IJCTA, 9(26) 2016, pp. 103-108

A Comparative Study of Polarity Classifiers on Movie Reviews

R. K. Renitaa[1] and M. Lovelin Ponn Felciah[2]

ABSTRACT

Sentiment analysis is the finding of attitudes such as enduring, effectively colored beliefs, disposition towards objects or persons. Sentiment analysis is known as opinion extraction, opinion mining, sentiment mining and subjective analysis. The sentiment analysis is to review positive or negative of the movie, for prediction, in politics to find what people think about this candidate or issue and as a public statement to know about the consumer confidence. This paper focus on document-level sentiment analysis for the Movie data set collected and experiment it. Whether the document reveals the positive, or negative about the Movie.

Keywords: Sentiment Analysis, Opinion Mining, Document Level, Movie Reviews.

1. INTRODUCTION

Sentiment analysis is a language processing task that used an algorithm formulation to classify narrow-minded content and sort out its positive, negative or neutral polarity. Sentiment analysis or opinion mining is the computational learning of people's opinions, attitudes and emotions toward a person[1]. It involves the discovery of opinions from a text segment at various levels of granularity as well as document-level, sentence-level and aspect-level.

- **Document-level:** It classifies the document as positive, negative or neutral. (E.g. a single product)
- **Sentence-level:** It classifies the sentences as positive, negative or neutral. (E.g. objective and subjective sentences)
- **Aspect-level:** It classifies the sentiment to the specific aspects of entities. (E.g. entity of product)

Sentiment analysis and opinion mining area along with the large amount frequently used applications of language technology, impacting together industries and a mixture of further intellectual principles. Yet sentiment analysis is still dominated by bag-of-words approaches, and attempts to contain extra linguistic framework normally prevent at the sentence level. Since document level opinion mining essentially involves multi-sentence texts, it seems to study of document-level formation must have a task to play[3]. This document-level classification aims to computerize the assignment of classifying a textual analysis, which is specified on a particular topic, as expressing a positive or negative statement.

Reviews are increasing in a faster rate day by day because every human being likes to present their view on the network. Hug records of reviews are presented for a particular product which makes complicated for a client to study all the reviews and come to a decision [2]. Thus, mining this record, identify the customer opinions and categorize them is a main task. Recent advances in the fields of text mining, information removal and information recovery have been forced by a related goal to utilize the secreted value protected in huge volumes of formless data [6].

[1] Research Scholar, Computer Science, Bishop Heber College, Tiruchirapalli, Tamil Nadu, India
[2] Assistant Professor, Computer Applications, Bishop Heber College, Tiruchirapalli, Tamil Nadu, India

2. RELATED WORK

[1] In this paper "Survey of Sentiment Analysis Algorithms and Applications" the author Walaa Medha *et al*, tackled the comprehensive overview of sentiment analysis in the text mining field. The opinion mining can be classified into three levels and it is a Natural Language Processing and Information Extraction task. The three levels stated on Document Level, Sentence Level and Aspect and Feature Level. They showed the effectiveness of the system by using experimental results of movie reviews.

[2] Research in the opinion mining of movie reviews at document level proposed by Richa Sharma *et al*, has mentioned the planned work is directly connected to the Minqing Hu and Bing Liu work on mining and shortening purchaser reviews. The proposed system divided into phases such as (i) Data Collection (ii) POS Tagging (iii) Extracting opinion words and seed list preparation and polarity detection and classification. The conduct experiment outcome shows that the Document based Emotion Orientation System at hand well between outlays to the picture meadow as compared to 'AIRC Sentiment Analyzer'. Upcoming structure makes the truth of 63%.

[3] In the paper "Document-level sentiment classification: Empirical comparison between SVM and ANN", the authors Rodrigo Moraes et al, supervised that the method consists of two stages and the reviews are classified by using learning models such as Support Vector Machines (SVM) and Naïve Bayes (NB) and the empirical comparison is represented between SVM and ANN and their adopted standard evaluation context in a traditional tag of words model.

[4] However "Better Document-Level Sentiment Analysis from RST Discourse Parsing" was published by Parminder Bhatia *et al.*, they stated that the document level sentiment analysis can be improved by Rhetorical Structure Theory (RST) parses by using composition of local information up the discourse tree.

[5] Zhaopeng Tu *et al.*, automatically identifying high impact sub-structures that are relevant to a given task a systematic study investigating sequence and convolution kernels using different types is presented by the paper titled "Identifying High-Impact Substructures for convolution kernels in document-level sentiment classification".

In the paper titled [7] "Which side are you on? Identifying Perspectives at the document and sentence levels" the author Wei-Hao Lin et al, investigated a fresh problem of routinely identifies the perception starting which a file is written by the perspective they intended a skewed assessment of virtual implication a point-of-view.

