

# SENTIMENT BASED RATING PREDICTION THROUGH TEXTUAL REVIEWS

<sup>1</sup>B.K.Shireesha, <sup>2</sup>S.Zahoor-Ul-Huq

<sup>1</sup>PG Student, Computer Science and Engineering Dept., GPREC Kurnool(District), Andhra Pradesh-518004, INDIA,

<sup>2</sup>Professor of Computer Science and Engineering Dept., GPREC Kurnool(District), Andhra Pradesh-518004, INDIA,

## Abstract

*Now a days, the Social Media has become very popular to share the users viewpoints to their friends by using various social networking platforms. It makes obligatory for the users to post their reviews for other users to know about the quality of the products. In this paper, information overloading problem are discussed. So, a Sentiment-based rating prediction method is proposed to improve the prediction accuracy in the traditional recommender systems. User trusted friend, Item reputation and User Sentiment similarity factors are introduced. In this, the three factors are fused into the recommender systems to make accurate rating prediction. The performance evaluation of three sentimental factors on the user datasets, product datasets are considered. As the result, it helps to improve the recommendation performance.*

**Keywords:** Sentiment analysis, User sentiment reviews, Recommender systems, Item reputation, Rating Prediction.

## 1 INTRODUCTION

In day to day life, customers mostly like to purchase the online products which have good reviews. Reviews contain enough detailed product information and user opinion. So, users buy the products based upon high rated reviews only. Item reputation is one of the important factors which reflects customer's comprehensive evaluation based on the intrinsic value of the specific product.

Sentiment analysis is the most fundamental work in extracting the user's interest preferences. By using sentiment analysis, each user's attitude can be acknowledged on the product items. Sentiment reviews are needed to obtain the reputation of the product based on the user's count. So, each customer have some positive and negative reviews and these will be taken as reference.

The advantages of the product can be known from the user positive reviews and disadvantages can be known from the negative reviews. The user's sentiment is difficult to predict interpersonal sentiment influence which makes difficult on exploring social users[1].

To report these problems, a sentiment-based rating prediction method is proposed. By this method the social user's sentiment will be used to infer ratings. This method approximates each user rating for target items to understand accurate recommendation in e-commerce. By using the previous ratings of the other items, the target user preferences can be estimated. There are some features that defines the user sentiments.

Initially, the product features are referred from user reviews. By examining these product features, the sentiment words can be filtered [2]. The sentiment dictionaries are used mainly to calculate the sentiment of a specific user on that item or product. For collecting trusted reviews, user friend's circle by using some sentiment words which are extracted from user reviews for recommending products to the recommender system are combined.

In this paper, mainly skilled information is used instead of other structured social factors. There is need to focus on the classifying users into binary sentiment (positive or negative) reviews and also further in mining user's sentiment. This method mainly focuses on user's sentiment and interpersonal sentiment influence and also item's reputation factors. Finally, all these terms will be taken into the recommender systems.

The rest of this paper is divided into VII sections. The problem statement and proposed systems are described in Section II, the recommender systems and filter techniques & their working process are detailed in Section-III, modules and factors are detailed in Section-IV after which implementation to estimate the performance of the rating prediction method based on users sentiment words in Section-V, some series of experiments are conducted to compare the rating prediction model based on the user's sentiment with few existing models in Section VI. Finally, conclusion and future work are introduced in Section VII.

## 2 PROBLEM STATEMENT

### 2.1 Existing System

Sentiment analysis can be conducted based on three levels. Review level analysis and Sentence level analysis is used to classify the sentiment of the whole review into one of the predefined sentiment Polarities including positive, negative and neutral reviews. Phrase level analysis is used to extract the sentiment polarity of each feature which is expressed by specific user.

### 2.2 Proposed System

To overcome phrase level analysis and also to mine the data which is related to user's products, the sentiment-based rating prediction method is introduced. It works on the framework of matrix factorization technique. A user sentimental measurement approach is proposed based on

mined sentiment words and similarly sentimental degree words from user reviews.

### 3 RELATED WORK

Initially Haraldsteck and Yong Liu[5] had proposed a novel approach to improve the recommendation accuracy by familiarizing the concept of inferred circles of friends in online social networks. This approach is mainly used to develop the circle-based recommender systems efficiently. Mohsen Jamali and Martin Ester[6] had explored a novel model based approach for recommendation in social networks which is based on matrix factorization model. By using this technique, the different ranges on the user ratings can be known.

In this section, the recommender systems and filter techniques & their working process are mentioned. E-commerce systems and Information system is important part in the recommender systems. In this paper, two different filtering techniques are proposed.

User-based Collaborative filtering[3] is used to solve the problem of scalability by isolating the user groups and calculating the recommended progression for each user individually.

Item-based Collaborative filtering[4] use many recommendations per second based on the users and items. This technique will produce the high quality recommendation.

