A Personality-aware Followee Recommendation Model Based On Text Semantics and Sentiment Analysis

Pan Xiao, YongQuan Fan*, and YaJun Du

School of Computer and Software Engineering, Xihua University, Chengdu 610039, China

Abstract. As the popularity of micro-blogging sites, followee recommendation plays an important role in information sharing over microblogging platforms. But as the popularity of microblogging sites increases, the difficulty of deciding who to follow also increases. The interests and emotions of users are often varied in their real lives. On the contrary, some other features of micro-blog are always unchangeable and they cannot describe the users characteristics very well. To solve this problem, we propose a personality-aware followee recommendation model(PSER) based on text semantics and sentiment analysis, a novel personality followee recommendation scheme over microblogging systems based on user attributes and the big-five personality model. It quantitatively analyses the effects of user personality in followee selection by combining personality traits with text semantics of micro-blogging and sentiment analysis of users. We conduct comprehensive experiments on a large-scale dataset collected from Sina Weibo, the most popular mircoblogging system in China. The results show that our scheme greatly outperforms existing schemes in terms of precision and an accurate appreciation of this model tied to a quantitative analysis of personality is crucial for potential followees selection, and thus, enhance recommendation.

Keywords: Followee recommendation, Personality traits, Semantic analysis, Sentiment analysis

1 Introduction

Since the emergence of social networks and microblogging sites, such as Twitter and Sina Weibo, hundreds of millions of users have become to use the microblogging service . In the context, finding high quality social ties becomes a difficult task due to the continuous expansion of microblogging communities. In microblogging systems, user follow or are followed by each other. In this regard, if user x follows user y, we can refer to x as y's follower, and y as x's followee[1].

* This work is supported by the National Nature Science Foundation (Grant No. 61472329 and 61532009), the Key Natural Science Foundation of Xihua University (Z1412620) and the Innovation Fund of Postgraduate, Xihua University.

There are many reasons for a user to follow some users because they publish interesting information, others because they have the same interests.

The so-called personality is not only the impact of each person's thinking, behavior and decision-making, and maintain its long-term stability, but also affect its social relations. Scientific research shows that there is a significant relationship between human personality and behavior in the real world[2]. Over the years, psychological researchers have been able to understand the personality traits and are committed to find a system.

Through the analysis of the personality of the social network users, we can achieve large-scale access to the user's personality data, which is conducive to the personality of the information in-depth research and extensive application. However, there are few researches on personality theory in the previous studies, but the research on the social network which is related to the social psychological characteristics is actually related to personality. There are many problems in the field of social computing, which are more or less related to personality theory. The potential value of personality prediction will be very helpful in solving these problems.

On the other hand, the user is susceptible to the outside world, this is because the user's needs are more extensive, the human emotional perception is relatively rich. Therefore, the user's textual and emotional expression largely reflects their personality traits, Particularly, text information and emotional information are also important factors to reflect user behavior.

The rest of this paper is organised as follows. In Section 2, we discuss related work. Section 3 presents the followee recommendation method we propose. In Section 4, We propose PSER model. In Section 5, we evaluate the performance of our design experiment and summarise the conclusions obtained from the performed experiment evaluation.

2 Related work

Personalized recommendation technology is the core and critical technology of E-commerce recommendation system. The recommendation technology based on the fusion link topology is to abstract the social information into a meaningful social network[3]. However, these traditional methods only focus on how to improve the accuracy of recommendation, while ignoring the inherent characteristics of the user behavior is determined by their personality characteristics[4].

With the rapid development of the Internet and the convenience of user network data acquisition, the researchers began to try to use the user's online behavior to predict the user's personality. In recent years, with the rapid growth of Internet users and the increasing coverage of users, real-time network data provides a new research perspective for the research of network users' personality. Hamburger et al.[5] analyzes the network behavior and the relationship between them and obtains the user's network behavior data by questionnaire.

