International Journal of Research in Engineering, Science and Management
Volume-2, Issue-1, January-2019
www.ijresm.com | ISSN (Online): 2581-5792

Product Purchase Recommendation of User by Data Analysis using Data Mining

Devhadkar Shivam Ramkisan[1], Kardile Suvarna Changdev[2], Pandit Nutan Dhanendra[3]
[1,2,3]*Student, Dept. of Computer Engg., Adsul Technical Campus Faculty of Engg. & Mgmt., Ahemadnagar, India*

Abstract**: With the boom of social media, it is a very popular trend for people to share what they are doing with friends across various social networking platforms. Nowadays, we have a vast amount of descriptions, comments, and ratings for local services. The information is valuable for new users to judge whether the services meet their requirements before partaking. In this paper, we propose a user-service rating prediction approach by exploring social users' rating behaviors. In order to predict user-service ratings, we focus on users' rating behaviors. In our opinion, the rating behavior in recommender system could be embodied in these aspects: 1) when user rated the item, what the rating is, 2) what the item is, 3) what the user interest that we could dig from his/her rating records is, and 4) how the user's rating behavior diffuses among his/her social friends. Therefore, we propose a concept of the rating schedule to represent users' daily rating behaviors. In addition, we propose the factor of interpersonal rating behavior diffusion to deep understand users' rating behaviors. In the proposed user-service rating prediction approach, we fuse four factors, user personal interest (related to user and the item's topics), interpersonal interest similarity (related to user interest), interpersonal rating behavior similarity (related to users' rating behavior habits), and interpersonal rating behavior diffusion (related to users' behavior diffusions), into a unified matrix-factorized framework. We conduct a series of experiments in Yelp dataset and Douban Movie dataset. Experimental results show the effectiveness of our approach.**

Keywords**: Data Mining, Natural Language Understanding Sentiment Analysis, Natural Language Processing, Centralized Data Availability, Statistic Report.**

1. Introduction

Recently people have been receiving more and more digitized information from Internet, and the volume of information is larger than any other point in time, reaching a point of information overload. To solve this problem, the recommender system has been created in response to the need to disseminate so much information. It does not only filter the noise, but also help to select attractive and useful information. Recommender system has achieved initial prescribed, although the various table text styles are provided. The formatter will need to creat success based on a survey that shows at least 20 percent of sales on Amazon's website come from the recommender system. These components, incorporating the applicable criteria that follow.

Social networks gather volumes of information contributed by users around the world. This information is versatile. It always contains item/services descriptions (including textual descriptions, logos and pictures), users' comments, moods and users' social circles, prices, and locations. It is very popular for recommending users' favorite services from crowd-source contributed information.

In 1994, the GroupLens system [1] utilized a CF (collaborative filtering) algorithm based on common users' preferences, known as user-based CF. The authors note that users will favor items recommended by users with similar interests. Sarwar et al. [2] proposed an item-based CF in 2001. The authors found that users favor items similar to those in which the user was previously interested. These are the most famous recommender system algorithms. The basic idea of CF is grouping users or items according to similarity. Most recent work has followed the two aforementioned directions (i.e., user-based and item-based). For example, Herlocker et al. [3] propose the similarity between users or items according to the number of common ratings. Deshpande and Karypis [4] apply an item-based CF combined with a condition-based probability similarity and Cosine Similarity. Collaborative filtering-based recommendation approaches [5]-[8], [13] can be viewed as the first generation of recommender system [10].

However, with the rapid increase in number of registered Internet users and more and more new products available for purchase online, the issue of cold start for users and sparsity of datasets has become increasingly intractable. Fortunately, with the popularity and rapid development of social networks, more and more users enjoy sharing their experiences, such as reviews, ratings, photos and moods. The interpersonal relationships have become transparent and opened up as more and more users share this information on social media websites such as Facebook, Twitter, Yelp, Douban, Epinions [10], etc. The circles of friends also bring opportunities and challenges for a recommender system to solve the issues of cold start and sparsity.

