Journal of Theoretical and Applied Information Technology
28th February 2013. Vol. 48 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# SOCITEMREC: A FRAMEWORK FOR ITEM RECOMMENDATION IN SOCIAL NETWORKS

**XING XING, WEISHI ZHANG, XIUGUO ZHANG, NAN XU**
School of Information Science and Technology, Dalian Maritime University, Dalian 116026, China

**ABSTRACT**

Collaborative filtering based recommendation methods focus on user-item information for modeling the user interest. However, in social networks the user interest is influenced by other user interests in the local social circle of the active user. In this paper, considering the homophily of relation to similar interests and similar friends, we propose a social item recommendation framework (SocItemRec). Our framework combines both global interest from the user-item information and local interest from social relation information for recommendations. We evaluate our framework on real world data from Sina Weibo, one of the most popular social network sites in China. The experimental results demonstrate that our framework leads to improved performance of top-k item recommendation.

**Keywords:** *Collaborative Filtering, User Interest Model, Social Recommendation, Recommender System*

## 1. INTRODUCTION

Social networks have changed the way by which we create, share and find content in the last decade of this century. Facebook, Twitter, Sina Weibo, Google+ – the number of social network sites available for individuals to create content seems infinite. The amount of information is increasing so quickly that users can't handle the information overload without the support of recommendation services in social networks.

One of the most promising technologies that address the information overload issue is collaborative filtering (CF). CF techniques have been successful in information filtering applications such as Amazon [1], Netflix [2]. CF based recommendation methods build on the intuition that if users *u* and *u'* have historically had similar interests on some items, they are likely to be interested in other items similarly. CF techniques exploit historical records of user-item data for future prediction. However, CF techniques could fail in the context of social networks [3], where user interests have high correlation between the active user and his or her friends with similar interests.

In social networks people with similar interest tend to connect to each other; moreover, people of similar interest are more likely to be friends. This social phenomenon of homophily [4, 5] along with the intuition of CF provides us a foundation to combine both user social features (e.g. user social relations) and user-item information (e.g. user like-dislike items, user click-through information [6]) for improving the quality of recommendation in the context of social networks.

In this paper, we focus on improving the performance of recommendation in social networks by introducing a recommendation framework SocItemRec, which characterizes both the global interest and local interest in social networks. In particular, we take use-item click information and user-user social relationship information for investigation.

Firstly, we use traditional CF techniques to model the global interest from use-item information. Second, we model the strength of social relationship by common friends between users in social networks. And then we model the local interest which characterizes interest influences in the local social circles of the active users. Finally, we combine the global interest and local interest for top-k item recommendation in social networks. Experiments are conducted on a real dataset collected from Sina Weibo, and results show that our framework outperforms the CF-based method for top-k item recommendation.

The rest of paper is organized as follows. Section 2 gives an overview of related work. In Section 3 we introduce the proposed framework. Section 4 describes the experiments. We make the conclusion in Section 5.

## 2. RELATED WORK

Collaborative filtering [7-9] techniques make predictions about the interests of active users with the assumption that those users, who had similar interests on some items, are likely to be interested

Journal of Theoretical and Applied Information Technology
28th February 2013. Vol. 48 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

in other items similarly. CF techniques exploiting historical records of user-item data for future prediction are based on neighborhood based [2] or latent variable based [10, 11] methods.

In neighborhood based CF methods, such as Tapestry [7], GroupLens [8] and Amazon [1], the recommendation for an item is computed as the weighted average of ratings given by a group of people called neighbors with similar interests to the active user. It relies on a few significant neighborhood relations, often ignoring the vast majority of ratings by a user.

In contrast to neighborhood based methods, which use the stored ratings directly in the prediction, latent variable based approaches use these ratings to learn a probabilistic latent model. Latent variable models such as Latent Semantic Models [10] and Latent Dirichlet Allocation [12] for CF uncover latent causes that explain observed ratings. However, the traditional CF techniques have limitations in the context of social recommendation [3, 13], for not effectively modeling the user interests and social relationship simultaneously for recommendation in social networks.

