DETECTION OF FAKE REVIEW AND BRAND SPAM USING DATA MINING TECHNIQUES

[1]Mr.V.AJAY KUMAR, [2]S.SHAREEN, [3]M.KIRANMAI,
[1]Assistant Professor, [2,3]UG Student, Department of Computer Science and Engineering,
Godavari Institute of Engineering & Technology, Rajahmundry, AP.

Abstract: Now a day's technology changes for publicity, way to traditional marketing also changes as person to-person communication to online reviews. As feedback these online reviews are important so customer and to companies or vendors. These reviews are helpful for making decisions regarding quality of products and services. Companies and vendors use opinions for making a decision for marketing strategies, performance to services or product, for improvement. However, the intentions to all customers of users are not true for writing reviews. This concept, changes the face of advertising to conventional, individual-to-individual correspondence to online audits. These online audits are important to client and to organizations or sellers. In this paper we proposed the method to recognizing the untruthful reviews that are given by the users which is having distinct semantic content based on sentiment analysis as the reviews of product. In this paper author represent to detect the spam untruthful reviews of product. For this classification we used J48 classifier. We generate ARFF from the distinct features to detecting the untruthful reviews. Using Support Count in Association Rules we further detect Brands in Fake Reviews.

Key words: Marketing Strategies, Semantic content, Sentiment analysis, ARFF, J48 classifier, etc.

INTRODUCTION

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. We 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 We 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 product 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 item. Hence, there are many unrated items in a user-item-rating matrix. It is inevitable in many rating prediction approaches. Review/comment, as we all know, is always available. In such case, it's convenient and necessary to leverage user reviews to help predicting the unrated items. The rise like DouBan1, Yelp2 and other review websites provides a broad thought in mining user preferences and predicting user's ratings. Generally, user's interest is stable in short term, so user topics from reviews can be representative. For example, in the category of Cups & Mugs, different people have different tastes. Some people pay attention to the quality, some people focus on the price and others may evaluate comprehensively. Whatever, they all have their personalized topics. Most topic models introduce users' interests as topic distributions according to reviews contents. They are widely applied in sentiment analysis, travel recommendation, and social networks analysis. 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 items. We observe 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. 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 people may have different sentimental expression preferences. For example, some users prefer to use "good" to describe an "excellent" product, while others may prefer to use "good" to describe a "just so" product.

DEFINITION OF FAKE REVIEW

Reviews which are designed specifically to give a false impression to consumers on the point of purchasing. They don't accurately reflect the product or service they're talking about because they are designed to go after your wallet in the same way as "fake news" aims to dupe your thought processes. You'll see both positive and negative fake reviews.

"**Fake positives**" are the glowing reports that bear no resemblance to reality. The e-book example above. The hotel you book only to find its "spacious bedrooms" mean you don't have room to unpack your suitcase and its "beachside location" has an inter-city train line running between it and the beach. Positive fake reviews are written to compel you to take action i.e. to buy the product or service.

"**Fake negatives**" are those that persuade you *not* to eat at a certain restaurant because it's "known to have cockroaches in the kitchen" or not to go to a particular counsellor because "she ruined my life forever". Negative fake reviews are written to persuade you not to take action on the product being reviewed. They are written to damage one company's reputation and generally go on to recommend a different product from a competitor. Neither positive nor negative fake reviews are based on reality. At best, they're based on shared misunderstanding; at worst, outright lies.

PROPOSED MODEL

This system proposes the method to recognize the untruthful reviews that are given by the consumers which is having different semantic content based on sentiment analysis as the reviews of particular product. This system proposes a behavioural approach to identify review spammers those who are trying to manipulate the ratings on some products. Here we derive an aggregated behaviour method for rank reviewers based on the degree that they have demonstrated the spamming behaviours. So as to verify our proposed methods, which conducts user evaluation on an Amazon dataset which contains reviews of different company's products? It is found the proposed method generally outperform the baseline method-based votes. Also, we learnt a regression model from the consumer ground truth spammers. The feedback and viewpoints are used by web users and companies for the manufacturing of new products.

The benefits of the proposed model are:

This paper mainly focuses on review centric spam identification which provides greater focus on feedback content. As part of future work, we can enhance review spammer identification into the review identification and vice versa. Exploring different ways to learn behaviour patterns which are related to the spamming so as to improve the accuracy of the current regression model is an interesting research direction in current era.