In order to study the personality traits of users on the social networking platform, Quercia et al.[6] conducted a large five personality test for some popular users on Facebook and an1alyzed the association between the number of interactions and personality traits. The study by Mairesse F et al.[7] Shows that linguistic features can be used for personality analysis. Hu and Pu[8] and Tkalcic et al.[3] presented approaches to include personality scores as complementary informationin traditional rating-based collaborative recommendation systems. Both relied on the explicit assessment of personality through the Big Five test and the IPIP questionnaire respectively. Wu et al.[9] aimed at adding personality scores to a content-based movie recommendation system in order to generate more personalised and diverse recommendations.

In the past, the recommended algorithm is often based on the user's registration information, labels. These factors which are not easily changed can calculate the user's similarity. The user's interests and emotions in the daily life will be subtle changes, and user needs have gradually become very wide, which makes the recommended algorithm cannot be guaranteed the effect. The literature[10] uses social networks that mark spanning trees and relational graphs to discover followee. However, the range of recommendations for this approach cannot be extended. The literature[11] translates the recommendation into the link forecasting problem. The social label of this method is more fixed. Therefore, the user is usually chosen so freely that the user's features are not accurately expressed and the user's emotional uncertainty in the face of different events is not taken into account.

In real life, the user is vulnerable to the outside world. The reason is that the various needs of a wide range of human, emotional perception is relatively rich, which led to the common recommendation model can not meet the needs of users in a timely manner. In the online social network, the information that appears in a posting, forwarding, or commenting not only contains the user's interest, but also implied the user's emotional characteristics, which can be recommended as a very important reference data.

All of the presented approaches share the same drawbacks. First, they included a relatively small number of users. Second, personality was self-assessed through questionnaires. The own view of themselves reported by users could not reflect their actual behaviour and, in turn, their real personality[12]. Finally, the approaches were tested in the context of item recommendation using collaborative filtering techniques, none of the works include personality in the context of user recommendation in social networks. In consequence, the impact of personality in social recommendation systems is yet to be proven.

3 Our method

In social networks, users generally have few changes to the registration information. As time goes on, the user's interest in labels and hobbies change a bit, so we focus on the user's recent behavior best reflects the user's interest. The literature[13] shows that different users in the online social network interact with each other through microblogging to influence emotions. Usually micro-blog information will contain two types of information, that is, text information and user emotional information. We mark it as TE={Text, Emotion}. Also, we also consider personality traits. The recommendation algorithm presented in this paper can be combined with the user's interests and emotions, as well as personality traits. The experimental results show that the proposed algorithm is more effective than the traditional recommendation algorithm in generating the followee who are interested in the user.

3.1 User Micgroblogging Text Analysis

In this section, we discuss the importance of micro-blog text in recommendation. Microblogging text content is basically more compact and important vocabulary and contains a lot of information. In the use of text similarity calculation method, it is necessary to extract the keywords from the text, which is used as a user's personalized label, and then the similarity of the user is calculated for recommendation. But in Chinese words, there are many synonyms, antonyms and similar words which are easy to be confused. Therefore, in view of the characteristics of short microblogging, the semantic analysis of the text which user published is based on the existing "synonym forest" in this paper. This method first calculates the text similarity, and then calculates the emotion similarity. Because the extracted user's text content contains a variety of different subject information, we compare the threshold after calculating the similarity of text content and filtering out the theme of the text to find the similar user, which can improve the computational efficiency of the recommendation and improve the accuracy of recommendation.

Text Similarity Analysis When calculating the user's text similarity, we first need to use vectorization to describe the Microblogging of the user U_i and U_j.

$$te_i = (te_{i1} = (T_{i1}, E_{i1}), ..., te_{in} = (T_{in}, E_{in})) \tag{1}$$

$$te_j = (te_{j1} = (T_{j1}, E_{j1}), ..., te_{jn} = (T_{jn}, E_{jn})) \tag{2}$$

The intersection of text similarity between user U_i and U_j is denoted as T_{com}.

$$T_{com} = T_{U_i} \cap T_{U_j} = \{t_{com1}, t_{com2}, ..., t_{comn}\} \tag{3}$$