Matrix factorization technique is the most prevalent approach for the low-dimensional matrix decomposition and it is a Probabilistic matrix factorization[PMF] technique. First, for the review of the Basic MF, with the use of Potential eigen vectors matrix for both users and items, the value of all the ratings are calculated.

The cold start problems are solved by the Social recommendations. But to reconnoiter the matrix factorization in social recommender systems it requires trust relations to achieve the trust circles in social networks. Even though if there is no relation exist between social users, it can be maintained by the sentiment analysis.

### 4 IMPLEMENTATION

**Product Features Using LDA:**[Latent Dirichletian Allocation]. By using LDA, the product features are extracted from the textual reviews. It contains product characteristics and named entities.

LDA trails generative process for each document.

1. Choosing  $N \sim \text{Poisson}(\xi)$ .
2. Choosing  $\theta \sim \text{Dir}(\alpha)$ .
3. For each of the  $N$  words  $w(n)$ .

(a) Choose a topic  $z(n) \sim \text{Multinomial}(\theta)$ .

(b) Choose a word  $w(n)$  from  $p(w(n) | z(n), \beta)$ , a multinomial probability conditioned on the topic  $z(n)$ .

**Data Pre-processing:**By initializing filtering process, some words are collected by extracting each user reviews. Positive words, Negative words and Sentiment degree of words are collected which contemplates prepositions, articles and pronouns etc. All the different words are involved in vocabulary.

**Generating The Process:** It considers all the users document as  $D$  and the number of topics represented as  $m$ . The output will be represented as topic preference distribution of each user and topic list consists of 10 features words.

The below are the step that have been maintained.

1. For each document  $d_j$ , chooses a random variable  $\theta_m \sim \text{Dirichletian}(a)$ .
2. Foreach and every topic  $z_k$ , where  $[1, T]$ , then choose  $\phi_k \sim \text{Dirichletian}(b)$ . For each topic  $z_k$  inference scheme upon the observation.

### THREE SENTIMENTAL FACTORS

In this paper, three factors are fused by applying the matrix factorization technique to implement sentiment-based rating prediction method.

**User's Trusted Friend:** In this factor, mining the data related to the different users ratings will be done. Users have different user id's. By using this method, first it extracts the interpersonal sentiment data and top-most interpersonal sentiment data from the user datasets. It means collecting the each user's friends interested items along with their users items and ratings will be done. In this user-based collaborative filtering technique will be implemented.

**Item Reputation:** In this factor, reputation and top-most reputation items are displayed by applying the item-based collaborative filtering technique. By this, top-most reputation items can be known easily.

**User Sentiment Similarity:** In this factor, five leverage levels in which different user reviews are displayed in the below table. The quality of the product can be represented by using sentiment words from the user reviews. Any item id can be considered. By that item id, they display how many users have purchased that product and that product features Like level 1 words represents the product is excellent. If level 2 words represents product is good. In this way, that exact product features can be known. Levels are unique based on their words.

Table 1

Levels	Representative words
Level 1	Best, Entirely, Superb, 100%, Highest, Sharply
Level 2	Very, Better, Lot, Greatly, Over
Level 3	Far, So, Even, More, Rather, Relatively
Level 4	Somewhat, Some, A Little, A Bit, Slight
Level 5	Less, Insufficiently, Passably, Less, Very Little Bit

### 5 EXPERIMENTAL RESULTS

To organize experiments, they have subsistent social relationships and reviews. The new factors like user's trusted friend, item reputation and user sentiment similarity of recommender systems are used to solve the cold start problems and sparsity problems.



Fig 1: Three factors

**Sentiment Evaluation:** The outline of phrase-level sentiment lexicon constructions is inheritably difficult. By using the steps of sentiment lexicon for RPS, the top-10 product features are used to avoid the negative effects. The five star Yelp reviews are represented as positive reviews and one-star Yelp reviews are represented as negative. Last three-star Yelp reviews as neutral ones.

#### Rating Prediction:

**Evaluation Metrics:** In this section, 80% of data from each dataset from the Yelp websites is taken as the training set and the remaining 20% as the test set.

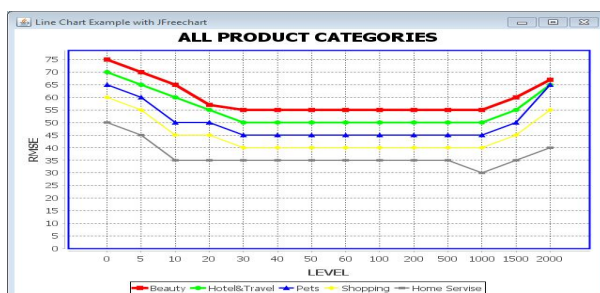


Fig 2: All product categories comparison

Root Mean Square Error(RMSE)[10] and Mean Absolute Error(MAE)[11] techniques are used in our experiments and the results of different product categories are represented in the graphical form.

### 6 ALGORITHMS

Experiments are conducted to compare the rating prediction model based on the user's sentiment with the following existing models.

