2015 IEEE International Conference on Multimedia Big Data

Prospects and Challenges of Deep Understanding Social Users and Urban Services – a position paper

Guoshuai Zhao
SMILES LAB
Xi'an Jiaotong University
Xi'an China
zgs2012@stu.xjtu.edu.cn

Xueming Qian*
SMILES LAB
Xi'an Jiaotong University
Xi'an China
qianxm@mail.xjtu.edu.cn

***Abstract*—With the boom of social media, it is a very popular trend for people to share their consumption experiences and rate the items on the review site. Users can share their experiences, reviews, ratings, photos, check-ins, moods, and so on. The information they shared is valuable for new users to judge whether the items have high-quality services. Nowadays, many researchers focus on personalized recommendation and rating prediction. They miss the significance of service objective evaluation. We can find the evaluations of services from large users by their ratings and comments. The more user ratings, the more easily we obtain the objective evaluation. But how does it work for new items? It is lack of objectivity if there are few users have rated the item, such as there are just two ratings. In this paper, we discuss the prospects and challenges of deep understanding social users and urban services, and propose some key problems for research by making full using of the big urban data generated by social users, including user rating behavior study, user sentiment study, spatial-temporal features study, and user social circle study. We focus on exploring user ratings confidence, which we propose to denote the trustworthiness of user ratings for service objective evaluation by deep understanding social users and urban services. We conduct some preliminary statistical analysis to demonstrate that these studies are necessary and potential for urban service objective evaluation.**

Keywords- urban computing; social media; service objective evaluation; user ratings confidence; big data

I. INTRODUCTION

In the past, when government wants to evaluate an urban service, such as a coffee bar, it required to ask citizens about the detail items and to require the attendees to provide some subjective scores to the services. However, this is very boring and time consuming, and sometimes the evaluations are very subjective. Recently people receive more and more digitized information from Internet. The volume of information is larger than any other point in time, reaching a point of information overload. With the explosion of social networks, users can share their experiences, reviews, ratings, photos, check-ins, and moods. It is possible to carry out urban service objective evaluation by exploring the users contributed information to urban services in social media sharing networks. Moreover, local urban service providers can get the feedbacks of their services from world-wide users, which are valuable for them to improve their services qualities. But urban service evaluation heavily relies on the already accumulating comments and ratings. But for new items, there are just few comments and ratings. Additionally, the objectivity of these comments and ratings cannot be guaranteed. We can suppose that, there are two restaurants, and one of the two restaurants is strictly better. However, for some reasons, the first customers give lower ratings to this high quality restaurant. Then other customers, who rely on ratings shown in website to make their choices, will make the wrong decision. From another aspect, it will make customers confused if there are only two reviews with different attitudes to an item. Factually, official website generally computes the average rating, and sets it as star level for each item. It is an apposite approach for the items those have been rated by a large number of users. But for a new item, we cannot straightforwardly to see the few ratings as the objective evaluation of this item. We can solve this kind of service objective evaluation by deep understanding social users and urban services.

*Corresponding author

There are some challenges of service objective evaluation of new items. The first big challenge is the sparsity of ratings. The average of user ratings cannot work well when there are only one or two ratings to an item. The second challenge is user confidence bias, because users have different patterns of giving ratings to the services. The third challenge is that users' tastes and habits are drifting over time. Users' preferences and ratings confidence are different in different places at different times. Additionally, sometimes users give high ratings but there are many negative comments in their reviews for some reasons. Thus, it is necessary for us to deep understand social users and urban services by exploring users' rating behaviors, sentiments, spatial-temporal contexts, and social circles.

Nowadays, many researchers focus on personalized rating prediction and commodity recommendation [1]-[12]. To the best of our knowledge, no work that focuses on urban service objective evaluation has been proposed so far. Our goals in this paper are to propose an issue about urban service objective evaluation and try our best to propose some key problems for the research of objective service evaluation by deep understanding social users and urban services.