The text contents which users published are successively in time. Therefore if the content of the text is closer to the current time, the ability to represent the user is stronger. In order to reduce the deviation caused by the time variation, this paper adds the corresponding weight value to reflect the influence of time change when calculating the similarity of microblogging text content. When T_{com} is empty, it means that there is no similarity between the two users' text, that is, the text similarity is 0, the formula is as follows:

$$Sim(te_i, te_j) = \begin{cases} \sum_{k=1}^{m}(wf_k \times Sim(T_{ik}, T_{jk})) & \text{if } T_{com} \neq \emptyset \\ 0 & \text{if } T_{com} = \emptyset \end{cases} \tag{4}$$

Where T_{ik} and T_{jk} represent the textual content of U_i and U_j in kth days respectively. wf_k is the weight assigned to the user's kth text information, and m represents the number of selected texts. Specially, we define $\sum_{k=1}^{m} wf_k = 1$. Also, with the increase of k, the value of wf_k becomes larger, which means that the textual content published time closer to the current time and the stronger the characterization of the user. In this paper, we use Jaccard distance to compute text similarity[14], as follows:

$$Sim(T_{ik}, T_{jk}) = \frac{N(T_{ik}) \bigcap N(T_{jk})}{\sqrt{|N(T_{ik})| \, |N(T_{jk})|}} \tag{5}$$

Where $N(T_{ik})$ and $N(T_{jk})$ represent the collection of keywords in the user U_i and U_j in kth day microblogging. In order to guarantee the efficiency of recommendation, we select the threshold $\delta = 0.4$ for this step to be recommended list and to be filtered.

3.2 User Sentiment Analysis

In this section, we discuss the role of user sentiment in the followee recommendation. In the "psychology dictionary", the emotion is described as a kind of attitude experience whether which meet their own needs. Therefore, whether to meet the needs of users, in essence, is the user's emotional analysis at the recommended.

In the social media, the user's character determines that he will have different emotional responses to different things and problems, which have a certain impact on their behavior. Microblogging information not only express the user's views on different issues, but also contains their emotional information, which are helpful to analyze the user's emotions. People with similar emotions have a certain cohesive force, and the emotion analysis of users has been used in product recommendation[15], which has remarkable effect.

Emotional Similarity Analysis The main task of calculating the similarity of emotion is to extract the emotional vocabulary generated by the user in the text. In particular, the extraction of emotional words is mainly based on corpus and dictionary[16]. In this paper, because the object of study is Chinese vocabulary, the dictionary based method is adopted. We have defined a dictionary, which is used to count and mark the common degree emotion words, including 19 adverbs of degree and set subscript according to the importance of degree's order. We compute the similarity of two emotional words by subscript distance. If the microblogging contains multiple emotional words, we choose the subscript of the largest keywords in the extraction, which can better reflect the user's emotions. Similar to the calculation of text similarity, the sentiment similarity calculation takes into account the temporal factors, in which the vector of the emotional word dictionary is defined as follows:

$$Emotion_{dict} = \{em_1, em_2, ..., em_i\} \tag{6}$$

Then, the method for calculating the emotional similarity of two users is as follows:

$$Sim(U, U_i) = \sum_{k=1}^{m} (wf_k \times \frac{1}{1 + \alpha \left| E - E_i \right|}) \tag{7}$$

Where α is the distance decay parameter, The distance between the two emotional degree words shows that the similarity between the two is smaller, and vice versa. If the emotion words of two users are in the same location, the similarity is 1. If the distance is far, you need to refer to the value of α, and then calculate. When we get the final list of followees recommended, we need to filter the list to ensure the recommended quality, which is similar to the threshold of the text similarity, and the threshold ε is 0.4 we experimentally obtained.

3.3 PERSONALITY-BASED FACTORS

THE BIG-FIVE PERSONALITY MODEL The "Big Five" model of personality dimensions has emerged as one of the most well-researched and well-regarded measures of personality structure in recent years. The model's five domains of personality, Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism, were conceived by Tupes and Christal as the fundamental traits that emerged from analyses of previous personality across age, gender, and cultral lines.

