends. It means that user i trusts other friends and other users who trust user i. User i's similar taste with friends are influenced by user i's selected items, and user i's taste also influences on others who are trusted by user i.

Social recommendation model tries to minimize the loss function given by equation 5. By performing gradient descent in vectors U_i and V_j

$$\frac{\partial L}{\partial U_i} = \sum_{j=1}^{n} I_{ij}(U_i^T V_j - R_{ij})V_j + \beta \sum_{f \in F^+(i)} Sim(i, f)(U_i - U_f) + \beta \sum_{g \in F^-(i)} sim(i, g)(U_i - U_g) + \lambda_1 U_i \tag{6}$$

$$\frac{\partial L}{\partial V_j} = \sum_{i=1}^{m} I_{ij}(U_i^T V_j - R_{ij})U_i + \lambda_2 V_j \tag{7}$$

4.3. Similarity Function

The definition of similarity between user A and user B in [20] is defined as the ratio between the amounts of information needed to state the commonality of A and B and the information needed to fully describe what A and B are. In our paper, we try to represent the similarity between users with social relation by similarity function Pearson Coefficient Correlation (PCC) which is shown as equation 8.

$$Sim(i,f)=\frac{\sum_{j\in I(i)\cap I(f)}(R_{ij}-\overline{R}_i)\times(R_{fj}-\overline{R}_f)}{\sqrt{\sum_{j\in I(i)\cap I(f)}(R_{ij}-\overline{R}_i)^2}\times\sqrt{\sum_{j\in I(i)\cap I(f)}(R_{ij}-\overline{R}_f)^2}} \tag{8}$$

In [6], the authors use user-item rating data to find the common rated items between two users and according to the rating similarity between the common rated items to calculate the similarity between two users with social relation. Suppose user u_i and u_j are friends. And each of them rate lots of items and have similar preference. But they rate no item in common, then, according to PCC, their similarity is 0. To solve this problem, we try to use review rating data, which is the user rating to reviews written by other users. By crawling the product rating and review site Epinions dataset, we find that review rating dataset is much larger than item rating dataset, and this dataset has not been utilized in previous researches. There are five kinds of rating to review. “not helpful”, “somewhat helpful”, “helpful”, “very helpful” and “most helpful”. By mapping these ratings to 1-5 scale. The crawled data and further analysis are summarized in table 4.1.

Table 4.1: Statistics of Similarity Number

	Similarity numbers
Item Rating	15718
Review Rating	45319
Item rating and Review Rating	7728

The number of similarity obtained from item rating data is 15718, and from review rating is 45319. The number of similarity that could calculate from both dataset is 7728.

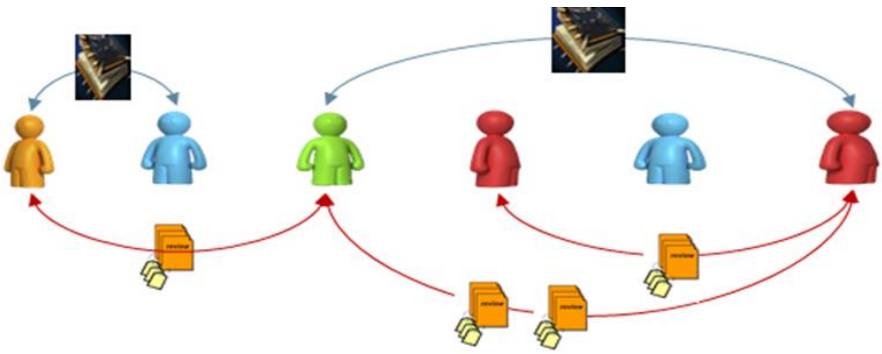

Figure 4.2: A Similarity Example

We utilize both item rating data and review rating data as an input to PCC, and get additional similarity values between users comparing with previous researches. Figure 4.2 describe a simple example of our approach. The blue link is the similarity between two users measured by item rating data, which the red link is the similarity between two users calculated by review rating data. Shown as the Figure 4.2 describes, much more similarity between users could be found. There could be a case that similarity between two users could be both measured by item rating data and review rating data. We count this case and try to find out the difference between this two values. We employ a mapping function $f(x) = (x + 1)/2$ to bound the range of PCC similarities into [0, 1]. So the difference range of similarity value measured form review rating data minus item rating data should be [-2, 2]. Figure 4.3 is the statics of the differences.

