A Survey On Sentiment Analysis And Techniques

[1]Jasneet Kaur, [2]Dr. Neelam Duhan
[1] PG Student, [2] Assistant Professor,
Department of Computer Engineering, YMCA University of Science & Technology, Faridabad, Haryana, India

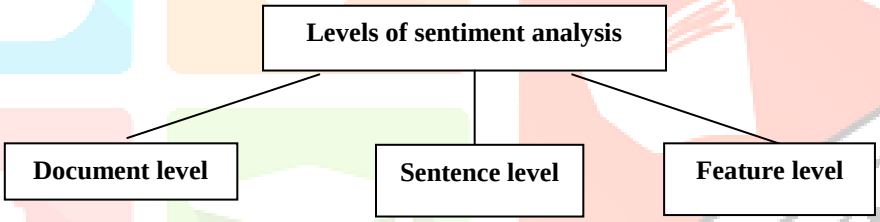
Abstract: Sentiment analysis, also known as opinion mining, is a way of analysing customer reviews on various social media or shopping sites, like Twitter, Facebook, Amazon etc, in order to depict the positive or negative sentiment of customer or blogger related to that particular person, article or product. Opinion mining (OM) is generally used by customers to decide whether to buy a product or not and by market analysers to decide whether & by how much sale of a product must be increased or decreased. This paper includes introduction to sentiment analysis, its techniques and challenges involved in it.

IndexTerms **-** Opinions, Sentiment analysis, Opinion mining, polarity**.**

- **Introduction**

People today give very much preference to the opinionated text available online on social media and e commerce sites. This has increased the importance of sentiment analysis or opinion mining which results in polarity of the text i.e if the given text has positive or negative polarity.

There are many challenges associated in sentiment analysis. A text might include some sort of sarcasm , it includes special symbols like # , %, @ etc. A review will also include some emoticons or smilies or it can include some numerical data. A review can also be in completely different language which is unknown to our system.

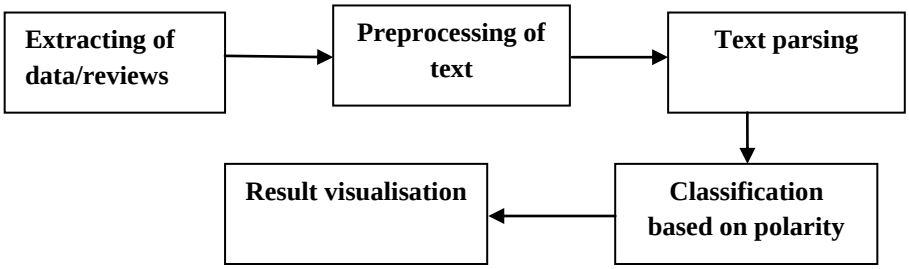


There are three levels of sentiment analysis[12]:

1. **Document level** : In this level of sentiment analysis, entire document is defined as of negative or positive polarity i.e. entire document is analysed at once. It doesn't work on different feature of the document which might have different polarities

2. **Sentence Level:** In this level of sentiment analysis, every sentence of the document is analysed separately and polarity of every sentence is depicted differently

3. **Feature level:** In this level of sentiment analysis, the words which consist of features related to the entity are extracted and analysis is done on the basis of them. These word are generally called phrases.

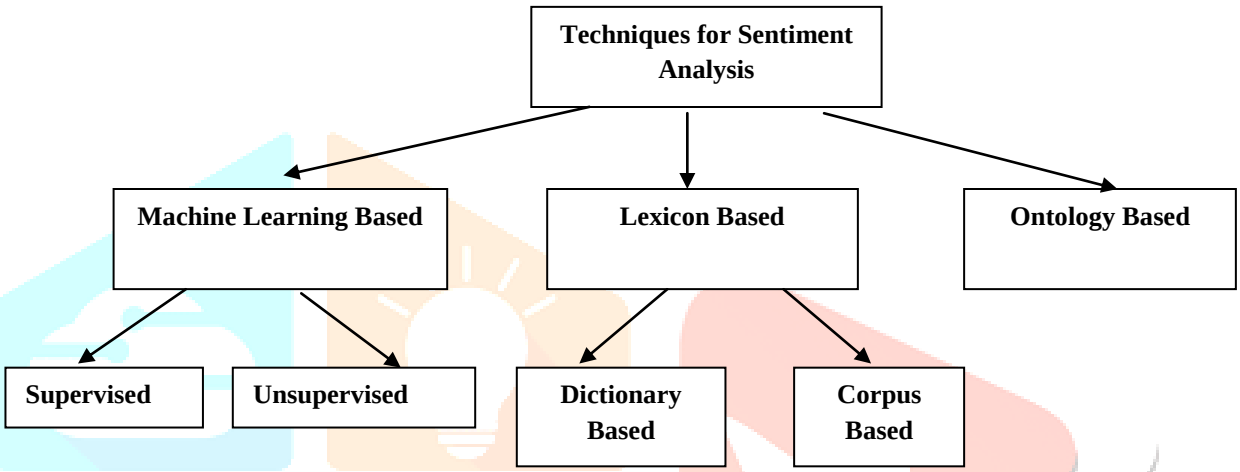


Workflow of sentiment analysis

The workflow of sentiment analysis goes as such:

- **Extraction of data/reviews:** First of all the dataset of the entity to be analysed is collected from the respected review site
- **Preprocessing:** Preprocessing involves removal of unwanted part of text i.e stop word removal such as pronouns(she/he) & articles(a,an,the) and stemming which involves reduction of a word to its base form for eg . studying turns to study, eating turns to eat etc[9]
- **Text parsing:** Parsing involves part of speech tagging in which every word is identifies by tagging the word as noun, pronoun, adjective etc.
- **Classification based on polarity:** In this step sentiments which the text bears is determined and classification is done on the basis of polarity
- **Result visualization:** The achieved result can then be displayed in desired graphical format

