An Informational Search for Review through Data Analytics

Preetha A. S[1], Swathi K[1], Jeya Selvi C[1], Mr. G. Elangovan[2]
[1]Student, [2]Assistant Professor
Department of Computer Science and Engineering, Velammal Institute of Technology,
Panchetti, Thiruvallur, Chennai, India

ABSTRACT

Spreading of large unstructured information generated by social networks provide useful results to the effective recommender. However, the high volume of reviews that are typically published for a single product makes harder for individuals as well as manufacturers to locate the best reviews and understand the true underlying quality of a product. This paper suggests an approach to identify product opportunities from customer reviews in social media. Our approach explores to identify multiple reviews and get processed by sentimental analysis. These extracted data may contain positive and negative sentimental reviews. The recommended product can be visible to all the users who are present in network. The sentimental words get ordered based on priority. With the help of this, the user can get the best product by comparing it with other products. Sentimental analysis plays an important role to decide a product and both the users and manufacturers get benefit through this.

Keywords: *Product Reviews, Sentimental Analysis, Recommended Product, Social Media Mining*

I. INTRODUCTION

Data mining is the process of sorting through large unstructured datasets to identify patterns and establish relationships to solve problems through data analysis [1].There is much personal information in online textual reviews, which plays a very important role in decision processes [2][3].The customer will decide what to buy through valuable reviews posted by others, especially user's trusted friends[4]. Customers may believe reviews and reviewers will do help to the rating prediction based on the idea of high-star ratings [5]. 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 [1][2]. We focus on the rating prediction task. However, star-based rating prediction is not always available on many review websites. Conversely, reviews contain enough detailed product information and user opinion information [6], which has great reference value for a user's decision. Most important of all, a given user on a website is not possible to rate every item. Hence, there are many unrated items in a user-item-rating matrix. It is inevitable in many rating prediction approaches [7].Sentiment analysis is the most fundamental and important work in extracting user's interest preferences. In general, the sentiment is used to describe user's own attitude on items. Generally, reviews are divided into two groups, positive and negative. 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 is good but also need to know how good the product is. It's also agreed that different users may have different sentimental preferences. Some users prefer to use "good" to describe an "excellent" product, while others may prefer to use "good" to describe a "just so "product. Customers are most likely to buy those products with highly-praised reviews. That is, customers are more concerned about item's reputation, which reflects consumers' comprehensive evaluation based on the intrinsic value of a specific product.

II. RELATED WORKS

In 2011, Anindya Ghose and Panagiotis G. Ipeirotis proposed an approach explores multiple aspects of

review text, such as subjectivity levels, various measures of readability and extent of spelling errors to identify important text-based features to re-examine the impact of reviews on economic outcomes like product sales and see how different factors affect social outcomes such as their perceived usefulness.

In 2012 the researcher Xiaohui Yu, Yang Liu, Jimmy Xiangji Huang and Aijun An proposed a technique to capture the complex nature of sentiment by using sentiment PLSA(probabilistic latent semantic analysis).Based on SPLSA they propose ARSA, an Autoregressive Sentiment-Aware model for sales prediction. Further, they improve the accuracy of prediction by considering the quality factor.

In 2013 the researchers X. Qian, H. Feng, G. Zhao, and T. Mei proposed a paper based on three social factors, personal interest, interpersonal interest similarity, and interpersonal influence. Interpersonal influence is a type of social influence in which people try to give their best reviews for the product what the user recommended. From this, the recommended product can be reviewed.

In 2015 the authors Shuhui Jiang, XuemingQian, JialieShen, Yun Fu, and Tao Mei proposed to facilitate comprehensive points of interest (POIs) recommendations for social users. The proposed technique was mainly focused on user preference level.

