

The Various Approaches for Sentiment Analysis: A Survey

M. Dorothy¹, S. Rajini²

¹PG Student, Department of Computer Science and Engineering, Kumaraguru College of Technology, Coimbatore, India

²Associate Professor, Department of Computer Science and Engineering, Kumaraguru College of Technology, Coimbatore, India

Abstract: *Sentiment analysis or Opinion mining is a machine learning approach in which machines analyze and classify the human's sentiments which are expressed in the form of either text or speech. The textual reviews available in the web are increasing day by day. Manual analysis of such large number of reviews is practically impossible. Therefore, sentiment analysis can be used to mine the overall sentiment or opinion polarity from the texts. The sentiment analysis finds its application in movie reviews, blogs, customer feedback, twitter etc. This literature survey is the study of various approaches/methods that can be used to classify the reviews.*

Keywords: Machine learning, Sentiment analysis, Opinion mining, polarity

1. Introduction

Sentiment analysis (often referred as opinion mining) is the computational study of people's opinions, attitudes and emotions towards an entity. It is very important in the current scenario because, lots of user opinionated texts are available in the web now. Hence, the usefulness of sentiment analysis is increasing day by day. Therefore, the machines must be made reliable and efficient in their ability to interpret and understand human emotions and feelings.

A. Different Levels Of Sentiment Analysis

1) Document Level Sentiment Analysis

In the document level classification, the whole document is classified into either positive or negative class. Here, a single review about a single topic is considered. Both supervised and unsupervised learning methods can be used for the document level classification. Supervised learning algorithms like naïve Bayesian, Support Vector Machine, can be used to train the system. The unsupervised learning can be done by extracting the opinion words inside a document.

2) Sentence Level Sentiment Analysis

In the sentence level classification, the sentences are classified into positive, negative or neutral class. The same document level classification methods can be applied to the sentence level classification problem. Here, the objective and subjective sentences must be found out because, the subjective sentences contain the opinion words which can help determine the sentiment about the entity and after this, the polarity classification is done into positive and negative classes.

3) Phrase Level Sentiment Analysis

In the phrase level sentiment classification, the phrases that contain opinions are determined and then a phrase level classification is performed. The phrase level sentiment classification is said to be called as pinpointed approach.

B. Sentiment Classification Techniques

Sentiment classification techniques can be mainly divided into two main categories. They are machine learning approaches and lexicon based approaches.

1) The Machine Learning Approach (ML)

This is the popularly used approach in sentiment analysis. The text classification methods using Machine learning approach can be divided into supervised, unsupervised and semi-supervised methods. The supervised methods use a large number of labeled training documents whereas the unsupervised methods are used when it is difficult to find these labeled training documents. Semi-supervised learning makes use of unlabeled data for training - typically a small amount of labeled data with a large amount of unlabeled data.

2) The Lexicon-based Approach

This approach completely relies on sentiment lexicon which is a collection of known and precompiled sentiment terms. The lexicon based approach is further divided into dictionary-based approach and corpus-based approach which use statistical or semantic methods to find sentiment polarity.

2. Literature Survey

A method of tweet sentiment analysis is specified in [1]. This work introduces an approach that automatically classifies the tweet sentiments using classifier ensembles which are formed by Multinomial Naïve Bayes, SVM, Random Forest and Logistic Regression. The authors showed that the use of ensembles of multiple base classifiers combined with scores obtained from lexicons can improve the accuracy of tweet sentiment classification. Tweets are classified as either positive or negative concerning a query term.

A novel ensemble approach in [2] has been introduced for the sentiment classification purpose to reduce the noise sensitivity related to language ambiguity and to improve the performance of the polarity classification task. The proposed

Bayesian ensemble learning approach makes use of multiple classifiers to predict the sentiment orientation of user-generated contents. The investigation results in an effective and efficient paradigm suitable for polarity detection both in user reviews and social media environments (twitter).

In [3], the authors analyzed the various methods that can be used to classify a given piece of natural language text according to the opinions expressed in it i.e. whether the general attitude is negative or positive. They implemented a set of techniques for aspect classification and polarity identification of product review using machine learning (SVM) combined with domain specific lexicons.

In [4], the sentiment analysis is carried out in unstructured reviews. The authors classified the most identified features present in the review using Naïve Bayes classifier and then determined their positive, negative and neutral polarity distribution.

In [5], the authors proposed an improved Naïve Bayes algorithm for the sentiment analysis of restaurant reviews. The result of the experiment showed that when this algorithm was used along with “unigrams + bigrams” feature, the gap between the positive accuracy and the negative accuracy was narrowed down when compared to the results obtained when the original Naïve Bayes was used. The improved Naïve Bayes algorithm proposed in this paper proves its effectiveness when the accuracy and balanced classification are considered. Moreover, the unigrams +bigrams feature can be seen as an effective feature for the sentiment analysis of the restaurant review.

In [6], Turney proposed a simple unsupervised learning algorithm for classifying the reviews as either recommended/positive (thumbs up) or not recommended/negative (thumbs down). The classification of the reviews is predicted based on the average semantic orientation of the phrases. A review is classified as recommended if the average semantic orientation of its phrases is positive else it is not recommended.

