

WDE-KNN BASED CLASSIFICATION TECHNIQUE FOR PRODUCT REVIEW SENTIMENT ANALYSIS- A REVIEW

¹Kanika Sharma, ²Akanksha Sambyal

¹Research Scholar, ²Assistant Professor

Department of Computer Science and Engineering

Sri Sai University Palampur (HP), India

Kanikasagar004@gmail.com er.akankshasambyal@gmail.com

ABSTRACT: The sentiment analysis is the technique which is applied on analyse sentiment of the input data. The technique of sentiment analysis is applied in various steps like feature extraction and classification. The classification techniques are applied which can classify data into certain number of classes. In this paper, various techniques of sentiment analysis are reviewed and analyse in terms of certain parameters.

KEYWORDS:

SVM, Naïve Bayes, KNN, Sentiment Analysis, Decision Tree, NLP.

1 INTRODUCTION

Sentiment analysis is a type of data mining that assess the inclination of people's opinion through NLP. It uses the NLP in order to categorize the opinions of people about the products or the reviews. Sentiment analysis sometimes known as opinion mining. Sentiment analysis is a technique which helps to grasp or understand the behaviour of user and speaker. It deals with opinions and perspective of human related to emotions and attitude about some occurrence or the event [1]. In social media sites it helps to determine whether the reviews, blog- posts, news, articles is positive, negative and neutral. Sentiment analysis helps merchants for improving their products and services. The two important tasks involved in Opinion Mining and Sentiment Analysis are 1) Opinion Extraction: extracting the opinionated phrases, in proper context, from free text and (2) Sentiment classification: classifying opinionated phrases based on sentiment orientation. It utilizes various machine learning techniques such as SVM, Naïve Bayes, Character Based N-gram model etc. for sentiment classification. It also help future customers to make decision such as they buy products or not. In this technique of sentiment analysis the features of the input data are extracted using pattern matching algorithm and for the sarcasm detection, classification techniques are applied. Sarcasm is a special kind of sentiment that have opposite means of what you really want to say (especially in order to insult or wit someone, to show irritation, or to be funny). People often express it verbally through the use of heavy tonal stress and certain gestural clues like rolling of the eyes. These tonal and gestural clues are obviously not available for expressing sarcasm in text, making its detection reliant upon other factors [12].

As sentiment analysis play an important role in today world (mainly in social media) instead of this there are lots of challenges and problems have been faced while using this.

1.2. ISSUES OF SENTIMENT ANALYSIS

As we know sentiment analysis helps to determine whether a review is positive, negative or neutral. It also helps to understand the behaviour of the user. Apart from this sentiment analysis serves with many problems.

1.2.1. INFORMATION OVERLOAD

Sentiment analysis helps to analyse the reviews i.e. whether a review is positive or negative. In another word sentiment analysis helps to determine the attitude, emotions etc. of user. It is an easy task for sentiment analysis to determine the reviews is positive, negative or neutral when data or text is small. But sometimes the data is too large than it serves with the problem of information overload.

1.2.2 ACCURACY

Accuracy is a measure concerned in sentiment analysis. Sometimes accuracy often referred to as precision. Accuracy alone does not tell us anywhere close to whole story. Therefore we need another metrics called recall. Recall is as such important as precision (accuracy). Finally, there is F-score also called as F-measure or F1, which is more holistic account of overall performance.

1.2.2.1 PRECISION/ACCURACY:

A measure of how often a sentiment rating was correct. In pattern recognition, information retrieval and binary classification, precision is the fraction of relevant instance among the retrieved instances.

1.2.2.2 RECALL:

A measure of how many documents with sentiments were rated as sentimental. In pattern recognition, information retrieval and binary classification, recall is the fraction of relevant instances that have been retrieved over the total amount of relevant instances.

1.2.2.3 F1 SCORE:

In statistical analysis of binary classification, the F1 score also known as F1 measure or F-score is a measure of test's accuracy. The F1 Score is very helpful, as it gives us a single metric that rates a system by both precision and recall. F1 is commonly used by experts and researchers in the linguistics and natural language processing fields to simply describe the performance of such systems [11].