The rest of this paper is organized as follows: A short overview of some works on recommender system and some definitions are presented in section 2. We can utilize biases and traditional rating prediction methods to simply address service objective evaluation, even the goals of these works are not to solve this problem. In section 3, some key problems for research on how to deep understand social

978-1-4799-8688-0/15 $31.00 © 2015 IEEE
DOI 10.1109/BigMM.2015.80

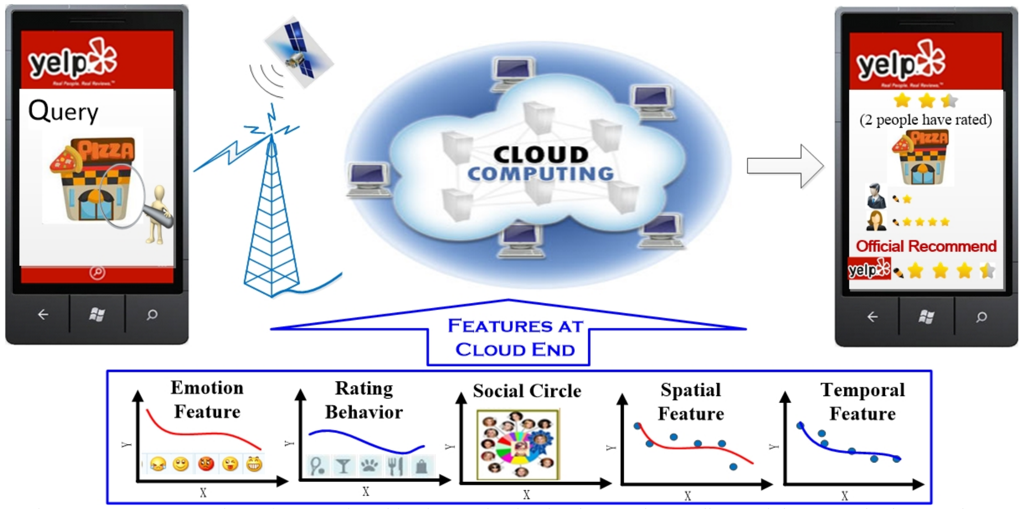

Figure 1. System overview of our service objective evaluation by deep understanding social users and urban services.

users and urban services are proposed, while some statistics and discussions are given. At last, conclusions and future prospects are drawn in section 4.

II. PRELIMINARY

A. Some Works on Recommender System

Even most of the researchers focus on personalized rating prediction and commodity recommendation, there are some basic traditional methods and techniques could be utilized to simply address service objective evaluation.

The basic method is that we can operate the average rating to evaluate items directly. Then we can consider users' rating biases as in [1] to overcome different rating criteria. Additionally, we can explore users' rating criteria with more refinements, such as the biases based on taxonomy as in [2]. For matrix factorization based rating prediction approaches [3]-[10], once we get the learned user and item features, we can use them to predict all users' ratings for each item. Then we evaluate star level of each item by averaging the predicted ratings. From another side, we cannot only utilize users' ratings to conduct service objective evaluation, but also exploit the similarity between items to predict evaluation directly, such as the adjusted cosine similarity algorithm proposed in [12]. These methods cannot work well service objective evaluation. Thus, we need to deep understand social users and urban service.

B. Definitions

Here, we will give some definitions about urban service objective evaluation. As we know, service objective evaluation is usually represented by star level (such as in Yelp[1], Dianping[2]), which is given by a large number of users. Thus, we can see that the ground truth of the star level of service *i*, represented by R_i, is difficult to obtain, because the ground truth heavily relies on the review count n_i. Official website generally computes the average rating, and sets it as the star level for each item. It is inappposite for new items, which have a few reviews. In order to address this problem, we propose the concept of rating confidence, which represents the reliability and credibility of a rating. We define the rating confidence user u to item i as $\Phi_{u,i}$. Then service objective evaluation can be represented by $\sum_{u=0}^{n_i}\left(\Phi_{u,i} r_{u,i} / \sum_{u=0}^{n_i} \Phi_{u,i}\right)$.

