IOSR Journal of Engineering
Mar. 2012, Vol. 2(3) pp: 389-391

Movie Review analysis using Rule-Based &Support Vector Machines methods

Swati A. Kawathekar[1], Dr. Manali M. Kshirsagar[2]
[1]*M.Tech. III[rd] Semester, Dept. of Computer Technology, Yashwantrao Chavan Collage of Engineering, Nagpur, Maharashtra, India*
[2]*Head, Dept. of Computer Technology, Yashwantrao Chavan Collage of Engineering, Nagpur, Maharashtra, India*

Abstract
Sentiment analysis (SA) is broad forte of Natural language processing which deals with the computational treatment of opinion, sentiment and subjectivity in text. Due to increased availability of online reviews, there is a growing need to organize them. Sentiment analysis is one present day solution for this issue. An important part of our information-gathering behavior has always been to find out what other people think and whether they have favorable (positive) or unfavorable (negative) opinions about the subject. This survey studies the role of negation in an opinion-oriented information-seeking system. We investigate the problem of determining the polarity of sentiments in movie reviews when negation words, such as not and hardly occur in the sentences. This paper combines rule-based classification, supervised learning and machine learning into a new combined method. This method is tested on movie review.

Keywords: ***sentiments analysis, movie review, rule base, support vector machines***

I. INTRODUCTION

Sentiment analysis of blog text, review sites and online forums has been a popular subject for several years in the field of natural language processing. Researchers have shown that several techniques can successfully estimate the opinion polarity of a given text. In many cases our decisions are influenced by the opinions of others. Before the internet awareness became widespread, many of us used to ask our friends or neighbors for opinion of an electronic good or a movie before actually buying it or going for it. With the growing availability and popularity of opinion-rich resources such as online review websites and personal blogs, new opportunities and challenges arise as people now can, and do, actively use information technologies to seek out and understand the opinions of others. Unfortunately, 85% of these opinion rich resources are available in unstructured format. It has encouraged the analysts to develop an intelligent system that can automatically categorize or classify these text documents. This paper is an overview of the area of Sentiment Analysis, which deals with subjective texts. This paper gives the approach how sentiments can be analyzed by using Rule Based approach and Support Vector Machines. This paper presents the empirical results of a comparative study that evaluates the effectiveness of different classifiers, and shows that the use of multiple classifiers in a hybrid manner can improve the effectiveness of sentiment analysis.

A. Applications to Review-Related Websites

Summarizing user reviews is an important problem. One could also imagine that errors in user ratings could be fixed: there are cases where users have clearly accidentally selected a low rating when their review indicates a positive evaluation. Moreover, as discussed later in this survey, there is some evidence that user ratings can be biased or otherwise in need of correction, and automated classifiers could provide such updates.

B. Applications as a Sub-Component Technology

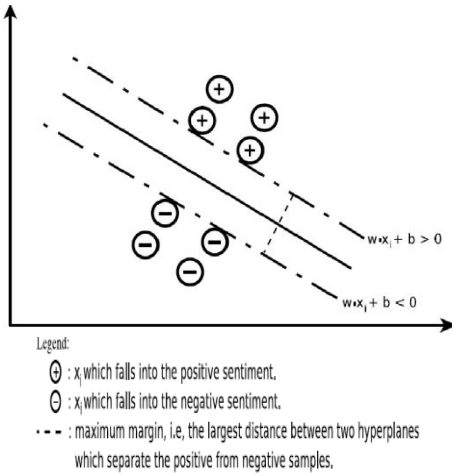
Sentiment-analysis systems also have an important potential role as enabling technologies for other systems. One possibility is as an augmentation to recommendation system, since it might behoove such a system not to recommend items that receive a lot of negative feedback. Detection of "flames" (overlay Sheated or antagonistic language) in email or other types of communication is another possible use of subjectivity detection and classification.

II. RULE BASED APPROACH

Using lexical rules, a baseline was created by tokenizing each sentence in every document and then testing each token, or word, for its presence within the compiled General Inquirer data set. If the word existed and was associated with a positive sentiment, a +1 rating was applied to the post's overall polarity score. Each post starts with a neutral score of zero, and was considered positive if the final polarity score was greater than zero, or negative if the overall score was less than zero. In the Rule-Based approach, rules are to be defined which will contain an antecedent and its associated consequent that have an if-else relation. In this methodology, certain rules are to be form and then the sentiments should be analyzed depending on it.

III. SUPPORT VECTOR MACHINES

A support vector machine (SVM) is for a set of related supervised learning methods that analyze data and recognize patterns, used for classification and regression analysis. The standard SVM takes a set of input data and predicts, for each given input, which of two possible classes the input is a member of, which class. More formally, a support vector machine constructs a hyperplane or set of hyperplanes in a high- or infinite- dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training data points of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier.



IV. PRE-PROCESSING OF TEXT

In the pre-processing of text, the words that cannot derive any sentiments are to be removed. The words like this, it, who, or, etc does not give any clue for analysis of sentiments. So these words must be discarded from the input. Along with the pre-processing of the text the tagging of the words to their relevant parts of speech. Suppose let's take an example, this is a good movie. In this sentence, this, is, a, will get removed as they are not deriving any sentiments.

After pre-processing of the text there are three tables for storing the refined input. The first table will be for the words that affect the analysis. These are the word that gives no meaning for deriving sentiments. The second table contains the general words and the third table will consist of the negative words.