- **Sentiment Analysis Techniques**



1. Machine learning based approach[4]: Machine learning is divided into two types: Supervised & Unsupervised learning approach

- **Supervised Learning**: In this type of technique classes of the features of provided entity are formed with predefined labels and the classification is done on the basis of these labels. Some of the supervised learning techniques are naive bayes, support vector machine and maximum entropy .
- **Unsupervised Learning:** Unlike supervised learning technique, in this techniques no predefined labels are created for the classes. These techniques are more complex and time consuming as compared to supervised learning techniques but are very much efficient from them. Neural networks is a type of unsupervised learning technique.

2. Lexicon based approach[4]: This type of approach focuses on finding the polarity of opinion words in the text. This is done after the text has been preprocessed and tokenisation has been done. After that each token is matched in the sentiment dictionary and appropriates score is assigned to the word which is then used to classify the text as positive or negative. Types of lexicon approaches are:

- **Dictionary based approach:** In this approach similar kinds of words are found using a small previously defined words. The newly found words are added to the dictionary till no new word could be found.
- **Corpus based approach:** This approach is practically not possible because it is not easy to create collection of each and every word of English to analyse the text. So this approach is not efficient as the one defined above.

3 Ontology based approach [1]: In the above defined approach, the sentiment analysis is done on the entire document indicating document as positive or negative on the basis of polarity of words in it. It does' nt classify the text on the basis of the sentiment score of the feature words in the text. This approach does so. It is done using OWL & RDF languages. In this approach first the ontology is created , then the opinions are extracted, the extracted opinions are tagged in the ontology , the sentiment score is assigned by creating dictionary indicating polarity of each phrase & sentiment score is calculated .

- **Challenges in Sentiment Analysis**

- Opinion text might sometimes contain some sort of sarcastic text in it, i.e. , a person states negative opinions but use positive words in the text. These sort of comments need special attention[3]

- A text will contain emoticons and special symbols such as %,*,&,# . A satisfied emoticons might contain negative words or vice versa.
- Opinions can be in languages other than English. Training the classifier to work on different language is a challenging task
- Opinion text might contain numerical value as well which also need special care while assigning polarities to the tokenised words
- Some users might post fake comments or reviews to degrade the reputation of any person or product. These are called spam reviews. Handling these reviews is very challenging task[2]
- Handling of implicit feedbacks is another big task. Implicit feedbacks are the feedbacks in which the entity is not mentioned directly. For eg: The comment "Oh! So bright" in a mobile phone feedback is mentioning about the screen resolution.
- Another area which needs research is sentiment analysis of textual images or audio clippings[7]
- Sentiment analysis is done on only positive or negative words. But an opinion can also be neutral , strongly positive or strongly negative as well[3]

• Conclusion & Future work

Sentiment analysis is one of the emerging research area in computer science. Our study concludes that the most effective machine learning classifier is Support vector machine, having 80% accuracy compared to naive bayes & maximum entropy[8]. Our work will be extended by working on complex , unlabelled & unstructured data set and improving the system accuracy using deep learning neural network. [8]
In future, research could be done in opinion mining in user level analysis, images, audio clips & videos using deep learning algorithms

• References

[1] **Katarzyna Wójcik**, **Janusz Tuchowski**," **Ontology Based Approach to Sentiment Analysis" June 2014, 6th International Scientific Conference Faculty of Management Cracow University of Economics, At Cracow University of Economics**
[2] Surya Prakash Sharma, Dr Rajdev Tiwari, Dr Rajesh Prasad," Opinion Mining and Sentiment Analysis on Customer Review Documents- A Survey" International Journal of Advanced Research in Computer and Communication Engineering- Vol. 6, Special Issue 2, February 2017

[3] Ms.A.M.Abirami, Ms.V.Gayathri," A SURVEY ON SENTIMENT ANALYSIS METHODS AND APPROACH" 2016 IEEE Eighth International Conference on Advanced Computing (ICoAC)

[4] Dudhat Ankit kumar M, Prof. R. R. Badre, Prof. Mayura Kinikar," A Survey on Sentiment Analysis and Opinion Mining" International Journal of Innovative Research in Computer and Communication Engineering Vol. 2, Issue 11, November 2014

[5] Harpreet Kaur, Veenu Mangat, Nidhi," A Survey of Sentiment Analysis techniques" International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)
(I-SMAC 2017)

[6] Manasee Godsay," The Process of Sentiment Analysis: A Study" International Journal of Computer Applications (0975 – 8887) Volume 126 – No.7, September 2015

[7] Jawed Khan, Byeong Soo Jeong," Summarizing customer review based on product feature and opinion" Proceedings of the 2016 International Conference on Machine Learning and Cybernetics, Jeju, South Korea, 10-13 July, 2016
[8] https://deeplearning4j.org/neuralnet-overview

[9] http://nlp.stanford.edu/software/corenlp.html

[10] Mitali Desai**,** Mayuri A. Mehta**,"** *Techniques for Sentiment Analysis of Twitter Data: A Comprehensive Survey***"** International Conference on Computing, Communication and Automation (ICCCA2016)

[11] Rekha,Dr. Williamjeet Singh," *Sentiment Analysis of Online Mobile Reviews* International Conference on Inventive Communication and Computational Technologies (ICICCT 2017)"

[12] Neha Raghuvanshi , Prof. J.M. Patil," A Brief Review on Sentiment Analysis" International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) - 2016