



International Journal of Control Theory and Applications

ISSN : 0974-5572

© International Science Press

Volume 10 • Number 19 • 2017

A Review on the Applications of Sentimental Analysis using Machine Learning

Sudharshan Duth P.¹ and Vinayak Ashok Kulkarni¹

¹ Dept. of Computer Science, Amrita School of Arts and Sciences Amrita Vishwa Vidyapeetham, Amrita University, Mysuru Campus, Karnataka, India, Emails: sudhisp@gmail.com, vinayakkulkarni6600@gmail.com

Abstract: Being human it requires little knowledge to classify the sentences by just understanding the tonality and usage of words in the text. The given text may be positive, negative or a neutral sentence which is very significant to know. Computer program cannot easily classify the text which is a challenging task. However, in some situations it is hard for a human to categorize the text. Opinion Mining can be used for the above stated problem which is a sub domain of Data Mining. It is observed that various Machine Learning techniques have supplemented towards Sentimental Analysis.

Keywords: Machine Learning, Data Mining, Sentiment Analysis, Opinion Mining.

1. INTRODUCTION

Opinions are very significant in each and every field of work, still it is crucial to classify the opinions into negative, positive or neutral type of opinion. It is very easy for a human to examine a text and to anticipate the sentiment of the applied text, by realizing how the words are used in the text or by realizing the circumstance of the applied text. In order to classify the opinions we can use a technique called Opinion Mining.

These opinions play a significant role in market for the product review analysis and also in understanding whether a given text is in positive or negative circumstance for sentimental analysis of the text which is available on internet. The opinion also plays a significant role in politics, where the politicians can understand the people's views on them, so that by analyzing the opinions of people, accurate future consequences can be predicted. Many different experiments are being carried out using Naive Bayes classifier, support vector machines and maximum entropy classifier, but still it is a complicated problem to classify the sentiments in given text.

2. LITERATURE SURVEY

In the literature survey that is done, it is found that many researchers proposed classification methods. Bo Pang et al., (2002) [1] examined about the effectiveness of applying machine learning methods to the sentiment

classification problem. For the experimentation, the movie reviews are considered. The stated domain is said to be experimentally convenient because of the presence of large on-line collections of such reviews, and because reviewers often summarize the overall sentiment with any of the machine-extractable rating indicator, such as a number of stars; the results which were produced by the machine learning techniques were stated to be quite good when compared with the human generated results. In terms of relative performance, it was stated that SVM was better in performance to Naive Bayes. Bo Pang et al., (2004) [2] proposed a novel machine learning methodology that considers subjective portions of the document and text categorization techniques. Extracting these portions, there is a mention that for finding minimum cuts in graphs. This paper [3] mainly concentrates on providing solutions to the new problems raised by sentiment applications. Rohini et al., (2009) [4] focus on opinions which are expressed by the reviewers on product and propose Opinion Miner system to analyze inspections on a product by the customers. Products are classified into negative and positive. The framework of lexicalized HMMs is a novel machine learning approach which is different from existing approaches. The online reviews from Amazon and publicly accessible datasets are processed in the estimation of the proposed method. Stefano Baccianella et al., (2010) [5] explains about SENTIWORDNET 3.0, is an enhanced lexical resource which is developed explicitly for supporting opinion mining and sentiment classification applications. Andrew et al., (2011) [6] presented a model that captures both sentiment and semantic similarities among words. Via an unsupervised probabilistic model of documents the word vectors are learnt and is stated in their model of semantic component. The model is extended with a supervised sentiment component that is capable of embracing many attitudinal and social aspects of meaning. To predict the sentimental observations on contexts in which the words appear this component of the modules are in the vector representation of words. This results in words expressing similar sentiment to have similar vector representations. Liu et al., (2012) [7] defined the opinion mining problem. The work describes diverse key mining tasks that have been studied in their representative techniques and their research literature. There is a discussion about the issue of detecting fake reviews. Finally, there is a mention of the research topic of assessing the quality or utility of online reviews. Trivedi Khushboo et al., (2012) [8] focus on sentence level classification to analyze and it is done through term counting based approach. The altered algorithm in terms of the parameter passing Naive Bayes is used for getting better accuracy. Abhishek Tiwari et al., (2012) [9] concentrates on a unique approach for opinion mining on comments section from blog, which uses the approach TF-IDF through RSS feed. This work proposes an altered TF-IDF algorithm for finding out relation of a comment with the subject and scoring is given for the terms which is based on opinion and their order in the comments. Hoeber et al., (2013) [10] explored that the Twitter has become precious resource of data. Twitter made the communication and collaboration between fans and organizers easy. In understanding the behavior of fans Twitter plays a vital role. This work is done to analyze temporally changing sentiment within Twitter, which is represented in visual analytics approach. Timeline is used to represent the individual tweets and interactional tools are used for giving analytical reasoning about the data.