V. EXPERIMENTAL SETUP

For the movie review analysis the rule based approach is used along with the Support Vector Machines approach. The only rule based approach does not provide the maximum efficiency. So it's necessary to have the supervised method so that the system can learn from the input. The above tables which will contain the stop word, general words and the negation words. Sometime it may happen negation word does not contribute for the negative sense but it gives the sense for positive review. So for such words, rules will be created and depending upon the context its analysis will be done. For this we have been using Java as a front end and My SQL as a back end. In our paper we have taken graphical user interface for the analysis.

VI. RESULT AND DISCUSSION

The rule based approach result creates the rules by taking the affecting words, inverted words, and negation words. After the output of rule based approach it will check or ask whether the output is correct or not. If the input sentence contains any word which is not present in the database which may help in the analysis of movie review, then such words are to be added to the database. This is supervised learning in which the system is trained to learn if any new input is given. This approach will always increase the efficiency of the system.

The result of the analysis of the movie review analysis can be shown as:

Input Statement: the movie is not only good the songs in the movie are awesome.

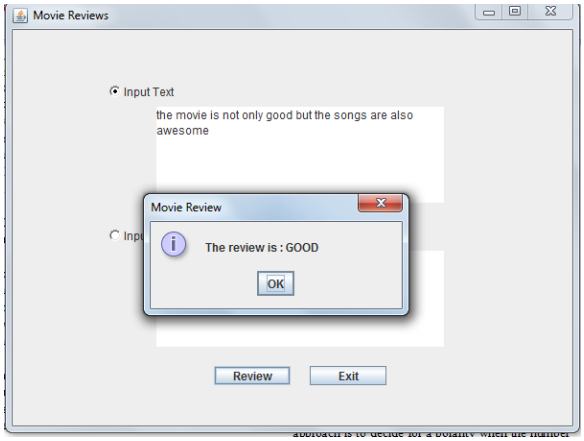
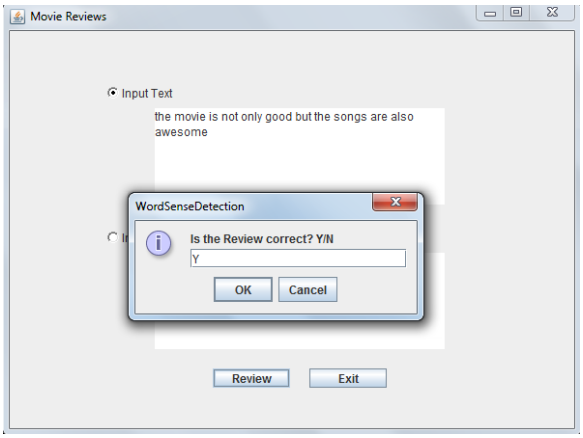


Fig 4:- Main Graphical User Interface (GUI)

Figure 4 shows the main graphical user interface via which input is given to the system. In this the processing of the text is been done. The movie review is: good.

After the review is given the system checks whether the output is correct or not. If the output is not correct then it will update the affecting word to the database.



Creating a domain independent sentiment classifier is not a simple task. This evaluation proposed two different approaches and found that each was only capable of accurately classifying documents across domains with a maximum accuracy. Alternatively, creating a sentiment classifier for a particular domain was capable of classifying documents. In contrast to the rule-based approach, the results of the machine-learning based classifier are significantly better. The benefit of the rule-based approach is that no training material is required. But, a problem for the rule-based approach is to decide for a polarity when the number of positive words equals the number of negative words.

REFERENCES

[1] Rudy Prabowo, Mike Thelwall, "Sentiment Analysis: A Combined Approach", School of Computing and Information Technology,University of Wolverhampton

[2] Mullen, T. and Collier, N. (2004). Sentiment analysis using support vector machines with diverse information sources. Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 412–418.

[3] Siva RamaKrishna Reddy V., D V L N. Somayajulu, Ajay R.Dani, "Classification of Movie Reviews Using Complemented Naive Bayesian Classifier", NIT Warangal, Prithvi Information Solutions Limited, India

[4] Maral Dadvar, Claudia Hauff, Franciska de Jong, " Scope of Negation Detection in Sentiment Analysis", Human Media Interaction Group University of Twente Enschede, Netherlands

[5] Bo Pang1 and Lillian Lee2, "Opinion mining and sentiment analysis", Yahoo! Research, 701 First Ave. Sunnyvale, CA 94089, U.S.A., Computer Science Department, Cornell University, Ithaca, NY 14853, U.S

[6] Alistair Kennedy and Diana Inkpen , "Sentiment Classification of Movie Reviews Using Contextual Valence Shifters",University of Ottawa, Ottawa, ON, K1N 6N5, Canada

[7] V. Suresh , Ashok Veilumuthu , Avanthi Krishnamurthy , C. E. Veni Madhavan , Kaushik Nath , Sunil Arvindam ,"A Non syntactic Approach for Text Sentiment Classification with Stopwords",

[8] Yelena Mejova, Computer Science Department, University of Iowa "Sentiment Analysis: An Overview"

[9] Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. Foundation and Trends in Information Retrieval, 2(1-2):1–135

[10] Pang, B. and Lee, L. (2002). Thumbs up?: sentiment classification using machine learning techniques. Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing, 10:79–86.