Many models based on social networks [10-14] have been proposed to improve recommender system performance. The concept of 'inferred trust circle' based on circles of friends was proposed by Yang et al. [10] to recommend favorite and useful items to users. Their approach, called the CircleCon Model, not only reduces the load of big data and computation complexity, but also defines the interpersonal trust in the complex social networks. Besides interpersonal influence, Jiang et al. [12]

prove that individual preference is also an important factor in social networks. The above algorithms are based on the probabilistic matrix factorization model [12]. Symeonidis et al. [11] propose Social-Union, a method which combines similarity matrices derived from heterogeneous (unipartite and bipartite) explicit or implicit social rating networks, and generalize their model for combining multiple social networks. Lee et al. [13] propose a recommender system that uses the concepts of experts to find both novel and relevant recommendations. Wang et al. [14] design a joint social-content recommendation framework to suggest users for videos that users are likely to import or re-share in the online social network. Meanwhile, there are some interesting works to infer social contexts. For example, Servia-Rodriguez et al. [12] proposed a model to infer social contexts by several Natural Language Processing and data mining techniques over users' interaction data on Facebook. Fang et al. [13] propose a relational latent SVM model to combine user features, attribute inference, and attribute relations. Mao et al. [6] propose to model user's vocal competence for personalized song recommendation.

Except for ratings prediction and products recommendations, location-based social networks (LBSNs) are attracting more and more users' attention [10]. Cho et al. [13] have developed a model of human mobility that combines periodic short range movements with travel based on the social network structure. Cheng et al. [14] fuse matrix factorization (MF) with geographical and social influence for POI (Point-of-interest) recommendations on LBSNs, and propose a novel Multi-center Gaussian Model to model the geographical influence of users' check-in behaviors. Jiang et al. [9] propose a user topic based collaborative filtering approach for personalized travel recommendation. Zahálka et al. [11] propose an interactive and multimodal content-based venue explorer based on location-based social networks. Chen et al. [13] propose to conduct personalized travel recommendation by taking user attributes and social information. Furthermore, there are some previous works [10], [11] focusing on objective evaluation in order to recommend the high-quality services by exploring social users' spatial-temporal information.

In this paper, we propose a user-service rating prediction model based on probabilistic matrix factorization by exploring rating behaviors. Usually, users are likely to participate in services in which they are interested and enjoy sharing experiences with their friends by description and rating. Like the saying "birds of a feather flock together," social users with similar interests tend to have similar behaviors. It is the basis for the collaborative filtering based recommendation model. Social users' rating behaviors could be mined from the following four factors: personal interest, interpersonal interest similarity, interpersonal rating behavior similarity, and interpersonal rating behavior diffusion. Why do we consider these four factors? In our opinion, the rating behavior in recommender system could be embodied in these aspects: when user rated the item, what the rating is, what the item is, what the user interest we could dig from his/her rating records is, and how user's rating behavior diffuse among his/her social friends. In this paper, we propose a user-service rating prediction approach by exploring social users' rating behaviors in a unified matrix factorization framework.

The main contributions of this paper are shown as follows:

- We propose a concept of the rating schedule to represent user daily rating behavior. We leverage the similarity between user rating schedules to represent interpersonal rating behavior similarity.
- We propose the factor of interpersonal rating behavior diffusion to deep understand users' rating behaviors. We explore the user's social circle, and split the social network into three components, direct friends, mutual friends, and the indirect friends, to deep understand social users' rating behavior diffusions.
- We fuse four factors, personal interest, interpersonal interest similarity, interpersonal rating behavior similarity, and interpersonal rating behavior diffusion, into matrix factorization with fully exploring user rating behaviors to predict user-service ratings. We propose to directly fuse interpersonal factors together to constrain user's latent features, which can reduce the time complexity of our model.