III. KEY PROBLEMS FOR RESEARCH

In this paper, we propose the issue of urban service objective evaluation, and intent to utilize the concept of rating confidence to distinguish ratings. But the rating confidence maybe related to user's rating behavior, sentiment, the spatial-temporal context and his/her social circle. Thus we should deep understand social user by exploring crowd source contributed urban services related big data. The main flowchart of our approach is shown in Figure 1. Hereinafter we turn to details of the key problems.

A. User Ratings Behavior Study

As mentioned before, we focus on user ratings confidence to discriminate their ratings to conduct service objective evaluation. Our basic idea is that user ratings have different confidence. Then how should we know which people are trustworthy? We have large records of users' historical ratings. We can exploit these large data to measure user ratings confidence. As we know, entropy is the measure of the disorder or randomness of energy and matter in a system. If a user's ratings are confident, his/her ratings must have little differences with real star levels of items. Thus, information entropy value of these differences can be used to represent the confidence value of user ratings. That is to say, we see the differences between user ratings and items' real star levels as the elements of an error value system, then entropy of this system can reflect user's rating habits and stability.

B. User Sentiment Features Study

On most of review sites, users cannot only rate the commodity, but also share their experiences and attitudes by

[1] http://www.yelp.com/
[2] http://www.dianping.com/

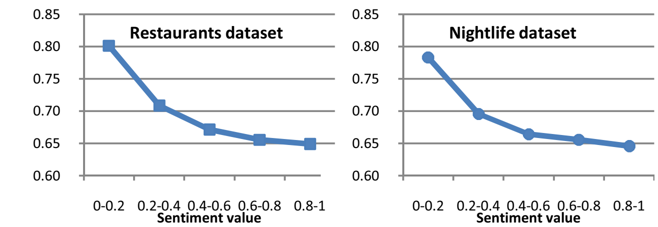

Figure 2. The distribution of average errors in different textural review sentiment values. The value of y-axis is the average error between ratings and corresponding real star levels of items. The value of x-axis is the normalized textural review sentiment value.

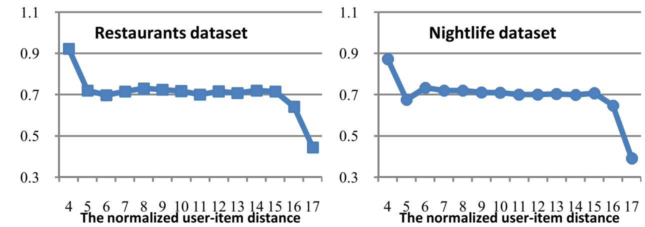

Figure 3. The distributions of ratings' confidence in different user-item geographic location distances based on Yelp Restaurants and Yelp Nightlife datasets. The value of x-axis denotes the user-item geographic distance which has been normalized by logarithm, and the value of y-axis denotes the average value of differences between user ratings and item real star levels.

reviewing. From their textual reviews, we can get more exact information, which always verifies and supports their ratings directly. In other words, to the same item, the rating with long and detailed comment may have more confidence. Because we can see user's exact attitude through his/her many positive or negative words. Thus, in this section, we propose to analyze the relevance between user confidence and textural review sentiment.

1) Utilizing Words to Compute Sentiment Score

We leverage words based sentiment analysis method to compute sentiment score. HowNet Sentiment Dictionary[3] is used to compute sentiment score, which is well known in the area of Chinese and English sentiment classification.

Then we analyze the relevance between user confidence and textural review sentiment in Yelp Restaurants and Nightlife datasets. The former dataset contains 263,124 ratings from 4,138 users who have rated a total of 62,221 items. The latter dataset contains 436,301 ratings from 11,152 users who have rated a total of 21,647 items. For each review, the sentiment score is normalized by min-max normalization method. Then we divide the ratings into five groups according to the normalized review sentiment scores as shown in Figure 2. For each group, we compute the average error between ratings and corresponding real star levels of items. We can see the error decreases with the review sentiment score increase. That is to say, user confidence is increasing with review sentiment score.