Another direction of related research has focused on social filtering [14-16], which utilizes user relationships by applying Random Walks [17] in social networks to obtain recommendations. Yildirim et al. [18] proposed a novel recommendation algorithm which performs Random Walks on a bipartite graph to represent the similarity between items. Craswell and Szummer [6] built two random walk processes to propagate query similarity along the clickthrough data graph and obtained a good performance of item recommendation. Deng et al. [19] proposed a generalized Co-HITS algorithm based on bipartite graph for recommendation.

## 3. OUR FRAMWORK

In this section, we propose a framework for top-k item recommendation (SocItemRec), which characterizes both the global interest of the active user from user-item information and the local interest throughout his or her local social circle.

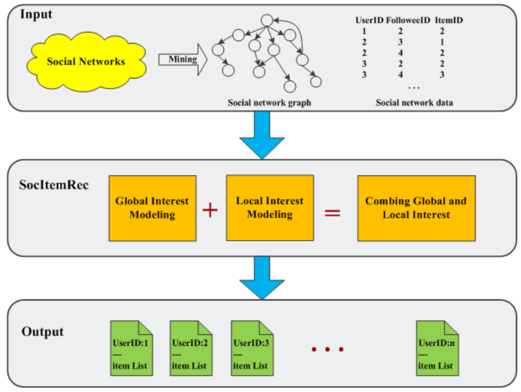


*Figure 1: The framework of SocItemRec*

As shown in Figure 1, SocItemRec has three components as follows.

- global interest modeling
- local interest modeling
- combing global and local interest

We describe the functionalities of each component in details in the following three subsections.

### 3.1 Global interest modeling

We firstly model the global interest of active users using the user-item click logs. Let us introduce a click indicator $click_k \in \{0,1\}$, such that $click_k = 1$ if the *kth* item is clicked by the user *u*, otherwise, $click_k = 0$. Then the user interest can be modeled by the user-item click logs with a form of $\mathbf{like}(u) = [0\ 1\ 0\ \ldots\ 1\ 0\ 1]^T$.

According to the *Jaccard similarity coefficient* [1], the similarity between items *i* and *j* is given by

$$sim(i,j) = \frac{|\mathrm{U}_i \cap \mathrm{U}_j|}{|\mathrm{U}_i \cup \mathrm{U}_j|} \qquad (1)$$

where $\mathrm{U}_i$ denotes the set of users who click item *i* and $\mathrm{U}_j$ denotes the set of users who click item *j*. The similarity matrix $\mathbf{Similarity} = \left(sim(i,j)\right)_{n\times n}$ is introduced to denote the similarity between all items to all items.

User interest can be computed from the matrix **Similarity**, which contains implicit user interest data and the vector **like**, which indicates the items the user clicked. The user global interest is calculated by

$$I^{global}(u) = \mathbf{Similarity} \times \mathbf{like}(u) \qquad (2)$$

The global interest modeling method is a CF-based user interest computation method. In addition, it provides a baseline for top-k recommendation if there is no social information available for active users in social networks.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

### 3.2 Local interest modeling

Social networks allow users to maintain a list of followees, and they can communicate with each other through social relationships. In this case, user interest propagates in the local social circle of the active users. Inspired by this intuition, we use the social action data to model the strength of social relationships and we build a weighted social network based on the scores of the pairwise social relationship between the active user and his followees.

To compute the pair-wise strength of followship between user *u* and his followee *u'*, we use a partial strength method as follow.

$$strength(u,u') = \frac{|F_u \cap F_{u'}|+1}{|F_u|} \quad (3)$$

where $F_u$ denotes the set of followees of the active user $u$ , and $|F_u \cap F_{u'}|$ denotes the number of common followees of *u* and *u'*.

And we normalize the strength of followship as

$$p(u'|u) = \frac{1}{\boldsymbol{Z}} strength(u,u') \quad (4)$$

where $\boldsymbol{Z}$ is a normalization factor, $\boldsymbol{Z} = \sum_{u' \in Followee(u)} strength(u,u')$ , and $p(u'|u)$ yields $\sum_{u' \in Followee(u)} p(u'|u) = 1$.