Hemalatha et al., (2013) [11] introduced a sentiment analysis tool, which comprises of three functions: sentiment analysis among tweets, finding neutral, positive and negative tweets from information resources. This tool focuses on analyzing tweets from the media sites, hence provide a way in the future to find out technology trends. The Naive Bayes classifier and Maximum entropy are used in the sentiment analysis tool. Arti Buche et al., (2013) [12] surveyed and analyzed, various techniques of opinion mining that have been developed for the primary tasks. There is a mention of an overall picture of what is involved in expanding a software system for opinion mining on the basis of their survey and analysis. Vivek Narayanan et al., (2013) [13] explains about Naive Bayes model which is more accurate and fast, sentiment classifier than other complex classifiers. Ritu Sindhu et al., (2014) [14] tried to demonstrate the idea of classifying and clustering opinion mining experiment on analysis of web blog posts on recent product policy and services reviews. The work focuses on a novel approach for analyzing the customer opinions. Minanovic et al., (2014) [15] explains about the text related data like word documents, email and posts on LinkedIn and Facebook. Twitter tweets and online reviews are analyzed in this work. Enforcement of sentiment analyses is more complicated on tweets when compared with the online

reviews. The tweets in twitter consists of twitter slang, sarcasm and analogy which is not clearly understood, where as online reviews will be written clearly and grammatically more precise manner. This work uses KNIME technique for analyzing online reviews and tweets.

Pablo Gamallo et al., (2014) [16] explains about detecting the polarity of English tweets based on Naive Bayes classifier. Classification is done with polarities like negative and positive. This work mainly focuses on determining the polarity of English tweets. Polarity lemmas are used within the tweet to determine whether the tweet is with or tweet is without polarity. Neutral value will come when there is no lemma in the tweet. Medhat et al., (2014) [17] explains about the feature selection and sentiment classification approaches depicting sentiment analysis. Feature selection is also explained in this paper. Features like Terms presence and frequency, Parts of speech(POS), Opinion words and phrases, Negations are used in sentiment classification. The techniques used for sentiment classification are Machine Learning, Lexicon based approach and hybrid approach. Ravi et al., (2015) [18] have found various applications in banking, financial, service, and insurance sector. In this regard, this work focuses on analyzing customer reviews on the basis of sentiment score. Based on analysis of a set of credible text reviews collected on 270 training programmes posted by 2688 participants in an organization. In order to evaluate the efficacy of the proposed approach, calculate correlation coefficient between sentiment score obtained from the unstructured reviews and the numerical rating assigned by the participants. Furthermore, the work employs the visualization analytics to get insights into different aspects of the programmes. Vinayak Hegde et al., (2015) [19] focuses on the improvement of the Naive Bayes classification algorithm with regards to the sentimental classification. It is clearly examined that the work concentrated on sentence level sentiment classification. The work focuses on the Naive Bayes model as the basis and on top of this model, there is inclusion of several other methodologies like effective negation handling, emphasizing words handling and feature extraction. There is a mention of the improvement of the overall accuracy measure of the sentimental classification algorithm. Aamera et al ., (2015) [20] focus on a new entity level sentiment analysis approach for Twitter. Lexicon based approach is adopted to perform entity level sentiment classification. This approach results in high precision and low recall. So, in order to improve the result of lexicon based approach additional tweets which are likely to be opinionated are recognized automatically by making use of information. Polarities are assigned for newly identified tweets through a trained classifier. Yadav (2015) [21] concentrates on sentiment classification based on polarity. Polarity are of three types positive, negative or neutral. And opinions are as classified into three types: direct opinions, comparative opinions and indirect opinions. There is a study of different levels of opinion mining like granularity, document level, and aspect level. Document level classification consists of Supervised and Unsupervised Methods in which supervised learning technique uses Naive Bayes and support vector machines. Dhanalakshmi et al., (2016) [22] focuses on the polarity of the student feedback using supervised learning algorithms. To process the student feedback, the application of a combination of natural language processing and machine learning are used. Various classification approaches have been compared for performance study. To train the algorithms for binomial classification data obtained from the survey is subjected to data preprocessing. Based on the obtained features like examination and teaching, the polarity of the student comment is predicted. It is then compared for different algorithms for better performance with respect to different evaluation basis. URL-1[23] Focuses on the problematic task of clustering of features regarding the online reviews. It is stated that simple methods for solving this problem are based on unsupervised learning using few of the distributional similarity. Then model for semi-supervised learning problem was created. To identify some labeled examples automatically the Lexical characteristics of the issue are exposed. Experimental evaluations show that by a large margin the intended method out performs the existing state-of-the-art methods.

