ISSN XXXX XXXX © 2017 IJESC

Research Article Volume 7 Issue No.5

Prediction of Rating by Using Users' Geographical Social Factors

Prof. Madhura .G. Sunil[1], Abhishek H .M.[2], Aditya Kumar Bari[3], Bharath Kumar Shetty B.R.[4]
Professor[1], Student[2, 3, 4]
Department of CSE
East West Institute of Technology, Bangalore, India

Abstract:
At a recent time, a development in mobile devices and positioning techniques has led to the growth of social networks, which allows users to share their photos, experiences, check-ins, ratings, reviews, etc. We use these social networks containing geographical data as location-based social networks. Such data gets opportunities and challenges for recommender systems to solve exiguity problem of datasets and rating prediction. In this paper, we mine the relevance between Geographical social factors. We conduct several experiments on the yelp dataset. Experimental results exhibits that the proposed model performs better than the existing models.

Index Terms: Geographical location, Android, rating prediction, recommender system

I. INTRODUCTION

Recently, with the huge growing development of mobile devices and Internet access, social network services, such as Twitter, Instagram, Yelp, Facebook, Foursquare, become dominant. In 2016, there were 2.1 billion Smartphone users in the world, and 70% of them had accessed to social network services. Through mobile devices or online Location-based social networks share our geographical position information or check-ins, reviews, moods, photos and ratings in location-based social networks with their friends. Such information brings challenges and opportunities for recommender systems. When users take a long journey, they try their best to have a nice trip. Most of the services they use are local featured things. They usually tend to give high ratings much easily than the local. These ratings help us to limit our rating prediction. In acquisition, when users take long distance travelling a far from their city, they tend to depend on their local friends. Hence, users and their local friend's ratings may be similar. These ratings help us to limit our rating prediction. Moreover, when we search the Internet for a travel, if the geographical location factor is neglected, recommender systems may suggest us a new spot. The recommendations may be more humanized and profound if recommender systems consider geographical location factors. This impulses us to employ geographical location information for rating prediction.
From the above motivations, the goals of this paper are: 1) to extract the relevance between user's ratings and user-item geographical location distance, termed as user-item geographical connection, 2) to extract the relevance between user's rating differences and user-user geographical connection, 3) to search the people whose interest is same as users. Three factors are taken into account for rating prediction: user-user geographical connection, user-item geographical connection, and interpersonal interest similarity. These factors are joined into location-based rating model. The main benefactions of this paper are as follows: 1) we extract the relevance between ratings and user-item geographical location distances. It's noted that a user normally gives high scores to the items which are very distant from their activity centers. 2) We extract the relevance between users' rating differences and user-user geographical distances. It is found that users and their geographical distant friends normally give the same scores to the similar item. 3) we combine three factors: user-user geographical connection, user-item geographical connection, and interpersonal interest similarity, to a location based rating prediction model. The proposed model is checked by experiments based on Yelp dataset. Experimental results bring out significant development compared with existing approaches.

II. RELATED WORK

In paper [1] also focusing on observations on ratings combining with geographical location information. They find that geographical neighborhood has influences on the rating of a business. They perform biases based matrix factorization model with their observations, but there are some differences between us: 1) we focus on the relevance between ratings and user-item geographic distances. They focus on item-item geographic location distances and the impact of items' neighborhoods. 2) We focus more on exploring social users' rating behaviors and social influence, i.e. the relevance between users' rating differences and user-user geographic distances. 3) They perform biases based matrix factorization model, but we perform our model with constraining user and item latent factor vectors. In [2], many researchers focus on personalized recommendation and rating prediction. They miss the significance of service objective evaluation, especially for the new services with few ratings. Additionally, local urban services providers can get the feedbacks of their services from world-wide users, which are valuable for them to improve their services qualities. In this paper, they propose an issue of service objective evaluation. To solve the problem of non-objectivity evaluation to items with few ratings, they propose a unified model to evaluate services by deep understanding social users with exploring user ratings confidence. They utilize entropy to evaluate user ratings confidence. Additionally, they find that the spatial-temporal features of users' ratings are helpful to constrain user ratings confidence. Service objective evaluation is usually represented by star level, which is given by a large number of users. The Proposed Service Objective Evaluation Model goals to learn temporal and spatial coefficient vectors of use ratings, because different users' ratings have different coefficient vectors. In [3], they have proposed a user-service rating prediction approach by exploring users' rating behaviors with considering four social network factors: user personal interest, interpersonal interest similarity, interpersonal rating behavior similarity, and interpersonal rating behavior diffusion. 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 social factors are fused together to improve the accuracy and applicability of predictions. We propose a Personalized Location Based Rating Prediction (LBRP) model by combining three factors: user-item geographical connection, user-user geographical connection, and interpersonal interest similarity.