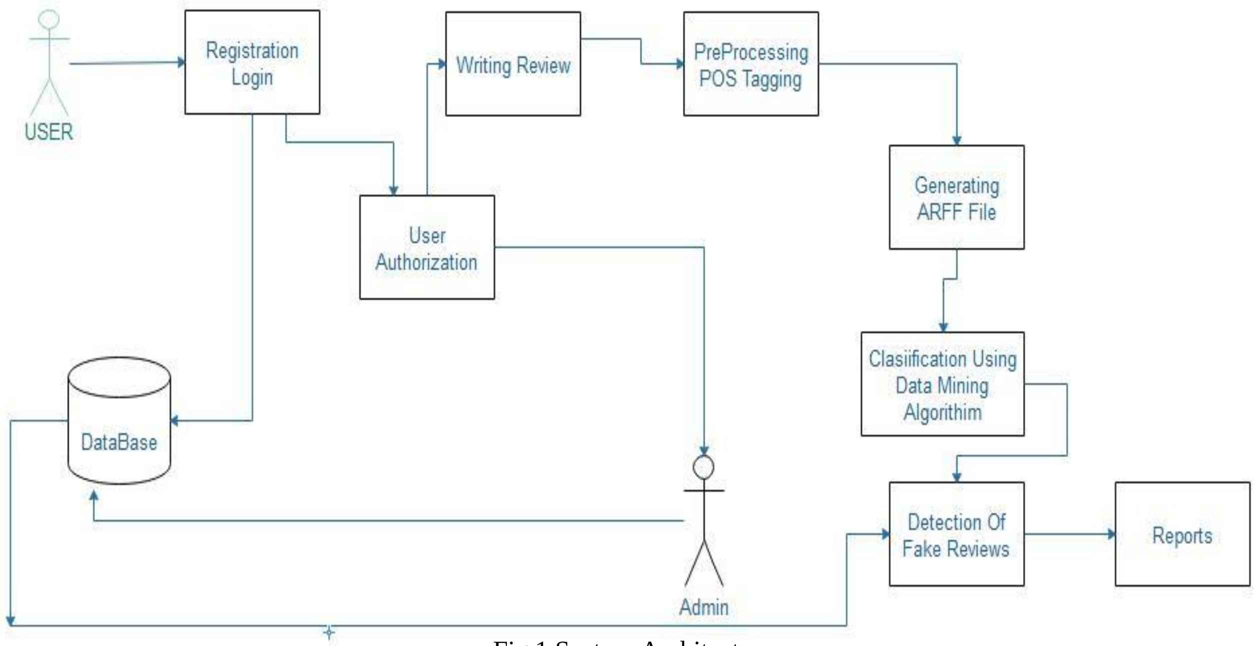

Fig 1 System Architecture

The above fig 1 gives an overview of System Architecture of our proposed system.

RESULTS

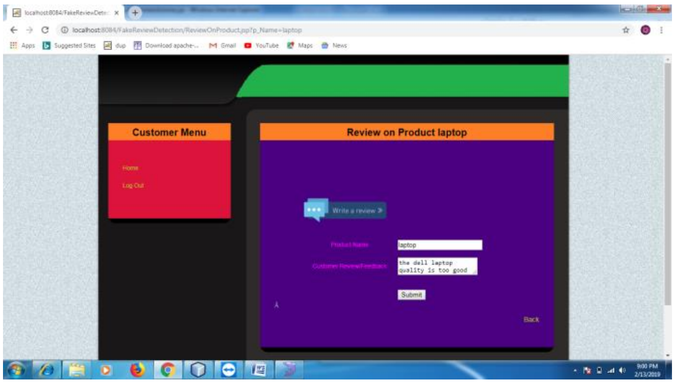

Fig 2-Writing review

Fig 3-Results using sentiment analysis

CONCLUSIONS AND FUTURE SCOPE

In this paper, a recommendation model is proposed by mining sentiment information from social 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. In particular, we use social users' sentiment to denote user preferences. Besides, we build a new relationship named interpersonal sentiment influence between the user and friends, which reflects how users' friends influence users in a sentimental angle. What is more, as long as we obtain user's textual reviews, we can quantitively measure user's sentiment, and we leverage items' sentiment distribution among users to infer item's reputation. The experiment results demonstrate that the three sentimental factors make great contributions to the rating prediction. Also, it shows significant improvements over existing approaches on a real-world dataset. In our future work, we can consider more linguistic rules when analysing the context, and we can enrich the sentiment dictionaries to apply fine-grained sentiment analysis. Besides, we can adapt or develop other hybrid factorization models such as tensor factorization or deep learning technique to integrate phrase-level sentiment analysis.

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.

ACKNOWLEDGEMENT

We, authors of this paper are very much grateful to management of GIET Institutions, Principal, Head of CSE department. Also we are thankful to all teaching and non-teaching staff.