A Novel Review of Various Sentimental Analysis Techniques

Anchal Kathuria[1], Dr. Saurav Upadhyay[2]

anchal9306@gmail.com [1], s4upadhyay@gmail.com [2]
Maharishi Markandeshwar University, India[1,2]

Abstract: Sentiment analysis (SA) is defined as an intelligent method of extricating various emotions and feeling of users. It is one of the major fields for researchers working in natural language processing. The evolution of Internet has become one of the biggest platform for users to exchange their ideas, share messages, post views etc. There also exists many blogs, Google+ which is gaining good popularity as they allow people to express their views. In this paper, the present state of various techniques of sentiment analysis for opinion mining like machine learning and lexicon-based approaches are discussed. The various techniques used for Sentiment Analysis are analysed in this paper to perform an evaluation study and check the usefulness of the existing literature. Our work will also help the future researchers to understand present gaps in the literature of sentiment analysis.

1.   Introduction

Sentiment analysis is often outlined as a method of mining opinions, views and emotions from written text, verbal speech or images taken from social media like Facebook, twitter and other information sources through natural language process (NLP) [1]. Sentiment analysis involves classification of data into various classes like positive i.e. good sense or negative i.e. bad sense or neutral i.e. non-effective. Therefore this classification plays a very important role in NLP[2]. For the processing of written text collections, user views we also sometimes divide the structure in steps with literal structure. The reviews that are composed of various sentences are treated as a single document. The designed corpus consists of many such documents. Such clear division answer, that is acceptable for automatic process in step with simple rules, is often simply and quickly used to get analysis objects in

succeeding steps. So to attain the correctness and to analyse information effectively the research is done in terms of machine learning i.e. training machine for effective and correct analysis [3] with natural language process (NLP).

The social networks today are serving as a medium where a user can post its views, comments and expressions in effective manner for any number of times. A lot of analysis work is being done in the field of sentiment analysis because of its significance in process of promotion and marketing level competition and also the dynamic desires of the individuals. Sentiment analysis requires a decent and correctly defined training set for performance, and dataset quality plays an exceptional role for correct analysis of the text. The linguistics analysis of the sentence also contributes in increasing the means and accuracy of the results[4]. The tagging will be useful to the people to understand whether the comment or tweet corresponds to the relevant subject.
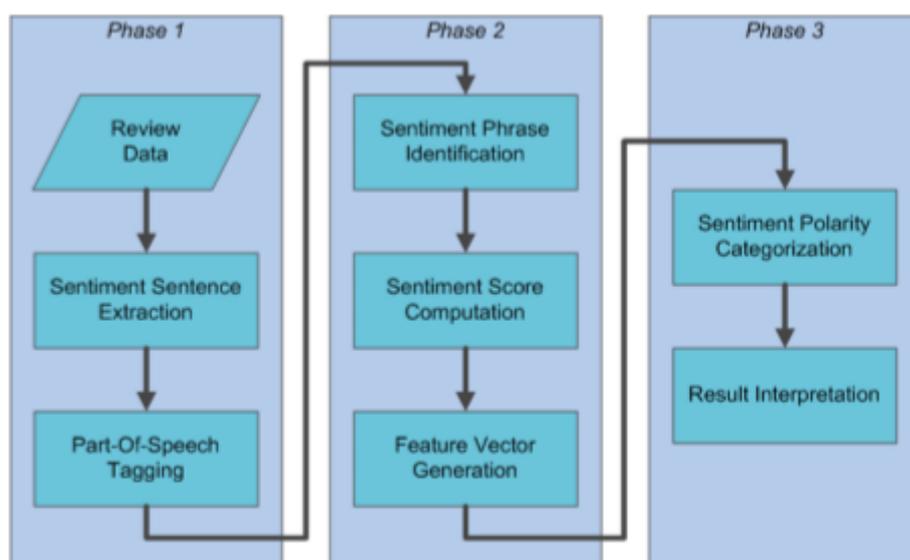
Figure 1: Sentiment Polarity Categorization Process [5]

2. Levels of Analysis

In general, sentiment analysis has been investigated primarily at three levels [4]. In document level, the major task is to classify whether an entire opinion document expresses, is a positive or negative sentiment. This level of study assumes that every document expresses opinions on one entity. In sentence level the basic task is to examine whether every sentence expressed a positive, negative, or neutral opinion. This level of study is closely associated with sentiment extraction, sentiment classification, and subjectiveness classification, report of opinions or opinion spam detection, among others [6]. It aims to investigate people's sentiments, attitudes, opinions emotions, etc. towards components like, products, people, topics, organizations, and services [7].

