Review of Sentiment Analysis using Naive Bayes and Neural Network Classifier

LINA L. DHANDE[1], DR. GIRISH K. PATNAIK[2]
[1]Research Scholar, Dept of Computer Engineering, SSBT's COET, Bambhori, Jalgaon, Maharashtra, India,
Email: dhanon19@gmail.com.
[2]Prof, Dept of Computer Engineering, SSBT's COET, Bambhori, Jalgaon, Maharashtra, India,
Email: sanjay_gharde358@yahoo.com.

Abstract: Nowadays sentiment analysis is active field of research, to extract people's opinion about particular product or service. The most useful application of sentiment analysis is the sentiment classification of product reviews. The task of sentiment classification is to classify reviews of user as positive or negative from textual information alone. For that purpose many researchers used machine learning technique such as Naive Bayes classifier with strong independence assumption. But Naive Bayes classifier lack in accuracy for many complex real-world situations where there exists dependency among features. Further, the Neural Network with appropriate network structure can handle the correlation/dependence between input variables. So using Naive Bayes classifier combining with Neural Network on standard dataset will improve accuracy and performance of sentiment classification. This implies that a highly accurate and fast sentiment classifier can be used for sentiment classification.

Keywords: Sentiment Analysis, Sentiment Classification, Naive Bayes, Neural Network Classifier.

I. INTRODUCTION

In the literature, varieties of approaches to sentiment analysis have been proposed. Sentiment analysis, which is also called opinion mining, is the field of study that analyzes people's opinions, sentiments, evaluations, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, movies and topics. In early research identifying the sentiment of overall document but recently social media provided large amount of data publically available on web, so organizations are increasingly interested in a population feelings toward its products. In simple words it is used to track the mood of the public. It uses natural language processing and data mining techniques to the problem of extracting opinions from text. Sentiment analysis can be performed at different levels of granularity with different levels of detail described in [1]. The level of detail typically goes into determining the polarity of a sentence.

The sentiment classification is most useful application of sentiment analysis for classify user reviews. This sentiment classification can be categorised into positive, negative. Users express opinions through review sites such as Amazon, Internet Movie Database(IMDB) and Epinions, as well as through blogs, discussion forums, peer-to-peer networks, user feedback and comments, and various types of social network sites. These kinds of online media have resulted in large quantities of textual data containing opinion and facts. This has led to the development of sentiment analysis and classification systems. Sentiment analysis and classification are technically challenging [2]. The web is a huge repository of structured and unstructured data. Most of the people give feedback or comment for particular site for example product or rating for movie. Also people giving comment for actor, director, film etc. The analysis of this data to extract latent public opinion and sentiment is a challenging task. Content growth in the Internet in recent years has made a huge volume of information available.

Measurement of sentiment in open-source information is currently an active area of research. This information is presented in different formats such as posts, news articles, comments, and reviews. Especially in the automotive, electronics and film sectors, customers have written reviews about products or their features. By collecting and analyzing these reviews, new customers find others opinion about different features of the product. They can compare the products to each other to find the best one that meets their needs. Moreover, Manufacturers will find out strengths and weaknesses of their products or those of their competitors. In this way, manufacturers will solve the reported problems and use the business intelligence behind the analysis for future investments. Movie sentiment classification has also drawn the interest of researchers [3] [4]. People express their views for the movies, for example I like movie and the story is fantasying. Research area of sentiment analysis motive for doing sentiment classification using new

combination of classifier for improving accuracy. The paper proceeds as follows. Related work is reviewed and discussed to provide the theoretical background and foundation for study. In next section describe the data and provide discussion on sentiment analysis. Finally conclude the paper by discussing study of sentiment analysis.

II. RELATED WORK

Sentiment Analysis is a task of natural language processing and information extraction that aims to obtain writer's feelings expressed in positive or negative comments, reviews, questions and requests, by analyzing a large numbers of documents. Generally, sentiment analysis aims to determine the attitude of a speaker or a writer with respect to some topic. In recent years, the exponential increase in the Internet usage and exchange of public opinion is the driving force behind Sentiment Analysis today. DENG et al.(2011) uses sentiment analysis for Stock Price Prediction in [5]. Sentiment classification is the task of determining whether a user reviews is positive or negative which is similar to the traditional binary classification problem. The classifier tries to classify the review into positive category or negative category. The classification result will be the basis of the rating. With the proportion of positive and negative reviews, the system could provide the rating information to end users. Some of the earliest works in Sentiment Classification were done by Pang and Lee[3].

