

The Rating Based Recommender System using Textual Reviews: A Survey

Mohd. Danish, Meharban Ali

Research Scholar, School of CS & IT, MANUU, Hyderabad, Telangana, India

ABSTRACT

Nowadays online shopping is emerging as a growth of business. Customers are getting used to purchasing the items online. Online reviews are an essential resource for users choosing to buy a product, watch a movie or go to a hotel. When it needs to decide the items/products through online, the opinions of other users through review matter a lot. It gives a good idea of the product to be purchasable or not. However, people face the information overloading problem. So the problem is as to how to get valuable information from user reviews so as to understand a user's preference and make an accurate recommendation. Recommender systems become risen as an essential tool to overwhelm the negative result of information overloading problem. The traditional recommendation system examines some factors like the user's buying records, product classification, and user's geographic location. This paper is an attempt to discuss the three social factors with some rating prediction algorithms based on user sentiment similarity, item reputation and user circle influence and review the applicable sentiment dictionary to the recommender system.

KEYWORDS: User sentiment reviews, Sentiment analysis, Recommender systems, Item reputation, Rating Prediction

1. INTRODUCTION

Today online textual reviews play very important role in the decision process. Most of the websites take reviews of the user so as to develop their business. For example, it will be easy for the user to purchase the products online as another user gives much more information about the product through reviews. When it comes to online shopping people pay much more attention to other user reviews, especially in user's circle friends. User's reviews will help in rating prediction as high-quality products will get attached with good user reviews. Hence, as to how extract useful reviews of the users' and the relationship among the reviewers in the social networking platform has broadly discussed in web mining.

Though, the user's rating is not always available on many websites. As reviews contain quite detail about the product information and the user opinion information, which have a great impact on a user's decision. There is a possibility that not every user prefers to rate every item on a website. Therefore, there are many unrated items still left in a user-item matrix. It is discussed in several rating prediction methods, e.g. [1], [4]. As we know user reviews are always available on the website. In such situation, it's useful and essential to parse the user's reviews to help to predict the unrated product/items for the user. The growth of review websites gives a view of mining the user choices and predicting user's ratings. Sentiment analysis is mostly used for extracting the user's interest. In general, the sentiment is practiced to describe user's own opinions about the items. Generally, reviews are categorized into three groups, positive, negative and neutral. However, it is quite difficult for the user's to take a decision when all candidate items show positive sentiment, negative sentiment or neutral sentiment. Customers not only pay attention towards better product but want to know how extremely good the product is. It is known that different users may have different opinions expression priority. For eg., users may use word "fine" to define a "splendid" product, while others may prefer to use "fine" to define a "wonderful" product [11]. That is

users are more concerned about item's reputation. For the product/item reputation, sentiment of the reviews is required. Normally, if the item's reviews show positive sentiment, then the item considered as a good reputation, but if the item's reviews show a negative opinion then the item considered to be a bad reputation. If we understand user sentiment, we can easily know the item reputation and even the user ratings. When we explore the internet for online purchasing, both positive and negative reviews are important to be a preference. It is observed that one user's review will influence the other users and, Users review gives more information about the product. Despite, it's hard to guess the user's textual sentiment. So there is much need to pay attention to the interpersonal influence so as to extract the user preferences. Most of the methods of the interpersonal social influence in social networking platform have shown good performance in the recommendation, which can appropriately solve the "cold start" problems and "sparsity problem". Despite, the existing approaches [2], [3], [5], [6], [9] mainly focus on product class information or tag information to study the interpersonal social influence of the user. These methods only use for structured data, which is not always available on every website. Though user reviews may be useful in mining the interpersonal social inference and user choices.

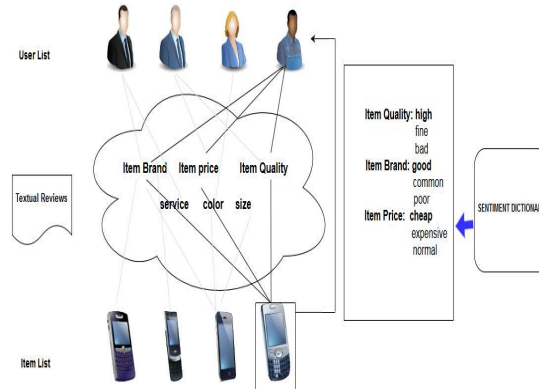


Fig-1 Sentiment-based recommendation

2. OBJECTIVE

- To improve the accuracy rating prediction.
- To perform a sentiment analysis on textual reviews.
- To overcome the sparsity problem.
- To implement a recommender system.