In [7], the authors evaluated and compared the performance of the three approaches i.e two Machine Learning based classifiers (Naïve Bayes and SVM), the Unsupervised Semantic Orientation approach (SO-PMI-IR algorithm) and the SentiWordNet approaches for sentiment classification on movie reviews. It is found that, the accuracy of classification by NB is marginally better than the SVM and is close to the SO-PMI-IR algorithm. The SO-PMI-IR algorithm has obtained impressive accuracy levels and seems the best choice due to its unsupervised nature. The authors concluded that NB performance can be comparable to the popularly believed superior performance of SVM, at least for sentiment classification.

The clustering-based approach [8] to SA focuses on the direction of positive/ negative polarity classification. It is suitable for document level analysis and is based on the basic k-means clustering algorithm. Documents are primarily clustered into two clusters which are expected to be positive

group and negative group. The proposed method involves three strategies. The technique of TF-IDF weighting will be firstly applied on the raw data to improve accuracy. Then, a voting mechanism will be used to extract a more stable clustering result. Finally, the term score which was mentioned in the introduction to symbolic techniques will be imported to further enhance the clustering result.

In [9], the authors proposed a new methodology for sentiment analysis called proximity-based sentiment analysis. They performed sentiment classification using only the proximity based features for the analysis of movie reviews. The experimental results showed that proximity-based sentiment analysis is able to extract sentiments from a specific domain, with performance comparable to the state-of-the-art approaches.

In [10], a new algorithm called Sentiment Fuzzy Classification algorithm with parts of speech tags is used to improve the classification accuracy on the benchmark dataset of Movies reviews dataset. To better handle intrinsic fuzziness in sentiment polarity, the authors applied the fuzzy set theory to sentiment classification. They redefined sentiment classes as three fuzzy sets, and then applied existing fuzzy distributions to construct membership functions for the three sentiment fuzzy sets.

In [11], the authors investigated the existing technology and their utility for sentiment classification on web log posts. They investigated the utility of Naïve Bayes and SVMs on a novel collection of datasets created from web log posts. The results showed that the Naïve Bayes classifier can on average correctly predict a posting's political category and it significantly outperforms Support Vector Machines at a confident level of 99%.

A lexicon-based method to sentiment classification was presented by Taboada et al.[12]. The author used dictionaries of positive or negative polarized words to do classification task. A semantic orientation calculator was built based on these dictionaries by incorporating intensifiers and negation words.

3. Conclusion

Sentiment Analysis problem has been a research interest for recent years. Through this literature survey, the relevant works done to solve this problem could be studied. Although several notable works have come in this field, a fully automated and highly efficient system has not been introduced till now. This is because of the unstructured nature of natural language. Several challenges still exist in the field of machine learning and some of them are Named entity Recognition, Co-reference Resolution and domain dependency. These problems have to be tackled separately and those solutions can be used to improve the methods to do sentiment analysis.

References

- [1] Nadia F.F. da Silva, Eduardo R.Hruschka, Estevam and R.Hruschka Jr, "Tweet sentiment analysis with classifier ensembles," *Decision support systems* vol.no.66, pp. 170-179, 2014.
- [2] E.Fersini, E.Messina, and F.A.Pozzi, "Sentiment analysis : Bayesian Ensemble Learning," *Decision Support Systems* vol.no. 68, pp. 26-38, 2014.
- [3] Chetashri Bhadane, Hardi Dalal, Heenal Doshi, "Sentiment analysis: Measuring opinions," *International Conference on Advanced Computing Technologies and Applications (ICACTA2015)*, *Procedia Computer Science* vol.no.45, pp. 808 – 814, 2015.
- [4] Nithya, R, Maheswari, D, "Sentiment Analysis on Unstructured Review," *International Conference in Intelligent Computing Applications (ICICA)*, pp.367-371, 6-7 March 2014.
- [5] Hanhoon Kang, Seong Joon Yoo, Dongil Han, "Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews," *Expert Systems with Applications*, Volume 39, Issue 5, Pages 6000–6010, April 2012.
- [6] P. Turney, "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews," *Proceedings of ACL-02, 40th Annual Meeting of the Association for Computational Linguistics*, pp. 417-424, Philadelphia, US, 2002.
- [7] Singh, V.K, Piryani, R., Uddin, A, Waila, P, Marisha, "Sentiment analysis of textual reviews; Evaluating machine learning, unsupervised and SentiWordNet approaches," *5th International Conference on Knowledge and Smart Technology (KST)*, pp.122-127, Jan. 31 2013-Feb. 1 2013.
- [8] Gang Li; Fei Liu, "A clustering-based approach on sentiment analysis," *2010 International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, pp.331-337, 15-16 Nov. 2010.
- [9] Hasan, S.M.S, Adjeroh, D.A., "Proximity-based sentiment analysis," *2011 Fourth International Conference in Applications of Digital Information and Web Technologies (ICADIWT)*, pp.106-111, 4-6 Aug. 2011.
- [10] Mouthami, K.; Devi, K.N, Bhaskaran, V.M., "Sentiment analysis and classification based on textual reviews," *International Conference in Information Communication and Embedded Systems (ICICES)*, pp.271-276, 21-22 Feb. 2013.
- [11] K.T. Durant & M.D. Smith, "Mining Sentiment Classification from Political Web Logs", *Proceedings of WEBKDD'06*, ACM, 2006.
- [12] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," *Computational Linguistics*. vol. 37, pp. 267-307, 2011.