

Personalized Microblog Recommendation Using Sentimental Features

Wenjuan Cui, Yi Du, Zhihong Shen, Yuanchun Zhou, Jianhui Li
Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China
Email:{wenjuancui, duy, bluejoe, zyc, lijh}@cnic.cn

Abstract—Microblogging services have been popular in recent years. There are a large number of real-time microblog messages generated in each day which results in the information overload problem especially for the users with many followees. Personalized microblog recommendation can help the users out of the trouble of information overload. It is an interesting and important research topic with wide applications. Many kinds of features are used in the microblog recommendations in the existing algorithms. In this paper, except for the features such as the user's microblog posting history, reposting history, the relations with other users, and the contextual knowledge, we also utilize the sentimental information to help with the microblog recommendation. We first build a sentiment classifier based on the contextual information of the microblogs and get the sentimental feature set. Then a latent factor model incorporating the sentimental features and other information in microblogs is designed. Finally, we develop an experiment plan to evaluate the performance of the method. We believe that the utilization of the sentimental features will improve the performance of microblog recommendation.

Index Terms—Microblogging; personalized recommendation; sentiment analysis.

I. INTRODUCTION

Microblogging services, such as Twitter, Weibo, have been very popular in recent years. People tend to express their opinions, feelings and tell the things they are doing on Twitter or Weibo. Microblog makes it possible that the users share information with their friends. When the users release one piece of message, we say that they post a status or they post a microblog. As the length of one particular microblog record is often limited to 140 characters and the users may post messages at any moment and in any place, the microblogs usually consist of tremendous messages, which may be short, real-time, and contain masses of acronyms.

The microblog users may have several or hundreds of friends in their relation network. The messages posted by these friends will be presented to the user in a particular order. As a result of the rapid increasing number of microblogs, the serious problem of information overload occurs to most of the microblog users. For the users with many friends, there may be hundreds or even thousands of statuses presented to them. The most valuable messages which the users are really interested in may be flooded by the vast amount of irrelevant messages. In this situation, the personalized microblog recommendation appears to be very important to the users.

Intuitively, a posted status is useful for a user if the user is interested in it. Due to the noisy properties of microblog, recommending useful microblogs is a task of challenge. There have been some related works on personalized microblog recommendation. Wu et al. extract keywords from the tweets and then design a system to generate personalized tags for Twitter users to label their interests[1]. However, it usually leads to poor results for recommendation if the text mining techniques for traditional corpus of long text are directly applied[2]. Chen et al. [3] recommend microblogs based on collaborative ranking. Their method considers tweet topic level factors, user social relation factors and explicit features such as authority of the publisher and quality of the tweet in the process of recommendation.

The previous works for microblog recommendation concentrate on analyzing the content of the microblogs or the other information introduced by the user reading activities. However, in reality, the microblog users may share similar interests in certain topics but have different opinions on them. Sentiment analysis is an important part in personalized microblog recommendation although few work has been done on it.

Identifying sentiments or opinions from microblogs can facilitate many disciplines. Xie et al. give a generic framework to incorporate various sentiment information to sentiment-based information for personalized search by user profiles and resource profiles[4]. Rao et al. propose a topic-level maximum entropy model for social emotion classification over short text[5]. Sentiment analysis is also used in twitter user recommendation which recommends new friends with similar interests[6]. In this paper, we will incorporate the sentimental features into the traditional content based microblog recommendation.

II. SENTIMENTAL FEATURES EXTRACTION

To extract the sentimental features of the microblogs, we have to carry out sentiment analysis. Microblog users often express their personal character and opinion bias in their posted messages. Due to the different language habits of different users, the same word may stand for different meanings. The method only using the content of the microblogs for analysis is not sufficient. As the microblog messages are short and contain many acronyms and informal words, the contextual knowledge will help improve the effectiveness of the sentiment analysis method.

Sentiment classification is often to identify the polarity of the entities which are the microblogs here. In this work, we focus on binary sentiment classification and define the sentiment polarities as positive or negative. Our task is to extract the sentiment polarity values for each microblog on certain topics. We use the similar idea with the work of Wu et al.[7] and a contextual knowledge based sentiment classification framework is built. In the framework, two kinds of contextual knowledge from large scale short messages, i.e., the word-word association and word-sentiment association are established. The word-word association is the information indicating that two words may share similar sentiment. We compute the co-occurrence frequency of the words by certain rules, and then utilize the pointwise mutual information (PMI) as the measure of the sentiment similarity between a pair of words. The prior knowledge about the sentiment polarity of words is indicated in the word-sentiment association. After we compute the word-sentiment association, we model the word-sentiment knowledge into a linear regularization term and model the word-word relation into a graph-guided fused lasso regularization term. A linear classification model for sentiment classification is built. An efficient optimization method is then designed to solve the regularized optimization problem. Through this method, we get the sentiment polarities for each microblog message and build the sentimental feature set.