3. RELATED WORK

In this paperwork, we survey some methods of collaborative filtering technique, the rating prediction approaches of matrix factorization (MF), review based approaches, and sentiment analysis.

3.1 Collaborative Filtering

The collaborative filtering scheme is used to predict the user preferences for the unrated items and after that, it recommends the most preferred items out of the list to the users. Nowadays it is mostly used recommender system technique. It provides the best preferences to the user. Many website using this technique like Amazon, twitter, etc.. As we are aware that 33% of sales of Amazon is just because of recommender system which uses the collaborative technique. Algorithms [18], [12], [24], [26], [35] have already been devised so as to get better the quality of the recommender system. An algorithm for Collaborative approach is CF method which is an old algorithm. The basic thought is that users prefer to get those products which they used to purchase after taking a look on their history preferences. Sutter [9] propose a method that enables tags to get fused into CF algorithms and combine the three-dimensional relationships between items, users, and tags. Moreover, in method [12] it provides a user's rating to an item by determining the average ratings of related or correlated items by the same user. Its performance enhances when determining the similarity between items. S.Gao [28] proposed a collaborative filtering recommendation scheme based on topic relationship, they assume that experts with related topics would possess similar feature vectors.

3.2 Matrix Factorization based Methods

3.2.1 Basic Matrix Factorization

The matrix factorization method used for low-dimensional matrix decomposition. It is the product of matrices of a matrix.. There are various distinct matrix decompositions, all find use amongst a selective class of queries. These methods have shown to be useful for predicting the user decisions from observed user rating matrix. A matrix is presumed by decomposing the user reviews what users assigned to the product. Matrix factorization approaches are recommended for social recommender system due to their capability to handle the large datasets. For collaborative scheme there are many matrix factorization schemes have been devised. From the Basic matrix factorization[1], a potential eigenvector matrix is used for both the Recommendation users and items, and it calculated all the rating value.

3.2.2 Social Recommendation

Some matrix factorization that are based on social recommendations is meant to resolve the "cold start" problems. In today life, people's decision is often influenced by the friends' action or recommendation. People tend to influence when it comes to buying the items. How to get Social information is broadly analyzed. Yang [2] Propose the "Trust Circles" in a social networking platform based on matrix factorization. Jiang [3] introduce another important factor, the personal preference. Some sites always not offer the structured information. These approaches are only

suitable for structured information, but not for the unstructured data(i.e. textual data). Hence, social information of each user is not available and it is quite complex to offer a reliable prediction for each user. To solve the issue, the sentimental factor is used to enhance social recommendation.

3.3 Applications based on Reviews

There are several recommendation tasks has been done on the basis of user reviews or comments. Most websites using user's reviews for the growth of their business. Qu [13] propose a method which predicts a user's numerical rating in an item review, and they use a constraint regression technique for determining the scores of sentimental opinions. Jang [10] propose a rating review prediction system by taking the social connections of a user or reviewer. It is used to analyze the social relation of user's/reviewers into strong and ordinary connection. Zhang [16] include various item review factors such as product quality, content, time of the review, durability of the item and positive reviews of the customer. They perform a product ranking model that applies weights to product/item review factors so as to calculate the ranking score. Ling et al. [23] proposes a model that fuses content-based collaborative filtering and examining the data of the ratings as well as the reviews. Luo [17] identify and resolve a new problem: aspect recognition and rating collectively with total rating prediction for unrated reviews. They propose an LDA-style topic design which makes ratable aspects of sentiment and associated modifiers with rating.

3.4 Sentiment-based Applications

It is used to check whether the text is negative, positive or neutral. This is also called as opinion mining, concluding the opinion or emotion of an announcer. Sentiment analysis opinion is basically directed on review-level, sentence-level, and phrase-level. Review-level [20], [21] and sentence-level analysis [22] tried to analyze the entire review as a positive, neutral and negative. Phrase-level [26], [24] try to extract the feature of the item rooted on the feature likings of the user. The main idea of using phrase-level sentiment analysis is the development of sentiment lexicon. Pang [20] propose a contextual-insensitive evaluative lexical approach. Though, it can't deal with the mismatch between the base valence of the term and the author's usage. Polanyi [18] explain that base valence of a lexical product is transformed by lexical and context, it also proposed a method for some contextual shifters. They measure user sentiment opinion based on a finer-grained approach. Taboada [19] exhibit a semantic familiarization calculator which uses the dictionaries of words interpreted with their semantic familiarization (polarity and strength) and incorporates intensification and converse. There are several schemes to sentiment analysis used to solve personalized suggestion [8], [25], [26], [27]. Zhang [8] propose a self-managed and lexicon-based sentiment distribution approach to find out the sentiment of a review that comprises both text words and emotions. They use the sentiment for the recommendation of the product. With analyzing ratings of a user, they can infer specific experts to a destination user based on the people population. The data held in the user-service cooperations can help in predict the friendship propagations. Then the data is collected from user-item interaction and user-user connection. Lee [25] introduce a recommendation system utilizing the idea of Experts to determine both novel and relevant recommendations. From parsing the ratings, of the

