REVIEW AUTOMATION AND BOGUS REPORT ANALYSIS

A.R.Avinash[1] | G.Gautham[2]

[1](Dept of CSE, UG Scholar, Anand Institute of Higher Technology, Chennai, India, Avinashrangarajan96@gmail.com)

[2](Dept of CSE, UG Scholar, Anand Institute of Higher Technology, Chennai, India, Gauthamking1997@gmail.com)

Abstract— *A big part of people rely on available content in social media in their decision. The possibility that anybody can leave a review provides a golden opportunity for spammers to write spam reviews about products and services for different interests. Online reviews have the potential to provide an insight to the buyers about the product like its quality, performance and recommendations. Identifying these spammers and the spam content is a hot topic of research. The methodologies put forth still barely detect spam reviews, and none of them show the importance of each extracted feature type. Both positive and negative reviews play a big role in determining the customer requirements and extracting consumer's feedback about the product faster. Website has opened up the avenues to smarter and informed decision making for large industries as well as the consumers. Online reviews on e-commerce giants like Amazon, Flipkart are one such paradigm which can be used to arrive at more profitable decisions. They are not only beneficial for the consumers but also for the product manufacturers. Online reviews have the potential to provide an insight to the buyers about the product like its quality, performance and recommendations; thereby providing a clear picture of the product to the future buyers. The usefulness of online reviews for manufacturers to realize customer requirements by analyzing helpful reviews is one such unrealized potential. Both positive and negative reviews play a big role in determining the customer requirements and extracting consumer's feedback about the product faster. Sentiment Analysis is a computational study to extract subjective information from the text. In this research, data analysis of a large set of online reviews for mobile phones is conducted. We have not only classified the text into positive and negative sentiment but have also included sentiments of anger, anticipation, disgust, fear, joy, sadness, surprise and trust. This delineated classification of reviews is helpful to evaluate the product holistically, enabling better-decision making for consumers*

Keywords— Product; Review; Positive; Negative; Sentiment Analysis

1. INTRODUCTION

The current widespread use of social media as well as web and mobile technologies, customer reviews are critical for Consumers decision making. As a recent survey has shown, 90% of consumers trust the recommendations of other consumers and 70% of customers read reviews or at least looked at ratings before deciding to buy certain products [1]. As a consequence of intense online market competition, the vital role of customer reviews has prompted a market for fraudulent reviews about real products and services [2, 3], whose production has cultivated an industry sustained by companies that write fake reviews [4]. Part of the industry's success stems from the fact that consumers – indeed, all humans – struggle to distinguish false reviews from true ones. Although some aspects tend to expose reviews as fake – e.g. the use of capital letters – since those aspects are not foolproof indicators, computer scientists have worked to construct automated approaches to detect fraudulent customer reviews on social media. In computer science, fake review detection is grouped with natural language processing, research on which not only applies language processing techniques, but also investigates machine learning (ML) methods, which can be used to build high performance predictors for fake review detection. Among the many classification, clustering, and regression problems often encountered in daily life that ML methods can resolve, fake review detection involves the dilemmas of classifying two types of review: ones that are fake and ones that are not. In our work, we aimed to develop a high-performance model using multiple classifier systems (MCSs) to detect fake negative reviews. Our research question is formulated as follows: RQ: Can MCS serve as an efficient learning approach to detect fake negative reviews and if so, to what extent? After investigating MCS, all of which can dramatically improve the performance of standalone machine learners, we built a model combining high- performance ML algorithms that we examined in detail and evaluated in combination. In our tests, to compare the performance of our model with others in the literature, we used Ott et al.'s dataset of customers' negative online reviews [5]. For their dataset, Ott et al. [5] used a support vector machine (SVM) ML technique with 5-fold cross-validation to attain 86% accuracy. However, since they applied only one classification algorithm for their model, we focused on MCS, which can contain multiple classifiers and use the majority voting combination rule to detect fake reviews. Accordingly, we tested many ML methods to assess their performance with the available dataset and ultimately identified five classification algorithms that could yield results better than those of other classification algorithms. At the same time, we sought to improve the performance of the algorithms identified by adjusting their parameters until reaching optimal values and, in turn, constructed many MCS using different combination rules and options. We ultimately identified two MCSs that can provide better results compared to the other models that we built. Whereas the first model uses three classifiers, the second applies five in conjunction with the majority voting combination rule According to our results, the second model, with the classifiers libSVM, libLinear, sequential minimal optimization (SMO), random forest (RF), and J48, performed better than the model described by Ott et al. We performed our tests with the Waikato Environment for Knowledge Analysis (Weka) open source ML tool and identified that using the classification algorithms libSVM, libLinear – both implementations of the SVMs algorithm –SMO, RF, and J48 could best detect fake reviews. Conclusions and makes recommendations for future research.

