A Novel Approach to Evaluate the Service Quality by Exploring Social User Contextual Information

E.Jai Vinitha¹, J.Maruthu Pandi², R.Divya Bharathi³

¹Department of IT, M.Tech, Mepco Schlenk Engineering College, Sivakasi, Tamilnadu, India
evinithait12@gmail.com

²Department of IT, Assistant Professor, Mepco Schlenk Engineering College, Sivakasi, Tamilnadu, India
jmaruthupandi@mepco.ac.in

³Department of IT, M.Tech, Mepco Schlenk Engineering College, Sivakasi, Tamilnadu, India
bharathikrishnan2.db@gmail.com

ABSTRACT

With the increase of social media and e-commerce, enormous people prefer to share their experience and rate on review websites. Existing research are mainly focused on personalized recommendation and rating prediction but evaluating the quality of service for recommender system is more important. The proposed approach focuses on service quality evaluation. There are some challenges that do not have enough review information for extracting opinion. In this paper, a Service Quality Evaluation model is proposed to evaluate the service quality. The proposed model can be done in three steps. First step is to calculate the entropy which is utilized in users’ confidence value. Second, to explore the contextual features of user rating in which the spatial-temporal features and sentimental features are reviewed. The final step is to fuse the above two steps into a unified model for calculating the overall confidence value to perform service quality evaluation. The experiments are implemented by using Yelp and Douban dataset.

Keywords : Spatio-temporal features, sentimental features, Data Mining, Contextual Information of User

I. INTRODUCTION

Nowadays, with the development of mobile devices and internet access, social network services have become popular. Users share their experiences like movie, ratings, and moods on internet. The first generation of recommender system [2]-[12] and social network based models [13]-[22] mainly focus on personalized recommendation and predicting users preferences but they all ignore the service quality. Thus, we mainly focus on quality of service.

When we choose an item, we mainly rely on users review and rating. Generally rating ranges from 1 to 5, the more user rating to the item there are more confidence in the overall rating. For example, consider for a movie the given rating is 4.5 by hundreds users we will assume that the movie is good. However, there will be some users who will not like the movie and the rating is two. The audiences will get confuse if the rating are contrary for same item. If the rating is contradiction then we take average for the rating and consider it as overall rating.

There are several challenges for service quality. The first challenege is the rating sparsity. The second challenge is user confidence bias. Users have different pattern for services. The third challenge is user confidence which is not isolated. In addition users may give high rating but there may be several negative reviews. So, we need to explore users’ rating confidence by closely examining social users’ contextual information.

In the proposed method we first utilize information entropy to calculate user rating confidence. Second from the users’ contextual information the spatial-temporal and sentimental features of rating are examined. Finally,
they are fused together to form the unified model. The main contribution of the paper is as follows:

- The issue of quality evaluation for service is addressed and the probabilistic linear model is proposed for exploring users’ contextual information.
- We use the user rating confidence to evaluate the quality of service because different users have different level of confidence. Further user profiles are changing at different places in different times. So, we implement probabilistic linear model with Gaussian observations.
- We find contextual information for constraining user rating confidence. User rating is higher when a user is very far away from the rated item.

The remainder of the paper is organized as follows. In Section 2 related work on recommender system. In Section 3 proposed model is presented. In section 4 the description about dataset is given. In Section 5 experiments are presented. Finally in Section 6 conclusion is explained.

II. RELATED WORKS

Koren [29] proposed temporal dynamics with the collaborative filtering and he considered only user and item time which changes and compared with various baselines.

Dror et al.[30] proposed a model that captures information from taxonomy of items and different temporal dynamics of music ratings and the idea can be used in user bias which convert the personalized rating prediction to service quality.

The matrix factorization model [6], [7], [21]-[27], [29]-[33] predicts the user’s ratings, in which the unknown ratings are predicted by using the latent features of users and items. In the previous work [21], [22], [25], [26], [27] considers the social factors in matrix factorization, including interpersonal influence, interest similarity, personal interest etc.

