

A Survey on Evaluating Sentiments by Using Artificial Neural Network

Pranali Borele¹, Dilipkumar A. Borikar²

¹ M.Tech. Student, ²Assistant Professor

CSE Department, Shri Ramdeobaba College of Engineering and Management, Nagpur, India

Abstract - Sentiment Analysis is the process of identifying whether the opinion or views expressed in a piece of work is positive, negative or neutral. It is also referred to as Opinion Mining. Opinion Mining and Sentiment Analysis have been sought after domains of research for past decade due to its potential to drive businesses. Sentiment Analysis has been widely used in classification of review of products and movie review ratings. The approaches to sentiment analysis can be categorized as lexicon-based, machine learning-based and hybrid. This paper reviews the machine learning-based approaches to sentiment analysis and brings out the salient features of techniques in place. The prominently used techniques and methods in machine learning-based sentiment analysis include - Naïve Bayes, Maximum Entropy and SVM. Naïve Bayes has very simple representation but doesn't allow for rich hypotheses. Also the assumption of independence of attributes is too constraining. Maximum Entropy estimates the probability distribution from data, but it performs well with only dependent features. For SVM may provide the right kernel, but lacks the standardized way for dealing with multi-class problems. For improving the performance regarding correlations and dependencies between variables, an approach combining neural networks and fuzzy logic is often used.

Key Words: Sentiment analysis, machine learning, neural network, natural language processing.

1. INTRODUCTION

Sentiment is a view or opinion that is held or expressed by the opinion holder. It involves use of Natural Language Processing (NLP), text mining and analysis, and computational linguistic to determine and retrieve subjective information from the given source of information (usually text). It reveals the attitude of a writer or a comment based on opinion polarity. The sentiments are usually classified under three types - positive, negative and neutral. Sentiment analysis is performed at different levels of granularity - word/word-phrase/aspect level, sentence level and document level.

Classification of product reviews based on sentiments of the buyer is among the practical applications of sentiment analysis. Users express opinions through review on social networking sites and product marketing sites, namely Amazon, Internet Movie Database (IMDB). Users also convey opinions through blogs, discussion forums, peer-to-peer, dialogs, user feedback, comments on social networking sites, viz., Twitter, Facebook, YouTube, . This has resulted in data majorly unstructured in huge quantity comprising of textual data containing opinion and facts. Text mining usually involves the process of structuring the input text, deriving patterns within the structured data, and finally evaluation and interpretation of the output. It is a real challenge to classify the review of users in text mining. Movie reviews are a good source for sentiment analysis and classification because authors can clearly express their opinion and authors can accompany by rating that makes it easier to train learning algorithms on this data.

The enormous upsurge of WWW has been driving the user communication through online reviews and feedback before the customer decides to buy a product. Thus WWW and online posting of opinionated text has become an integral part of online transactions and business. It has resulted in WWW being the huge repository of highly unstructured data from different sources like blogs, videos, forums, etc. comprising the views of the opinion holder.

Numerous web sites offer reviews of such items like books, cars, snow tires, vacation destinations, movies, etc. Consumer usually checks opinions about the product before buying the product or checks reviews about that product. Manufacturer can get details about its products strength and weaknesses based on the sentiment of the customers. These opinions are very helpful for both business purpose and individuals. To analyze and summarize the opinions expressed by an author about particular product or review data is an appropriate domain for researchers. This new research approach is called Sentiment Analysis or Opinion Mining. It is also frequently referred to as subjectivity analysis, and appraisal extraction.

Sentiment classification is the part of the opinion mining and that is used for identification of opinions and disagreement in a given text. Its main aim is to find the "like" or "dislike" statements dealing with "positive", "negative" or "neutral" sentiment in the comments or review. Many of the existing research based on mining and retrieval of factual information not on opinions. Sentiment analysis is carried out through classification by using new combination of

classifier for improving more accuracy. A modern approach towards sentiment classification employs machine learning techniques. These include a classification model of a given set of categories by training several sets of labeled document. Popular machine learning methods are Naïve Bayes, K-Nearest Neighbor, Support Vector Machines and Neural Network. Among these, Naive Bayes have been experimented heavily. Besides being simple in implementation and high on accuracy, Naïve Bayes classifier has major limitation that the real-world data may not always satisfy the independence assumption among attributes. This affects the accuracy of Naïve Bayes classifier.