Inspired by the homophily [2, 3] of social networks, the conditional probability of *u'* given *u* above can be interpreted as the interest propagation from followee *u'* to active user *u*, where the direction of interest propagation is from followees to the active user. This measure gives us a chance to investigate the interest propagation throughout the local social networks.

The interest of the active user is influenced by the interest propagation from his followees and followees of followees in his social circles. And different strength of social relationship between active user and the followees plays a different role. In order to compute the influence of interest propagating in the local followship networks, the local interest propagation of active user $u$ is calculated by

$$\lambda^{local}(u) = \sum_{i=1}^{d} \lambda^{(i)}(u) \quad (5)$$

where

$$\lambda^{(1)}(u) = \sum_{u' \in Followee(u)} \lambda(u') w(u,u')$$

$$\lambda^{(2)}(u) = \sum_{u'' \in Followee(u')} \lambda(u'') w(u,u'')$$

$$\vdots$$

$$\lambda^{(d)}(u) = \sum_{u^{(d)} \in Followee(u^{(d-1)})} \lambda(u^{(d)}) w(u,u^{(d)})$$

and

$$w(u,u^{(d)}) = \min(w(u,u'),...,w(u^{(d-1)},u^{(d)})) \quad (6)$$

where $u^{(0)} = u$ , $w(u,u') = p(u'|u)$ in Eq. (4), *d* is the maximum depth for searching the social graph.

The local interest propagation models all the influences between the active user and the followees within the depth *d* in his local followship network. This guarantees the high quality estimates of the user local interest in social recommendation scenarios.

### 3.3 Combing global and local interest

To consider both the global interest generated from user-item data and local interest from the influents of user local social networks, we combine the user interest as follows.

$$\lambda(u) = \lambda^{global}(u) + \lambda^{local}(u) \quad (7)$$

Based on the scores obtained by Eq. (7), we sort the user interest in a descending manner, and then we recommend a top-k item list that active users like most.

The idea of SocItemRec is to integrate both interests and followship networks to connect a user to both items of potential interest and other users with similar interests. SocItemRec provides a single unified framework to encode both user global interest information and local interest information influenced by his or her local social circles in social networks.

## 4. EXPERIMENTS

In this section, we evaluate our method on Sina Weibo and report our experimental results of the performance for top-k recommendations.

### 4.1 Dataset

Sina Weibo is one of the most popular SNS in China, which has more than 300 million registered users. Like Twitter and Facebook, it contains rich user interest data and social features for mining and analysis. Moreover, it provides mechanism for researchers and developers to access the public data of the platform. We collect a subset of features (e.g. followees, recommended items, items clicked by active users) from the public data on Sina Weibo. We start with a small seed set of 5 random users, and expand the user base according to their followee lists in a breadth-first manner. We stop searching at the depth of 6 according to *Six degrees of separation*.

The dataset has 108 items, 12,156 users, 98,093 followship links, 175,431 recommended items and 26,357 item click interactions. The user-item data is very sparse, with the sparsity level of 0.9799. In

Journal of Theoretical and Applied Information Technology
28th February 2013. Vol. 48 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

contrast, the followship network is relatively dense as each user has 8 followees on average, and the users are high correlated with their followees. The sparsity [4] is calculated by

$$\text{sparsity}=1-\frac{\#\text{nonzero elements}}{\#\text{total elements}} \tag{8}$$

To evaluate the performance of the top-k recommendations, we divide the dataset by 80% dataset used as training set and 20% test set. We conduct our experiments on the training set, and evaluate the performance of top-k recommendation on the test set.

### 4.2 Experimental settings

In our experiments, we use four evaluation metrics, which are recall-at-*k* (recall@*k*), precision-at-*k* (precision@*k*), F1-at-*k* (F1@*k*) and average-precision-at-*k* (AP@*k*), to evaluate the performance of top-k recommendation methods [20].