THE MATCHING CALCULATION OF THE PERSONALITY TRAITS TextMind is a Chinese language psychological analysis system developed by Computational Cyber-Psychology Lab, Institute of Psychology, Chinese Academy of Sciences. TextMind provides easy access to analysis the preferences and degrees of different categories in text, which provides an all-in-one solution from automatic Chinese words segmentation to psychological analysis. The processing to dictionary, text and punctuation is optimized to simplified Chinese, and the categories is compatible to LIWC[7]. The TextMind is the same as the SC-LIWC dictionary. Through the relationship between the function of these words and the text, in this paper, we can obtain the relationship between each word in TextMind Chinese psychological analysis system dictionary and each specific factor in the big five personality. Because many words have multiple functions, in this content, we match these functional words with the big five personality, and then, the value of the comprehensive calculation and the average calculation is taken as the relationship between the term and each factor in the big five personality. The character factor values for the ith dimension in word w are defined as follows:

$$BFM(\omega_i) = \frac{\sum_{j=1}^{n} P_{ij}}{n} \tag{8}$$

In Eq.(8), ω_i denotes the ith personality factor in word w. n indicates that word w has n functional nouns. P_{ij} represents the relationship between the jth word function and the ith personality factor. The functional word of TextMind and BFM corresponding a correspondence table can be divided into five groups.

First of all, we filter the number of micro-blogging published more than 10, and then the user all the microblogging text is assembled into a large text. We put each user's micro-blogging text information into the TextMind and statistical processing in frequency. In this paper, we calculate the personality score vector of each dimension of the big five personality according to the BFM correspondence table[7]. The user u's personality score in the ith dimension is calculated as follows:

$$Score(u)_i = \frac{\sum_{j=1}^{N_i} k_{\omega_j} \cdot BFM(\omega_{ji})}{N_i} \tag{9}$$

Where $BFM(\omega_{ji})$ denotes the personality factor value of the jth-class word in the ith dimension in the micro-blogging text published by the user u. k_{ω_j} represents the word frequency of the jth-class word ω of the microblogging text published by the user u. N_i is the total number of functional words that are statistically relevant under the ith personality dimension. We use the Eq.(9) to calculate the average score between five dimensions in the table, which were μ_E, μ_A, μ_C, μ_N, μ_O, and then we put the user scores compared with the average correlation score. If it is higher than the average, it shows that the user has the personality characteristics of the dimension.

MATCHING THE PERSONALITY TRAITS BETWEEN USERS The purpose of this study is to recommend followee, so the personality matching score between a user u and the potential blogger pf is expressed as follows:

$$TPM(u, pf) = \mu(\sum MS(u, pf, dim)) \tag{10}$$

In Eq.(10), $TPM(u, pf)$ is the total personality matching(TPM) score between user u and potential followee pf. μ is the average value of each dimension. $MS(u, pf, dim)$ indicates the personality matching score of the user u and the potential recommendation followee pf in a certain dimension.

Next, the matching score calculation formula for each dimension is defined as follows[1]:

$$scoreAgreement(u, pf, dim) = \begin{cases} 0.5 & \text{both u and pf are dimension} \\ 0.25 & \text{either u or pf are dimension} \\ 0 & \text{None is dimension} \end{cases} \tag{11}$$

4 Personality-Aware Followee Recommendation Model Based On Text Semantics And Sentiment Analysis

Personality-aware followee recommendation model(PSER) based on text semantics and sentiment analysis proposed in this paper is shown in Fig.1..