This work proposes to effectively classify movie review in positive and negative polarities and to increase the accuracy of sentiment analysis. It incorporates neural network and fuzzy logic. Neural network is well trained for handling the correlations an inter dependencies.

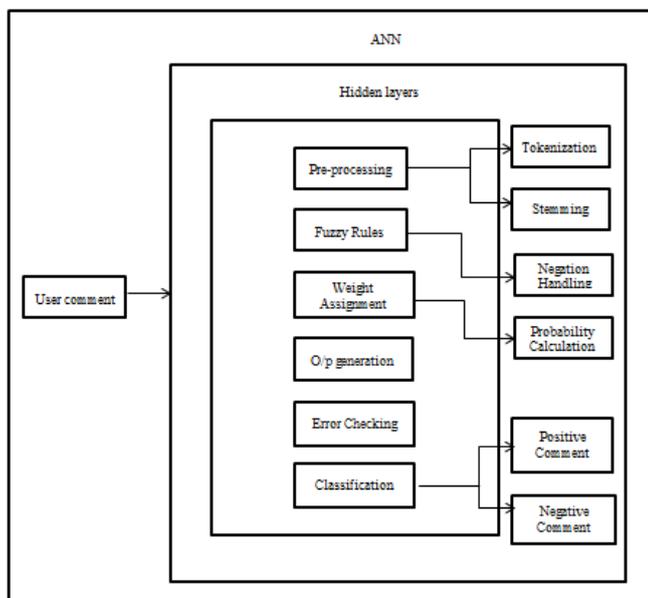


Fig 1. Framework of Sentiment Analysis

The figure 1 shows the framework for Sentiment Analysis. The input is the comment entered by the user for the movie review. The dataset used here is IMDB review dataset. Following sub-section gives the dataset description. Section 2 discusses the current system of Sentiment Analysis. Section 3 describes different approaches for Sentiment Analysis. Section 4 consist the Literature survey. Section 5 gives the brief idea about the proposed system.

Dataset Description

This dataset contains movie reviews along with their associated binary sentiment polarity labels. It is intended to serve as a benchmark for sentiment classification. The core dataset contains 50,000 reviews split evenly into 25k train and 25k test sets. The overall distribution of labels is balanced (25k positive and 25k negative). We also include an additional 50,000 unlabeled documents for unsupervised learning.

2. AN OVERVIEW OF SENTIMENT ANALYSIS OF CURRENT SYSTEM

Sentiment Analysis plays an important role in opinion mining. It is generally used when consumers have to make a decision or a choice regarding a product along with its reputation which is derived from the opinion of others. Sentiment analysis can reveal what other people think about a product. According to the wisdom of the crowd sentiment analysis gives indication and recommendation for the choice of product. A single global rating could change perspective regarding that product. Another application of sentiment analysis is for companies who want to know the review of customers on their products. Sentiment analysis can also determine which aspects are more important for the customers. Knowing what people think provides numerous possibilities in the Human/Machine interface domain. Sentiment analysis for determining the opinion of a customer on a product is a non-trivial phase in analyzing the business activities like brand management, product planning, etc.

Detecting author’s opinion is concerned with identifying pretend opinion from reviews. Sentiment analyses classify the polarity of a given text by expressing the author opinion as positive or negative. The sentiment classification process is carried out at following three levels.

1. Document Level: In this level the whole document is consider and is then classify text as positive or negative. But in the case of forums or blogs, comparative sentences may appear. Customers may compare one product with another that has similar feature and hence document level analysis is not desirable in forums and blogs. The challenge in the document level classification is that the entire sentence in a document may not be relevant in expressing the opinion about an entity. Therefore subjectivity/objectivity classification is very important in this type of classification where unrelated sentences must be eliminated from the document.

