

An Overview: Rating Prediction System From Textual Review

Miss. Snehal Manmohan Rathi

Computer Science and Engineering

H.V.P.M.'s College of Engineering & Technology, Amravati.

Prof. P. L. Ramteke

Information Technology

H.V.P.M.'s College of Engineering & Technology, Amravati.

Abstract— In recent years, shopping online is becoming more and more popular. When it need to decide whether to purchase a product or not on line, the opinions of others become important. It presents a great opportunity to share our viewpoints for various products purchase. However, people face the information overloading problem. How to mine valuable information from reviews to understand a user's preferences and make an accurate recommendation is crucial. Traditional recommender systems consider some factors, such as user's purchase records, product category, and geographic location. In this work, it propose a sentiment-based rating prediction method to improve prediction accuracy in recommender systems. Firstly, it propose a social user sentimental measurement approach and calculate each user's sentiment on items. Secondly, it not only consider a user's own sentimental attributes but also take interpersonal sentimental influence into consideration. Then, consider item reputation, which can be inferred by the sentimental distributions of a user set that reflect customers' comprehensive evaluation. At last, by fusing three factors-user sentiment similarity, interpersonal sentimental influence, and item's reputation similarity into recommender system to make an accurate rating prediction. It conduct a performance evaluation of the three sentimental factors on a real-world dataset. Experimental results show the sentiment can well characterize user preferences, which help to improve the recommendation performance.

Index Terms— Item reputation, Reviews, Rating prediction, Recommender system, Sentiment influence, User sentiment.

I. INTRODUCTION

With the development of Web, more and more people are connecting to the Internet and becoming information producers instead of only information consumers in the past, resulting to the serious problem, information overloading. There is much personal information in online textual reviews, which plays a very important role on decision processes. For example, the customer will decide what to buy if he or she sees valuable reviews posted by others, especially user's trusted friend. People believe reviews and reviewers will do help to the rating prediction based on the idea that high-star ratings may greatly be attached with good reviews. Hence, how to mine reviews and the relation between

reviewers in social networks has become an important issue in web mining, machine learning and natural language processing. It focus on the rating prediction task. However, user's rating star-level information is not always available on many review websites. Conversely, reviews contain enough detailed food information and user opinion information, which have great reference value for a user's decision. Most important of all, a given user on website is not possible to rate every product or item. Hence, there are many unrated products or items in a user-item-rating matrix. In such case, it's convenient and necessary to leverage user reviews to help predicting the unrated items.

Sentiment analysis is the most fundamental and important work in extracting user's interest preferences. In general, sentiment is used to describe user's own attitude on product or items. It is observed that in many practical cases, it is more important to provide numerical scores rather than binary decisions. Generally, reviews are divided into two groups, positive and negative.

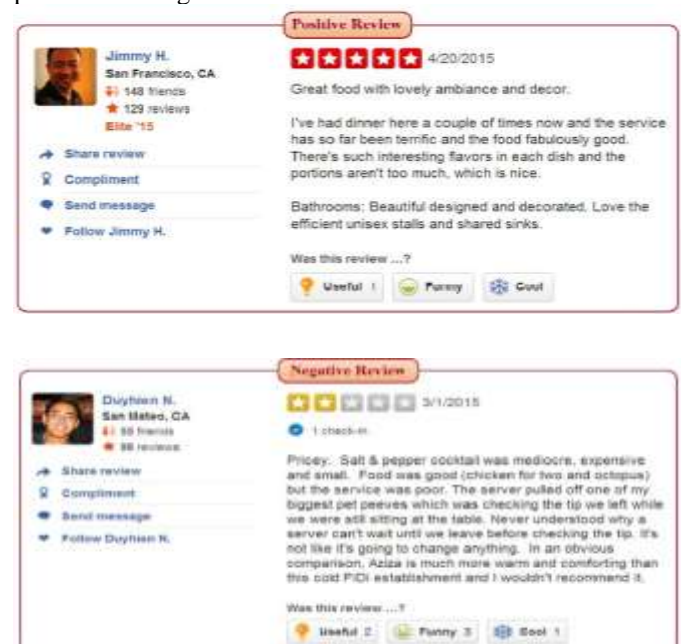


Fig. 1. An example of positive review and negative review on websites.