Bo Pang et al.,[3] present sentiment classification using machine learning techniques. For the effectiveness of classification of documents by overall sentiment used learning methods Naive Bayes, maximum entropy classification and support vector machines. The human produced baseline for sentiment analysis was outperformed by the machine learning techniques, which experimentally demonstrated on movie review data. The unigrams and bigrams features were used for classification. The movie-review corpus with randomly selected positive sentiment and negative sentiment reviews were used for experimental setup. Whereas the accuracy achieved in sentiment classification is much lower when compared to topic based categorization. Claster et al., [4] proposed that microblogs that can be used as a source of sentiment expression which analyze twitter posts. Recently released movies for sentiment towards were collected. The sentiment analyzed on an organic multidimensional space derived from Self Organizing Maps (SOM) instead of evaluating the sentiment across a unidimensional scale (positive/negative). In experiment calculated a SOM and use it as a model to visualize sentiment.

Pimwadee Chaovalit and Lina Zhou, [6] proposed movie review mining using machine learning and semantic orientation. To classify the movie reviews supervised classification and text classification techniques are used in proposed machine learning approach. A corpus is formed to represent the data in the documents and all the classifiers are trained using this corpus. Thus, the proposed technique is more efficient. Though, the machine learning approach uses supervised learning, the proposed semantic orientation approach uses "unsupervised learning" because it does not require prior training in order to mine the data. Experimental results showed that the supervised approach achieved 84.49% accuracy in three-fold cross validation and 66.27% accuracy on hold-out samples. William Simm et al., [7] investigated short text comments classification by sentiment and actionability. Experimentally contributes an independent comparative analysis of sentiment and actionability estimation accuracy, applying the chosen methods without additional data preprocessing. VoiceYourView is novel in that allows unstructured voice and text input so that users can comment on anything. VoiceYourView shares some similarities with social media commenting systems such as Twitter and Facebook, where users are free to comment on anything and everything. Four automatic methods of analysis are discussed for sentiment analysis including a tagger based approach, a naive Bayesian classifier, the ReadMe tool and a lexicon and rule method.

Tao Xu et al., [8] used online forum for sentiment analysis. The online forum like BBS provides a communication platform for people to discuss and express their views. Back-Propagation Neural Network (BPNN) used for extracted feature and Vector Space Model (VSM) used for represented text documents as vectors which is an algebraic model. Experiment had shown three layer neural networks to predict hotness of a topic. The neural network consisted the input layer has four neurons, hidden layer has eight neurons, and output has one neuron. Approach is divided into data collected and preprocessing, neural network training and prediction. SINA reading forum is used for collected data. Experiment effectively predicts the hotness of topics. Duan et al., [9] presented mining online user reviews for both quantitative aspects and textual content from multi-dimensional perspectives. In recent years online user generated content exploded which revolutionizing the hotel industry. Online user generated reviews for the hotel industry used as data source. Experiment had shown the results of sentiment analysis with increasing accuracy in measuring and capturing service quality dimensions. That was compared with existing text mining studies. After used econometric modeling technique to examine the potential differential effect of different service quality dimensions.

III. PROPOSED SYSTEM

A modern approach towards sentiment classification is to use machine learning techniques which inductively build a classification model of a given set of categories by training several sets of labeled document. Popular machine learning methods include Naive Bayes, K-Nearest neighbour, and Support Vector Machines. Among them, Naive Bayes classifier is more appropriate to be extended. The basic of Naive Bayes and Neural Network Classifier as discuss below.

A. Naive Bayes Classifier

Naive Bayes Classifier is a simple model for classification. It is simple and works well on text classification. It is a probabilistic classifier based on applying Bayes' theorem with strong independence assumptions. This is the simplest form of Bayesian Network, in which all attributes are independent given the value of the class variable. This is called conditional independence. It assumes each feature is conditional independent to other features given the class.