2. Sentence Level: In this level sentences are classified as positive, negative or neutral. In case of simple sentences, a single sentence indicates a single opinion about an entity. But in presence of complex sentences in the opinionated text, sentence level sentiment classification is not appropriate. The advantage of sentence level analysis lies in the subjectivity/ objectivity classification.

3. Word or Phrase Level: Analysis of features of product for sentiment classification is usually called as word or phrase

or feature based sentiment analysis. It is having fine-grained analysis model among all other models.

Prominently approaches sentiment analyses are clustered into machine learning-based approaches and lexicon-based approaches. Although the hybrid approaches using earlier two approaches and statistical model is often used.

3. DIFFERENT APPROACHES FOR SENTIMENT ANALYSIS

Techniques for sentiment analysis may be categorized under machine learning methods, lexicon-based methods, rule-based methods, methods based on statistical model and others. These methods are elaborated as below –

The machine learning method: It incorporates machine learning algorithms to deduce the sentiment by training on a known dataset. This approach to sentiment classification is supervised and allows effective text classification. Machine learning classification necessitates two different sets of documents, namely for training and testing. A training set is used by an automatic classifier to learn and differentiate attributes of documents, and a test set is used to check the performance of the automatic classifier. There are many machine learning techniques adopted to classify the reviews. Machine learning techniques like Naïve Bayes (NB), maximum entropy (ME), and support vector machines (SVM) have achieved great success in text categorization.

The lexicon-based approach: It deals with computation of sentiment polarity for the input text (such as blog, review, comment, etc.) using subjectivity and opinion orientation of the text. By using the semantic orientation of words or sentences in the review, the lexicon-based approach calculates sentiment polarity for the review. Lexicon Based techniques mostly work on an assumption that the polarity of a document or any sentence is the sum of polarities of the individual words and/or phrases.

The rule-based approach: These approaches incorporate semantic dictionaries. The process runs into creating dictionary for polarity, negation words, booster words, idioms, emoticons, mixed opinions etc. The rule-based approach focuses on opinion words in a document and then classifies the text as positive or negative. The machine-learning based classifier is significantly better than rule based approach. The main advantage of the rule-based approach is that no training phase is required. The rule-based approach fails to ascertain the polarity of the text

when the number of positive words and the number of negative words are equal.

Statistical model: Statistical models treats each review as a mixture of latent aspects and ratings. It is assumed that aspects and their ratings can be represented by multinomial distributions and the head terms may be grouped into aspects and sentiments providing proper ratings. Many approaches rely on statistics-based machine learning techniques for sentiment analysis. A multiclass sentiment analysis problem can be addressed by combining statistics-based method with sentiment lexicon.

4. LITERATURE REVIEW

Artificial Neural Network with NLP focuses on sentiment analysis using MLPs. It includes the removal of human efforts in sentiment analysis. It eliminates manual pre-processing or feature selection work. The artificial neural network gives good speed and accuracy [1].

Sentiment analysis involves machine learning, semantic orientation, negation handling and feature-based sentiment classification. The Semantic orientation approach to Sentiment analysis is an unsupervised learning technique as it does not require any prior training in order to mine the data. It measures the likelihood of an aspect word of being positive or negative. When compared to the semantic orientation approach with the N-gram model-based technique for movie review, it is observed that the machine learning approaches are more accurate, although they require a significant amount of time to train the model [2]. The semantic orientation approach is less accurate but is more efficient to use in real-time applications. The performance of semantic orientation also depends on the performance of the underlying POS tagger. Negation is a very common grammar construct that affects polarity and therefore, should be adequately dealt with in sentiment analysis.

The notion of context may assist in classification of text to verify whether the expressed opinion in the text is positive or negative or neutral. It determines the overall topic of the given text. Context which do not depend on sentences and have implicit meaning in the text are also considered in polarity classification. The framework using context is elaborated in [3] and provides for better accuracy.