3. Methods for Sentiment Analysis

There exist many algorithms, methodologies for sentiment analysis. Still many researchers are working of developing new effective methods or improving existing methodologies. There are three main techniques:

3.1 Machine learning Approach

Machine learning approach is used to train an algorithm with a predefined dataset before applying it to actual dataset. Machine learning techniques first trains the algorithm with some particular inputs with known outputs so that later it can work with new unknown data. Some of the most renowned works based on machine learning are as follows:

3.1.1 Support Vector Machines (SVM)

A standard SVM takes a collection of large input data and predicts, for every given input, there are some attainable classes which forms the output. When given a collection of training examples, every marked as belonging to a selected class, an SVM training rule builds a model which will be used to assign new examples into a class[8]. An SVM model may be a representation of the examples as points in area, mapped such as the members of the separate classes are divided by a gap as wide as attainable. New examples are then mapped into that very same area and expected to belong to at least one of the classes supported that aspect of the gap they fall in. Defining very formally, a support vector machine constructs a hyperplane or a collection of hyperplanes in an infinite dimensional area, which may be used for classification. Naturally, an effective separation is achieved by the hyperplane that has the most important distance to the closest training information of any category. Larger the margin, lower would the generalization error of the classifier be[9].

3.1.2 Naive Bayes

This approach presupposes the supply of at least a set of articles with pre-assigned opinion and reality labels at the document level [10]. They used single words, while not stemming or stop word removal as options. Naive Bayes assigns a document d to the category c, that maximizes P (c/d) by applying Bayes' rule,

$$P\left(\frac{c}{d}\right) = \frac{P(c)P\left(\frac{d}{c}\right)}{P(d)}$$

3.1.3 Feature Driven Sentiment Analysis

The product feature extraction plays a key role within the analysis of the product, since we are able to see the importance of the information of the options and their relationships for the improved promoting plan. In [1], it is done by Fuzzy Domain metaphysics Sentiment Tree (FDOST). In FDOST, the basis node represents the product, the leaf nodes represent the polarity and also the non-leaf nodes represent the sub options of corresponding parent features.

3.2 Rule Based Approach

Rule based approach is employed by shaping various rules for obtaining the opinion, created by tokenizing every sentence in each document then testing every token, or word, for its presence. If the word is there and has a positive sentiment, a +1 rating was applied to that. Every post starts with a neutral score of zero, and was considered positive. If the ultimate polarity score was bigger than zero, or negative if the score was less than zero [11] once the output of rule based approach it will check or raise whether the output is correct or not. If the input sentence contains any word that isn't present within the database which can facilitate within the analysis of moving picture review, then such words are to be added to the database. This is often supervised learning within which the system is trained to learn if any new input is given.

3.3 Lexical Based Approach

Lexicon based techniques work on an assumption that the collective polarity of a sentence or documents is total of polarities of the individual phrases or words. In the seminar ROMIP 2012 the lexicon based technique planned in [12] was used. This methodology relies on emotional analysis for sentiment analysis dictionaries for every domain. Next, every domain lexicon was replenished with appraisal words of applicable training collection that have the best weight, calculated by the strategy of RF (Relevance Frequency) [8]. The word-modifier changes (increases or decreases) the weight of the subsequent appraisal word by an exact share. Word-negation shifts the load of the subsequent appraisal word by an exact offset: for positive words to decrease, for negative to extend. The procedure of the text sentiment classification was dispensed as follows. 1st weights of all coaching texts of the classified text is calculated. All the texts are placed into a one dimensional emotional area. The proportion of deletions was determined by the cross-validation technique. Then the common weights of training texts for every sentiment category were found. The classified text was referred to the category that was situated nearer within the one dimensional emotional space.

