International Journal of Multidisciplinary and Current Research

Research Article

ISSN: 2321-3124
Available at: http://ijmcr.com

Customer Reviews Mining using Machine Learning Technique

Pawan Shankar Sutar

Department of Computer Engineering, Shah and Anchor College of Engineering, Mumbai

Accepted 08 Oct 2015, Available online 10 Oct 2015, **Vol.3** (**Sept/Oct 2015 issue)**

Abstract

Nowadays large numbers of customers are choosing online shopping because of its convenience, reliability, and cost. As the number of products being sold online increases, it is becoming difficult for customers to make purchasing decisions based on only short product descriptions. For this reason customer's product reviews are a crucial and very important part of ecommerce. Reviews are the opinion of the customer who bought product and other customers really take them into consideration before making a decision. As e-commerce is becoming more and more popular, the number of customer reviews that a product receives grows rapidly. This makes it difficult for a potential customer to read them in order to make a decision on whether to buy the product. Customer reviews mining, particularly the text describing the features, comparisons and experiences of using a particular product provide a rich source of information to compare products and make purchasing decisions Product reviews are classify as either positive review or negative review. The purpose of this system is to classify reviews either positive review or negative review. To achieve these we used semantic analysis. It also includes pre-processing, POS tagging and feature extraction.

Keywords: *POS tagger, Pre-processing, Features selection etc.*

1. Introduction

This research paper is to classify customer reviews. To find people's opinions on a topic and its different aspects, which we call *evaluative opinions*, those irrelevant sentences should be removed. The goal of this research is to identify evaluative opinion sentences [3]. This paper proposes the problem of identifying evaluative sentences from online discussions. In paper [1] combination of both supervised machine learning and rule-based approaches are proposed for mining feasible feature-opinion pairs from subjective review sentences. In the first phase of the proposed approach, a supervised machine learning technique is applied for Classifying subjective and objective sentences from customer reviews. In the next phase, a rule based method is implemented which applies linguistic and semantic analysis of texts to mine feasible feature-opinion pairs from subjective sentences retained after the first phase. The effectiveness of the proposed methods is established through experimentation over customer reviews on different mobile products. I took three mobile models which are in same range. I have selected three models. They are as follow:

- Nokia Lumia 730
- Samsung Galaxy Grand Prime
- Sony Xperia M

For development of our project we used preprocessing technique, which include stop word removal, It compressed our data and removed unwanted words for further processing. POS tagger assign part of speech to each word, which help for feature extraction. Each feature is co-related to opinion words, which define polarity of that word. It helps to decide whether review is positive or negative.

2. Data Source

As I mentioned in introduction for our system I took mobile products reviews (Nokia Lumia 730, Samsung Galaxy Grand Prime and Sony Xperia M). After I selected mobile models, I catch reviews of those mobile models from online shopping sites like snapdeal, amazon etc. Those reviews are randomly selected for analysis of my system.

3. Pre-processing

The main objective of pre-processing is to obtain the key features or key terms from selected reviews. It is important to select the significant keywords that carry the meaning and discard the words that do not contribute to decide whether review is positive or negative.

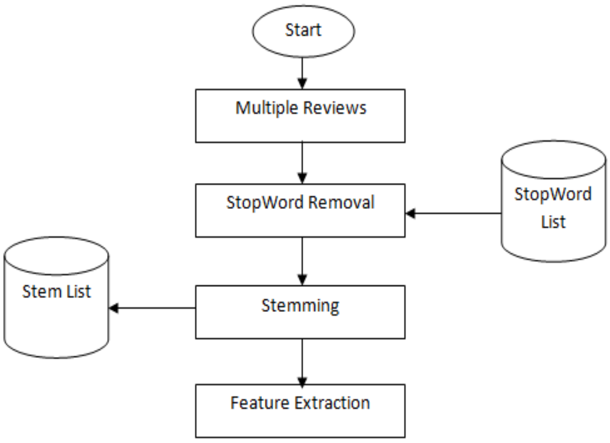

Fig.1 Pre-Processing

3.1 Stop Word Removal

A stop words is a commonly occurring grammatical word that does tell us anything about documents content. In English language words such as "a‟, ‟an‟, "the", "and" etc. are stop words. Stop Words are words which do not contain important significance to be used in Search Queries. Usually these words are filtered out from search queries because they return vast amount of unnecessary information

3.2 Stemming

In Stemming process we find out the root/stem of a word. The purpose of this method is to remove various suffixes, to reduce number of words, to have exactly matching stems, to save memory space and time. For example, the words producer, producers, produced, production, producing can be stemmed to the word "Produce".