Self-Organizing Maps (SOM) have been incorporated in novel models for analyzing sentiment based on visualization and classification of online review. SOM algorithm has been

implemented for both supervised and unsupervised learning from text documents [4]. The supervised learning algorithms (LVQ, OLVQ1, and Multi-pass LVQ) have found to perform better than unsupervised learning algorithm for sentiment analysis. The results with online movie review dataset show that SOMs approaches are better candidates for sentiment based classification and sentiment polarity visualization.

Sentiment classification involving recognition of the emotion of text based communication from the text with negative sense is vital preventing the harmful damage [5]. Neural Network (NN) based index combines the advantages of machine learning techniques and information retrieval (semantic orientation indexes). Such an approach will help companies in detecting harmful negative bloggers comments quickly and effectively.

An approach to prediction of future value on the basis of available previous values is discussed in [6]. It uses ANN algorithm to train the available previous day's data and predict the value for the next day. This prediction has been done for Bombay Stock Exchange. In the recent years Genetic Algorithm and Rough Set Theory have been used for sentiment analysis along with NLP techniques. Machine Learning Algorithms are independent of the filtration process. For evaluating the features wrapping technique can be used along with learning algorithms [13, 14].

The major challenge in sentiment analysis in social media is scalability. The micro blogging which helps to resolve the problem of scalability An approach implementing SSSA (Semantic Scoring Sentiment Analysis) service is proposed by [15] to address the problem of scalability and to efficiently process large volume of micro blogs [15].

Deep Learning is an important aspect of sentiment analysis. It shows better results which helps to gain the efficiency in Sentiment Analysis. The approach based on deep Learning uses a Convolutional Neural Network (CNN) inspired through Deep Learning Algorithm. CNN uses *word2vec()* to get the vector representation of word as an input. This improves the performance of the network and helps in gaining the efficiency [16]. It also uses Parametric Rectified Linear Unit (PReLU), Normalization and Dropout technology which will increase the accuracy and maintain the generalizability of the model.

A model using Artificial Neural Network for sentiment analysis of movie reviews is proposed. This method classified the review into positive, negative and fuzzy tone [18]. The proposed approach combines the advantages of the machine learning techniques and the information retrieval techniques [19].

The experimentation using Naïve Bayes classifier and the ANN classifiers have revealed that although Naïve Bayes

classifiers are simple and sought after models for implementation, the ANN based classifiers outperforms them both on accuracy and precision as depicted in Table-1.

TABLE 1. ANN CLASSIFIER VERSUS NAÏVE BAYES [21].

Approach	Accuracy	Precision	Recall
ANN	93	93.2	93
Naïve Bayes	88	88.2	88

6. CONCLUSIONS

Sentiment Analysis plays an important role formulating the buyer's decision. It finds its scope in the product or movie review domains. This paper introduces the breadth of sentiment analysis and brings out the specific techniques, methods and models to ascertaining sentiment orientation and its classification. In particular the focus is on the machine learning approaches and use of artificial neural networks (ANN) in sentiment classification and analysis. The review suggest that the ANN implementations would result in improved classification, combining the best of supervised and unsupervised methods.

REFERENCES

- [1] M. Ghiassia, J. Skinnerb, D. Zimbraa (2013), "Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network", Journal of Expert Systems with Applications, ELSEVIER Volume 40, Issue 16, 6266-6282.
- [2] Arturo Montejó-Ráez, Eugenio Martínez-Cámara, M. Teresa Martín-Valdivia, L. Alfonso Ureña-López (2014), "Ranked WordNet graph for Sentiment Polarity Classification in Twitter"- Journal Computer Speech and Language archive Volume 28 Issue 1, 93-107.
- [3] S.Suruthi, M.Pradeeba, A.Sumaiya, J.I Sheeba "Neural Network Based Context Sensitive Sentiment Analysis", International Journal of Computer Applications Technology and Research Volume 4- Issue 3, 188 - 191, 2015, ISSN:- 2319-8656.
- [4] Anuj Sharma, Shubhamoy Dey, "Using Self-Organizing Maps for Sentiment Analysis", CoRR (2013).
- [5] Soumith Chintala, "Sentiment Analysis using neural architectures", New York, NY 10012-2012.
- [6] G.Vinodhini, R. M.Chandrasekaran, "Sentiment Analysis and Opinion Mining: A Survey", Volume 2, Issue 6, June 2012, International Journal of Advanced Research in Computer Science and Software Engineering.
- [7] S.Suruthi, M. Pradeeba, A.Sumaiya, J.I Sheeba "Neural Network Based Context Sensitive Sentiment Analysis", International Journal of Computer Applications Technology and Research Volume 4- Issue 3, 188 - 191, 2015, ISSN:- 2319-8656.
- [8] Emma Haddia, Xiaohui Liua, Yong Shib (2013)" The Role of Text Pre-processing in Sentiment Analysis. International Conference on Information Technology and Quantitative Management" 27 Emma Haddi et al. / Procedia Computer Science 17 (2013) 26 - 32.