Multimedia recommendation is addressed in [19], [20], [34], [35], [36]. Lee et al. [34] proposed the recommendation concepts for both novel and relevant recommendations. Wang et al. [19] proposed a framework for suggesting the videos that users import in the online social network.

Existing works focus on personalized recommendation or rating prediction but we focus on service quality evaluation by exploring social users’ contextual information. For predicting the users’ rating matrix factorization can also be used [23], [24], [25], [26], [27], [37]. Sarwar et al. [2] proposed an algorithm for item collaborative filtering in which they focus on predicting the users’ rating of an item by calculating the average ratings of similar items.

The sentiment analysis method [38], [39], [40] focus on social networks, public sentiment, and web queries. Zhang et al. [38] proposed self-supervised emotion-integrated sentiment classification results into collaborative filtering in which user-item rating matrix is inferred by decomposing item reviews that user give for item.

Tan et al. [39] proposed a model for text collection in comparison with another background text collection. The cold start problem is the item with only few ratings and research is focused on cold start problem [41]-[43]. Leroy et al. [41] focused on cold start link prediction.

Jiang et al. [43] proposed a user topic based on collaborative filtering approach for personalized travel recommendation which is the improved version of traditional for collaborative filtering by fusing user information in social media.

III. PROPOSED APPROACH

The proposed service quality evaluation is done by three different steps,

- Users’ confidence value
- Contextual Features of User Ratings
- Service Quality Evaluation Model

In the users’ confidence value information entropy is calculated and in the Contextual user information the features like spatial-temporal features and sentimental features are obtained. Finally in the service quality evaluation model the above two steps are fused together to obtain the overall confidence rating of an item.
3.1 Users’ Confidence Value

Different users’ have different contribution in quality evaluation. In this paper, user rating confidence is leveraged to conduct service quality. Entropy is a measure of uncertainty. The information entropy is used to calculate the confidence value. The difference between the user rating and overall rating reflects the stability of the system. Additionally, the coefficient is added to distinguish the weights because entropy cannot make any difference. If the entropy value is low then the system is more stable so the reciprocal of entropy value is taken.

\[
E_u = -\frac{1}{\sum_i (d_i \times p(d_i)) \log_2 p(d_i)}
\]

(1)

\[
d_i = r_{u,i} - r_i
\]

(2)

Where \(E_u\) denotes user \(u\)’s confidence value. \(d_i\) is the difference between user rating \(r_{u,i}\) and overall rating \(r_i\). \(p(d_i)\) indicates the probability of the value \(d_i\).

3.2 Contextual Features of User Ratings

The entropy calculation of user ratings confidence is based on the ratings of the user. User profile changes constantly so that their rating’s confidence may be different at different places and different time. Sometimes, user gives high rating but there are many negative words in their review. Thus, we further constrain each rating’s confidence by its spatial-temporal features and review sentimental features.

3.2.1 Spatial Features

In large network people are living and so they may be influenced by others easily. We start by analyzing the distribution of rating’s confidence in different user-item geographic location distances. The user-item geographic distance is calculated by following algorithm:

\[
x = \ln D(u, i)
\]

(3)

Where \(D(u, i)\) denotes the geographical distance between user \(u\) and item \(i\). If the users are close to rated item then the rating’s confidence is low. The users may be influenced by their friends or some discounts for services. In terms of items, most of them have competitors.

Generally, competitors are close and mostly native geographically. For different datasets the spatial features are different. Therefore, curve fitting is conducted to learn rating’s spatial features. The curve fitting model is based on the 4th Gaussian degree of model. The curve fitting model formula is:

\[
y = \sum_j a_j \times \exp\left(-\frac{(x - b_j)^2}{c_j^2}\right)
\]

(4)

Where \(a_i\), \(b_i\), and \(c_i\) are the coefficients which is learned in curve fitting. Rating confidence is inversely proportional to \(y\). The rating confidence based on spatial features is represented by:

\[
G_{u,i} = \sum_j a_j \exp\left(-\left(\ln D(u, i) - b_j\right)/c_j^2\right)
\]

(5)

Where \(G_{u,i}\) denotes rating confidence user \(u\) to item \(i\) based on spatial features. \(a_i\), \(b_i\), and \(c_i\) are the coefficients which is learned in curve fitting. \(D(u,i)\) denotes the geographical distance value between user \(u\) and item \(i\).