1.2.3. SARCASM AND EMOTICONS

A sentence sometimes has two meaning, in that case it is difficult to find positive and negative. Sentiment analysis sometimes not able to find such sentences as positive or negative.

1.2.4. FAKE REVIEWS DETECTION

The trend of fake news in social media is increasing day by day. These news spreads as fast as fire. Social media is one of the fastest way of spreading these types of fake news. Because In today world a lots of people are active on social media .It is more difficult task for sentiment analysis to determine whether a review is fake or not.

1.2.5. STANCE DETECTION

Stance Detection is a task which can automatically determining from text whether the author is in favour of or against towards proposition or tar Mohammad, Sobhani, and Kiritchenko (2016b) created the first dataset of tweets labelled for both stance and sentiment [10].

2 LITERATURE SERVEY

TABLE 1: COMPARISON OF TECHNIQUES

S.NO	AUTHOR NAME	PAPER NAME	DISCRIPTION	OUTCOMES
1	J. Zhu, H. Wang, M. Zhu, B. K. Tsou, and M. Ma.	Aspect-based opinion polling from customer reviews	Opinion mining or sentiment analysis is the computational field of study of people's opinions, emotions, and attitude towards particular aspect [1].	In this paper, we are going to study in depth opinion mining and survey the existing methods that are being used for opinion mining.
2	X.Ding, B.Liu, and P. S. Yu	A holistic lexicon-based approach to opinion mining	In this paper, we focus on customer reviews of product. In particular we study the problem of determining the semantic orientations (positive, negative, or neutral) of opinion expressed on product features in reviews.	In this paper, we propose a holistic lexicon-based approach to solving the problem [2]. In this work we considered both implicit and explicit opinion. The result shows that proposed method performs better than previous method.
3	NarendraAndhale, et.al	An Overview of Text Summarization Techniques, 2016	The process is which the condensed type of document can be generated that can help in recording the significant information and provides importance to the source text is known as text summarization. An important method through which the related information can be identified from huge documents is known as automatic text summarization method.	The comprehensive survey of both of the techniques present within the text summarization is presented in this paper [4]. The testing for hybridization is studied within this paper which helps in generating the information which is compressed and readable by the users.
4	ShwetaRana et el.	Comparative analysis of sentiment orientation using SVM and Naive Bayes techniques.2016	With the expanding web and social networking people starts to share data and information online. This social media data can be used for Sentiment analysis [7]. In this research paper the author analysed the sentiment of movies reviews. Three different algorithms Naïve Bayes, Synthetic words and Linear SVM have used and compared.	The results generated by these algorithms indicate that Linear SVM algorithm provides the best accuracy than Naïve Bayes. The author suggests for future to identify accuracy rate of different products.

5	LameiXu , Jin Lin, Lina Wang, Chunyong Yin, Jin Wang	Deep Convolutional Neural Network based Approach for Aspect-based Sentiment Analysis	Aspect level sentiment classification is a fundamental task in the field of sentiment analysis, this task aims at extracting aspects from the review text and then inferring the sentiment polarity (e.g. positive, negative) of the aspect[2]. Deep convolutional neural networks (CNN) utilize layers with convolving filters that are applied to local features, and CNN models have demonstrated remarkable results for text classification and sentiment analysis [3].	In this paper, we described our novel approach for aspect-based sentiment analysis used for aspect category identification, extraction of opinion target expression and polarity identification, which employs a convolutional neural network for aspect extraction and sentiment analysis.
6	Vishal A.kharde and S.S sonawane	Sentiment analysis of twitter data.	In this paper, we have a comparative analysis of techniques for opinion mining such as machine learning and lexicon-based approaches using various machine-learning algorithm like Naïve Bayes, Maximum entropy and SVM [9].	The experimental result shows that SVM has the highest accuracy than Naive Bayes and Max entropy.
7	Manoj Y. Manohar Prof. Pallavi Kulkarni	Improvement Sarcasm Analysis using NLP and Corpus based Approach [13].	Sarcasm is most important part of social network websites. It is the best way to convey message to end user. In this paper ,author propose a NLP and CORPUS based approach to detect sarcasm on twitter and also compare data with Ontology based emotion detection to classify tweets as sarcastic or non-sarcastic.	The experimental results shows that the propose technique gives good result on real time database of twitter.
8	Wei Zhao, Ziyu Guan* , Long Chen, Xiaofei He, Fellow, IAPR, Deng Cai, Beidou Wang and Quan Wang.	Weakly- Supervised Deep Embedding For Product Review Sentiment Analysis [14].	The role of opinion mining is to judge whether a review is positive or negative. In this paper, researcher propose a novel framework for product review sentiment analysis. They test two low level network for modelling review sentence such as Convolutional feature extractor [16][17][18] and long- short- term – memory [15].	The experiment result shows that WDE-LSTM perform better than WDE –CNN.