In Fig.1, we intuitively show an example of positive reviews and negative reviews on website. From Fig.1, there

are many positive words in a 5-star review, such as “great”, and “lovely”. But in a 2-star review we find negative words, such as “expensive”, and “poor”. That means a good review reflects a high star-level and a bad review reflects a low-level. When we know the advantages and disadvantages from the two kinds of reviews, we can easily make a decision. Normally, if item’s reviews reflect positive sentiment, then the item may be with good reputation. Oppositely, if item’s reviews are full of negative sentiment, then the item is most likely with bad reputation. So based on users’ reviews sentiment, we can infer users’ comprehensive ratings on items.

However, it is difficult for customers to make a choice when all candidate products reflect positive sentiment or negative sentiment. To make a purchase decision, customers not only need to know whether the product or item is good, but also need to know how good the item is. It’s also agreed that different people may have different sentimental expression preferences.

To address these problems,[12] propose a sentiment-based rating prediction method in the framework of matrix factorization. In our work, we make use of social users’ sentiment to infer ratings. Fig. 2 is an example that illustrates our motivation. First, we extract product features from user reviews. Then, we find out the sentiment words, which are used to describe the product features. Besides, we leverage sentiment dictionaries to calculate sentiment of a specific user on an item/product. What is more, we combine social friend circle with sentiment to recommend. In Fig.2, the last user is interested in those product features, so based on the user reviews and the sentiment dictionaries, the last item will be recommended. Compared with previous work [6-8], the main difference is that: we use unstructured information to recommend instead of other structured social factors. Compared with [3], [5], [9], the main difference is that: their work mainly focuses on classifying users into binary sentiment (i.e. positive or negative), and they do not go further in mining user’s sentiment. In our paper, we not only mine social user’s sentiment, but also explore interpersonal sentimental influence and item’s reputation. Finally, we take all of them into the recommender system.

The main contributions of our approach are as follows: 1) we propose a user sentimental measurement approach, which is based on the mined sentiment words and sentiment degree words from user reviews. Besides, some scalable applications are proposed. For example, we explore how the mined sentiment spread among users’ friends. What is more, we leverage social users’ sentiment to infer item’s reputation, which showed great improvement in accuracy of rating prediction. 2) We make use of sentiment for rating prediction. User sentiment similarity focuses on the user interest preferences. User sentiment influence reflects how the sentiment spreads among the trusted users. Item reputation similarity shows the potential relevance of items. 3) We fuse the three factors: user sentiment similarity, interpersonal sentimental influence, and item reputation similarity into a probabilistic matrix factorization framework to carry out an accurate recommendation. The experimental results and discussions show that user’s social sentiment that we mined is a key factor in improving rating prediction performances.

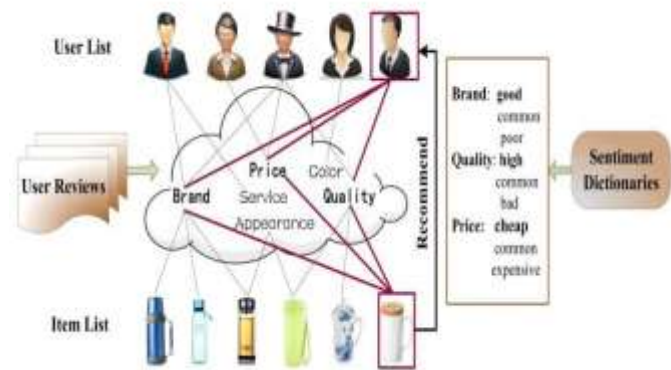


Fig. 2. The product features that user cares about are collected in the cloud including the words “Brand”, “Price”, and “Quality”, etc. By extracting user sentiment words from user reviews, we construct the sentiment dictionaries. And the last user is interested in those product features, so based on the user reviews and the sentiment dictionaries, the last item will be recommended.

II. LITERATURE REVIEW & RELATED WORK

In this section, we survey recent work related to our approach. Firstly, we review some approaches based on collaborative filtering (CF). Then, we review the often utilized rating prediction/recommendation methods based on matrix factorization.

A. Collaborative Filtering

Collaborative filtering (CF) is an important and popular technology for recommender systems. The task of CF is to predict user preferences for the unrated items, after which a list of most preferred items can be recommended to users. The methods are classified into user-based CF and item-based CF. The basic idea of user-based CF approach is to find out a set of users who have similar favour patterns to a given user (i.e., ‘neighbours’ of the user) and recommend to the user those items that other users in the same set like, while the item-based CF approach aims to provide a user with the recommendation on an item based on the other items with high correlations (i.e., ‘neighbours’ of the item). In all collaborative filtering methods, it is a significant step to find users’ (or items’) neighbours, that is, a set of similar users (or items). Currently, almost all CF methods measure users’ similarity (or items’ similarity) based on co-rated items of users (or common users of items). Collaborative filtering and content based filtering have been widely used to help users find out the most valuable information.