user it can suggest special experts to a target user based on the people population. Lei [27] work on phrase-level opinion analysis to conclude a particular item's reputation. They also propose the concept of "virtual friends" to model items' relations, which can reduce the time complexity during training. Zhang [26] recommend an EFM model to make an understandable recommendation, they derive explicit item features and user choices by phrase-level sentimental analysis of user reviews.

4. LDA APPROACH

4.1 Product Features Extraction

Product features mean discussing the issues of a product. In this paperwork, we review the extracted item features of textual reviews utilizing LDA [7]. LDA is a generative topic pattern extractor. This technique applied to statistics, pattern recognizing and machine learning so as to get a linear sequence of characteristics that defines or classifies two or more classes of objects or events. The features are named entities, product attributes etc. As LDA is an analytical model which is used to exhibit the relationship between topics, reviews, and words. LDA describes documents as mixtures of topics eject out words with several probabilities. It works as following steps when writing the documents:

- I. Decide the number of words for the N document (using Poisson distribution).
- II. Pick a topic mix for the record/document with the help of the Dirichlet distribution fixed set of L topics.
- III. Produce each word w in the document by:
 - A. Selecting a topic (by using multinomial distribution)
 - B. Produce the word utilizing the topic (topic's multinomial distribution).

4.2 Data Pre-processing

By filtering process, words are gathered from the user's reviews positive words, neutral words, negative words and the sentiments degree of words. It also filters out "Stop Words" [14, 15] like a pronoun, article, etc. All the different words of the reviews are included in vocabulary.

4.3 Generating The Process

It studies all the user's document as D and the number of topics expressed as m. The output will be represented as topic preference distribution of each user and topic list consists of 10 features words.

Table1 : Sentiment Dictionary

Levels	Product Features
Level 1	Best, Awesome, Superb, Amazing
Level 2	Very, Better, Fine, Slightly
Level 3	Comparative, Further, Rather, Few
Level 4	Some, Just, A little bit, Quite
Level 5	Less, Little, Insufficient, Not very

5. RATING PREDICTION ALGORITHMS

There are many algorithms has been offered for rating prediction. In this review paper, we study some algorithms of rating prediction.

- A. Basic MF: It is baseline matrix factorization method approach [1]. This method doesn't contain any the social factors.
- B. Circle Con: It is proposed in [8], this method uses the interpersonal trust factor in the social networking platform. By using matrix factorization it suggests the

trust circles.

- C. Context MF: This method used the social factors; interpersonal influence and individual preferences. This method [3] enhances the performance of traditional item-based collaborative filtering technique [12], [24].
- D. PRM: This approach proposed in [5]. It gets to consider the three social factors, i.e. interpersonal interest similarity, interpersonal influence, and personal interest. It also used the matrix factorization for prediction of the user's ratings.
- E. EFM: This approach [26], builds two specific matrixes: the first one is user-feature concentration matrix and another one is an item-feature quality matrix. User-feature matrix measure the product feature of the item which the user cares. The item-feature matrix contains the quality of an item of the product.

6. ACKNOWLEDGEMENT

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions, which helped a lot in improving the quality of the paper. Last but not the least I am very thankful to my guide to encourage and assist me to write this paper.

7. CONCLUSION

Recommender system has highly influenced the business. Every website getting use of recommender system so as to enhance their business. As more and more data are used in the form of reviews so this is the best idea to utilize that Reviews data. Textual reviews easily help to identify with what the user wants to say about a product. Collaborative filtering gives the information about the user preference to the other user. In other words, it recommends the item to the user. In this review paper, three social factors have been discussed which can be used in recommender systems. For product feature, sentiment dictionary is needed for storing the information. So in our future work we are going to make a rating prediction recommender system by using three social factors i.e. user sentiment similarity, user circle influence, and item reputation. We will modify the sentiment dictionaries by applying the fine-grained sentiment analysis, so based on these it will predict the users' rating.