CONCLUSION

In this paper, it is concluded that sentiment analysis is the technique which can analyse sentiments of the twitter data, social media sites data (Facebook, Amazon, Flip kart) etc. The technique of sentiment analysis contains various steps like pre-processing, feature extraction and classification. In this review paper, various techniques of sentiment analysis are reviewed and analysed in terms of certain parameters.

REFERENCES

1. J. Zhu, H. Wang, M. Zhu, B. K. Tsou, and M. Ma. Aspect-Based Opinion Polling From Customer Reviews. *IEEE Transactions on Affective Computing*, 2(1):37–49, 2011.
2. X. Ding, B. Liu, and P. S. Yu. A Holistic Lexicon-Based Approach To Opinion Mining. In *WSDM*, pages 231–240, 2008.
3. NarendraAndhale, L.A. Bewoor. An Overview Of Text Summarization Techniques.2016, IEEE.
4. ShwetaRana, Archana Singh. Comparative Analysis Of Sentiment Orientation Using SVM And Naive Bayes Techniques.2016 2nd International Conference on Next Generation Computing Technologies (NGCT).
5. M. Hu and B. Liu. Mining And Summarizing Customer Reviews. In *SIGKDD*, pages 168–177, 2004.
6. L. Dong, F. Wei, C. Tan, D. Tang, M. Zhou, and K. Xu. Adaptive Recursive Neural Network For Target-Dependent Twitter Sentiment Classification. In *ACL*, pages 49–54, 2014.
7. Pang B, Lee L. Opinion Mining And Sentiment Analysis [J]//Foundations And Trends In Information Retrieval, 2008, 2(1-2):1–135.
8. Kim, Y. (2014). Convolutional Neural Networks for Sentence Classification. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, 1746–1751. Retrieved from <http://emnlp2014.org/papers/pdf/EMNLP2014181.pdf>
9. Vishal A.Khrade and S.S.Sonawane. “Sentiment Analysis Of Twitter Data: A Survey Of Twitter”. *International Journal of Computer Applications* (0975 – 8887) Volume 139–No.11, April 2016 Retrieved from <https://arxiv.org/ftp/arxiv/papers/1601/1601.06971.pdf>
10. <http://www.saifmohammad.com/WebDocs/StarSem2016-stance-tweets.pdf>
11. <http://blog.infegy.com/understanding-sentiment-analysis-and-sentiment-accuracy>
12. Shruti Kaushik, Prof.Mehul P. Barat” Sarcasm Detection In Sentiment Analysis published in *IJARIE-ISSN (0)-2395-4396 VOL-2 ISSUE-6 2016*.
13. Manoj Y. Manohar Prof. Pallavi Kulkarni “Improvement Sarcasm Analysis Using NLP And Corpus Based Approach. In *ICICCS*, pages 618-622, 2017.
14. Wei Zhao, Ziyu Guan*, Long Chen, Xiaofei He, Fellow, IAPR, Deng Cai, Beidou Wang and Quan Wang” Weakly- Supervised Deep Embedding For Product Review Sentiment Analysis. In *IEEE*, pages 1-12, august 2017.
15. S. Hochreiter and J. Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, 1997.
16. X. Qiu and X. Huang. Convolutional Neural Tensor Network Architecture For Community-Based Question Answering. In *IJCAI*, pages 1305–1311, 2015.
17. Y. Kim. Convolutional Neural Networks For Sentence Classification. In *EMNLP*, pages 1746–1751, 2014.
18. N.Kalchbrenner,E.Grefenstette, andP.Blunsom. A Convolutional Neural Network For Modelling Sentences. In *ACL*, 2014.