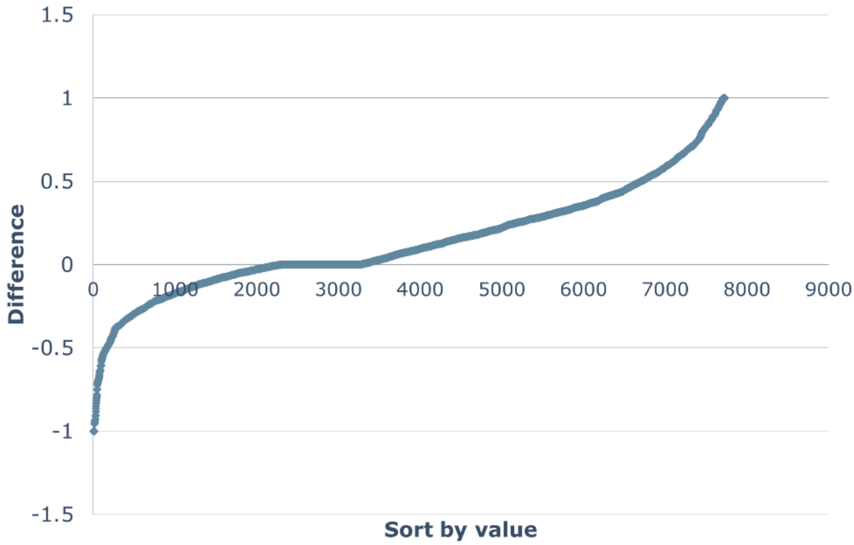

Figure 4.3: The Difference of Similarity Measured by Two Dataset

From figure 4.3, we can find that the difference of similarity measured by the two dataset exists, and the similarity value from review rating is higher than item rating got similarity value in many cases.

Chapter 5 Experiments

5.1 Dataset

The dataset we crawled for experimental evaluation is Epinion [33], which is a well-known product review site. Users could write reviews about items after registration. And rate the items based on the scale from 1 to 5 stars. Every Epinions user could make “trust” relation with other users which can be regarded as a social network between users. We express the trust relation in binary. For example, If user A trust user B, then user A’s trust value toward B is 1 and otherwise it is 0. The dataset from Epinions consists of 28,827, users who have rated a total of 96,406 different items. The total number of ratings is 96,380 and the number of review rating is 484,282. The total number of issued trust statements is 125,669.

Table5.1: Statistic of Crawled Dataset

User	28827
Item	96406
Item rating	96380
Review rating	484282
Trust relation	125669

5.2 Metrics

To compare our approach and other previous approach, we use Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) to measure the predict accuracy of these methods.

The metrics MAE is defined as,

$$MAE = \frac{1}{T}\sum_{i,j} | R_{ij} - \widehat{R}_{ij} | \ , \tag{9}$$

Where T denotes the number of tested ratings and R_{ij} denotes the rating user i gave to item j, $\widehat{R_{ij}}$ denotes the rating user i gave to item j as predicted by a method. RMSE is defined as

$$RMSE = \sqrt{\frac{1}{T}\sum_{i,j} (R_{ij} - \widehat{R}_{ij})^2} \ , \tag{10}$$

From the definition, we could know that the smaller MAE or RMSE value, the better performance is achieved.

5.3 Experimental results

For the Epinions dataset, we randomly choose 80% of dataset as training data and the remaining 20% of dataset as test data. The training set is used as a input to the model and the test set is used to check the accuracy of the training model. We experiment random selection 5 times independently when selecting the training and test dataset and get the average value. We use the dimension K=20 and all of the regularization parameters λ_1, λ_2 is 0.01.

1. PMF: This method is proposed in [21]. It only considers user-item matrix for recommendations
2. SR_I: Social regularization method in [6]. It considers not only user-item matrix but social trust information. It also uses similarities of users with trust relation for recommendations.
3. SR_I+R: This is our expanded model which utilized review rating dataset when calculate the similarity between users. From section 4.3, we know that there are some situations that two users' similarity could be got from item rating data and review rating data. Taking consideration of the situation, we choose a naïve approach. We choose the similarity got from item rating or review rating individually when there are two similarity value between two users. SR_I(100%)+R means considering the similarity from item rating and SR_I+R(100%) means taking the similarity from review rating.

SR_I+R is our proposed method, which use review rating additionally to calculate the similarity between users. From section 4.3, we know that

3times much similarity values obtained from review rating dataset, and we use the enlarged similarity to the social recommendation model and do experiment on crawled Epinions dataset.