The Bayes rule is

$$P(c|t) = P(c) * P(t/c) / P(t) \qquad (1)$$

Where, c is a specific class and t is text want to classify. P(c) and P(t) is the prior probabilities of this class and this text. P(t|c) is the probability the text appears given this class. Here, the value of class c might be POSITIVE or NEGATIVE, and t is just a sentence. The goal is choosing value of c to maximize P(c|t). From numerical based approach group, Naive Bayes has several advantages such as simple, fast and high accuracy. Naive Bayes classifier for classification or categorization of text using word attributes that appear in a document as a basis for classification. Some research are done that although the assumption independence between words in a document is not fully met, it is obvious that the conditional independence assumption is rarely true in most real-world applications.

B. Neural Network Classifier

Neural networks have emerged as an important tool for classification. The recent vast research activities in neural classification have established that neural networks are a promising alternative to various conventional classification methods. The Neural Network with appropriate network structure can handle the correlation/dependence between input variables. The advantage of neural networks lies in the following theoretical aspects. First, neural networks are data driven self-adaptive methods in that they can adjust themselves to the data without any explicit specification of functional or distributional form for the underlying model. Second, they are universal functional approximates in that neural networks can approximate any function with arbitrary accuracy. Since any classification procedure seeks a functional relationship between the group membership and the attributes of the object, accurate identification of this underlying function is doubtlessly important.

Neural networks [4] have been successfully applied to a variety of real world classification tasks in industry, business and science. Many researchers have done sentiment classification by using Naive Bayes classifier. But Naive Bayes classifier has major limitation that the real-world data may not always satisfy the independence assumption among attributes. Hence, it affects the accuracy of Naive Bayes classifier. Combining Naive Bayes Classifier with Neural Network will improve the accuracy and performance of sentiment classification in real world

dataset. Naive Bayes Nerual Classifier is proposed for classify the standard movie review dataset. In which, work on the Sentence Level Analysis that means classify sentence in more than one dimension. For example Movie was very good and music so rocking. This sentence cover two dimension first one is movie and other is music. Also user searches particular film by querying in given dataset. For example query on search best movie with rocking songs. The main objective of paper is

- To analyze the movie review datasets.
- To extract features based on standard dataset and use it for sentiment classification.
- To classify movie review uses Naive Bayes Neural Classifier.

A straightforward approach to overcome the limitation of Naive Bayes is to combine with Neural Network to represent explicitly the dependencies among attributes.

IV. DISCUSSION

To make a decision, opinions are very important for anyone. From others user reviews about something like movie or product it will be easy to take decision. Analysis the sentiment of review is classified in positive or negative polarities. Sentiment analysis plays vital role in Businesses and organizations. The overview of the work done in the task of Sentiment Analysis with Naive bayes and Neural Network is shown in Table I. This table represents a sample of work done and some works published on the topic of Sentiment Analysis. It is evident from the Table I, as far as the data source is concerned, a lot of work has been done on movie and product reviews. Internet Movie Database (IMDb) is used for movie reviews. Movie review mining is a more challenging application than many other types of review mining. The challenges of movie review mining lie in that factual information is always mixed with real-life review data and ironic words are used in writing movie reviews.

As per summarized survey in Table I, Pang et al. classify movie review at document level which sometimes not sure review is positive or not. Here proposed sentence level analysis which will analyze positive or negative review and also classify more dimensions present in one sentence of movie review. Naïve Bayes is a very simple probabilistic model that tends to work well on text classifications and usually takes orders of magnitude less time to train when compared to models like support vector machines. But in real world data not always independence assumption is succeed there are some cases where real world complex relationship is required. Neural Networks are powerful techniques for representing complex relationships between inputs and outputs.