III. PROPOSED SYSTEM
1. DESCRIPTION
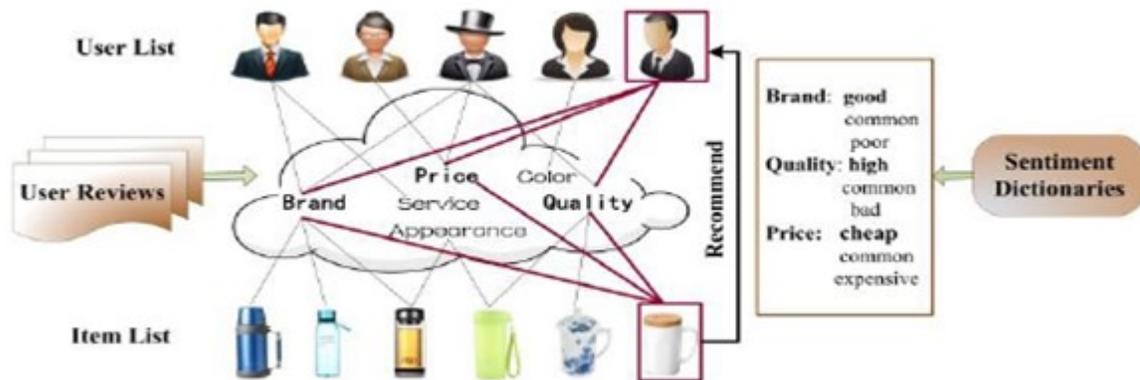
The main contributions of this approach are as follows: 1) A user sentimental measurement approach, which is based on the mined sentiment words and sentiment degree words from user reviews. Besides, some scalable applications are proposed. The user explores how the mined sentiment spread among users' friends. What is more, we leverage social users' sentiment to infer item's reputation, which showed great improvement in accuracy of rating prediction. 2) The approach makes 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) The approach fuses 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.

2. SYSTEM DESIGN
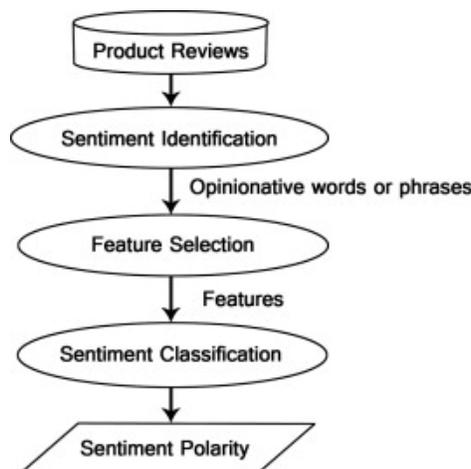
Product reviews are used on shopping sites to give customers an opportunity to rate and comment on products they have purchased, right on the product page. Other consumers can read these when making a purchase decision. A customer review is a review of a product or service made by a customer who has purchased and used or had experience with, the product or service. Customer reviews are a form of customer feedback on electronic commerce and online shopping sites.

Sentiment identification: The reviews may have both positive and negative words. Customer buys the product based on the review given by the user one who used that product. Mostly they focus on positive comments because it describes the sentimental feelings of the user. A review may have sentimental words like "good", "bad", "excellent", "poor", etc., Through this the customer can identify the reviews which are used to buy the product.

Sentiment classification: Sentiment analysis or opinion mining is a field of study that analyses people's sentiments, attitudes, or emotions towards certain entities. User classifies the words based on the sentimental words used. The proposed system used the sentimental words such as "good", "bad", "excellent", "poor", etc., based on the positive and negative reviews the words get classified.

SENTIMENT ANALYSIS

Discovering people opinions, emotions and feelings about a product or service

Sentiment polarity: A basic task in sentiment analysis is classifying the polarity of a given text at the document, sentence, or feature/aspect level—whether the expressed opinion in a document, a sentence or an entity feature/aspect is positive, negative, or neutral.

Sentimental analysis: Opinion mining (sometimes known as sentiment analysis or emotion AI) refers to the use of natural language processing, text analysis to systematically identify, extract, quantify, and study effective states and subjective information. Sentiment analysis is widely applied to the voice of the customer materials such as reviews and survey responses, online and social media, and healthcare materials for applications that range from marketing to customer service to clinical medicine.

3. ALGORITHM USED
In this paper, the following algorithms are applied:

A. Decision Trees
Decision tree builds classification or regression models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes.

B. Naive Bayesian Algorithm – For Product Pattern Learning
Naive Bayes is one of the simplest density estimation methods from which one can form standard classification methods in machine learning. Even if we are working on a data set with millions of records with some attributes, it is suggested to try Naive Bayes approach. Naive Bayes classifier gives great results when we use it for textual data analysis, such as Natural Language Processing. To understand the naive Bayes classifier we need to understand the Bayes theorem.

