Sentiment Analysis on Product Reviews Using Hadoop

Chaudhari Devendra[1], Patharkar Shubham[2], Sayyed Sharif[3]
[1, 2, 3] Dept of Computer Engineering
[1, 2, 3] Suman Ramesh Tulsiani Technical Campus: Faculty of Engineering

Abstract- *Now a day's use of ecommerce is increased. Peoples buy products and post their opinions, suggestions about topic or product. Also before buying a product they may need to go through those reviews before making any decision. This will help them finalize their choice and take a decision whether or not to purchase the product. If the number of reviews is more than going through all reviews will become time consuming process for users. They will not able to interpret all product reviews and might get confused. We are proposing a method based on sentiment analysis than can be used to interpret reviews and summarize of reviews in the user suitable format. It focuses on specific words or attributes that the customer will be interested about product. Product reviews will be classified based on emotions extracted from reviews. We will be using machine learning based approach and neural network based language model for sentiment analysis of product reviews. To process product reviews, we are proposing the use of Map Reduce technique. We proposing a system which efficiently summarizes the reviews posted by customer to help other for deciding about product. This Hadoop based system significantly reduces time needed for the customer for going through intensive process of reading multiple reviews.*

Keywords- Sentiment analysis; Sentiment polarity categorization; Natural language processing; Product reviews

I. INTRODUCTION

Now in order to increase in sales merchants enabled customer to share their opinions about product. With rapid growth in social media people are sharing their views and opinions online. So, the reviews are generated at enormous rate. The merchandise provides rating system to reduce time required for selecting the product. Data mining technique can be applied to assess the information and their classification. Text mining consists of techniques to analyze human language. So, sentiment analysis can automate the process of rating based on summarization of reviews. For classification of reviews written by single user or of one product are considered as temporal ordered sequence. By assessing the reviews of same person it is easy to classify them because reviews from same person are more consistent than that of the others. Neural network has been used for distributed representation learning and can be used in sentiment analysis.

In this it is possible to learn distributed representation of a product, which captures the semantic information contained in review posted by the user is represented as a vector.

II. SIGNIFICANCE OF STUDY

Now to manage large review data, we have to distribute data and by assigning priority to words which makes larger impact on rating of product. By storing those in vector and calculating the rating about product based on the reviews. This reduces quality time of going through reviews and comments for selection of product. This increases usability of product reviews in customer's perspective. It will be beneficial for taking quick overview of product reviews.

III. METHODOLOGY

A. **Naive Bays Classifier**

Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. Naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable.

The corresponding classifier, a Bayes classifier, is the function that assigns a class label $\hat{p}=C_k$ for some k as follows:

$$\hat{y} = \underset{k \in \{1,\dots,k\}}{argmax}\ p(C_k) \prod_{i=1}^{n} p(X_i|C_k)$$

B. **Method for Feature Extraction from product review**
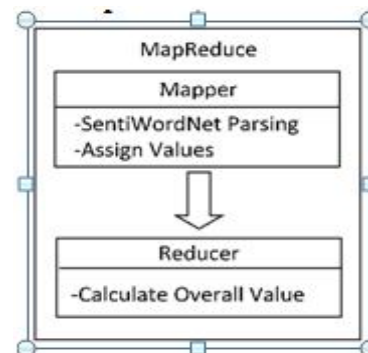
It represents searching for features related words in the sentence and then classifying in the same feature cluster. After finding such word in the sentence, the sentence is classified in the cluster of only those features (such that in display cluster, only sentence relevant to display will be stored).

IV. ADVANTAGES

- Improved Productivity of Users Sales and Services.
- Improved Productivity of Users Sales and Services.
- Process Automation It helps fast decision making
- Classification of product reviews
- Minimize required time for analyzing product reviews.
- Extracting important information from review

V. FUTURE WORK

In future work, these techniques and rating process can be improved by taking into consideration the usage of slangs term and smiley symbols used by people. Features can also be clubbed together according to the score as good, neutral, and bad. Spam reviews can be detected and removed from the list to increase the overall efficiency. Algorithm can be developed to check whether features are present in the reviews posted or not.

Goals

Our goal is to achieve proper classification of product reviews using machine learning approach and summarizing given set of product reviews. Also, to recognize different features of given product from the product reviews.

APENDIX

Sentiment analysis; Sentiment polarity categorization;

VI. CONCLUSION

Sentiment analysis or opinion mining is a field of study that analyzes people's sentiments, attitudes, or emotions towards certain entities. This paper tackles a fundamental problem of sentiment analysis, sentiment polarity categorization. Online product reviews from Amazon.com are selected as data used for this study. A sentiment polarity categorization process (Figure 2) has been proposed along with detailed descriptions of each step. Experiments for both sentence-level categorization and review-level categorization have been performed.

REFERENCES

[1] R. Collobert and J. Weston,"A unified architecture for natural language processing: deep neuralnetworks with multitask learning," in Proc. 25th Int. Conf. Machine Learning, 2008, pp. 160–167.

[2] F. Li, N. Liu, H. Jin, K. Zhao, Q. Yang, and X. Zhu, "Incorporating reviewer and product information for review rating prediction," in Proc. 22th Int. Joint Conf. Artifical Intelligence, 2011,vol. 11, pp. 1820–1825.

[3] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in Proc. Workshop Int. Conf. Learning Representations, 2013.

[4] W. Gao, N. Yoshinaga, N. Kaji, and M. Kitsuregawa, "Modeling user leniency and product popularity for sentiment classification," in Proc. IJCNLP, Nagoya, Japan, 2013.

[5] D. Tang, F. Wei, N. Yang, M. Zhou, T. Liu, and B. Qin, "Learning sentiment-specific word embedding for twitter sentiment classification," in Proc. 52nd Annu. Meeting of the Association for Computational Linguistics, Baltimore, MD, June 2014, pp. 1555–1565.

[6] Y. Zhang, H. Zhang, M. Zhang, Y. Liu, and S. Ma, "Do users rate or review? Boost phrase level sentiment labeling with review-level sentiment classification," in Proc. 37th Int. ACM SIGIRConf. Research & Development in Information Retrieval, 2014, pp. 1027–1030.

[7] D. Tang, B. Qin, and T. Liu, "Learning semantic representations of users and products for document level sentiment classification," in Proc. 53rd Annu. Meeting of the Association for Computational Linguistics, 2015.