III. PROPOSED SYTEM

In this paper, three factors are taken into consideration for rating prediction:

user-item geographical connection, user-user geographical connection, and interpersonal interest similarity. These factors are fused into a location-based rating prediction model. The novelties of this paper are user-item and user-user geographical connections, i.e., we explore users' rating behaviors through their geographical location distances. Here, we mine the relevance between users' rating differences and user-user geographical distances. It is discovered that users and their geographically far away friends usually give the similar scores to the same item. It can help us to understand users' rating behaviors for recommendation. Accuracy of information can be improved by using several algorithms. We integrate three social factors into the Location-based rating prediction (LBRP) model.

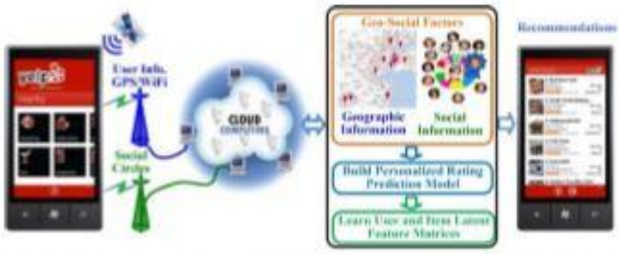

Fig. 1. System overview of our personalized recommendation via geographical social networking, including smart phone user of mobile social network services, cloud computing, rating prediction, and the recommendation lists.

IV. GEOGRAPHICAL SOCIAL FACTORS

Geographical social factors include interpersonal interest similarity, user-item geographical connection and user-user geographical connection. The user-item and user-user geographical connections are measured by ratings through diverse geographical distances. Interpersonal interest similarity is measured by the similarity between user's interest vector and friend's interest vector [4]. Note that, the geographical distance between two latitude/longitude coordinates is calculated by using the Haversine geodesic distance equation proposed in [5].

User-Item Geographical Connection

As mentioned before, mobile social network services have pervasive influence on users' daily life. Based on the analysis of data of Foursquare, users tend to activities in nearby areas. The researchers find that the activity radius of 45% users is no more than 10 miles, and the activity radius of 75% users is no more than 50 miles. The relevance of users' rating number and the distances of user-item is shown in Fig. 1. It can be seen that about 45% of the items users have rated are in the radius of 20 km. It is reasonable that people's activity centers are close to their residences or companies. It can be used to solve the cold start problem, especially when users travel to a new city.

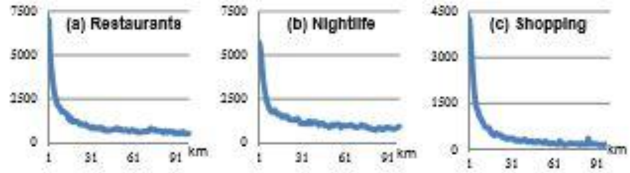

Figure.2. the distributions of the number of ratings in different distances (km)

User-user Geographical Connection

As mentioned before, user-item geographical connection is mined. Therefore, the user-user geographical connection can be learned in the same way. In this section, we analyze the relevance between users' rating differences and user-user geographical distances. For each user, the difference between his/her rating and his/her friends' to the same item is calculated. Meanwhile, we compute the geographical distance between them.