Let $n$ be the number of active user, $hits(u)$ be the number of items clicked by the active user $u$ in the top-*k* recommended items and $n^{test}(u)$ be the number of items clicked by $u$ in the test set, the **recall**@*k*, **precision**@*k* and **F1**@*k* are computed respectively by

$$\mathbf{recall}@k=\frac{1}{n}\sum_u r(u)@k \tag{9}$$

where $r(u)@k=\dfrac{hits(u)}{n^{test}(u)}$;

$$\mathbf{precision}@k=\frac{1}{n}\sum_u p(u)@k \tag{10}$$

where $p(u)@k=\dfrac{hits(u)}{k}$;

$$\mathbf{F1}@k=\frac{1}{n}\sum_u \mathbf{F1}(u)@k \tag{11}$$

where $\mathbf{F1}(u)@k=2\times\dfrac{r(u)@k\times p(u)@k}{r(u)@k+p(u)@k}$.

For the proposed method that recommends a top-*k* ranked list of items, we care much more about the accuracy of the top *m* ranked items and we prefer to the algorithms that detect more hits earlier on. Inspired by this expectation, we use the evaluation metric **AP**@*k*, which is computed by

$$\mathbf{AP}@k=\frac{1}{n}\sum_u \mathbf{AP}(u)@k,\ \ \mathbf{AP}(u)@k=\sum_{m=1}^{k} p(u)@m\times\Delta r(u)@m \tag{13}$$

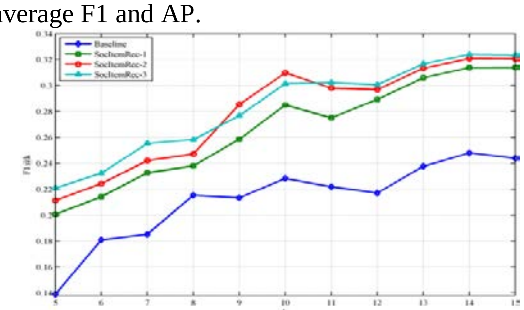
where $n$ is the number of active user and $\Delta r(u)@m$ is the change in the recall from items *m-1* to *m*.

We compare the following methods for top-k item recommendation in social networks.

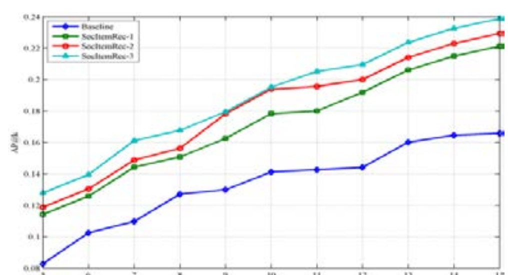
- **Baseline**: a traditional CF recommendation method for item recommendation. In particular, we use the item-based CF recommendation method [21] as the baseline for comparison.
- **SocItemRec-*d***: our framework combing both the global interest and local interest for top-k item recommendation in social networks, and *d* is the searching depth, which is determined in the experiments.

### 4.3 Experimental results

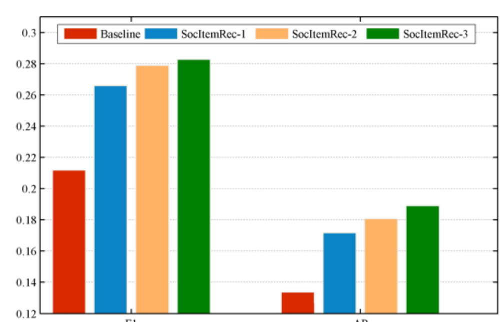
The performance of different version of our method and comparison baseline method for top-k item recommendation in social networks is shown in Table 1 and Table 2. The results show that SocItemRec-3 outperforms the baseline method and the other two version of SocItemRec in terms of average F1 and AP.



(a) F1@k



(b) AP@k




(c) Average F1 and AP

*Figure 2: Performance Of Different Recommendation Methods*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The baseline method [21] ignores the impact of interest propagation in the local social circles of the active user, thus the performance of the baseline method is relative low comparing with SocItemRec in Figure 2 (c). By taking user local interest into account, SocItemRec improves the performance in terms of average F1 and AP.

