



DETECTION OF FRAUDULENT REVIEW AND RATING IN E-COMMERCE SERVICE

Ms.S.P. Vidhyapriya¹, Safia mahmoodha. S², Saranya.P³, Boopathi.A⁴,Sabarisan.G⁵
^{1,2,3,4,5} Department of IT, Kathir College of Engineering

Abstract: E-Commerce presents a greater opportunity for people to purchase products online at cheaper price. In this E-Commerce service there exists few problems in return of payment and breakage issues. It also provides a platform for users to share their consumption experience and quality of product. However, at the same time we face the information overloading problem and existence of fake contents. How to filter trusted information from these reviews and make an accurate recommendation and to solve issues in services is difficult for us. Here, we firstly propose analyze scheme which eliminates fake contents and it calculates each user's score based on real reviews. In addition to this, we consider rating scores from user based on the specifications of product and its performance. Secondly, a platform for users which allows users to purchase based on their geographical location. At last, we combine both factors into our recommender system to make an accurate rating prediction, which is based on probabilistic matrix factorization. We conduct a series of experiments that shows accurate results and needed platform for both customers and sellers.

Keywords: Fake contents, reviews, rating, ranking, probabilistic matrix factorization, summarization

I. INTRODUCTION

Data mining, *the extraction of hidden predictive information from large databases*, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. Data mining tools can answer business questions that traditionally were too time consuming to resolve. Data mining techniques can be implemented rapidly on existing software and hardware platforms to enhance the value of existing information resources, and can be integrated with new products and systems as they are brought on-line. Data mining techniques are the result of a long process of research and product development. Data mining is ready for application in the business community because it is supported by three technologies that are now sufficiently mature:

- Massive data collection
- Powerful multiprocessor computers
- Data mining algorithms

II. LITERATURE SURVEY

The related work indicates the previous sentiment analysis done with the social user's sentiment in buying products. Those reviews can be made useful by mining the data contents to provide useful ranking for the users. This helps to rank the products according to the favorite list of user. In present, the E-Commerce just provides a platform to the users to share their experience about the products. To make this useful, a trusted recommendation system is proposed to filter the review contents and a strong entry authentication is provided for the users. This recommendation system only permits certain authorized users to enter the review contents and they are permitted to enter the review only after the usage of Consumers of product provide various reviews at the E-Commerce platform. They provide their aspect value about the product

and suggestions and queries after using the product. To avoid the market research for the firms, we recommend filtering this useful information and making high authorization to restrict fake users. Fake users are getting chance to increase their product ranking, this may increase their productions. So high authorization is provided along with the platform which collects user comments and provides a trusted ranking system.

2.1 Ranking

This section is used for the user to rank the product after visiting the ranking is done with the help of email id that the person provided while booking the product. With this ranking is placed and next time the new user will get listed with products buy and helpful review the particular website that are mostly ranked with high value.