The rest of this paper is organized as follows: In Section II, we define the problem we focus on in this paper, and briefly introduce comparable algorithms. In Section III, the proposed user-service rating prediction approach combining with personal interest, interpersonal interest similarity, interpersonal rating behavior similarity, and interpersonal rating behavior diffusion is introduced in detail. The experimental results and some discussions are given in Section IV. The main conclusions are drawn in Section V.

2. Preliminary

In this paper, we focus on probabilistic matrix factorization. Thus, in this section, we first define the notations which are utilized in this paper, and then review the compared approaches in this domain.

A. Problem formulation

We define the notations which are utilized in this paper. The proposed model aims to predict unknown ratings in social rating networks (like Yelp1, Epinions2). We utilize latent feature vectors to predict user ratings.

- *Notations and their descriptions*

We extract a set of users U=$\{u_1,\cdots,u_M\}$ and a set of items P=$\{i_1,\cdots,i_N\}$ from our dataset, which we collect from Yelp and Douban Movie3 website. We set a rating matrix $\boldsymbol{R}=[R_{u,i}]_{M\times N}$ which represents ratings matrix, where $R_{u,i}$ denotes the rating of user u to item i. The ratings may be any real number in different rating networks, but in the Yelp dataset they are integers ranging from 1 to 5.

There are four significant parameters which represent the factors we consider. The interest similarity values are given in matrix $\boldsymbol{W}=[W_{u,v}]_{M\times M}$, where $W_{u,v}\in[0,1]$ denotes the interest similarity between user u and user v. The rating behavior similarity values are given in matrix $\boldsymbol{E}=[E_{u,v}]_{M\times M}$, where $E_{u,v}\in[0,1]$ denotes the rating behavior similarity between user u and user v. The smooth degree of interpersonal rating behavior diffusions between users is represented by matrix $\boldsymbol{D}=[D_{u,v}]_{M\times M}$. The last factor of users' personal interest is represented by matrix $\boldsymbol{Q}=[Q_{u,i}]_{M\times N}$, where $Q_{u,i}\in[0,1]$ denotes the relevance between user *u*'s interest and the topic of item *i*.

The task of the proposed algorithm is as follows: Given a user $u\in$U and an item $i\in$P, whose rating $R_{u,i}$ is unknown, we predict the rating of *u* to *i* using *R*, *W*, *E*, *D* and *Q* based on the probabilistic matrix factorization model.

We train the latent features of users and items with matrix factorization techniques [11]-[13], [14], [10] in this paper, and predict the unknown ratings using these latent features. We set $\boldsymbol{U}\in\mathbb{R}^{M\times k}$ and $\boldsymbol{P}\in\mathbb{R}^{N\times k}$ as user and item latent features matrices, in which row vectors $\boldsymbol{U}_u$ and $\boldsymbol{P}_i$ represent *k*-dimensional user and item latent feature vectors. Certainly *k* is much less than *M*and *N*. Moreover, $\boldsymbol{U}_u$ and $\boldsymbol{P}_i$ can be seen as the brief characterization of user *u* and item *i*. The goal of matrix factorization is to learn these latent features and exploit them to predict user-service ratings.

B. Compared algorithms

In this subsection, we will review some major approaches about social factors in this domain, and all of them focus on probabilistic matrix factorization. The basic matrix factorization model [14] without any social factors, the CircleCon model [10] with the factor of interpersonal trust values, the Social Contextual (Context MF) model [12] with interpersonal influence and individual preference, and the PRM model [11], [12] with more factors will be outlined.