From the Figure 5.1 and Figure 5.2, we could find out that our proposed method is better than other methods. PMF method only use user-item rating matrix to predict the missing value, so its MAE and RMSE value is higher than other schemes. However, in SR_PCC, the author use social relations among users. By both consider the user-item matrix and social network among users, the performance is better than PMF. This results prove that user's friendship or trust relationship contributes to the performance of recommender system as being more accurate. In SR_PCC method, they only get similarity from item rating data, and many similarity between users are missed just because they do not have the same rating items. Taking consideration of this problem, we try to user review rating data to find more similarity between users. By applying the similarity got from item rating and review rating to the social recommendation model, we could achieve a better performance comparing with previous approaches.

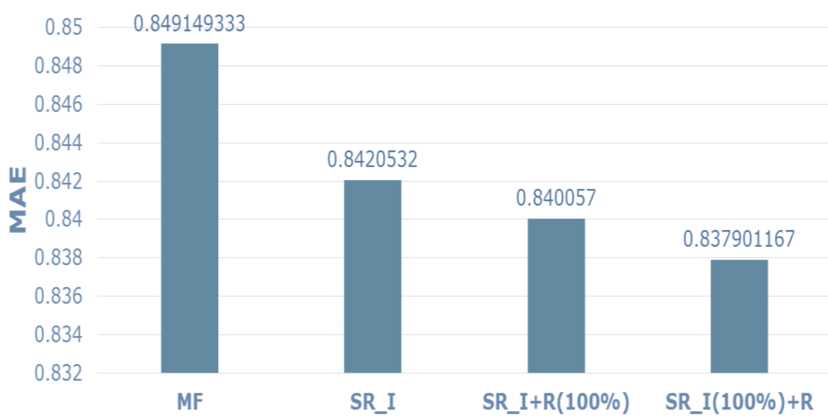

Figure 5.1: MAE Comparison

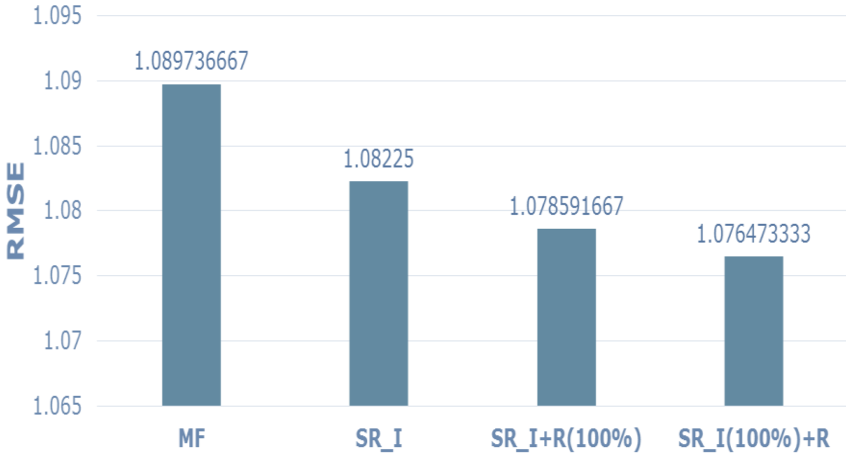

Figure 5.2: RMSE Comparison

Chapter 6 Conclusion

In this paper, we try to make use of a new data sauce, review rating, in social recommender system. We notice the problem existing in previous work that the number of similarity between users got from item rating is limited because of data sparsity. To solve this problem, we additionally consider review rating data to find other similarity between users, and we could confirm from the crawled Epinions dataset that the number of similarity got from review rating is almost 3 times more than the number of similarity got from item rating. We implement both similarity got from item rating and review rating to social recommendation model, and by experiment, we observe that our proposed method achieves a better performance than previous work.

Bibliography

[1] G. Linden, B. Smith, and J. York. Amazon.com recommendations: Item-to-item collaborative filtering. IEEE Internet computing, 7(1):76–80, 2003

[2] R. Pan, Y. Zhou, B. Cao, N. N. Liu, R. M. Lukose, M. Scholz, and Q. Yang, One-class collaborative Filtering, In IEEE International Conference on Data Mining (ICDM 2008), Pages 502-511.

[3] X. Su and T. M. Khoshgoftaar. A survey of collaborative filtering techniques, Advances in Artificial Intelligence, 2009.

[4] Y. Zhang, B. Cao, and D. Y. Yeung. Multi-domain collaborative filtering, In Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence (UAI), Catalina Island, California, USA, 2010.

[5] D. Crandall, D. Cosley, D. Huttenlocher, J. Kleinberg, and S. Suri. Feedback effects between similarity and social in online communities. In KDD '08: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 160-168, 2008. ACM

[6] H. Ma, D. Zhou, C. Liu, M. R. Lyu, I. King, Recommender Systems with Social Regularization, In Proc. WSDM '11

[7] H. Ma, H. Yang, M. R. Lyu, and I. King. Social recommendation using probabilistic matrix factorization, CIKM 2008, pages 931-940.