3. POSSIBLE SOLUTIONS

In literature, sentiment analysis is done using different approaches for better results. In the work [1] SVM is used for sentiment classification and focuses on machine learning technique SVM which gave better results than Naive Bayes. Text classification [2] for the subjective portions was done using novel methodology of machine

learning that is applied for text categorization technique for the subjective parts of the document. The sentiment classification can be classified as positive, negative and neutral. There is a provision of polarity values [23] for the words as positive and negative polarities. One of the ways, to classify is through the framework of lexicalized HMMs [4] which was done to classify products as negative and positive product. Sentiment classification can classify words by capturing both semantic and sentiment similarities [6]. Opinion mining, which is also called as sentiment analysis, used to detect fake and spam reviews [7]. Opinion mining also uses Naive Bayes classifier and Maximum entropy [11] for analyzing tweets. For sentiment classification, altered algorithms are also used for improving accuracy. The parameter passing Naive Bayes algorithm [8] and optimized Naive Bayes [13] are altered for improvement in the accuracy. Another modified algorithm is TF-IDF algorithm [9], is used for detecting the opinion based on appearance of opinion based terms and thus, resulting in considerable amount of improvement in the accuracy.

After analyzing the literature, it is found that there is need of improvement in sentiment classification approaches which is quiet an open area for research. In our literature, it is found that SVM and Naive Bayes gave better results among all the mentioned algorithms. There is considerable improvement in the accuracy when SVM and Naive Bayes used for sentiment classification.

4. CONCLUSION

Sentiment analysis is wide area of research with plenty of challenges. Every word in a given text is associated to the alongside or any other associated words. Sentiment analysis assists in classifying, condensing reviews and in real time applications. This paper mainly focuses on review of sentiment classification and classification approaches. As noticed in the literature done on the topic sentiment classification, we can infer that the SVM and Naive Bayes can be used as the classification algorithms. It is observed that the accuracy of sentiment classification can be improved using feature selection. For the easy understanding of output, data visualization techniques can be included.