1) Basic matrix factorization

As a basic model, the basic probabilistic matrix factorization (BaseMF) approach [12] will be reviewed first, without any social factors taken into consideration. They learn the latent features by minimizing the objective function on the observed rating data $\boldsymbol{R}$:

$\Psi(\boldsymbol{R},\boldsymbol{U},\boldsymbol{P})=12\Sigma(Ru,i-\hat{R}u,i)2u,i+\lambda 2(\|\boldsymbol{U}\|F2+\|\boldsymbol{P}\|F2)$ (1)

Where $\hat{R}u,i$ denotes the ratings predicted by:

$\hat{\boldsymbol{R}}=r+\boldsymbol{UP}T$ (2)

Where *r* is an offset value, which is empirically set as users' average rating value in the training data. Ru,i is the real rating values in the training data for item *i* from user *u*. *U* and *P* are the user and item latent feature matrices which need to be learned from the training data. $\|\boldsymbol{X}\|F$ is the Frobenius norm of matrix *X*, and $\|\boldsymbol{X}\|F=(\Sigma xi,j2i,j)12/$. It is used to avoid over-fitting [13]. This objective function can be minimized efficiently using the gradient descent method as in [13]. Once the low-rank matrices *U* and *P* are learned, rating values can be predicted according to (2) for any user-item pairs.

2) CircleCon model

This approach [10] focuses on the factor of interpersonal trust in social network and infers the trust circle. The trust value of user-user is represented by matrix *S*. Furthermore, the whole trust relationship in social network is divided into several sub-networks *Sc*, called inferred circle, and each circle is related to a single category *c* of items. The basic idea is that user latent feature *Uu* should be similar to the average of his/her friends' latent features with weight of $Su,vc*$ in category *c*. Once the model is trained in *c*, the rating value in *c* can be predicted according to (2).

3) Context MF

Besides the factor of interpersonal influence, Jiang *et al*. [12] propose another important factor: the individual preference. Interpersonal preference similarity is mined from the topic of items adopted from the receiver's history. The basic idea is that user latent feature *Uu* should be similar to his/her friends' latent feature with the weight of their preference similarity in social networks.

4) PRM

In our previous work [11], [14], we consider three social factors to constrain user and item latent features, involving interpersonal influence, interpersonal interest similarity, and personal interest. The basic idea of interpersonal interest similarity is that user latent feature *Uu* should be similar to his/her friends' latent feature with the weight of interpersonal interest similarity $Wu,v*$. The factor of personal interest $Qu,i*$ focuses on mining the degree of user interest to an item.

5) Differences

In this paper, we consider four factors, personal interest $Qu,i*$ (related to user and the item's topics), interpersonal interest similarity $Wu,v*$ (related to user interest), interpersonal rating behavior similarity $Eu,v*$ (related to users' rating behavior habits), and interpersonal rating behavior diffusion $Du,v*$ (related to users' behavior diffusions), to explore users' rating behaviors.

The differences between our work and previous works are:

1) We focus on exploring user rating behaviors. A concept of the rating schedule is proposed to represent user daily rating behavior. The factor of interpersonal rating behavior diffusion is proposed to deep understand users' rating behaviors. We consider these two factors to explore users' rating behaviors.
2) We fuse three factors, interpersonal interest similarity, interpersonal rating behavior similarity, and interpersonal rating behavior diffusion, together to directly constrain users' latent features, which can reduce the time complexity.

3. The approach

In this paper, in order to predict user-service ratings, we focus on users' rating behaviors. We fuse four factors, personal interest, interpersonal interest similarity, interpersonal rating behavior similarity, and interpersonal rating behavior diffusion, into matrix factorization. Among these factors, interpersonal rating behavior similarity and interpersonal rating behavior diffusion are the main contributions of our approach. Hereinafter we turn to the details of our approach.

International Journal of Research in Engineering, Science and Management
Volume-2, Issue-1, January-2019
www.ijresm.com | ISSN (Online): 2581-5792

Table 1
An example of the rating schedule. The schedule shows the statistic of the rating behavior given by user's rating historical records

Day / Rating	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
1				1			
2	3		2		1		
3		7		1	6	4	2
4		2	4		3	5	2
5	9		2				1

A. User rating behavior exploration

The factors of interpersonal interest similarity $W_{u,v}^{*}$ and personal interest $Q_{u,i}^{*}$ proposed in [13], [12] have been proved effective. Thus, in this subsection, we turn to the details of our proposed interpersonal rating behavior similarity and interpersonal rating behavior diffusion.