In this paper, the text similarity of two users can be calculated by formula (5) get the first recommended list. Because different users in the expression of things will produce different emotional differences, so we need to continue to analyze the user's emotions. Microblogging text often contains some emotional vocabulary. According to the custom emotional dictionary, we extract the user's

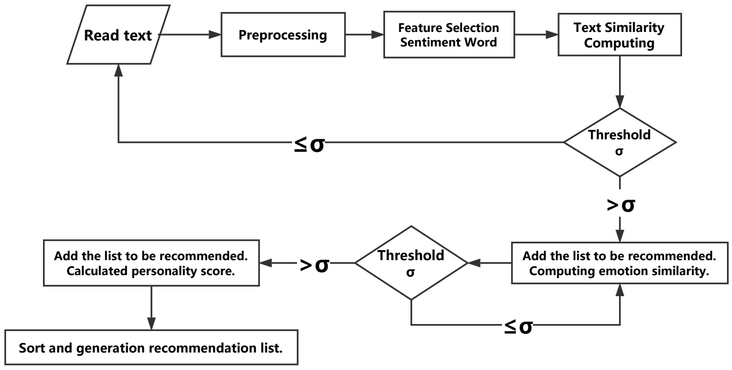

Fig. 1. PSER model

emotional word, and then calculate the emotional similarity, in order to filter the first recommendation list.

The PSER algorithm steps are in Algorithm 1:

Algorithm 1 The PSER algorithm

Input: The ID of user U
Output: Recommended list RFL

initial $R_m List = \emptyset, R_m = \emptyset$
for $U_i \in U_{list}$ **do**
  $Sim(U, U_i) = \sum_{k=1}^{m}(wf_k \times Sim(T_k, T_{ik}))$
  if $Sim(U, U_i) > \sigma$ **then**
    Add U_i to $R_m List$
  end if
end for
for $U_i \in R_m List$ **do**
  $Sim(U, U_i) = \sum_{k=1}^{m}(wf_k \times \frac{1}{1+\alpha|E-E_i|})$
  if $Sim(U, U_i) > \varepsilon$ **then**
    Add U_i to RL
  end if
end for
for $U_i \in RL$ **do**
  $TPM(U, U_i) = \mu(\sum MS(U, U_i, dim))$
end for
Add to RFL in descending order
return RFL

---

5 EXPERIMENT

In this section, we evaluate our design using experiment. We use the large-scale dataset crawled from Sina Weibo, the most popular microblogging system in China. The dataset contains 256.7 million users' social link information and 550 million tweets. The dataset include tweets, user relations and user background information.

5.1 Experimental Evaluation Criteria

In the recommender system, the most commonly used criteria is the precision and recall rate of Top-K recommendation. At the same time, we have added the evaluation criteria P@N because we need to consider the order of users in the recommendation results.

The precision can be expressed as the ratio of the number of followees to the total number of followees recommended, the formula is as follows:

$$Precision = \frac{N_{hit}}{N_r} \tag{12}$$

Where N_{hit} indicates the number of followees recommended correctly. N_r is the total number of recommended followees.

The recall rate can be expressed as the correct recommendation of the number of followees and the total number of followees. The formula is as follows:

$$Recall = \frac{N_{hit}}{N_A} \tag{13}$$

Where N_{hit} indicates the number of followees recommended correctly. N_A is the total number of followees of the user.

5.2 Experimental Design

In experiment, We designed two sets of experiments to show our results.

Exp 1. In the analysis of the user's emotional information, the formula (7) is used to calculate. This experiment selects the best value of α through the change of F-measure.

Exp 2. Generally, the traditional recommendation algorithm of followee and friend are the same. Therefore, the contrast methods used in the experiment are as follows:

FOF+:Calculating followee's recommendation based on users' common followees[17].

FOF+Tag:This method performs tag similarity matching according to the user's common followees[18].

UMFR(Unified Micro-blog Followees Recommendation):This method combines information about labels, location, attendance, and hot topics[19].