REFERENCES

- [1] Bo Pang and Lillian Lee ShivakumarVaithyanathan, "Thumbs up?Sentiment Classification using Machine Learning Techniques," Conference on Empirical Methods in Natural Language Processing (EMNLP), Volume 10, pp. 79-86, 2002.
- [2] Bo Pang and L. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," In Proceedings of the ACL, pp. 271, 2004.
- [3] Bo Pang, "Opinion Mining and Sentiment Analysis," pp. 1-135, 2008.
- [4] Jin, Wei, Hung Hay Ho, and Rohini K. Srihari, "OpinionMiner: a novel machine learning system for web opinion mining and extraction," Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp. 1195-1204, 2009.
- [5] Stefano Baccianella, Andrea Esuli, FabrizioSebastiani, "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining," In Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10), Vol. 10, pp. 2200-2204, 2010.
- [6] Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts, "Learning Word Vectors for Sentiment Analysis," The 49th Annual Meeting of the Association for Computational Linguistics (ACL 2011), Volume 1, pp. 142-150, 2011.
- [7] Liu, B., & Zhang, L., "A survey of opinion mining and sentiment analysis," In Mining text data, Springer US, pp. 415-463, 2012.
- [8] Khushboo, T. N., Vekariya, S. K., & Mishra, S., "Mining of Sentence Level Opinion Using Supervised Term Weighted Approach of Naïve Bayesian Algorithm," International Journal of Computer Technology and Applications, 2012.
- [9] C AbhishekTiwari, KshitijPathak, Upasaniwari, Rupam das, "Opinion Polarity Detection in Blog Comments from Blog

- Rss Feed by Modified TF-IDF Algorithm,” *International Journal of Engineering Research and Applications (IJERA)*, Vol. 2, pp. 412-416, 2012.
- [10] Hoeber, L., Wood, L., Snelgrove, R., Hugel, I., & Wagner, D., “Visual twitter analytics: exploring fan and organizer sentiment during Le Tour de France,” 2013.
- [11] Hemalatha, Dr. G. P SaradhiVarma, Dr. A.Govardhan, “Sentiment Analysis Tool using Machine Learning Algorithms,” *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*, Volume 2, pp. 105-109, 2013.
- [12] ArtiBuche, Dr. M. B. Chandak, AkshayZadgaonkar, “OPINION MINING AND ANALYSIS: A SURVEY,” *International Journal on Natural Language Computing (IJNLC)*, Vol. 2, 2013.
- [13] Narayanan, V., Arora, I., & Bhatia, A., “Fast and accurate sentiment classification using an enhanced Naive Bayes model,” In *International Conference on Intelligent Data Engineering and Automated Learning*, Springer Berlin Heidelberg, pp. 194-201, 2013.
- [14] Minanovic, A., Gabelica, H., & Krstic, Ž., “A Novel Approach for Sentiment Analysis and Opinion Mining,” *International Journal of Emerging Technology and Advanced Engineering*, Volume 4, 2014.
- [15] Gabelica, H., Krstic, Ž and Minanovic, A., “Big data and sentiment analysis using KNIME: Online reviews vs. social media,” In *Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pp. 1464-1468, 2014.
- [16] Pablo Gamallo, Marcos Garcia, “A Naive-Bayes Strategy for Sentiment Analysis on English Tweets, *Proceedings of the 8th International Workshop on Semantic Evaluation*,” pp. 171–175, 2014.
- [17] W. Medhat, A. Hassan, and H. Korashy, “Sentiment analysis algorithms and applications: A survey,” *Ain Shams Eng. J.*, 2014.
- [18] Ravi, K., Siddeshwar, V., Ravi, V., & Mohan, L., “Sentiment analysis applied to Educational Sector,” In *2015 IEEE International Conference on Computational Intelligence and Computing Research*, pp. 1-6, 2015.
- [19] Vinayak Hegde and Narendra Kamath G, “Prediction of Sentimental Analysis using Machine Learning Technique,” *International conference Trends in Technology for convergence TITCON15, International Journal of Engineering Research & Technology*, 2015.
- [20] Khan, A. Z., Atique, M., & Thakare, V. M., “Combining lexicon-based and learning-based methods for Twitter sentiment analysis,” *International Journal of Electronics, Communication and Soft Computing Science & Engineering (IJECSCE)*, 89, 2015.
- [21] Yadav, S. K., “ Sentiment analysis and classification: a survey,” *International Journal of Advance Research in Computer Science and Management Studies*, 3(3), 113-121, 2015.
- [22] Dhanalakshmi, V., Bino, D., & Saravanan, A. M., “Opinion mining from student feedback data using supervised learning algorithms,” In *2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC)*, pp. 1-5, 2016.
- [23] <http://sentiwordnet.isti.cnr.it/> as on 09/01/2017