It demonstrates that user sentiment study is necessary for deep understanding social users and urban services.

2) Utilizing LDA to Build Classification

In our sentiment feature representation part, we can study the relevance between user sentiment and ratings in deeper. We can classify textual reviews into 5 categories (similar to the rating from 1 to 5) by utilizing LDA to build their classification models. There are many more studies on sentiment analysis [13]-[15]. We can combine these studies with user rating confidence to mine more interesting information. We will go in more details in this research.

C. Spatial-temporal Features of User Ratings

We know that users' profiles are changing constantly. That is to say user ratings confidences are different in different places at different times. Thus we suppose that spatial-temporal features are beneficial to the research of deep understanding social users and urban services.

1) Spatial Features

We suppose that user-item geographic location distance may influence user ratings' confidence. Thus, we conduct a statistic to prove it. Figure 3 is the distribution of ratings' confidence in different user-item geographic location distances based on Yelp Restaurants and Nightlife datasets. The horizontal axis represents user-item geographic distance, which has been operated by logarithm. The ordinate axis represents the difference between user ratings and item real star levels, which is an absolute value here. From Figure 3, we can see user ratings are mostly unreliable if the user is much near to the rated item geographically. As the distance increase, user ratings' confidence is stable. When the distance becomes very large, user ratings are very reliable correspondingly. Why does this happen? We suppose that, users may be influenced by their friends or some discounts for services. In addition, in terms of items, most of them have their competitors. Inevitably there may be some unfair ratings and comments appear on the Internet. Even it seems reasonable that geographic distance can distinguish different ratings' confidence to a certain extent, we think there are much more interesting latent disciplines needed to be dug.

2) Temporal Features

We show the statistics of temporal features based on our datasets in Figure 4. In Figure 4, we show the distribution of ratings' temporal features in different times. The x-axis represents the rating time, and y-axis represents the difference between user ratings and item real star levels, which is also an absolute value. We can see the difference between user ratings and item real star levels is decreasing over time. We suppose that the number of ratings is increasing constantly for each item, which will become a reference to the item for other customers. As time passed by, users may get more useful information from former ratings and comments, and give a suitable rating. That is to say, when we search the Internet, we will be unconsciously influenced by the ratings and comments, because the external environment can affect a person's views, especially on the fields he/she doesn't know well. Maybe there are some more interesting latent disciplines we have not mined, but at least the basic statistic demonstrates that temporal features study is necessary for deep understanding social users and urban services.

[3] http://www.keenage.com/download/sentiment.rar

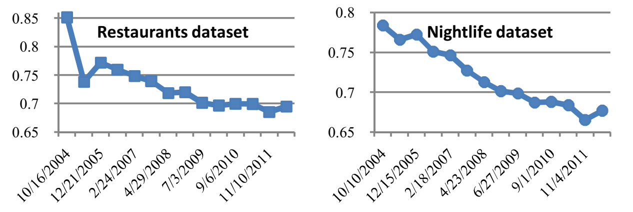

Figure 4. The distributions of ratings' confidence in different periods based on Yelp Restaurants and Yelp Nightlife datasets. The value of x-axis denotes the day time user rated item, and the value of y-axis denotes the average value of differences between user ratings and item real star levels.

D. User Social Circle Study

Social factors are proved to be an effective clue in services recommendation. In some previous work, authors have shown that user's rating is closely affected by their friends [6]-[10]. In our future work, we will analyze this factor in deeper to understand social users.

E. The Basic Preliminary Model

The purpose of features study is to compute the rating confidence by deep understanding social users and urban services. Thus the goal of the rest work is to build a unified model to fuse these features to compute the rating confidence. Each feature, presented hereinbefore, has a certain capacity of distinguishing ratings. Thus, the goal of basic preliminary model is to learn the weights of these features. We represent the rating confidence as $\Phi_{u,i} = \sum_k W_k^{(u,i)} F_k^{(u,i)}$ by the linear model, where k is the number of considered features, W denotes the weight, F denotes the feature. We can learn the weight W in training set. Furthermore, we can also perform the learning model by non-linear models, probabilistic models, and other cross-modal approaches. There are many potentials to develop urban service objective evaluation by deep understanding social users and urban services.