3.3 Feature Extraction

It is very difficult task to process text which contains millions of different unique words, so it makes text analytics process difficult. Therefore, feature-extraction is one of the approaches used when applying machine learning algorithms like Support Vector Machine for text categorization. With the survey, it has been found that a feature is a combination of keywords. (Attributes), which captures essential characteristics and sentiment of the text. A feature extraction method detects and filters only important features which a far smaller set than actual number of attributes and make them a new set of features by decomposition of the original data.

4. Sentiment Analysis

Analyzing Sentiment of the text is itself a challenging task in Natural Language Processing. There are many approached studied by me while finding solution for sentiment classification of reviews.

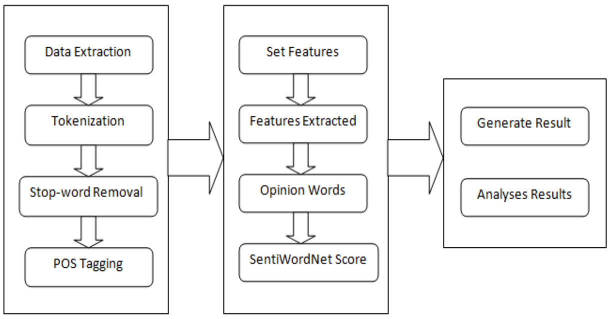

Fig. 2 Proposed System

4.1 POS tagger

A Part-Of-Speech Tagger (POS Tagger) is a piece of software that reads text in some language and assigns parts of speech to each word (and other token), such as noun, verb, adjective, etc., It provides a way to ―tag the words in a string. That is, for each word, the ―tagger gets whether it‟s a noun, a verb, etc. and then assigns the result to the word.

For example:
―This is a mobile data
The output of POS tagger is
―This/DT is/VBZ a/DT mobile/NN data/NN

To get this, the tagger needs to load a "trained" file that contains the necessary information for the tagger to tag each word in a string. This "trained" file is called a model and has the extension "tagger". Here we are interested in grammatical rules for English language only.

4.2 Feature Selection

The features of mobile phones are the set of capabilities, services and applications that they offer to their users. All mobile phones have a number of features in common, but manufacturers also try to differentiate their own products by implementing additional functions to make them more attractive to consumers. Most common features which I selected for mobile models Nokia Lumia 730, Samsung Galaxy Grand Prime, Sony Xperia M are display, screen size, internal memory, battery, processor, operating system, app, camera, Bluetooth, cost, GPS, voice, sound, RAM. Identify features of the products that customers have expressed their opinions on.

Given a set of customer reviews of a particular product, we need to perform the following tasks:
1. Identifying product feature that customer commented on;
2. Extracting opinion words or phrases through adjective, adverb, verb, and noun and determining the orientation;
3. Generating the summary.

We use a part-of-speech tagger to identify phrases in the input text that contains adjective or adverb or verb or nouns as opinion phrases. A phrase has a positive semantic orientation when it has good associations (e.g., “awesome camera”) and a negative semantic orientation when it has bad associations (e.g., “low battery”).

After POS tagging is done, we need to extract features that are nouns or noun phrases using the pattern knowledge.

For opinion words extraction, we used extracted features that are used to find the nearest opinion words with adjective/adverb. To decide the opinion orientation of each sentence, we need to perform three subtasks. First, a set of opinion words (adjectives, as they are normally used to express opinions) is identified. If an adjective appears near a product feature in a sentence, then it is regarded as an opinion word. We can extract opinion words from the review using the extracted features

Moreover, both adjective and adverb are good indicators of subjectivity and opinions. Therefore, we need to extract phrases containing adjective, adverb, verb, and noun that imply opinion. We also consider some verbs (like, recommend, prefer, appreciate, dislike, and love) as opinion words. Some adverbs like (not, always, really, never, overall, absolutely, highly, and well) are also considered. Therefore, we extract two or three consecutive words from the POS-tagged review if their tag conforms to any of the patterns.