3. PROPOSED WORK

The support vector machine method and naïve based technique are used in this system. The dictionary used to determine the opinion words is Wordnet, and is used to find their synonyms and antonyms. The proposed work is related to the work of Richa Sharma [2] on the paper of opinion mining of movie reviews at document level. The proposed system differentiates the positive, negative and neutral documents and states the total account of positive, negative and neutral documents in the output. The output is helpful in decision making for the users and they can easily conclude how many optimistic and pessimistic documents are present. The majority of opinion words determine the polarity of a given document.

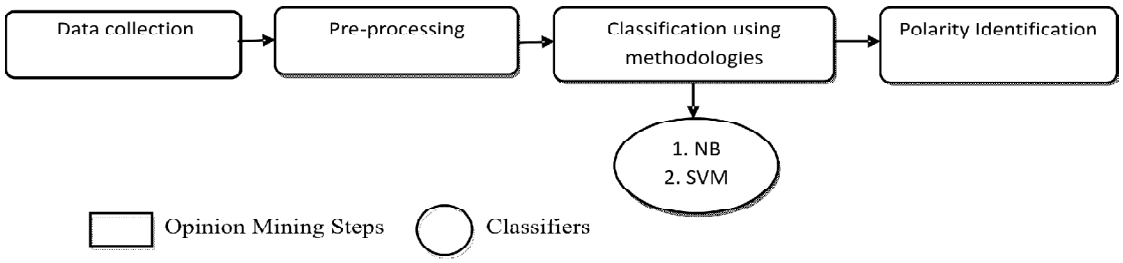

Figure: Flowchart for Proposed Work

3.1. Preprocessing Work

Movie Reviews are collected in large numbers from different-different websites. Various websites are available on the web which contain movie reviews such as www.movie-review-data.com, www.imdb.com etc. The reviews are collected to determine the polarity and are necessary to get the cleared reviews. These preprocessed reviews are applied as input.

3.2. Comparison of Algorithm

3.2.1. Naïve Bayes

Naïve Bayes is a supervised probability machine learning classifier method that assumes terms occur independently. This can be used to in classifying textual documents in simplest method[9]. The Naïve Bayes model computes the posterior probability of a class, based on the allocation of words in the document this illustration works with the BOWs feature extraction which ignores the situation of the word in the document. It uses Bayes Theorem to calculate the probability to facilitate a certain characteristic set belongs to an exacting label[1]. It is easy and fast to predict the class of the test data set. It also performs well in multiclass prediction. When the assumption of independence holds, a naïve Bayes classifier performs better evaluate to other models and you need less training data.

3.2.2. Support Vector Machine

It is a method learning in supervision through various attractive qualities that make at a popular algorithm[3]. The SVM is the main principle which can best separate the different classes. It is used in the search space to determine the linear separators. The SVM has the capability to build a non-linear assessment surface with the unique attribute freedom by mapping the statistics instances non linearity[1]. It is another well-known and wide margin machine learning based classifier. It is vector space model based classifies that needs feature vectors transformed into numerical[9]. SVM is also used to evaluate the quality of information in reviews and has Information Quality (IQ) frameworks which are used to find information oriented feature set. It provides a compact numeric summarization of opinions on micro-blog platforms[1].

3.3. Polarity Identification

The polarity is based on opinion. If the positive words are more, then it is a positive polarity document and if negative words are more than the document shows negative polarity. The three classifications of polarity are positive, negative and neutral opinion.

Positive Opinion: If the document has more number of positive opinion words than the negative opinion words, then the system gives the positive opinion to the document.

(E.g) The movie portrays, “a dreamy love story, sweet romance, realistic approach to life’s problem, **technical brilliance**”.

In this positive polarity document, the positive words are shown in bold.

Negative Opinion: The system gives the negative opinion, if the numbers of negative words are more in the document.

(E.g) “The film has moved out a little **overboard** in the picture of characters which can make one **cringe** at times.

This document has number of negative words than positive words, therefore it shows negative polarity.

Neutral Opinion: If the numbers of positive opinion words are equal to the negative opinion words then the document shows neutral opinion.

(e.g) "The preeminent part of the movie is that the movie explores a wide range of emotions but it **never *gets*** too ***melodramatic.***

Never gets too melodramatic shows positive polarity. Here to get melodramatic is a negative opinion word, but it is preceded by the word never so the polarity of the sentence is reversed.