IV. CONCLUSIONS AND FUTURE PROSPECTS

In this position paper, we propose to focus on the issue of urban service objective evaluation by deep understanding social users and urban services. Too many researchers pay more attention to personalized rating prediction and recommendation without considering the significance of service objective evaluation, especially for the new services with few ratings. Additionally, the 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. But there are many more challenges to address this issue for the new services, such as the sparsity of ratings, user confidence bias, users' tastes and habits, users' social influence, the conflicting ratings and reviews. Thus, it is necessary for us to deep understand social users and urban services.

We propose some key problems for research, including user social circle study, user rating behavior study, user sentiment study, and spatial-temporal features study. We conduct some preliminary statistical analysis, and try to utilize these features to deep understand social users and urban services. Preliminary statistics demonstrate these features are necessary for deep understanding social users and urban services.

Service objective evaluation by deep understanding social users and urban services is related to the hot research areas, including Geo-life, urban computing, and Travel Guide, etc. It will benefit the government, service providers, and users to know the quality of the urban services more objectively, and the local urban services providers can get the feedbacks of their services from word-wide users, which are valuable for them to improve their services qualities.

ACKNOWLEDGMENT

This work is supported in part by the Program 973 No. 2012CB316400, by NSFC No.60903121, 61173109, 61332018, and Microsoft Research Asia.

REFERENCES

[1] Y. Koren, "Collaborative filtering with temporal dynamics," KDD'09, pp. 447-456, 2009.

[2] G. Dror, N. Koenigstein, and Y. Koren, "Yahoo! Music recommendations: modeling music ratings with temporal dynamics and item taxonomy," in ACM Recsys, 2011.

[3] R. Salakhutdinov, and A. Mnih, "Probabilistic matrix factorization," in NIPS, 2007.

[4] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," Computer, Aug.2009, pp. 30-37.

[5] Y. Koren, "Factorization meets the neighborhood: a multifaceted collaborative filtering model," KDD'08, 2008.

[6] M. Jamali, and M. Ester, "A matrix factorization technique with trust propagation for recommendation in social networks," in ACM RecSys, 2010.

[7] X. Yang, H. Steck, and Y. Liu, "Circle-based recommendation in online social networks," in KDD'12, 2012, pp. 1267-1275.

[8] M. Jiang, P. Cui, R. Liu, Q. Yang, F. Wang, W.-W. Zhu, and S.-Q. Yang, "Social contextual recommendation," in CIKM'12, Oct.2012, pp. 45-54.

[9] 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.

[10] H. Feng, and X. Qian, "Recommendation via user's personality and social contextual," ACM CIKM 2013.

[11] G. Adomavicius, and A. Tuzhilin, "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions," IEEE Transactions on Knowledge and Data Engineering, pp. 734-749, Jun. 2005.

[12] B. Sarwar, G. Karypis, J. Konstan, and J. Reidl, "Item-based collaborative filtering recommendation algorithms," In Proceedings of the 10th International Conference on World Wide Web (WWW), 2001, pp. 285-295.

[13] T. Zhao, C. Li, Q. Ding, L. L, "User-Sentiment Topic Model: Refining User's Topics with Sentiment Information," In Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics, 2012.

[14] S. Tan, Y. Li, H. Sun, Z. Guan, X. Yan, et al. "Interpreting the Public Sentiment Variations on Twitter," IEEE Trans. Knowledge and Data Engineering, vol.26, no.5, pp.1158-1170, 2014.

[15] W. Zhang, G. Ding, L. Chen , C. Li ,and C. Zhang, "Generating virtual ratings from Chinese reviews to augment online recommendations," ACM Transactions on intelligent systems and technology (TIST), vol.4, no.1, 2013.