1) Interpersonal rating behavior similarity

The behavior habit is essential. It could not be separated from temporal information. Thus, we define rating behavior in this paper as what the user has done and when it happened. This kind of behavior presentation arouses us to the curriculum schedule. The schedule arranges which course would we take and when we should go to class. From the schedule it can be sensed that the student's daily study behavior. Therefore, we put forward a concept of the rating schedule shown in Fig. 1.

We leverage a rating schedule for the statistic of the rating behavior given by user's rating historical records. For example, the user has rated an item 1 star and another 3 stars on Thursday. It can be seen that the user has little possibility to take rating behavior on Thursday. We leverage this kind of rating schedule to represent users' rating behaviors. The behavior similarity could embody user latent features similarity to some extent. For example, a student's curriculum schedule could represent his/her study behavior to a certain degree. If the student's curriculum schedule is similar with another student, we could infer that they have similar study behaviors, and furthermore, they may be classmates. Thus, we could extend it to the rating schedule to calculate interpersonal rating behavior similarity.

We set a rating behavior matrix $B^u = [B^u_{r,d}]_{X\times Y}$, which represents user u's rating behavior, where $B^u_{r,d}$ denotes the behavior count that user u has rated r stars in day d. In this paper, we set the rating schedule in the type of the week from Monday to Sunday, and the rating is integer in the range of 1 to 5. That is to say, X and Y are set as 5 and 7 respectively in this paper. where $E_{u,v}$ denotes the rating behavior similarity between user u and his/her friend v. The basic idea of interpersonal rating behavior similarity is that user u's rating schedule should be similar to his/her friend v to some extent. In order to be fair in measuring the similarity degree, each row of $\boldsymbol{E}$ is normalized to unity $\Sigma E_{u,v} c{*}v=1$.

In this paper, we consider the factor of social users' rating behavior diffusions. We explore the diffusion of user rating behavior by combining the scope of user's social network and the temporal information of rating behaviors. For a user, we split his/her social network into three components, direct friends, mutual friends, and the indirect friends.

4. Experiments

Yelp is a local directory service with social networks and user reviews. It is the largest review site in America. Users rate the businesses, submit comments, communicate experience, etc. It combines local reviews and social networking functionality to create a local online community. Yelp dataset4 contains eight categories, including Active Life, Beauty & Spas, Home Services, Hotels & Travel, Night Life, Pets, Restaurants, and Shopping. More details are shown in our previous work [10]. We experiment with 80% of each user's rating data randomly as the training set and the rest 20% of each user's rating data as the test set in each category, to ensure all users' latent features are learned.

5. System diagram

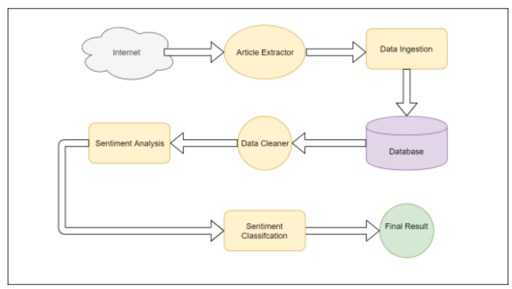

Fig. 1. System diagram

Purpose:

We are developing a web application User Review Based Recommendation System in which user will get sentiment analysis of the article. Application will provide interface to the user by which user can get product review article's sentiment.

Scope:

The main objective behind this paper is to design the web application which would provide an effective and easier way to evaluate product quality. This paper is designed keeping in mind the cost, ease of use, less overhead for target users .To reduce the cost, we are designing this web application for all smart devices.

6. Conclusion

This paper presented an application of data mining for product purchase recommendation of user by data analysis.