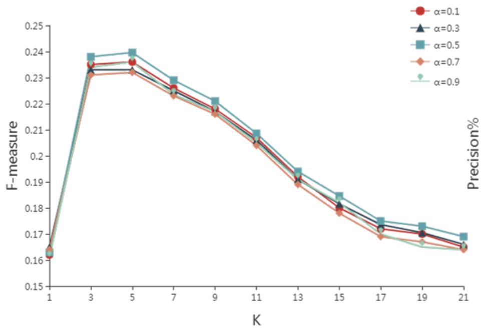

Fig. 2. Comparison of F-measure Values at Different Values

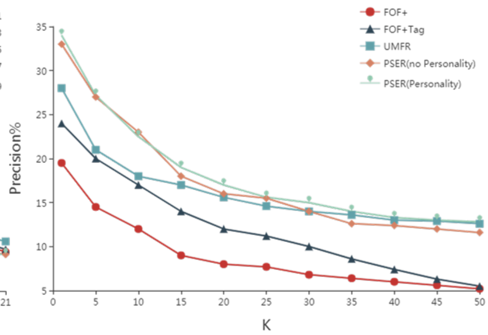

Fig. 3. Comparison of Precision Values at Different Values

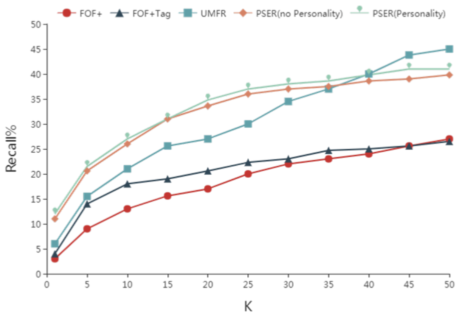

Fig. 4. Comparison of F-measure Values at Different Values

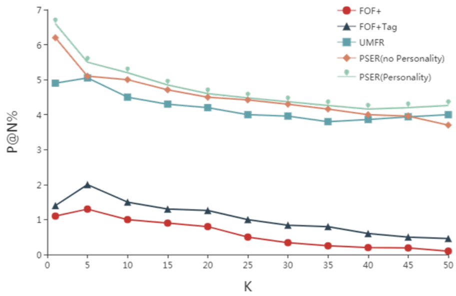

Fig. 5. Comparison of P@N Values at Different Values

5.3 Results

We compare our design with several existing schemes including FOF+, FOF+Tag, UMFR. In the experiment, the value of α is calculated by formula (7) and the experimental results show that the Fig.2. Fig.2 shows that the recommended effect can be the best when $\alpha = 0.5$. Therefore, in the user's emotional word similarity calculation, we take $\alpha = 0.5$ into equation(3).

Firstly, we use the randomly stratified sampling strategy to sample in the raw data. According to the number of users followees for the division, we randomly sampled in each layer of the user. Our experiments were conducted on sampled users, and the result can be calculated for the final precision, recall rate, and P@N value. The experimental results of the precision rate, recall rate and P@N value of the experimental data are shown in Fig.3, Fig.4, Fig.5.

Fig.3 shows the precision rate contrast diagram under different algorithms. As can be seen from the Fig.3, the algorithm we proposed has higher recommendation effect when the number of recommendation is less than 35 in this paper. Moreover, after adding personality traits, the precision rate has obviously improved.

Fig.4 shows the recall rate contrast diagram under different algorithms. It can be seen from the Fig.4 that the algorithm proposed in this paper can maintain good results when the recommended number is less than 40, and the addition of personality traits can also improve the recall rate.

Fig.5 shows the P@N contrast diagram under different algorithms. As can be seen from the Fig.5, the algorithm proposed in this paper is that the effect of the other algorithm is better in P@N. After adding personality traits, the algorithm slightly improved the P@N without considering the personality traits, and P@N tended to be steady as the number of recommended followees increased.

Through these experiments, we can see that the user's emotional analysis and personality traits can improve the recommended precision rate and recall rate to a certain extent. However, in the experiment without personality, the recommendation model proposed in this paper will decline with the increase of the recommended users.

6 Conclusions

In this paper, we propose a personality-aware followee recommendation model based on text Semantics and sentiment analysis(PSER), a novel personality followee recommendation scheme over microblogging systems based on text Semantics and sentiment analysis and the big-five personality model. This paper analysed how user personality conditions the followee selection process by combining a quantitative analysis of personality traits with the most commonly used predictive factors for followee recommendation.

Experiments show that the combination of the text semantics, emotional analysis and personality traits of followees recommended algorithm precision is higher than the traditional recommended algorithm. Also, this model can improve the recommendation quality and improve the novelty of the recommendation.