Bayes theorem named after Rev. Thomas Bayes worked on conditional probability. Conditional probability is the probability that something will happen, given that something else has already occurred. Using the conditional probability, we can calculate the probability of an event using its prior knowledge.

Below is the formula for calculating the conditional probability.

$$P(H \mid E) = \frac{P(E \mid H) * P(H)}{P(E)}$$

where
- P (H) is the probability of hypothesis H being true. This is known as the prior probability.
- P (E) is the probability of the evidence (regardless of the hypothesis).
- P (E|H) is the probability of the evidence given that hypothesis is true.
- P (H|E) is the probability of the hypothesis given that the evidence is there.

IV. RESULT

The experimental results and discussions show that user's social sentiment that is mined is a key factor in improving rating prediction performances.
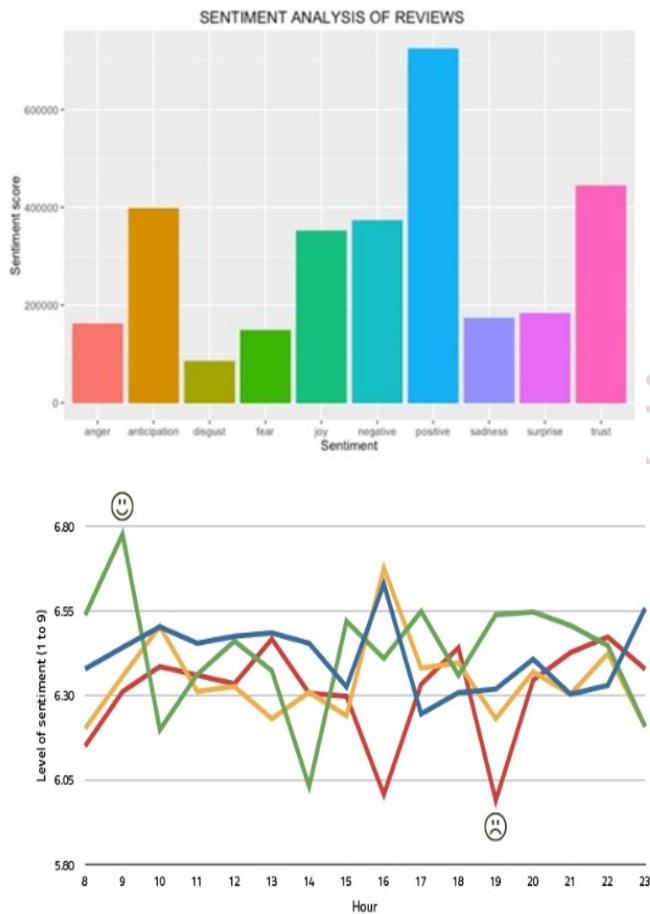
V. CONCLUSION

This paper tries to provide the best product by comparing all the features of other product through user sentimental words. So that the user can be benefited, and feel happy. The user can get the review for every product. Credible because the prediction is based on majority reviews shared by social media users. The biggest reason why online reviews are important to businesses is that ultimately it increases sales by giving consumers the information they need to make the decision to purchase a product or service from a business. In the future work, the reviews are to be implemented with enhanced retrieval rate and throughput.

VI. REFERENCES

1. L. Cao, Y. Zhao, H. Zhang, D. Luo, C. Zhang, and E.K. Park, "Flexible Frameworks for Actionable Knowledge Discovery," IEEE Trans. Knowledge and Data Eng., vol. 22, no. 9, pp. 12991312, Sept. 2009.

2. N. Archak, A. Ghose, and P.G. Ipeirotis, "Show Me the Money! Deriving the Pricing Power of Product Features by Mining Consumer Reviews," Proc. 13th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD), pp. 56-65, 2007.

3. J. A. Chevalier and D. Mayzlin, "The Effect of Word of Mouth on Sales: Online Book Reviews," J. Marketing Research, vol. 43, no. 3, pp. 345-354, Aug. 2006.

4. 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

5. G. Ganu, N. Elhadad, A Marian, "Beyond the stars: Improving rating predictions using Review text content," in 12th International Workshop on the Web and Databases (WebDB 2009). pp. 1-6.

6. Yuan Zuo, Junjie Wu, Hui Zhang, Deqing Wang, KeXu, "Complementary Aspect-based Opinion Mining,"in IEEE Transaction on Knowledge And Data Engineering , vol 30, oct 2016.

7. 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.