- [9] Long-Sheng Chen, Hui-Ju Chiu, "Developing a Neural Network based Index for Sentiment Classification", Proceedings of the International MultiConference of Engineers and Computer Scientists 2009 Vol I IMECS 2009, March 18 - 20, 2009, Hong Kong.
- [10] Sunil Kumar Khatri, Himanshu Singhal, Prashant Johri, "Sentiment Analysis to Predict Bombay Stock Exchange Using Artificial Neural Network" 978-1-4799-6896-1/14/\$31.00 ©2014 IEEE .
- [11] Asad Bukhari, Um-e-Ghazia and Dr. Usman Qamar "Critical Review of Sentiment Analysis Techniques", Proceeding of the International Conference on Artificial Intelligence and Computer Science(AICS 2014) 15 - 16 September 2014, Bandung, INDONESIA.
- [12] Azuraliza Abu Bakar, Mohd Ridzwan Yaakub "Metaheuristic Algorithms for Feature Selection in Sentiment Analysis", Science and Information Conference 2015 July 28-30, 2015 | London, UK .
- [13] H. Liu and L. Yu, "Toward Integrating Feature Selection Algorithms for Classification and Clustering," IEEE Trans. Knowl. Data Eng., vol. 17, no. 4, pp. 491-502, 2005.
- [14] Y. Saeys, I. Inza, and P. Larrañaga, "A Review of Feature Selection Techniques in Bioinformatics," Bioinformatics, vol. 23, no. 19, pp. 2507-2517, 2007.
- [15] Asso Hamzehei, Mohammad Ebrahimi, Elahe Shafiei Bavani, "Scalable Sentiment Analysis for Microblogs based on Semantic Scoring" , IEEE International Conference on Services Computing (IEEE SCC), At New York, USA
- [16] Xi Ouyang, Pan Zhou, Cheng Hua Li, Lijun Liu "Sentiment Analysis Using Convolutional Neural Network" ,2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and computing .
- [17] Anais Collomb , Crina Costea , Damien Joyeux and Omar Hasan and Lionel Brunie- "A Study and Comparison of Sentiment Analysis Methods for Reputation Evaluation".
- [18] G. Vinodhini and RM. Chandrashekharan, "Sentiment Analysis and Opinion Mining : A Survey" ,International Journal of Advanced Research in Computer Science and Software Engineering, Volume2, Issue 6, June 2012.
- [19] Dudhat Ankitkumar, "A Survey on Sentiment Analysis and Opinion Mining", International Journal of Innovative Research in Computer and Communication Engineering Vol. 2, Issue 11, November 2014.
- [20] Zohreh Madhoushi, Abdul Razak Hamdan and Suhaila Zainudin "Sentiment Analysis Techniques in Recent Works", Science and Information Conference (SAI), 2015. IEEE, 2015.
- [21] Amit Kumar Sharma, Sudesh Kumar Prajapat, "A Comparative Study between Naïve Bayes and Neural Network for Spam Email Detection", International Journal of Computer Applications@ (IJCA) (0975 - 8887) National Seminar on Recent Advances in Wireless Networks and Communications, NWNC-2014.