Study Of Sentiment Analysis On Product Features

[1]Snehal Kakad, [2]Pooja Kudtarkar, [3]Kajal Parmar
[4]Prof. Vaishali Salvi, [5]Prof. Reena Mahe.
[1, 2,3]Student, Department of Information Technology
[4,5]Professor, Department of Information Technology
AET's Atharva College of Engineering, Malad(W), Mumbai, India

Abstract—Sentiment Analysis has grabbed attention of many researchers around the world in the last few years, since subjective texts are used for various applications. Specifically, sentiment analysis of Online Reviews is becoming increasingly popular area. This is said so because understanding the point of view of customers efficiently is valuable for both companies as well as other buyers. This paper focuses on methods used for extraction of characteristics, classification according to the polarity, and generation of summary of such reviews put by the customers, which can help users in making a decision to buy a product.

I.Implementation

The Internet comprises of several websites that let the users write variety of blogs devoted to range of topics like travel, entertainment, history, finance and so on where people share their perspectives and ideas. Many websites allow users to express their views about the products that are purchased or sold. Such reviews can be referred by new customers prior to investing in a product.

Mostly, people tend to inspect the number of reviews about the product given by different users. This search helps them to understand if a particular product satisfies their needs. Some of the customers look for a particular feature in the product while buying it. This paper provides the ways to do feature wise analysis of the sentiment for the product.

In past few years, researchers have put numerous methods for understanding opinions behind the reviews from areas, for example, political debates [3], movie reviews [2], food reviews [4], and product reviews [1,7]. However this task of analyzing reviews is not as simple as it seems. This is because the reviews may contain use of slang or non dictionary words, erroneous punctuations, spelling mistakes, and unspecified abbreviations. The reviews may also have grammatical mistakes. Thus, the job of summarizing such inappropriate reviews requires some amount of preprocessing [5]. In this paper, we put forward an approach that incorporates the stages like Preprocessing, Extraction of features, Classification and Summarization.

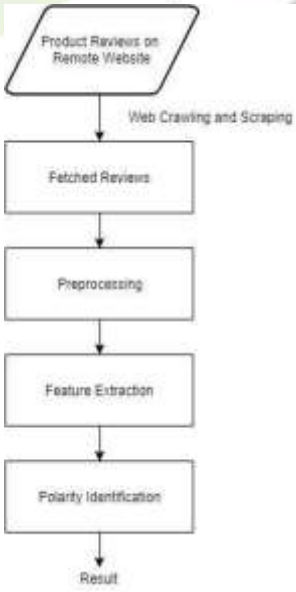

Fig. System Block Diagram

II.Related Work

Feature based classification of online product reviews is very important to the growth of E-commerce and social-networking applications. Nowadays, it has become a huge research activity. The sentiment words present in the free text can be identified by considering the following: adjectives or adverbs, uni-grams [9] or n-grams [10] with their frequency of occurrence, features extracted efficiently by using parts-of-speech (POS) tagging [2, 7,11], the negation of words [9]. The automatic text classification can be done through different machine learning techniques. Normally, people prefer to write reviews in natural language format and hence the spelling and grammatical mistakes are quite common. Hence, the documents are needed to be cleaned before processing [7]. It is language-dependent semantic process, and it tags each word by considering its position in the grammatical context. Usually linguistic parsers such as Link grammar parser [7] and Stanford parser [5] are used for POS tagging of sentences present in user-generated contents.

Yun-Qing Xia in the paper describes a method of mining the opinions [12]. The goal of this system is to extract and summarize the opinions) and reviews, and determine whether these reviews and opinions are positive or negative. This study divides the whole task into four subtasks: expression identification, opinion determination, content-value pair identification, and sentiment analysis.

Adjectives or adverbs are main indicators of polarity orientation of an opinionated phrase. Another approach is by considering synonyms and antonyms of opinionated word in which first start the polarity determination with an initial known seed list of sentiment words and then gradually expand the seed list by considering synonyms and antonyms from lexical resource like WordNet. Since the semantic orientations of sentiment words are depend closely with its usage, it is not accurate to assign fixed polarity orientation to each sentiment descriptor. Hence, most of the recent researches are utilized SentiWordNet API to determine semantic polarity orientation of opinionated words[7].The SentiWordNet assigns a score for each sentiment descriptor in the form of a triplet which represents its positivity/objectivity/negativity scores.

III.Implementation

A review database will be created by gathering product reviews posted by genuine users. This data can be acquired by crawling popular websites. The crawled data then goes through various phases, which are:

A. Preprocessing Phase

Online reviews posted by usersoften contain spelling errors and false punctuations. It is important to detect these flaws to avoid complication in the next phase. So, in this phase basic cleaning task will be performed which include removing of unimportant or disturbing elements for the next phases of analysis, spell-error correction and sentence boundary detection. Sentences end with punctuations like period (.), exclamation mark (!) or question mark (?). Sometimes users use “!” and “?” symbols multiple times for emphasis. These multiple occurrences have to be replaced by single occurrence of that symbol. Tabs and line breaks in the text need to be replaced with a blank and single quotation mark. After sentence boundary detection, we will carry out spell-error correction. The stop words such as articles, pronouns, etc. are filtered out. Finally, all the text is converted to lower case, and extra blank spaces shall be removed.

B. Feature Extraction Phase

In this phase we will be extracting product features from the pre-processed text obtained from the previous phase. We will be considering the frequently occurring nouns and noun phrases as product features and associated adjectives describing them will indicate the sentiment about that particular feature.