Acknowledgement

In this paper, we propose a user-service rating prediction approach by exploring users' rating behaviors with considering

four social network factors: user personal interest (related to user and the item's topics), interpersonal interest similarity (related to user interest), interpersonal rating behavior similarity (related to users' rating habits), and interpersonal rating behavior diffusion (related to users' behavior diffusions). A concept of the rating schedule is proposed to represent user daily rating behavior. The similarity between user rating schedules is utilized to represent interpersonal rating behavior similarity. The factor of interpersonal rating behavior diffusion is proposed to deep understand users' rating behaviors. We explore the user's social circle, and split the social network into three components, direct friends, mutual friends, and the indirect friends, to deep understand social users' rating behavior diffusions. These factors are fused together to improve the accuracy and applicability of predictions. We conduct a series of experiments in Yelp and Douban Movie datasets. The experimental results of our model show significant improvement.

References

[1] H. Feng, and X. Qian, "Recommendation via user's personality and social contextual," in *CIKM*, 2016.

[2] G. Zhao, X. Qian, and H. Feng, "Personalized Recommendation by Exploring Social Users' Behaviors," in *Proc. MMM*, 2015.

[3] P. Symeonidis, E. Tiaksa, and Y. Manolopoulos, "Product recommendation and rating prediction based on multi-modal social networks," in *RecSys*, 2017.

[4] Z. Chen, J. R. Jang, and C. Lee, "A Kernel Framework for Content-Based Artist Recommendation System in Music," *IEEE Trans. Multimedia*, vol. 13, no. 6, pp.1371-1380, 2011.

[5] J. Huang, X. Cheng, J. Guo, H. Shen, and K. Yang, "Social recommendation with interpersonal influence," in *ECAI*, 2010, pp. 601–606.

[6] Z. Wang, L. Sun, W. Zhu, S. Yang, H. Li, and D. Wu, "Joint Social and Content Recommendation for User-Generated Videos in Online Social Network," *IEEE Trans. Multimedia*, vol. 15, no. 3, pp. 698-709, 2013.

[7] S. Servia-Rodriguez, A. Fernandez-Vilas, R. P. Diaz-Redondo, and J. J. Pazos-Arias, "Inferring Contexts From Facebook Interactions: A Social Publicity Scenario," *IEEE Trans. Multimedia*, vol. 15, no. 6, pp. 1296-1303, 2015.

[8] S. Jiang, X. Qian, J. Shen, Y. Fu, and T. Mei, "Author Topic Model-based Collaborative Filtering for Personalized POI Recommendations," *IEEE Trans. Multimedia*, vol. 17, no. 6, pp. 907-918, 2015.

[9] J. Bian, Y. Yang, H. Zhang, and T. Chua, "Multimedia Summarization for Social Events in Microblog Stream," *IEEE Trans. Multimedia*, vol. 17, no. 2, pp. 216-228, 2015.

[10] X. Wang, Y. Zhao, L. Nie, Y. Gao, W. Nie, *et al*., "Semantic-Based Location Recommendation With Multimodal Venue Semantics," *IEEE Trans. Multimedia*, vol. 17, no. 3, pp. 409-419, 2015.

[11] Q. Fang, J. Sang, C. Xu, and M. Hossain, "Relational User Attribute Inference in Social Media," *IEEE Trans. Multimedia*, vol. 17, no. 7, pp. 1031-1044, 2015.

[12] W. Min, B. Bao, C. Xu, and M. Hossain, "Cross-Platform Multi-Modal Topic Modeling for Personalized Inter-Platform Recommendation," *IEEE Trans. Multimedia*, vol. 17, no. 10, pp. 1787-1801, 2017.

[13] J. Zahálka, S. Rudinac, and M. Worring, "Interactive Multimodal Learning for Venue Recommendation," *IEEE Trans. Multimedia*, vol. 17, no. 12, pp. 2235-2244, 2015.

[14] R. Salakhutdinov and A. Mnih, "Probabilistic matrix factorization," in *NIPS*, 2008.