B. Matrix Factorization based Approaches

1) Basic Matrix Factorization

Matrix factorization is one of the most popular approaches for low-dimensional matrix decomposition. Matrix factorization based techniques have proven to be efficient in recommender systems when predicting user preferences from known user-item ratings. Matrix can be inferred by decomposing item reviews that users gave to the items. Matrix factorization methods have been proposed for social

recommendation due to their efficiency to dealing with large datasets. Several matrix factorization methods have been proposed for collaborative filtering. The matrix approximations all focus on representing the user-item rating matrix with low-dimensional latent vectors.

2) Social Recommendation

In real life, people's decision is often affected by friends' action or recommendation. How to utilize social information has been extensively studied. Yang et al. [6] propose the concept of "Trust Circles" in social network based on probabilistic matrix factorization. Jiang *et al.* [7] propose another important factor, the individual preference. Some websites do not always offer structured information, and all of these methods do not leverage users' unstructured information, i.e. reviews, explicit social networks information is not always available and it is difficult to provide a good prediction for each user. For this problem the sentiment factor term is used to improve social recommendation.

C. Reviews based Applications

There are also many reviews based work for the task of recommendation. Wang et al. [1] propose a review rating prediction method by incorporating the social relations of a reviewer. In addition, they classify the social relations of reviewers into strong social relation and ordinary social relation. In addition, they classify the social relations of reviewers into strong social relation and ordinary social relation. Luo *et al.* [10] define and solve a new problem: aspect identification and rating, together with overall rating prediction in unrated reviews.

D. Sentiment based Applications

Sentiment analysis can be conducted on three different levels: review-level, sentence-level, and phrase-level. Review-level analysis, [2] and sentence-level analysis [11] attempt to classify the sentiment of a whole review to one of the predefined sentiment polarities, including positive, negative and sometimes neutral. While phrase-level analysis [3], attempt to extract the sentiment polarity of each feature that a user expresses his/her attitude to the specific feature of a specific product.

There are many approaches leveraging sentiment analysis for personalized recommendation [4], [3], [5]. Zhang et al. [4] propose a self-supervised and lexicon-based sentiment classification approach to determine sentiment polarity of a review that contains both textual words and emotions. And they use sentiment for recommendation. By analyzing the user ratings, they can recommend special experts to a target user based on the user population. The information contained in user-service interactions can help predict friendship propagations and vice versa. They use data from both user item interactions and user-user relations.

III. THE PROPOSED APPROACH

The proposed method comprises main components: Identify social relation between users, sentiment dictionaries, Recommendation system and User.

Proposed method

The purpose of approach is to find effective clues from reviews and predict social users' ratings. We firstly extract product features from user review corpus, and then we introduce the method of identifying social users' sentiment. At last we fuse all of them into our sentiment-based rating prediction method. It proposed a Highest rating recommendation system for products and items. The contributions can be summarized as follows:

- I. It propose a recommendation system for food items. To develop the recommendation system, rating data sets of products and items in the particular category which is used to read the textual reviews given by the users. The main categories which are used in the application are nothing but Lectures & books, Fashions, Food & Drink, Sports, Kids & Family, Electronic appliances. The datasets used in this recommendation system are "DouBan" and "Yelp" and other review websites provides a broad thought in mining user preferences and prediction user's ratings. And other dataset used is nothing but "Online Product Rating" Dataset.
- II. Textual reviews obtained from data sets is categorised into three types: To identify positive reviews, To identify negative reviews and To identify neutral reviews. With the help of these types of reviews we can identify the social relation between users which will help to categories the item. Fig 3 shows how review analysis is done form the original reviews on the websites.
- III. Sentimental dictionaries will give the information of brands, quality and price on the basis of matrix factorization. This matrix factorization can be performed by using two types of methods which are by applying conjunctive rules and another is by comparing product feature and sentiment words.
- IV. This matrix factorization method will ultimately give the highest rating product recommendation for all types of products and items to the user.
- V. This recommendation system can be used by the user to select which items to be ordered or purchased and which are not. This recommendation system will help to take any decisions for any type of product.