We will perform parts-of-speech tagging on the sentences Link Grammar Parser. We will first extract all nouns and noun phrases and identify the frequently occurring ones as product features. The frequently occurring noun and noun phrases are those who occur more number of times in the user's reviews. We won’t use Apriori based approach to extract frequent items from review sentences since this method uses bag of words approach which just records the words and does not keep track of the order in which the words occur in a phrase. Using this approach will require ordering and redundancy pruning. We will use multi-word approach to generate a frequent feature set. A multi-word is an ordered sequence of words that has a higher significance in comparison to the individual words involved in it.

Single words occurring constantly should also be added to the feature set if they are not already subsets of existing multi-words. Features written in plural have to be converted into singular, by method of stemming, so that chances of matching improve.

The features which are semantically same should be considered as one in order to improve the significance of the feature set. Final feature set is enhanced by adding synonyms of the extracted features using WordNet available online.

C. Classification and Summarization Phase.

In this phase we will classify the reviews into positive and negative polarity using the information gained from previous phase. In the previous phase we extracted features as well as adjectives describing them.

To determine the sentiment polarity of an adjective describing a feature we will make use of SentiWordNet which is a lexical resource for sentiment analysis.

If a user's review has more features with positive polarity than the features with negative polarity, it should be classified as positive. Similarly user's review having more features of negative polarity should be classified as negative.

We will generate a feature orientation table that records the features and their corresponding descriptors of positive and negative polarities. The feature oriented table enables us to create feature-wise summary of a particular product using the reviews generated by several online users.

IV.Expected Output

A website that provides a user login and provide user authority to give their review about a particular product. These reviews will then be classified according to the polarity and a graphical representation for polarity of each feature of the product will be given. The visitor can see positive as well as negative reviews entered by users.

V.Conclusion and Future Scope

Classifying reviews has several interesting and commercially significant applications but this task is difficult than classifying regular text and requires intensive preprocessing. The success of the sentiment analysis task is mainly dependent on the efficiency of the Preprocessing and feature extraction phases. The feature based analysis approach used in this project shall identify opinions in the sentences of user reviews and classify them with great accuracy.

In the future, sentiment analysis can be performed on larger data sets. The classification technique can be extended to support slang words by building more efficient datasets. A classifier can be build that understands sarcasm in a sentence and can overcome linguistic barriers.

Acknowledgment

It gives us great pleasure in presenting this paper titled: "Study of Sentiment Analysis on Product Features" and we are thankful to IJCRT (International Journal of Creative Research Thoughts) for giving us the opportunity to research and publish this paper.
We express our gratitude to our project guide Prof. Reena Mahe and Prof. Vaishali Salvi, who provided us with all the guidance and encouragement and helped us in finding our mistakes and improving it. We also would like to express our sincere gratitude to Project Coordinators. We are thankful to the Head of the Information Technology Dept. Prof. Neelima Pathak, for her approval of this project. We are also thankful to her for providing us the needed assistance, detailed suggestions and also an encouragement to do the project.
We would like to deeply express our sincere gratitude to our respected principal Prof. Dr. Shrikant Kallurkar and the management of Atharva College of Engineering for providing such an ideal atmosphere to work on this project with all necessary reference materials

REFERENCES

[1]Mr. Nilesh Shelke, Dr. Shrinivas Deshpande, Dr. Vilas Thakare, "Statistical Feature based Approach for Aspect Oriented Sentiment Analysis", *published in IEEE 2017.*

[2]L. Zhao and C. Li, "Ontology based opinion mining for movie reviews," in *Proceedings of the 3rd International Conference onKnowledge Science, Engineering and Management*, pp. 204–214,2009.

[3]A. Balahur, Z. Kozareva, and A. Montoyo, "Determining the polarity and source of opinions expressed in political debates," in *Proceedings of the 10th International Conference on IntelligentText Processing and Computational Linguistics*, vol. 5449 of *Lecture Notes in Computer Science*, pp. 468–480, Springer, 2009.

[4]Y. H. Gu and S. J. Yoo, "Mining popular menu items of a restaurant from web reviews," in *Proceedings of the InternationalConference on Web Information Systems and Mining (WISM '11)*, vol. 6988 of *Lecture Notes in Computer Science*, pp. 242–250, Springer, 2011.

[5] L. Dey and S. M. Haque, "Opinion mining from noisy text data," *International Journal on Document Analysis and Recognition*,vol. 12, no. 3, pp. 205–226, 2009.

[6]Seyed-Ali Bahrainian, Andreas Dengel, "Sentiment Analysis and Summarization of Twitter Data", *published in IEEE 2013.*

[7]Mita K. Dalal, Mukesh A. Zaveri, "Semi Supervised Learning Based Opinion Summarization and Classification for Online Product Reviews", *published in Hindawi Publishing Corporation 2013.*

[8]S. Bao, S. Xu, L. Zhang, R. Yan, Z.Su, D. Yan and Y. Yu, "Mining Social Emotions from Affective Text", *published in IEEE 2012.*

[9] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?sentiment classification using machine learning techniques", Proceedings of the ACL-02 conference on Empirical methods in natural language processing, vol.10, 2002, pp. 79-86.

[10]K.Dave, S.Lawrence, and D. M. Pennock, "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews", Proceedings of WWW, 2003, pp. 519–528.

[11] R. Prabowo and M. Thelwall, "Sentiment analysis: A combined approach", Journal of Informetrics, vol. 3, pp.143-157, 2009.

[12] Yun-Qing Xia, The Unified collocation Framework for Opinion Mining, Proceedings of the Sixth International Conference on Machine Learning and Cybernetics, Hong Kong, 19-22 August 2007.