3.2.2 Temporal Features

In the same way we calculate the rating’s confidence based on temporal features. For a single item there are more and more ratings and reviews which result in getting more and more information from former ratings reviews, and then give a suitable rating.

Curve fitting is conducted based on 4th degree Gaussian model. Rating’s temporal features can be represented by:
\[ T_{u,i} = 1/\sum_j a_j \exp\left( -\left( \frac{\text{Day}(u,i) - b_j}{c_j} \right)^2 \right) \] (6)

Where \( T_{u,i} \) denotes rating confidence user \( u \) to item \( i \) based on temporal features. \( a, b, \) and \( c \) are the coefficients which is learned in curve fitting. \( \text{Day}(u,i) \) denotes the rating time of user \( u \) to item \( i \).

### 3.2.3 Sentimental features

In most review web sites users not only rate the commodity but also share their experiences and attitude by reviewing. From the textual reviews, we can get exact information, which verifies and supports the rating directly. It is necessary to analyze the relevance between user confidence and textual review sentiment.

First, the method of sentiment analysis is used to calculate sentiment scores. Second, the relevance between user rating confidence and review sentimental is mined. Last, we learn sentimental features to constrain user’s confidence. The overall rating of service decrease with the sentiment score. The user confidence increases with review sentiment score. The sentimental features can be represented by:

\[ S_{u,i} = 1/\sum_j a_j \times (RS(u,i))^2 \] (7)

Where \( S_{u,i} \) denotes rating’s confidence user \( u \) to item \( i \) according to review sentimental features. \( RS(u,i) \) is the normalized sentiment score user \( u \) to item \( i \).

### 3.3 Service Quality Evaluation Model

In this model, we fuse user’s confidence with contextual features, including spatial-temporal features and review sentimental features to calculate the overall confidence value of rating. The coefficient is defined in such a way that the sum of coefficient is one. By using probabilistic unified model the spatial-temporal and sentimental features are calculated. The overall confidence of the rating that user \( u \) to item \( i \) as follows:

\[ D_{u,t(u,i),g(u,i),s(u,i)} = 1 - A_{u,t(u,i)} - B_{u,g(u,i)} - C_{u,s(u,i)} \] (9)

Where \( t(u,i) \) denotes the time user \( u \) rated item \( i \). \( g(u,i) \) denotes the geographic distance between user \( u \) to item \( i \). \( s(u,i) \) denotes the sentimental value of the review. \( T_{t(u,i)} \) denotes the temporal value of the review. \( A, B, C, D \) are the corresponding coefficients matrices.

### 3.3.1 Model Inference

The Gaussian with the probabilistic linear model is chosen [23], [25] and [31]. The conditional probability of the observed rating is as follows:

\[ p(R|A, B, C, D, G, E, T, S, \sigma_R^2) = \prod_i N(R_i | \sum_{u=0}^{N_i} \phi_{u,i} \cdot \frac{r_{u,i}}{\sigma_R^2}, \sigma_R^2) \] (10)

Where \( N(x|\mu,\sigma^2) \) denotes the probability density function with mean \( \mu \) and variance \( \sigma^2 \). \( A, B, C, D \) is user’s temporal, spatial, sentimental and coefficients matrix. If there is only one user having rated item \( i \), quality evaluation of the service cannot be performed.

According to [31], zero mean Gaussian priors are assumed for user’s spatial-temporal and sentimental coefficients vectors:

\[ p(A|\sigma_A^2) = \prod_u N(A_u|0, \sigma_A^2) \] (11)

\[ p(B|\sigma_B^2) = \prod_u N(B_u|0, \sigma_B^2) \] (12)

\[ p(C|\sigma_C^2) = \prod_u N(C_u|0, \sigma_C^2) \] (13)

### 3.3.2 Model Training

The gradient, we update the coefficient matrices as follows:

\[ A_{u,t(u,i)} = A_{u,t(u,i)} - \alpha \frac{\partial \psi}{\partial A_{u,t(u,i)}} \] (14)

\[ B_{u,g(u,i)} = B_{u,g(u,i)} - \alpha \frac{\partial \psi}{\partial B_{u,g(u,i)}} \] (15)

\[ C_{u,s(u,i)} = C_{u,s(u,i)} - \alpha \frac{\partial \psi}{\partial C_{u,s(u,i)}} \] (16)
Where $\alpha$ is the learning rate.