Fig. 3. An example of review analysis for identifying user's sentiment. Product features are denoted in red font, the sentiment words are denoted in green font, the sentiment degree words are denoted in blue font, the conjunction words like "and", "but" are denoted in blank font, and the negation words are denoted in bright green font.

IV. APPLICATIONS

- I. Recommend the people or user what is best for them to purchase or order product from textual reviews.
- II. The rating prediction system will give the rating to the all products which are popular and which are not even used.
- III. System help the user to do their work in less time and with good quality product for healthy life.

V. CONCLUSION

In this paper, a recommendation model is proposed by mining sentiment information from social users' reviews. We propose social user's sentiment measurement approaches based on the mined sentiment words and sentiment degree words from users' reviews. We fuse user sentiment similarity, interpersonal sentiment influence, and item reputation similarity into a unified matrix factorization framework to achieve the rating prediction task. This rating product recommendation system which can be also used as the social relation collaboration model which can be used to identify the social relation between the users. The genuine reviews will give the rating prediction easy and user will easily get the result in the desired time. This prediction based on rating also decides the product's or item's purchasing quality whether it is good or bad. This recommendation will also help us to identify the products reputation on the basis of good and bad reviews. This type of accurate recommendation system can be used to identify the items or products on mobile also. Also the products or items which do not have any type of rating or do not have any reviews will also be recommended to the person or user for the good decision. This use will help the users to do their work in less time and with good quality products for healthy life.

REFERENCES

- [1] B. Wang, Y. Min, Y. Huang, X. Li, F. Wu, "Review rating prediction based on the content and weighting strong social relation of reviewers," in *Proceedings of the 2013 international workshop of Mining unstructured big data using natural language processing*, ACM, 2013, pp. 23-30.
- [2] D. Tang, Q. Bing, T. Liu, "Learning semantic representations of users and products for document level sentiment classification," in *Proc. 53th Annual Meeting of the Association for Computational Linguistics and the 7th*

International Joint Conference on Natural Language Processing, Beijing, China, July 26-31, 2015, pp. 1014-1023.

- [3] Y. Zhang, G. Lai, M. Zhang, Y. Zhang, Y. Liu, S. Ma, "Explicit factor models for explainable recommendation based on phrase-level sentiment analysis," in *proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, 2014.
- [4] W. Zhang, G. Ding, L. Chen, C. Li, and C. Zhang, "Generating virtual ratings from Chinese reviews to augment online recommendations," *ACM TIST*, vol.4, no.1, 2013, pp. 1-17.
- [5] X. Lei, and X. Qian, "Rating prediction via exploring service reputation," *2015 IEEE 17th International Workshop on Multimedia*
- [6] X. Yang, H. Steck, and Y. Liu, "Circle-based recommendation in online social networks," in *Proc. 18th ACM SIGKDD Int. Conf. KDD*, New York, NY, USA, Aug. 2012, pp. 1267-1275.
- [7] M. Jiang, P. Cui, R. Liu, Q. Yang, F. Wang, W. Zhu, and S. Yang, "Social contextual recommendation," in *proc. 21st ACM Int. CIKM*, 2012, pp. 45-54.
- [8] Z. Fu, X. Sun, Q. Liu, et al., "Achieving Efficient Cloud Search Services: Multi-Keyword Ranked Search over Encrypted Cloud Data Supporting Parallel Computing," *IEICE Transactions on Communications*, 2015, 98(1):190-200.
- [9] Y. Ren, J. Shen, J. Wang, J. Han, and S. Lee, "Mutual Verifiable Provable Data Auditing in Public Cloud Storage," *Journal of Internet Technology*, vol. 16, no. 2, 2015, pp. 317-323.
- [10] W. Luo, F. Zhuang, X. Cheng, Q. H. Z. Shi, "Ratable aspects over sentiments: predicting ratings for unrated reviews," *IEEE International Conference on Data Mining (ICDM)*, 2014, pp. 380-389.
- [11] T. Nakagawa, K. Inui, and S. Kurohashi, "Dependency tree-based sentiment classification using CRFs with Hidden Variables," *NAACL*, 2010, pp.786-794.
- [12] Xiaojiang Lei, Xueming Qian, Member, IEEE, and Guoshuai Zhao, "Rating Prediction based on Social Sentiment from Textual Reviews," *IEEE Transactions On Multimedia*, MANUSCRIPT ID: MM-006446