**CircleCon:** This method [5] mainly focuses on the factors of interpersonal trust in social networks and infer the trust circles on the matrix factorization. It improves the accuracy of Basic MF.

**Basic MF:** This is the baseline matrix factorization approach without deliberation of social networks[7].

**Context MF:** This method[8] takes both interpersonal influence and item preferences into consideration and improves the accuracy of traditional item-based collaborative filtering[4].

**EMF:** This method labels two different characteristics matrixes: user-feature attention matrix and item-feature quality matrix in which each element of the quality-feature measures the quality of an item for the consistent product feature. Each element in the user-feature attention measures an extent feature that user cares about the consistent product feature.

**Sentiment-Based Rating Prediction:** Compared with other models like EFM[9], three sentiment dictionaries are considered and two linguistic rules are added to calculate user's sentiment. It combines user sentiment preferences and social networks.

### 7 CONCLUSION AND FUTURE WORK

In this paper, a recommendation model is proposed by mining sentiment information from the users reviews. In this, all the three factors are associated. i.e., user sentiment similarity, interpersonal sentiment influence and items reputation factors are fused into matrix factorization to achieve the rating prediction task. By using interpersonal sentimental influence, new relationship is maintained between the user and their friends. It shows enhancement in the existing approach on the real-world dataset. For the future enhancement, the sentiment dictionaries can be enriched to apply the fine-grained sentiment analysis.

### References

- [1] Z. Fu, K. Ren, J. Shu, et al. "Enabling Personalized Search over Encrypted Outsourced Data with Efficiency Improvement," IEEE Transactions on Parallel and Distributed Systems, 2015.
- [2] T. Kawashima, T. Ogawa, M. Haseyama, "A rating prediction method for e-commerce applications using ordinal regression based on LDA with multi-model features" IEEE 2nd Global Conference on Consumer Electronics (GCCE). 2013, pp 260-261.
- [3] G. Zhao, X. Qian, X. Xie, "User-service rating prediction by exploring social user's ratings behavior,"

(frontl.in) IEEE Transactions on Multimedia, 2016, 18(3):496-506.

- [4] B. Sarwar, G. Karypis, J. Konstan, and J. Reidl, "Item-based collaborative filtering recommendation algorithms" in Proc.10th International Conference on World Wide Web, ([www.ijedr.org](http://www.ijedr.org)) 2001, pp. 285-295.
- [5] X. Yang, H. Steck, and Y. Liu, "Circle-based recommendation in online social ([www.meng-jiang.com](http://www.meng-jiang.com)) networks, " in Proc. 18th ACM SIGKDD Int. Conf. ([www.ijedr.org](http://www.ijedr.org)) KDD, New York, NY, USA, ([idb.snu.ac.kr](http://idb.snu.ac.kr)) Aug. 2012, pp. 1267-1275.
- [6] M. Jamali and M. Ester ([www.meng-jiang.com](http://www.meng-jiang.com)) : A matrix factorization technique with trust propagation for recommendation in social networks. In proceedings of the (frontl.in) fourth ACM conference on Recommender systems, 2010. ([www2013.org](http://www2013.org))
- [7] R. Salakhutdinov, and A. Mnih, "Probabilistic matrix factorization," in NIPS, 2008.
- [8] M. Jiang, P. Cui, R. Liu, Q. Yang, F. Wang, W. Zhu, and S. Yang, "Social contextual ([reconcile.pjwstk.edu.pl](http://reconcile.pjwstk.edu.pl)) recommendation," in Proc, 21st ACM Int. CIKM, 2012, pp, 45-54.
- [9] X. Qian, H. Feng, G. Zhao, T. Mei, "Personalized recommendation combining user interest and social circle, " IEEE Trans. ([www.meng-jiang.com](http://www.meng-jiang.com)) Knowledge and data engineering, 2014, pp. 1763-1777.
- [10] J. Huang, X. Cheng, J. Guo, H. Shen, and K. Yang, "Social recommendation with interpersonal influence," in Proc. ECAI, 2010, pp, 601-606.
- [11] Y. Lu, M. Castellanos, U. Dayal, C. Zhai, "Automatic construction of a context-aware sentiment lexicon: an optimization approach," ([yongfeng.me](http://yongfeng.me)) World Wide Web ([www.meng-jiang.com](http://www.meng-jiang.com)) Conference series, 2011, pp. 347-356.

## Author



**Shireesha** was born in Andhra Pradesh, India. She received the B.Tech degree in Computer Science and Engineering from Jawaharlal Nehru Technological University, Anantapur Branch, India in 2015 and M.Tech Degree also in the same branch and University. Her research interests area is Data mining.



**S. Zahoor-Ul-Huq** M.Tech [Ph.D] was born in Andhra Pradesh. He is working as Professor in Computer Science and Engineering Dept. GPREC Kurnool(district), AP 518007 India. His research interest area is in Data mining.