4. EXPERIMENTAL RESULTS

The collected reviews are applied to perform the experiment using the proposed system to classify the reviews into positive, negative and neutral. All the documents are manually calculated and it is compared with the result of the proposed system to check the performance efficiency of the system. The reviews are applied to the system by using two techniques such as NB and SVM. Finally the results of two techniques of the proposed system are compared and the results have shown that the accuracy of NB is better than SVM. The system is compared with three evaluation measures,

- Precision
- Recall
- Accuracy

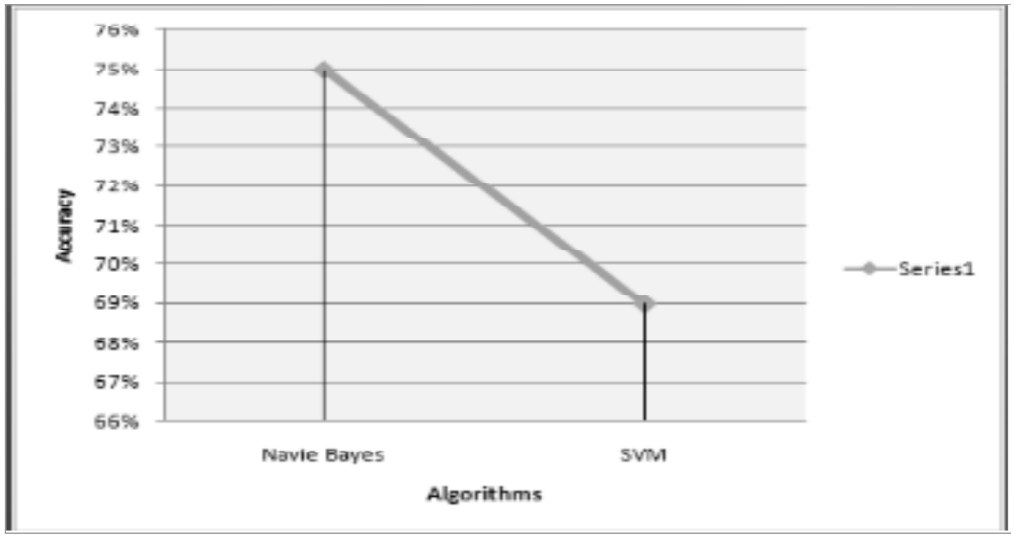

Figure 1: Analysis of Accuracy

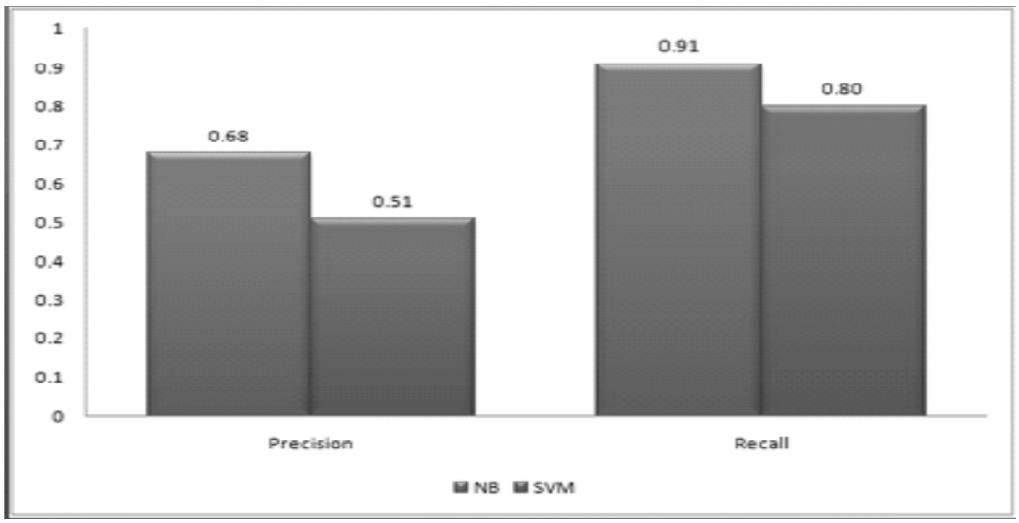

Figure 2: Comparison between Precision and Recall of NB and SVM

The table gives the accurate results of NB and SVM based system. Then the graph explains the pictorial representation of accuracy for both the system. The chart shows the recall and precision of the NB for the clear study of the performance of the system.

Table 1
The overall results rates of NB and SVM

Algorithms	*Accuracy*	*Precision*	*Recall*
NB	0.75	0.68	0.91
SVM	0.69	0.51	0.8

The experimental results show that SVM has 69% and NB has 75%. Thus, by analyzing the table and chart values it is evident that NB has more accuracy and better performance than SVM.