V. CONCLUSION

In this paper it is seen that sentiment analysis play vital role to make decision like movie, product etc. Sentiment analysis can be applied to a broad domain to classifying and

summarizing review and prediction. However, finding sentiment rating can still be a difficult task because there are a large number of diverse opinions of user. Due to its tremendous value for practical applications, there has been an explosive growth of both research in academia and applications in the industry. Various works done in sentiment analysis are discussed with techniques. Naïve bayes is simple and fast classifier but some dependence assumption are not covered, so the accuracy of sentiment analysis can be improved by combining neural network classifier with naïve bayes classifier.

TABLE I: LITERATURE REVIEW OF SENTIMENT ANALYSIS WITH NAIVE BAYES AND NEURAL NETWORK CLASSIFIER.

Studies	Technique	Data Source	Limitation
Pang et al.,2002 [3]	Naive Bayes, SVM, Maximum Entropy	Movie Review	It works only on document level analysis
Chaovalit and Zhou,2005 [6]	N-gram Classifier and Semantic Orientation	Movie Review	Semantic orientation gives result for small sample data size
Simm et al.2010 [7]	Naive Bayes, ReadMe, Tagger	VoiceYourView	Sentiword dictionary is required
Xu et al., 2010 [8]	Backpropagation Network	BBS Hotness Topic	It calculates Sentiment value for whole document
Claster et al., 2011 [4]	Artificial Neural Network	Movie Review	SOM require more visualizing location for accurate result
Duan et al., 2012 [9]	Naive Bayes, SERPERF Model	Hotel Review	Sentence level analysis may not cover more than one dimension
Mesuf KAYA et al. 2012 [10]	Naive Bayes, SVM,Maximum Entropy	Turkish Political News	Less accuracy
Tseng et al. 2012 [11]	Naive Bayes Classifier	Twitter	Independence assumption for real world data

VI. REFERENCES

[1] Nilesh M. Shelke, Shriniwas Deshpande, Vilas Thakre, "Survey of techniques for opinion mining", International Journal of Computer Applications, vol. 57, no. 13, pp. 30-35, November 2012.

[2] Bakhtawar Seerat, Farouque Azam, "Opinion mining: Issues and challenges(A survey)", International Journal of Computer Applications, vol. 49, no. 9, pp. 42-51, July 2012.

[3] Bo Pang, Lillian Lee, Shivakumar Vaithyanathan, "Thumbs up?Sentiment classification using machine learning techniques", Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 79-86, July 2002.

[4] William B. Claster, DINH Quoc Hung and Subana Shanmuganathan, "Unsupervised Artificial Neural Nets for Modeling Movie Sentiment", Second International Conference on Computational Intelligence, Communication Systems and Networks, pp. 349-354, 2010.

[5] Shangkun DENG, Takashi MITSUBUCHI, Kei SHIODA, Tatsuro SHIMADA and Akito SAKURAI, "Combining Technical Analysis with Sentiment Analysis for Stock Price Prediction", Ninth IEEE International Conference on Dependable, Autonomic and Secure Computing, pp. 800-807, 2011.

[6] 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, pp. 1-9, 2005.

[7] William Simm, Maria-Angela Ferrario, Scott Piao, Jon Whittle and Paul Rayson, "Classification of Short Text Comments by Sentiment and Actionability for VoiceYourView", IEEE International Conference on Social Computing / IEEE International Conference on Privacy, Security, Risk and Trust, pp. 552-557, 2010.

[8] Tao Xu, Ming Xu and Hong Ding, "BBS Topic's Hotness Forecast Based on Back-Propagation Neural Network", International Conference on Web Information Systems and Mining, pp. 57-61, 2010.

[9] Wenjing Duan, Qing Cao, Yang Yu and Stuart Levy, "Mining Online User-Generated Content: Using Sentiment Analysis Technique to Study Hotel Service Quality", 46th Hawaii International Conference on System Sciences, pp. 3119-3128, 2013.

[10] Mesut KAYA, Guven FYDAN, Ismail H. Toroslu, "Sentiment Analysis of Turkish Political News", IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, pp. 174-180, 2012.

[11] Chris Tseng, Nishant Patel, Hrishikesh Paranjape, T Y Lin and SooTee Teoh, "Classifying Twitter Data with Naive Bayes Classifier," IEEE International Conference on Granular Computing, 2012.