2. RELATED WORKS

[1] A simple yet efficient model, called dual sentiment analysis (DSA), to address the polarity shift problem in sentiment classification. By using the property that sentiment classification has two opposite class labels (i.e., positive and negative), we first propose a data expansion technique by creating sentiment reversed reviews. The original and reversed reviews are constructed in a one-to-one correspondence

[2] We review the development of Sentiment Analysis and Opinion Mining during the last years, and also discuss the evolution of a relatively new research direction, namely, Contradiction Analysis. We give an overview of the proposed methods and recent advances in these areas, and we try to layout the future research directions in the field.

[3] We can train linear support vector machines that achieve high classification accuracy on data that present classification challenges even for a human annotator. We also show that, surprisingly, the addition of deep linguistic analysis features to a set of surface level word n-gram features contributes consistently to classification accuracy in this domain. Automatic sentiment classification addresses the second question. Text mining tools can help make large quantities of feedback more manageable by splitting Authors and Affiliations them into clusters based on keywords or topics. Sentiment analysis, which is the focus of this paper, adds a second dimension to the analysis.

[4] An important issue that has been neglected so far is opinion spam or trustworthiness of online opinions. Reviews based on duplicate and voting system for spam identification does not work well so we try to introduce something new based on delivery of product approach.

[5] The datasets acquired through crawling the iOS App Store, compare a baseline Decision Tree model with a novel Latent Class graphical model for classification of app spam, and analyze preliminary results for clustering reviews.

3. PROPOSED SYSTEM

Online reviews have the potential to provide an insight to the buyers about the product like its quality, performance and recommendations. The opportunity that anybody can leave a review provide a golden possibility for spammers to write spam reviews about products and services for different interests, which enables other buyers to glance through the reviews written by fake reviewers since most of the online buyers go through the ratings and reviews as their guide. Now most of the online websites have minimized the use of bad words in the comment/reviews using Stop word Elimination system, but there is chance for mentioning other product, which makes reviewers to write about different brand details in some other product. This will ultimately divert the buyers/customers to look towards the mentioned product. Some fake reviewers put reviews without buying a product the only thing needed for them is the login ID for the specific website and some of them are working for other companies to degrade the product buying hit. Since authentication process for each customer/buyer is at backend and not visible to customers, which makes the customer to know whether the reviewer is an authenticated person or not and there are no facility to make a communication between the buyers/customers/reviewers. We are going to do three things in the proposed system:

First the chat option between the customer and the reviewer which is for knowing more about the product, this is for the customer's satisfaction.

Second the reviewer in the particular site should first get the particular product before reviewing about the product and the product delivery id will be displayed near the review so that the customer can understand.(i.e.)The review updated by the users order id has to be displayed.

Third the review must be based on the particular product manufacturers alone the other manufacturers review will be deleted if reviewed. Stop words can be used to identify the similar word or repeated words present in the review list. Eg: Lenovo laptop review should not contain review about other laptop brands like HP, DELL etc.

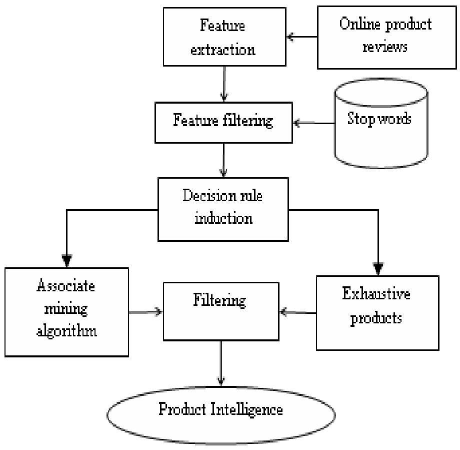

Figure 1.Architecture Diagram

4. WORKING PRINCIPLE

1. Order id can be used to identify the correct user details that they are displayed in the review. The customer initially logged in to particular website to preview products which is mandatory. Order id is displayed only for the purchased customer the rest of the reviews are not displayed.
2. Maintaining order id: if the customer buys the product, the product id was maintained and displayed in the review page. if they cancel the order , the order id for particular customer was removed and review was blocked
3. If any user tries to post review about another product, then the specified review will be removed and the review will not be displayed in the review forum.
4. Chat feature: the communication facility between the reviewer and customer so that the customer can know more about the product.