The combined attributes were insert into a recommendation algorithm that computed the similarity among target users and potential followees. We conduct experiments using large-scale traces form Sina Weibo to evaluate our design. Results show that PSER model greatly outperforms existing recommendation schemes.

References

1. Tommasel A, Corbellini A, Godoy D, et al. Personality-aware followee recommendation algorithms: An empirical analysis[J]. Engineering Applications of Artificial Intelligence, 2016, 51(C):24-36.
2. Selfhout M, Burk W, Branje S, et al. Emerging late adolescent friendship networks and Big Five personality traits: a social network approach[J]. Journal of Personality, 2010, 78(2):509C538.
3. Tkalcic M, Kunaver M, Tasic J, et al. Personality based user similarity measure for a collaborative recommender system[C]//Proceedings of the 5th Workshop on Emotion in Human-Computer Interaction-Real world challenges. 2009: 30-37.

4. Gao R, Hao B, Bai S, et al. Improving user profile with personality traits predicted from social media content[C]// ACM Conference on Recommender Systems. ACM, 2013:355-358.
5. Hamburger Y A, Benartzi E. The relationship between extraversion and neuroticism and the different uses of the Internet.[J]. Computers in Human Behavior, 2000, 16(4):441-449.
6. Quercia D, Lambiotte R, Stillwell D, et al. The personality of popular facebook users[C]// Acm Conference on Computer Supported Cooperative Work. DBLP, 2012:955-964.
7. Mairesse F, Walker M A, Mehl M R, et al. Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text[J]. Journal of Artificial Intelligence Research, 2007, 30(1):457–500.
8. Hu R, Pu P. Enhancing collaborative filtering systems with personality information[C]//Proceedings of the fifth ACM conference on Recommender systems. ACM, 2011: 197-204.
9. Wu W, Chen L, He L. Using personality to adjust diversity in recommender systems[C]// ACM Conference on Hypertext and Social Media. ACM, 2013:225-229.
10. Gou L, You F, Guo J, et al. SFViz: interest-based friends exploration and recommendation in social networks[C]// Visual Information Communication - International Symposium. ACM, 2011:15.
11. Yang T, Cui Y D, Jin Y H. BPR-UserRec: a personalized user recommendation method in social tagging systems[J]. Journal of China Universities of Posts & Telecommunications, 2013, 20(1):122-128.
12. Godin F, Slavkovikj V, Neve W D, et al. Using topic models for Twitter hashtag recommendation.[J]. 2013:593-596.
13. Xiong X B, Zhou G, Huang Y Z, et al. Dynamic evolution of collective emotions in social networks: a case study of Sina weibo[J]. Science China Information Sciences, 2013, 56(7):1-18.
14. Xin S, Lee T H, Whang J J, et al. Parallel Clustered Low-Rank Approximation of Graphs and Its Application to Link Prediction[C]// International Workshop on Languages and Compilers for Parallel Computing. Springer Berlin Heidelberg, 2012:76-95.
15. Bao S, Xu S, Zhang L, et al. Joint Emotion-Topic Modeling for Social Affective Text Mining[C]// IEEE International Conference on Data Mining. IEEE, 2009:699-704.
16. Silva N F F D, Hruschka E R, Jr E R H. Tweet sentiment analysis with classifier ensembles[J]. Decision Support Systems, 2014, 66:170-179.
17. Hsu W H, King A L, Paradesi M S R, et al. Collaborative and Structural Recommendation of Friends using Weblog-based Social Network Analysis [C]// AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs. USA: AAAI, 2006: 55-60.
18. Chu C H, Wu W C, Wang C C, et al. Friend Recommendation for Location-Based Mobile Social Networks[C]// Seventh International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing. IEEE Computer Society, 2013:365-370.
19. Feng S, Zhang L, Wang D, et al. A Unified Microblog User Similarity Model for Online Friend Recommendation[J]. Communications in Computer & Information Science, 2014, 496:286-298.