Interpersonal Interest Similarity

User interest is a representative and prevalent factor in recommender system. It is necessary to represent user interest vector. In this paper, we replace topic distribution with category distribution. The basic idea is that user latent feature vector should be similar to his/her friends' latent feature vector based on the similarity of their interest.

STATISTIC OF OUR YELP DATASETS

Dataset	Number of users	Number of items	Number of ratings	Sparsity
Active Life	6152	6390	48803	1.24E-03
Arts & Entertainment	11182	5221	108861	1.86E-03
Automotive	1351	2523	6213	1.82E-03
Beauty & Spas	5529	7323	36845	9.10E-04
Event Planning & Services	11447	6028	98491	1.43E-03
Food	9770	21370	341573	1.64E-03
Hotels & Travel	4897	2146	31833	3.03E-03
Restaurants	10,449	67,857	321,551	4.54E-04
Nightlife	11,152	21,647	436,301	1.81E-03
Shopping	8,121	15,460	112,844	8.99E-04

V. PROPOSED RATING PREDICTION MODEL

The proposed LBRP model contains the following three factors: 1) user-item geographical connection, which denotes the relevance between rating and user-item geographical distance, 2) user-user geographical connection, which denotes the relevance between user-user rating difference and user-user geographical distance, 3) interpersonal interest similarity, which means whose interest is similar to yours. We combine these three factors with the rating matrix R to decrease the rating prediction errors.

Algorithm of Proposed LBRP:

1) Initialization: Q (t)=Q(U(t),P(t),t=0.
2) set parameters: k, l, n, y1, y2
3) Iteration:
while (t<n)
Calculate dQ/dU dQ/dP
U (t) =U (t)-l dQ/dU P(t)=P(t)-l dQ/dP

t++
4) Return: U, P<-- U (n), P(n)
5) Prediction: R'=r + Ut P
6) Errors: RMSE, MAE

Parameter Settings

Here we focus on parameter settings. First, the meaning of each parameter is explained as follows.

- k: The dimension of the latent vector. If k is too small, it is difficult for the model to make a distinction among users or items. If k is too large, the complexity will considerably increase. the changes of performance with different k. But whatever the k is, it is fair for all compared algorithms when we set it as an invariant.
- y1 and y2: The parameters of trading-off over-fitting factor).

VI. CONCLUSION

The proposed Personalized Location-based rating prediction model (LBRP) has three main steps. The first is, to obtain three geo-social factors through smart phone with WIFI technology and Global positioning System (GPS). Secondly, build up personalized rating prediction model combining with the three factors in the cloud. Lastly, Train the model in the cloud to learn user and item latent feature matrices for rating prediction to recommend suitable item of users' interest. In this paper, when the geo-social data through smart phone is given, the model is built up combining geo-social factors to learn user and item latent features. User and item latent feature matrices can be calculated by machine learning methods for rating prediction. Once the ratings are predicted, the items can be ranked by ratings.

VII. REFERENCES

[1]. L. Hu, A. Sun, Y. Liu, "Your Neighbors Affect Your Ratings: On Geographical Neighborhood Influence to Rating Prediction," ACM SIGIR'14, 2014. W.-K. Chen, *Linear Networks and Systems* (Book style). Belmont, CA: Wad sworth, 1993, pp. 123–135.

[2]. G. Zhao, X. Qian, "Service Objective Evaluation via Exploring Social Users' Rating Behaviors," in Proc. BigMM, 2015, pp. 228235. B. Smith, "An approach to graphs of linear forms (Unpublished work style)," unpublished.

[3]. G. Zhao, X. Qian, and X. Xie, "User-Service Rating Prediction by Exploring Social Users' Rating Behaviors," IEEE Trans. Multimedia, 2016, vol.18, no.3, pp.496-506.

[4]. X. Qian, H. Feng, G. Zhao, and T. Mei, "Personalized Recommendation Combining User Interest and Social Circle," IEEE Trans. Knowledge and Data Engineering, vol. 26, no. 7, pp. 1763– 1777, 2014.

[5]. R. Sinnott, "Virtues of the haversine", Sky & Telescope, vol. 68, no. 2, 1984, pp. 159.