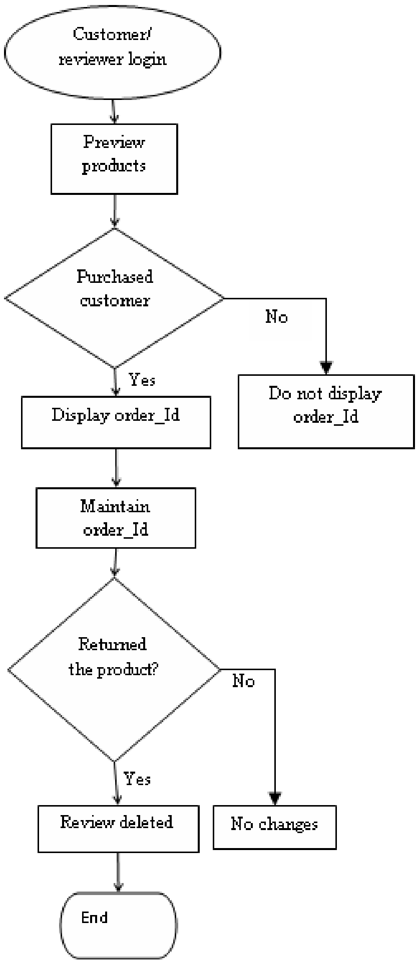

Figure 2.Order id and maintaining order

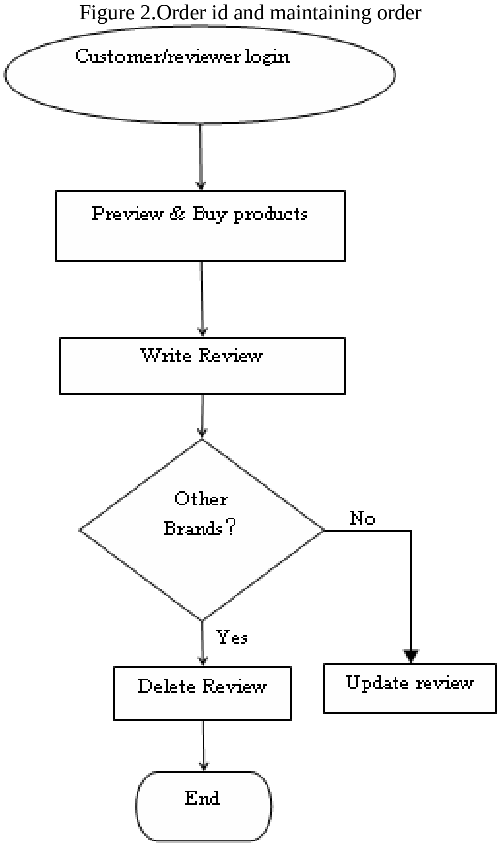

Figure 3.verifying review

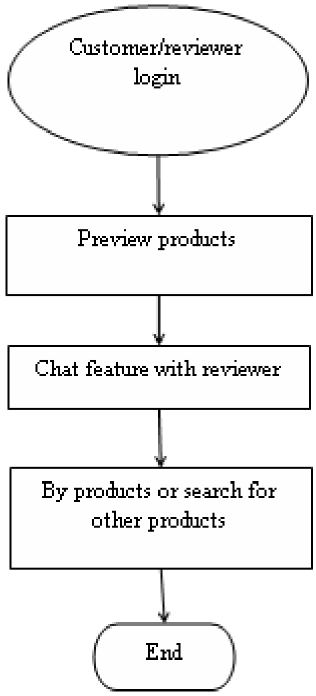

Figure 4.Chart Option

5. CONCLUSION

The aim of this study was to investigate the influence of review valence and concreteness in online consumer reviews on trust in the online review, product attitude and purchase intention of the consumer. Regarding trust in online reviews, this study shows that consumer trust towards an online review is lower for positive online reviews than for negative online reviews. The findings of this study however, contrary to these expectations. This study disproves the expectations that reading a concrete online review leads to more trust in the online review and that negative concrete online reviews lead to more perceived trust by consumers compared to negative abstract online reviews. One reason for this might be, that the reviews were or very positive or very negative, which could drown out the level of concreteness.

REFERENCES

[1] Jindal, N., Liu, B.: 'Review spam detection'. 16th Int. Conf. on World Wide Web, New York, NY, USA, 2007, pp. 1189–1190 J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.

[2] Broder, A.: 'On the resemblance and containment of documents'. Compression and Complexity of Sequences 1997: Proc., Salerno, Italy, 11–13 June 1997

[3] Jindal, N., Liu, B.: 'Opinion spam and analysis'. 2008 Int. Conf. on Web Search and Data Mining, New York, NY, USA, 2008, pp. 219–230

[4] Ott, M.: 'Deceptive opinion spam corpus v1.4', Available at: http:// myleott.com/op_spam, accessed: 03 August 2015.

[5] Ott, M., Choi, Y., Cardie, C., et al.: 'Finding deceptive opinion spam by any stretch of the imagination'. 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 2011.

[6] Ott, M., Cardie, C., Hancock, J.: 'Estimating the prevalence of deception in online review communities'. 21st Int. Conf. on World Wide Web, New York, NY, USA, 2012, pp. 201–210

[7] Ott, M., Cardie, C., Hancock, J.: 'Negative deceptive opinion spam'. Proc. Of the 2013 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Atlanta, GA, USA, 2013

[8] Lau, R.Y.K., Liao, S.Y., Kwok, R.C.W., et al.: 'Text mining and probabilistic language modeling for online review spam detection', ACM Trans. Manage. Inf. Syst., 2012, 2, (4), pp. 1–30, (Article no 25)

[9] Sharma, K., Lin, K.-I.: 'Review spam detector with rating consistency check'. 51st ACM Southeast Conf., New York, NY, 2013

[10] Morales, A., Sun, H., Yan, X.: 'Synthetic review spamming and defense'. 22nd Int. Conf. on World Wide Web. Int. World Wide Web Conf.s Steering Committee, Geneva, Switzerland, 2013, pp. 155–156