8. REFERENCES

- [1] R. Salakhutdinov, and A. Mnih, "Probabilistic matrix factorization," in *NIPS*, 2008.
- [2] 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.
- [3] 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
- [4] M. Jamali and M. Ester, "A matrix factorization technique with trust propagation for recommendation in social networks," in *Proc. ACM conf. RecSys*, Barcelona, Spain. 2010, pp. 135-142.
- [5] X. Qian, H. Feng, G. Zhao, and T. Mei, "Personalized recommendation combining user interest and social circle," *IEEE Trans. Knowledge and data engineering*. 2014, pp. 1763-1777.
- [6] H. Feng, and X. Qian, "Recommendation via user's personality and social contextual," in *Proc. 22nd ACM*

- international conference on information & knowledge management*. 2013, pp. 1521-1524.
- [7] D.M. Blei, A.Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of machine learning research* 3. 2003, pp. 993-1022.
- [8] 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.
- [9] K.H. L. Tso-Sutter, L. B. Marinho, L. Schmidt-Thieme, "Tag-aware recommender systems by fusion of collaborative filtering algorithms," in *Proceedings of the 2008 ACM symposium on Applied computing*, 2008, pp. 1995-1999.
- [10] 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.
- [11] F. Li, N. Liu, H. Jin, K. Zhao, Q. Yang, X. Zhu, "Incorporating reviewer and product information for review rating prediction," in *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence*, 2011, pp. 1820-1825.
- [12] B. Sarwar, G. Karypis, J. Konstan, and J. Reidl, "Item-based collaborative filtering recommendation algorithms," in *Proc.10th International Conference on World Wide Web*, 2001, pp. 285-295.
- [13] L. Qu, G. Ifrim, G. Weikum, "The bag-of-opinions method for review rating prediction from sparse text patterns," in *Proc. 23rd International Conference on Computational Linguistics*, 2010, pp. 913-921.
- [14] S. Jiang, X. Qian, J. Shen, Y. Fu, T. Mei, "Author topic model based collaborative filtering for personalized POI recommendation," *IEEE Transactions. Multimedia*, vol. 17, no. 6, June 2015, pp. 907-918.
- [15] V. Suresh, A. Veilumuthu, A. Krishnamurthy, "A Non-syntactic approach for text sentiment classification with stop words," *ACM WWW 2011*, March 28-April 1, Hyderabad, India. pp. 137-138.
- [16] K. Zhang, Y. Cheng, W. Liao, A. Choudhary, "Mining millions of reviews: a technique to rank products based on importance of reviews," in *Proceedings of the 13th International Conference on Electronic Commerce*, Aug. 2011, pp. 1-8.
- [17] 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
- [18] L. Polanyi, A. Zaenen, "Contextual Valence Shifters," *Information Retrieval* 20(2006): pp. 1-10.
- [19] M. Taboada, J. Brooke, M. Tofiloski, K.Vol, M.Stede, "Lexicon-based methods for sentiment analysis," *Computational Linguistics*, 2011, 37(2). pp. 267-307.
- [20] B. Pang, Bo, L. Lee, and S. Vaithyanathan, "Thumbs up? sentiment classification using machine learning techniques," in *Proc. EMNLP*, 2002, pp. 79-86.
- [21] 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.
- [22] T. Nakagawa, K. Inui, and S. Kurohashi, "Dependency tree-based sentiment classification using CRFs with Hidden Variables," *NAACL*, 2010, pp.786-794.
- [23] G. Ling, Michael R. Lyu, and Irwin King, "Ratings meet reviews, a combined approach to recommend," in *RecSys'14. ACM*, New York, NY, USA, 105-11.
- [24] H. Ma, H. Yang, M. R. Lyu, and I. King, "SoRec: Social recommendation using probabilistic matrix factorization," in *Proc.17th ACM CIKM*, Napa Vally, CA, USA, 2008, pp.931-940
- [25] K. Lee, and K. Lee, "Using dynamically promoted experts for music recommendation," *IEEE Trans. Multimedia*, vol. 16, no. 5, 2014.
- [26] 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.
- [27] X. Lei, and X. Qian, "Rating prediction via exploring service reputation," *2015 IEEE 17th International Workshop on Multimedia Signal Processing (MMSP)*, Oct 19-21, 2015, Xiamen, China. pp.1-6.
- [28] S. Gao, Z. Yu, L. Shi, X. Yan, H. Song, "Review expert collaborative recommendation algorithm based on topic relationship," *IEEE/CAA Journal of Automatica Sinica*, 2015, 2(4): pp. 403-411.