As shown in Table 1, Table 2 and Figure 2, we see that different settings of *d* have different impacts on the performance of SocItemRec. Small values for *d* tend to overly reward the influences of close followees, while ignore the influences of remote ones. Hence, SocItemRec-3 obtains the best performance and the parameter *d* determines the performance of SocItemRec.

In top-k item recommendation scenarios, the number of items for recommendation is also an important factor for evaluations. We vary *k* from 5 to 15 to evaluate the sensitivity of performance. The results in Figure 2 (a) and (b) show that when *k* increases, the performance in terms of F1@k and AP@k increases. The reason is that more items for recommendation, more potential items are clicked by users, higher scores of F1@k and AP@k obtain.

*Table 1: Performance In Terms Of F1@K*

| *k* | **F1@*k*** | | | |
|---|---|---|---|---|
| | Baseline | *d* = 1 | *d* = 2 | *d* = 3 |
| 5 | 0.1390 | 0.2006 | 0.2113 | 0.2208 |
| 6 | 0.1809 | 0.2142 | 0.2243 | 0.2324 |
| 7 | 0.1852 | 0.2326 | 0.2421 | 0.2553 |
| 8 | 0.2154 | 0.2380 | 0.2471 | 0.2582 |
| 9 | 0.2134 | 0.2585 | 0.2853 | 0.2766 |
| 10 | 0.2283 | 0.2849 | 0.3097 | 0.3014 |
| 11 | 0.2219 | 0.2750 | 0.2980 | 0.3022 |
| 12 | 0.2170 | 0.2892 | 0.2968 | 0.3006 |
| 13 | 0.2377 | 0.3059 | 0.3132 | 0.3166 |
| 14 | 0.2480 | 0.3137 | 0.3207 | 0.3238 |
| 15 | 0.2438 | 0.3138 | 0.3205 | 0.3233 |
| **Average** | **0.2119** | **0.2660** | **0.2790** | **0.2828** |

*Table 2: Performance In Terms Of AP@K*

| *k* | **AP@*k*** | | | |
|---|---|---|---|---|
| | Baseline | *d* = 1 | *d* = 2 | *d* = 3 |
| 5 | 0.0827 | 0.1141 | 0.1187 | 0.1278 |
| 6 | 0.1025 | 0.1258 | 0.1304 | 0.1395 |
| 7 | 0.1097 | 0.1442 | 0.1487 | 0.1611 |
| 8 | 0.1271 | 0.1507 | 0.1562 | 0.1676 |
| 9 | 0.1299 | 0.1625 | 0.1781 | 0.1794 |
| 10 | 0.1412 | 0.1782 | 0.1938 | 0.1951 |
| 11 | 0.1426 | 0.1800 | 0.1956 | 0.2051 |
| 12 | 0.1441 | 0.1919 | 0.1999 | 0.2095 |
| 13 | 0.1601 | 0.2060 | 0.2141 | 0.2236 |
| 14 | 0.1646 | 0.2149 | 0.2229 | 0.2325 |
| 15 | 0.1657 | 0.2212 | 0.2292 | 0.2387 |
| **Average** | **0.1337** | **0.1718** | **0.1807** | **0.1891** |

## 5. CONCLUSION

In this paper, we study the top-k item recommendation in a real scenario on Sina Weibo. Considering the homophily of relation to similar attributes (interests) and similar nodes (friends), we propose an social item recommendation framework (SocItemRec), which incorporates the social relation features into the recommendation method. Our framework combines the user global interest and local interest influenced by similar friends in social networks, improving the performance of top-k item recommendation. We conduct the experiments to compare different versions of our methods based on SocItemRec with the CF-based method. The experimental results demonstrate that SocItemRec outperforms the CF-based method.

## ACKNOWLEDGEMENTS

This research is supported by the National Natural Science Foundation of China under Grant No.61272172 and No.60973013, and the Fundamental Research Funds for the Central Universities in China under Grant No.2011QN027.

## REFERENCES:

[1] G. Linden, B. Smith, J. York, "Amazon.com recommendations: item-to-item collaborative filtering", *Internet Computing, IEEE*, Vol. 7, No. 1, 2003, pp. 76-80.