4. Comparison of three major techniques of sentimental analysis

The three major techniques used in sentimental analysis are analysed based on their performance and accuracy. The major advantages and disadvantages of using any approach are also discussed. The comparison of all these techniques is shown in tabular form below:

Table 1 Comparison of various sentimental analysis approaches

Approach	Classification	Advantage	Disadvantage	Methods
Machine Learning Method	It is classified as both supervised and unsupervised learning	Support feature learning and parameter optimization for best results	Large data requirement and works on single domain	SVM Naïve Bayes FDOSA
Rule based approach	It is classified as both supervised and unsupervised learning	Higher accuracy, require lesser data but need expert human labour	Rules must need to define accuracy as performance is highly rule dependent	Fine-grained dictionary Booster words
Lexicon Based Approach	It is classified under unsupervised learning	Labelled data and the procedure of learning is not required	Excessively rely on emotional dictionary	Corpus Dictionary

5.   Conclusion

This paper attempts to provide a survey and comparative study of existing techniques for opinion mining as well as machine learning, rule based approach and lexicon-based approaches with some analysis metrics. The performance of machine learning strategies, like SVM and naive bayes have the best accuracy and may be considered the baseline learning strategies, whereas lexicon-based strategies are terribly effective in some cases, which need few efforts in human-labelled document . The rule based approach is highly dependent on rule process for performance, therefore mostly this methodology underperforms in contrast with machine learning and lexicon methodology. Study additionally shows that more the cleaner knowledge, more correct results are often obtained. Research work is carried out for higher analysis strategies in this area, as well as the semantics by considering higher rule definition to reinforce rule based approach. In the world of web, majority of individuals depend upon social networking sites to urge their valued data, analysing the reviews from these blogs can yield a higher understanding and facilitate in their decision-making.

References

[1]   M. D. Devika, C. Sunitha, and A. Ganesh, "Sentiment Analysis: A Comparative Study on Different Approaches," in *Procedia Computer Science*, 2016, vol. 87, pp. 44–49.

[2]   V. A. Kharde and S. S. Sonawane, "Sentiment Analysis of Twitter Data: A Survey of Techniques," *Int. J. Comput. Appl.*, vol. 139, no. 11, pp. 975–8887, 2016.

[3]   Z. Hu, J. Hu, W. Ding, and X. Zheng, "Review Sentiment Analysis Based on Deep Learning," in *2015 IEEE 12th International Conference on e-Business Engineering*, 2015, pp. 87–94.

[4]   D. M. E.-D. M. Hussein, "A survey on sentiment analysis challenges," *J. King Saud Univ. - Eng. Sci.*, vol. 34, no. 4, 2016.

[5]   X. Fang and J. Zhan, "Sentiment analysis using product review data," *Springer J. Big Data*, vol. 2, no. 1, p. 5, 2015.

[6]   S. K. Dwivedi and B. Rawat, "A review paper on data preprocessing: A critical phase in web usage mining process," in *2015 International Conference on Green Computing and Internet of*

Things (ICGCIoT), 2015, pp. 506–510.

[7]   V. M. Pradhan, J. Vala, and P. Balani, "A Survey on Sentiment Analysis Algorithms for Opinion Mining," *Int. J. Comput. Appl.*, vol. 133, no. 9, pp. 7–11, 2016.

[8]   C. Bhadane, H. Dalal, and H. Doshi, "Sentiment analysis: Measuring opinions," *Procedia Comput. Sci.*, vol. 45, no. C, pp. 808–814, 2015.

[9]   A. Tripathy, A. Agrawal, and S. K. Rath, "Classification of Sentimental Reviews Using Machine Learning Techniques," *Procedia Comput. Sci.*, vol. 57, pp. 821–829, 2015.

[10]   Q. Rajput, S. Haider, and S. Ghani, "Lexicon-Based Sentiment Analysis of Teachers ' Evaluation," *Hindawi Appl. Comput. Intell. Soft Comput.*, vol. 2016, no. 6, 2016.

[11]   R. Nithya and D. Maheswari, "Sentiment analysis on unstructured review," in *IEEE Proceedings - 2014 International Conference on Intelligent Computing Applications, ICICA 2014*, 2014, pp. 367–371.

[12]   K. Ahmed, N. El Tazi, and A. H. Hossny, "Sentiment Analysis over Social Networks: An Overview," in *2015 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2015, no. October, pp. 2174–2179.