The service quality evaluation is conducted by the following coefficient matrix as follows:

$$\hat{r}_i = \sum_{u=0}^{\hat{n}_i} \frac{\hat{p}_{u,i}}{\sum_{u=0}^{\hat{n}_i} \hat{p}_{u,i}} r_{u,i} \quad (17)$$

### Algorithm 1. Service Quality Evaluation (SQE)

**Model**

**Input:** Rating matrix $R$ in training dataset

- User confidence $E$ calculated by equation (1)
- Spatial bias $G$ calculated by equation (5)
- Temporal bias $T$ calculated by equation (6)
- Sentimental bias $S$ calculated by equation (7)

**Output:** Quality evaluation of test services.

1. Calculate entropy using difference between user rating and overall rating.
2. Compute the distance for user $u$ and item $i$.
3. Evaluating the temporal features by using curve fitting.
4. Then performing sentimental review using normalized sentiment score.
5. Fuse all the steps from (1)-(4) into a model to calculate overall rating.
7. **for** $t=1:T$ **do**
   8. **for** each element of coefficients matrices $A$, $B$, $C$, **do**:
   9. Using equation (14), (15), (16) the coefficient matrices are updated.
10. **end for**
11. **end for**
12. **for** each test item **do**
13. **for** each rating of this item **do**
   - Calculate the overall confidence by Equation (8);
14. **end for**
15. calculating the overall rating of the item
16. **end for**
17. **Return:** The overall rating of services

### IV. DATASET DESCRIPTIONS

Yelp dataset and Douban dataset is introduced in this section. The dataset can be downloaded from web site of SMILES LAB.

#### 4.1 Yelp Dataset

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. In our work, we utilize two categories: Restaurant and Nightlife. Moreover, it is proved by the data of Yelp that users are more willing to visit places or to consume items that his/her friends have visited or consumed before. We analyze the relevance between user ratings and user-item location distances.

#### 4.2 Douban Dataset

Douban is one of the most popular social networks in china. It includes several parts: Douban Movie, Douban Read and Douban Music, etc. We crawled the ratings from the Douban Movie websites. The dataset consists of 2,968,648 ratings from 8,226 users who have rated 14,715 movies. Note that there is no geographic location information and reviews in Douban dataset. We perform our model on Douban dataset by fusing user ratings' confidence and temporal features.

#### 4.3 Pre-processing

The issue proposed in this paper is quality evaluation for services with very few ratings. The ratings in our dataset are split according to preselected items. Every tested item will not have more than five ratings. Some items are used for training and some item are used for testing.

### V. EXPERIMENTS

The predicted overall ratings of services, the performance of methods will be embodied by the errors. The differences between the prediction and the overall rating of services can be leveraged to measure the model. The real overall ratings of services are discrete as [1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, and 5.0], while the predictions are in decimals. The predicted decimals can be rounded into discrete quantities. Then precision, Recall and AUC (Area under Curve) measures [47], [48], [49] are utilized to evaluate the proposed model. The proposed service quality is implemented in NetBeans using JAVA language.

### VI. CONCLUSIONS

Many researches are focused on rating prediction and personalized recommendation. So, it is important to conduct service quality evaluation. In this paper, we propose service quality evaluation by exploring user’s contextual information. We focused on exploring user rating’s confidence. The spatial-temporal and sentimental features are calculated. Finally all are fused together to calculate the overall rating confidence. We
use a few ratings to predict the overall ratings of the services.

REFERENCES


be published. Doi: 10.1109/TBDATA.2016.2552541.