ACM, 2008

[8] H. Ma, I. King, and M. R. Lyu. Learning to recommend with social trust ensemble, In SIGIR 2009, pages 203-210.

[9] J. S. Breese, D. Heckerman, and C. Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In Proc. of UAI '98, 1998.

[10] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl. Grouplens: An open architecture for collaborative filtering of netnews. In Proc. of CSCW '94, 1994.

[11] R. Jin, J. Y. Chai, and L. Si. An automatic weighting scheme for collaborative filtering. In Proc. of SIGIR '04, pages 337-244, Sheffield, United Kindom, 2004.

[12] M. Deshpande and G. Karypis. Item-based top-n recommendation. ACM Transactions on Information Systems, 22(1):143-177, 2004.

[13] J. Canny. Collaborative filtering with privacy via factor analysis. In Proc. of SIGIR '02, pages 238-245, Tampere, Finland, 2002.

[14] Y. ZHANG, J. Koren. Efficient bayesian hierarchical user modeling for recommendation system. In Proc. of SIGIR '07, pages 47-54, Amsterdam, The Netherlands, 2007.

[15] A. Kohrs and B. Merialdo. Clustering for collaborative filtering applications. In Proceedings of CIMCA, 1999.

[16] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques

for recommender systems. Computer, 42(8):30–37, 2009.

[17] Au Yeung, C., Iwata, T.: Strength of social influence in trust networks in product review sites. In: WSDM 2011, ACM Press, New York, USA.

[18] J. Tang, H. Gao, X. Hu, H. Liu: Exploiting homophily effect for trust prediction. In: WSDM 2013, ACM Press, Rome, Italy.

[19] M.Jamali, M.Ester: A matrix factorization technique with trust propagation for recommendation in social networks. In proc. RecSys‘10,pages 135-142. 2010

[20] Dekang Lin. An information-theoretic definition of similarity. In ProcICML,pages 296–304, 1998.

[21] R. Salakhutdinov and A. Mnih.Probabilstic matrix factorization. In advances in Neural Information Processing Systems, volume 20, 2008

초　　록

사용자의 역사 평점기록을 통하여 추천을 해주는 협업 필터링 방법이 광범위하게 적용되고 있다. 그런데 현실 생활에서 사람들은 흔히 친구나 지인의 의견을 고려하여 판단을 내린다. 이러한 소셜 관계를 고려한 소셜 추천기법에서 친구에게 받는 영향력을 정확하게 반영하는 사용자 사이의 유사도 정보가 필요된다. 기존 방법에서는 사용자 사이의 유사도를 사용자가 아이템에 대한 평점만 고려하여 계산하였는데 본 연구는 사용자가 작성한 평론에 대한 평가를 사용자 사이의 유사도 추출에 적용하여 기존 추천 기법보다 시스템의 성능을 향상시켰다. Epinions에서 크롤링을 통하여 얻은 데이터로 소셜 추천 시스템 모델에 적용하여 실험한 결과 성능향상을 증명하였다.

주요어: 추천 시스템, 소셜 네트워크, 평론에 대한 평가, 협업 필터링, 매트릭스 인수분해

학 번: 2011-24086

감사의 글

한국으로 유학하여 김종권 교수님 연구실에 들어와서 많은 지식도 학습할 수 있었고 교수님 그리고 선배님들 또한 후배들과 인연을 맺을 수 있어서 너무 좋은 경험이 되였습니다. 처음 와서 많은 부족한 점이 있어서 교수님과 선배님들께 실수를 많이 한 것 같은데 다 저를 이해해주시고 도와주셔서 감사합니다. 연구실 2년 생활을 통하여 지식뿐만 아니라 한국 문화와 생활에 대하여 보람찬 경험을 하였습니다. 연구실에서 얻은 이러한 경험이 저의 나중의 삶에 중요한 작용을 하리라 생각합니다.

저의 연구를 지도해 주신 교수님께 진심의 감사를 전합니다. 그리고 저의 연구에 조언을 주시고 도와주신 감영명, 오하영, 최한, 박근모, 임상순, 이수철, 이규행, 노기섭, 이재훈, 윤효진 선배님들 감사합니다. 또한 처음 연구실에 들어와서 열심히 하고 있는 노태완, 김서향, 노폰 후배들께 도움을 많이 주지 못해서 아쉽습니다.

끝으로 제가 지금까지 무사히 성장할 수 있도록 항상 저를 격려하고 챙겨주신 어머니, 할머니, 할아버지 고생 많으셨습니다.