5. CONCLUSION

In order to conclude, the NB based system has better accuracy and performance efficiency in document level classification. This Naïve Bayes model is easy to build and it is used in large data sets. It is known to outperform even highly sophisticated classification methods. This Bayes Algorithm is used in Real Time Prediction, Multiclass Prediction, Recommendation system and Text Classification or Spam Filtering or Sentiment Analysis. Hence the NB is performed in Document Level sentiment analysis. Nowadays, the document level opinion mining is widely used to perform movie reviews. It is good at its performance and accuracy is better. Hence the movie reviews and other documents, can be processed and can be classified by their polarity as positive, negative and neutral using NB based document level sentiment analysis.

REFERENCES

[1] Walaa Medhat, Ahmed Hassan, Hoda Korashy, "Sentiment analysis algorithms and applications: A survey", *Ain Shams Engineering Journal* **5(4)**, 1093-1113.2014.

[2] Richa Sharma, Shweta Nigam and Rekha Jain, "Opinion mining of Movie Reviews At Document level", *Internatinol Journal on Information Theory (IJIT),* **3(3)**, 2014.

[3] Rodrigo Moraes, Joao Francisco Valiati, Wilson P.Gaviao Neto, "Document-level sentiment classification: An empirical comparison between SVM and ANN", *Expert Systems with Applications,* **40(2),**621-633, 2012.

[4] Parminder Bhatia and Yangfeng Ji and Jacob Eisenstein, "Better Document-level Sentiment Analysis from RST Discourse Parsing", *School of Interactive Computing, Georgia Institute of Technology, Atlanta,* GA 30308.

[5] Zhaopeng Tu, Yifan He, Jennifer Foster, Josef van Genabith, Qun Liu, Shouxun Lin, "Identifying High-Impact Sub-Structures for Convolution Kernels in Document-level Sentiment Classification", *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics,* **8(14)**, 338-343, 2012.

[6] Matthew Hurst and Kamal Nigam, "Retrieving Topical Sentiments from Online Document Collections", *Intelliseek Applied Research Center, Pittsburgh. 27-14, 2004.*

[7] Wei-Hao Lin, Theresa Wilson, Janyce Wiebe, Alexander Hauptmann, "Which Side are You on? Identifying Perspectives at the Document and Sentence Levels", *In Proceedings of the Tenth Conference on Natural Language Learning (CoNLL'06),* 109-116, 2006.

[8] Ainur Yessenalina, Yisong Yue, Claire Cardie, "Multi-level Structured Models for Document-level Sentiment Classification", *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing,* **9(11),** 1046-1056, 2010.

[9] Vivek Kumar Singh, Rajesh Piryani, Pranav Waila and Madhavi Devaraj, "Computing Sentiment Polarity of Texts at Document and Aspect Levels", *Ecti Transaction on Computer and Information Technology*, **8(1),** 67-79, 2014.

[10] Neha Nehra, "A Survey on Sentiment Analysis of Movie Reviews", *International Journal of Innovative Research in Technology (IJIRT),* **1(7),** 36-40, 2014.

[11] Li Zhuang, Feng Jing, Xiao-Yan Zhu, "Movie Review Mining and Summarization", *CIKM'06***, 5(11),** 43-50, 2006.

[12] Jyotika Yadav, "A Survey on Sentiment Classification of Movie Reviews", *International Journal of Engineering Development and Research (IJEDR)* **3(1)**, 2321-9939, 2014.

[13] Pimwadee Chaovalit and Lina Zhou, "Movie Review Mining: a Comparison between Supervised and Unsupervised Classification Approaches", *Proceedings of the 38th Hawaii International Conference on System Sciences, 1-9,* 2005.

[14] Abd. Samad Hasan Basari, Burairah Hussin, Gede Pramudya Ananta, Junta Zeniarja, "Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization", *Malaysian Technical Universities Conference on Engineering and Technology (MUCET),* 53, 453-462, 2012.

[15] Cristianini, N. and Shawe-Taylor, J., "An introduction to support vector machines and other Kernel-Based learning method 1st ed", *Cambridge University Press*, 2000.

[16] Bo Pang and Lillian Lee, "Opinion mining and sentiment analysis", *Foundations and Trends in Information Retrieval,* **2(12),** 1-135, 2008.

[17] Sunidhi Dwivedi and Shama Parveen, "Document Level Opinion Mining of Reviews of Mobile Phone Companies", *International Journal of Innovative Research in Science,Engineering and Technology,* **5(6)**, 9718-9725, 2016.

[18] B. Pang, L. Lee, and S. Vaithyanathan,(2002),"Thumbs up? Sentiment classification using machine learning techniques" *In Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing* (EMNLP), 79-86, 2002.