[2] Y. Koren, "Factorization meets the neighborhood: a multifaceted collaborative filtering model", Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2008 of Conference, pp. 426-434.

[3] G. Groh, C. Ehmig, "Recommendations in taste related domains: collaborative filtering vs. social filtering", Proceedings of the 2007 international ACM conference on Supporting group work, ACM, 2007 of Conference, pp. 127-136.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[4] A. Anagnostopoulos, R. Kumar, M. Mahdian, "Influence and correlation in social networks", Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2008 of Conference, pp. 7-15.

[5] M. McPherson, L. Smith-Lovin, J. M. Cook, "Birds of a Feather: Homophily in Social Networks", *Annual Review of Sociology*, Vol. 27, No. 1, 2001, pp. 415-444.

[6] N. Craswell, M. Szummer, "Random walks on the click graph", Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, ACM, 2007 of Conference, pp. 239-246.

[7] D. Goldberg, D. Nichols, B. M. Oki, et al., "Using collaborative filtering to weave an information tapestry", *Commun. ACM*, Vol. 35, No. 12, 1992, pp. 61-70.

[8] J. A. Konstan, B. N. Miller, D. Maltz, et al., "GroupLens: applying collaborative filtering to Usenet news", *Commun. ACM*, Vol. 40, No. 3, 1997, pp. 77-87.

[9] B. Sarwar, G. Karypis, J. Konstan, et al., "Item-based collaborative filtering recommendation algorithms", Proceedings of the 10th international conference on World Wide Web, ACM, 2001 of Conference, pp. 285-295.

[10] T. Hofmann, "Latent semantic models for collaborative filtering", *ACM Transactions on Information Systems*, Vol. 22, No. 1, 2004, pp. 89-115.

[11] R. Jin, L. Si, C. Zhai, "A study of mixture models for collaborative filtering", *Information Retrieval*, Vol. 9, No. 3, 2006, pp. 357-382.

[12] D. M. Blei, A. Y. Ng, M. I. Jordan, "Latent dirichlet allocation", *J. Mach. Learn. Res.*, Vol. 3, 2003, pp. 993-1022.

[13] S.-H. Yang, B. Long, A. Smola, et al., "Like like alike: joint friendship and interest propagation in social networks", Proceedings of the 20th international conference on World wide web, ACM, 2011 of Conference, pp. 537-546.

[14] M. Roth, T. Barenholz, A. Ben-David, et al., "Suggesting (more) friends using the implicit social graph", 28th International Conference on Machine Learning, ICML 2011, June 28, 2011 - July 2, 2011, Association for Computing Machinery, 2011, pp. lxvii-lxxiv.

[15] E. Bakshy, J. M. Hofman, W. A. Mason, et al., "Everyone's an influencer: quantifying influence on twitter", Proceedings of the fourth ACM international conference on Web search and data mining, ACM, 2011 of Conference, pp. 65-74.

[16] R. Xiang, J. Neville, M. Rogati, "Modeling relationship strength in online social networks", Proceedings of the 19th international conference on World wide web, ACM, 2010 of Conference, pp. 981-990.

[17] M. Jamali, M. Ester, "TrustWalker: a random walk model for combining trust-based and item-based recommendation", Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2009 of Conference, pp. 397-406.

[18] H. Yildirim, M. S. Krishnamoorthy, "A random walk method for alleviating the sparsity problem in collaborative filtering", Proceedings of the 2008 ACM conference on Recommender systems, ACM, 2008 of Conference, pp. 131-138.

[19] H. Deng, M. R. Lyu, I. King, "A generalized Co-HITS algorithm and its application to bipartite graphs", Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2009 of Conference, pp. 239-248.

[20] G. Karypis, "Evaluation of Item-Based Top-N Recommendation Algorithms", Proceedings of the tenth international conference on Information and knowledge management, ACM, 2001 of Conference, pp. 247-254.

[21] M. Deshpande, G. Karypis, "Item-based top-N recommendation algorithms", *ACM Transactions on Information Systems*, Vol. 22, No. 1, 2004, pp. 143-177.