We collect all opinionated phrases of mostly 2/3 words like (adjective, noun), (adjective, noun, noun), (adverb, adjective), (adverb, adjective, noun), (verb, noun), and so forth from the processed POS-tagged review.

4.3 SentiWordNet

SentiWordNet is a lexical resource for opinion mining. SentiWordNet assigns to each synset of WordNet two sentiment scores: Positivity and negativity. After sentiwordNet 3.0 downloaded, we store all sentiwords of sentiwordNet in our database.

Hence we can we get sentiment polarity of each opinion word. On the basis of that sentiment polarity and mobile features we make analysis of mobile products (Nokia Lumia 730, Samsung Galaxy Grand Prime and Sony Xperia M). Once the model has been trained and tested, we need to measure the performance of the model. For this purpose we used six features of mobile model camera, battery, memory, cost, hang and display.

Figure 3 shows overall performance of mobile models (Nokia Lumia 720, Samsung grand and Sony experia M). For overall performance we consider all crucial features of mobile models (Camera, battery, memory, cost, hang and display). If we consider these features then we came into the analysis which shows that Nokia Lumia 730 have highest positive sentiment polarity with no negative polarity.

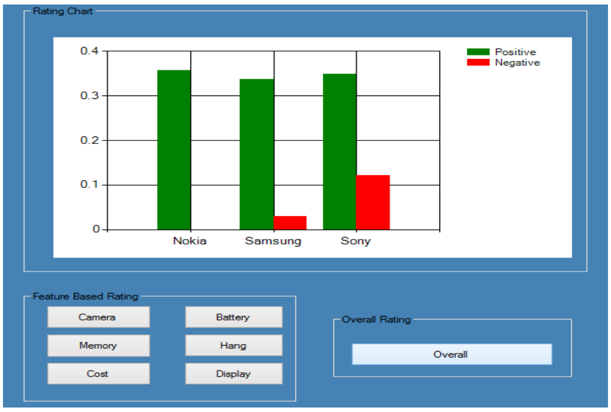

Fig. 3 Overall Reviews Analysis

Conclusions

Thus we have successfully loaded the data of the reviews in the system and classify review as on their polarity (Positive review and Negative review). The proposed solution and implementation of the system justify that we build our model on dynamic data set of mobile review and we are getting satisfactory performance analysis for reviews data set which is very crucial step of classification.

In the end we have also created some of the charts to get both the summarized result as well as the detailed result which based on the crucial features of mobile model. Thus we have achieved our objective of showing some meaningful patterns which is useful for online shopping. These patterns can be used effectively to guide the customers to make decision while doing online shopping. This kind of model can be used in various e-learning systems. We can use this proposed system in the area where product reviews are necessary in online shopping sites.

References

[1] Subjectivity Classification using Machine Learning Techniques for Mining Feature- Opinion Pairs from Web Opinion Sources Ahmad Kamal Department of Mathematics Jamia Millia Islamia (A Central University).

[2] Mining Twitterspace for Information: Classifying Sentiments Programmatically using Java
Jinan Fiaidhi.

[3] Identifying Evaluative Sentences in Online Discussions Zhongwu Zhai Bing Liu Lei Zhang Hua Xu Peifa Jia State Key Lab of Intelligent Tech. & Sys., Tsinghua National Lab for Info. Sci. and Tec.

[4] Mining Opinion Features in Customer Reviews Minqing Hu and Bing Liu Department of Computer Science University of Illinois at Chicago.

[5] Importance of Online Product Reviews from a Consumer's Perspective Georg Lackermair1;2, Daniel Kailer1;2;_, Kenan Kanmaz1 1Munich University of Applied Sciences, Germany.

[6] A Survey of Text Mining Techniques and Applications Vishal Gupta Lecturer Computer Science & Engineering, University Institute of Engineering & Technology.