III. COLLABORATIVE RECOMMENDATION

Collaborative filtering is a commonly used recommendation method by learning users' preferences based on the community data. In collaborative filtering, the similar users are considered to have similar interests on certain subjects. Unlike the content based recommendation method, it does not require the explicit user profiles or domain knowledge. As the microblog message may be very short, it is difficult to do content analysis. The collaborative filtering method is more suitable than the content based recommendation method. There are two kinds of frequently used collaborative filtering methods which are neighborhood-based approaches and model-based approaches.

In this work, we utilize the model-based collaborative filtering method for recommendation. We build a latent factor model and the features of microblog posting history of the users, the reposting history and the friend relation of the users are all used in the model. Besides, the sentimental features obtained earlier will also be incorporated into the model. Suppose the number of users is U , the number of microblogs is I and there are K features for each microblog. We build a matrix $P_{U,K}$ in which $p_{u,k}$ stands for the preference of user u for feature k , and a matrix $Q_{K,I}$ in which $q_{k,i}$ stands for the value of feature k for the i th microblog. The latent factor model is defined in formula (1).

$$R_{U,I} = \sum_{k=1}^K P_{U,K} Q_{K,I} \quad (1)$$

$r_{u,i}$ stands for the interest value of user u on the i th microblog. The model helps us to solve the problem of information sparsity in users' reposting actions.

IV. EXPERIMENT PLAN

In the experiment part, we aim to use the real Weibo data to evaluate the performance of our method. Weibo is a popular social microblogging service in China, which is similar to twitter. It has a large number of users and tremendous messages are generated in each second. Therefore, using the Weibo data makes sure that the model in this work will be evaluated sufficiently. As the Weibo data is public to its users, we first crawl the Weibo data via Weibo APIs. Then we preprocess the data for extracting features. Utilizing the contextual information, we build the sentiment classifier and get the sentimental feature set for the microblogs. The other features of the microblogs and the users along with the sentimental features are incorporated into the latent factor model. The results produced by the model will then be ranked for recommendation. We will compare the results with the case where sentimental features are not used in the recommendation.

V. CONCLUSION

In this paper, we focus on the personalized microblog recommendation with the help of sentimental information. We first design a sentiment classifier based on the contextual knowledge of the microblogs and get the sentimental feature set. Then a latent factor model is built with the sentimental features incorporated. An experiment is planned using the real Weibo data for evaluation. In the future, we will consider more microblog features in the model and design more specific experiment to evaluate the method.

ACKNOWLEDGMENT

This work was supported by the National key Research Program of China (2016YFB1000600, 2016YFB0501900), the National Natural Science Foundation of China under Grant No. 61402435, 41371386, and the Strategic Priority Research Program of the Chinese Academy of Sciences (No.XDA06010307).

REFERENCES

- [1] Wu, W., Zhang, B., and Ostendorf, M.(2010). Automatic generation of personalized annotation tags for twitter users. In Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pages 689-692.
- [2] Duan, Y., Jiang, L., Qin, T., Zhou, M., and Shum, H.(2010). An empirical study on learning to rank of tweets. In Proceedings of the 23rd International Conference on Computational Linguistics, pages 295-303.
- [3] Chen, K., Chen, T., Zheng, G., Jin, Ou., Yao, E., and Yu, Y.(2012). Collaborative Personalized Tweet Recommendation. SIGIR12, August 12-16, 2012, Portland, Oregon, USA.
- [4] Xie, H., Li, X., Wang, T., Lau, R. Y., Wong, T. L., Chen, L., Wang, F. L., and Li, Q. (2016). Incorporating sentiment into tag-based user profiles and resource profiles for personalized search in folksonomy. Information Processing & Management, 52(1), 61-72.
- [5] Rao, Y., Xie, H., Li, J., Jin, F., Wang, F. L., and Li, Q. (2016). Social emotion classification of short text via topic-level maximum entropy model. Information & Management, 53(8), 978-986.
- [6] Gurini, D., Gasparetti, F., Micarelli, A., and Sansonetti, G.(2013). A Sentiment-Based Approach to Twitter User Recommendation. RSWE-B@RecSys.
- [7] Wu, F., Song, Y., Huang, Y.(2015). Microblog Sentiment Classification with Contextual Knowledge Regularization. In AAAI 2015, 2332-2338.