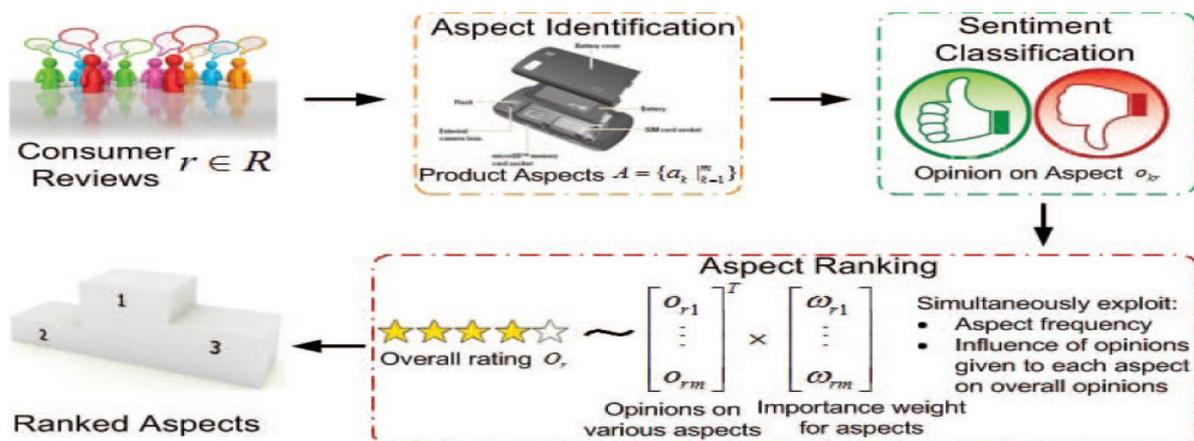
III. PROPOSED SYSTEM

In this, a recommendation model is proposed by mining review comments from social users' reviews and their ratings. We propose social user's sentiment measurement approaches which predict ranking of product based on the review words and rating score from users' reviews. Besides, we build algorithm to combine and evaluate those scores into a unified matrix factorization framework to predict users' ratings. In particular, as long as we extract user's reviews from his/her rating histories, we can measure user's sentiment. And also we demonstrate user's review can really denote what user's interest preferences. In our future work, we will explore social users' sentiment deeply by analyzing whether the comments of user are deserved well or not and also check right or wrong. The framework contains three main components, i.e., product review analysis, rating classification, and evaluating ranking. First, we exploited the *Pros* and *Cons* reviews to improve aspect identification and sentiment classification on free-text reviews. We then developed a probabilistic aspect ranking algorithm to infer the importance of various aspects of a product from numerous reviews. The algorithm simultaneously explores aspect frequency and the influence of consumer opinions given to each aspect over the overall opinions. The product are finally ranked according to their importance scores. We have conducted extensive experiments to systematically evaluate the proposed framework. The ranking given is based only by the present review and rating we propose to improve it further by providing changes in review and how valuable the product for that review. A deep analysis is further proposed for future work.

IV. ARCHITECTURAL DESIGN

4.1 Nearest neighbour method

A technique that classifies each record in a dataset based on a combination of the classes of the k record(s) most similar to it in a historical dataset (where $k \geq 1$). Sometimes called the k-nearest neighbour technique.



V. ALGORITHM

5.1 Summarization Process

Step1: An extractive summarization method consists of selecting important sentences, paragraphs etc. from the original document and concatenating them into shorter form.

Step 2: Output should be a meaningful summary

Step 3: Pre-Processing step

Step 4: Processing step

Step 5: Pre Processing is Structured Representation of the Original text.

It Usually includes:

a) Sentences Boundary identification. In English, Sentence Boundary Is identified with Presence Of dot at the end of sentence.

b) Stop-Word Elimination:

examples:

the, a, an, another, for, an, nor, but, or, yet, so, in, under, towards, before

Step 6: The Purpose of Stemming is to Obtain the Stem each word.

Step 7: The weight is calculated as follows:

The weight, $w_t = \text{frequency of the term} / \text{Total no. of terms in the document}$

VI. ADVANTAGE

The proposed framework and its components are domain-independent. Fake reviews will not be permitted to enter and system will not permit a user to enter more than one review. The proposed system will predict the ranking of product based on the previous reviews and ratings. This also helpful in determining cost analysis and performance analysis. If item's reviews reflect positive sentiment, then the item may be with good reputation, if item's reviews are full of negative sentiment, then the item is most likely with bad reputation. The summarization algorithm is helpful with prediction of user's sentiment reviews and mining words will evaluate the value of entered comments.

VII. FUTURE WORK

In Future, we plan to consider implement a method of comparison to find the trustiness exist in review and rating entered by the product users based on the key specifications. Also a platform is required for user to change their review or to additional content for user's better convenience. A graphical approach on rankings will also be helpful for user to choose their product based on either cost, performance or something else.

VIII. CONCLUSION

E-Commerce acts as a stoppage point which protects and blocks un authorized persons from entering reviews and rating. The problem of existing is to detect the Fraudulent Review and Rating. It can be solved here by usage of probabilistic matrix factorization also, a review can be evaluated and estimated by Summarization Algorithm. Fake reviews can be blocked with the authorization process. Only permitted users who really used the product can only be able to enter the review and rating. These two working principles are evaluated with supervised approach which predicts the ranking of products that helps buyers to easily choose their products.

REFERENCES

1. Discovery of Ranking Fraud for mobile apps
2. L. Azzopardi, M. Girolami, and K. V. Risjbergen, "Investigating the relationship between language model perplexity and ir precision-recall measures," in Proc. 26th Int. Conf. Res. Develop. Inform.Retrieval, 2003, pp. 369–370.
3. D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," J. Mach. Learn. Res., pp. 993–1022, 2003.

4. Y. Ge, H. Xiong, C. Liu, and Z.-H. Zhou, "A taxi driving fraud detection system," in Proc. IEEE 11th Int. Conf. Data Mining, 2011, pp. 181–190.
5. D. F. Gleich and L.-h. Lim, "Rank aggregation via nuclear norm minimization," in Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2011, pp. 60–68.
6. T. L. Griffiths and M. Steyvers, "Finding scientific topics," Proc. Nat. Acad. Sci. USA, vol. 101, pp. 5228–5235, 2004.
7. G. Heinrich, Parameter estimation for text analysis, "Univ. Leipzig, Leipzig, Germany, Tech. Rep., <http://faculty.cs.byu.edu/~ringger/CS601R/papers/Heinrich-GibbsLDA.pdf>, 2008.
8. N. Jindal and B. Liu, "Opinion spam and analysis," in Proc. Int. Conf. Web Search Data Mining, 2008, pp. 219–230.
9. A. Klementiev, D. Roth, and K. Small, "An unsupervised learning algorithm for rank aggregation," in Proc. 18th Eur. Conf. Mach. Learn., 2007, pp. 616–623.