

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology



ISSN 2320-088X

IMPACT FACTOR: 6.017

IJCSMC, Vol. 8, Issue. 2, February 2019, pg.75 – 79

Analysis of Various Sentiment Analysis Techniques

Neha Rajput

neharajput22may@gmail.com

Research Scholar, Bhagwant Institute of Technology, Muzaffarnagar UP India

Mrs. Shivani Chauhan

Shivanichauhanbit@gmail.com

Assistant Professor, Department of Computer, Bhagwant Institute of Technology, Muzaffarnagar UP India

Abstract: Any kind of attitude, through or judgment that occurs due to any feeling is known as a sentiment which is also known as opinion mining. The sentiments of individuals towards particular elements are analyzed in this approach. To gather sentiment information, web or internet is the best known source. A platform that is accessed socially by various users to post their views is known as Twitter. The messages that are posted by these users are known as tweets. The properties of Tweets are highly unique due to which new challenges have raised. In comparison to several other domains, the sentiment analysis requires higher analysis studies. In the form of data source, the tweets are used from Twitter. The twitter public API is given for the extraction of tweets on a large scale. The "twitteroauth" version of the public API by Williams (2012) is utilized in this situation. It is possible to run this version on local host or web servers and the implementation is done on PHP. For performing feature analysis in sentiment analysis, the N-gram approach is used here.

KEYWORDS: *Twitter, Sentiment Analysis, Classification*

Introduction

In order to mainly, determine whether a product is satisfactory for the users before it is delivered to them, the sentiment analysis (SA) has been developed. On the basis of the requirements of the users, the markets and industries develop their products and services. The factual data is processed, searched or analyzed with the help of textual information retrieval methods. The subjective properties of the components can be presented on the basis of various textual contents within the actualities [1]. The base of sentiment analysis (SA) includes opinions, attitudes, emotions, appraisals and so on. In order to develop new applications, various challenges have been faced while applying these techniques. The major reason due to which issues arise is the regular generation of huge types of data on various online platforms [2]. The different types of positive or negative opinions are given by the users related to various objects which can help organizations in providing feedbacks that can be used in enhancing the quality of those objects. With the utilization of Natural Language Processing (NLP), various tweets, speech or text available on sources can be processed in sentiment analysis [3]. There are three broader categories into which the sentiments are classified. They are "positive" or "negative" or "neutral". There is a need to assign a

class of category to any kind of input that is given to us on the basis of the manner in which it is generated. The voices, faces, emails, grades and so on can be classified on the basis of various their characteristics. An automated process through which the units of texts can be divided and labeled amongst different categories which are called classes, is known as text classification method. The topic can be extracted from the text or can be involved within sentiment classification through this process [4]. The text available online can be classified as positive, negative or neutral with the help of NLP within sentiment analysis. The application which considers every individual word as a separate class and predicts the next work on the basis of previous content available is known as language modeling which is also a method of classification. The extraction and classification of useful features on the basis of their properties into discrete classes is the prior objective of classification. The utilization of hand-written rules is one method through which the text can be classified. Various applications are present that utilize hand-written rule-based classifiers as classification methods. The manner in which the humans can learn from the experiences they had from their pasts is followed by the machine learning techniques. The knowledge is gathered and applied in order to make any kinds of decisions in future [5]. Within the applications of artificial intelligence and document classification, such various types of learning methods are utilized. There are mainly two steps that are followed which classification applied in machine learning. Utilizing the training dataset for learning the model is the first step. In the second step, the trained model is applied to the test dataset. There are various supervised classification methods amongst which any can be used within sentiment analysis method as it the part of text classification problem. A classifier that comprises of Bayes theorem is known as Naïve-Bayes classifier. It is a simple probabilistic classifier in which it is assumed that the presence or absence of one feature in a document is not reliable on any other feature present within that document [6]. In order to provide document classification, Support vector machines (SVM) has been proved to be the most efficient technique that can be used. The identifying of maximum margin hyperplane within the document which can separate the document vector in one class from that of another document which also has maximum margin is the main objective of SVM.

Literature Review

Mondher Bouazizi, et.al (2017) presented in this paper [7], a novel technique to classify the texts that are gathered from Twitter in detailed manner. Amongst multiple sentiment classes, the text is classified through this technique. In order to assist the users for choosing appropriate features from various features provided, SENTA tool is introduced in this study. The identification of such features will help in making it easy to run the classification with the help of a readily available graphical user interface. In order to evaluate the experiments related to multi-class classification SENTA is utilized within this paper. On the basis of experimental results achieved it is seen that the multi-class classification is achieved with 60.2% of accuracy. Thus, by providing various comparisons it is seen that the performance of proposed technique is better than the existing approaches.

Ankit Kumar Soni, (2017) presented in this paper [8] the objective of developing a system which will help in extracting the useful information from raw data available in the application by using classification methods. This extracted data, which is mainly present in twitter micro blogging applications, has various sentiments present within it which need to be analyzed in order to analyze the views of users. The dataset which has Multilanguage tweets is not very easy to be handled by the sentimental classification. There is no such technique proposed which can help in handling the multi-language data. In this paper, Naïve Bayes and Maximum Entropy classifiers are combined to generate one algorithm. Amongst various algorithms, the results are compared which can help in analyzing the performance of various algorithms amongst each other and show which has provide to be better. It is seen through the results achieved that the proposed technique has provided better results in comparison to other existing approaches.

Aldo Hernández, et.al (2016) presented in this paper [9], a sentiment analysis technique that can help in predicting any kinds of future attacks that can possible arise within the web applications. From two sets of users, the tweets are gathered on daily basis which are used by this method. These users express their various issues in the form of content in the web applications. In order to predict the chances of attack, the daily gathered data is presented in statistic manner which can help in detecting the chance of any attack to occur. The sentiments of users and groups are analyzed collectively by the hackers here once the data is extracted by them from sources. Various experiments are conducted and it is seen that the proposed technique can help in providing an estimate of number of actual attacks and negative sentiments provided by the users in the form of tweets.

Venkata Sasank Pagolu, et.al (2016) proposed in this paper [10], the utilization of sentiment analysis and supervised machine learning principles in combined manner to analyze the sentiments of users on twitter platform. The data is extracted from twitter and the relation between the stock market movements of a company as well as the sentiment of the tweets are analyzed here. The users are motivate to invest in the stocks of a company when the tweets available on twitter are positive

which might results in increasing the stock value of the organization in market. There is a direct relation seen amongst the tweets present by public and rise and fall of stock prices. The judgment of the types of sentiments present within the tweets posted by user is the prior object of the sentiment analyzer. On the basis of various experiential results achieved at the end, it is seen that the proposed technique provides better evaluation results in comparison to existing techniques.

Chhaya Chauhan, et.al, (2017) presented that major researches in the field of current time is natural language processing, text analysis, text preprocessing, stemming etc. to analyze the unstructured data. In order to generate desired results, different techniques and tools are used. Sentiment of a text or sentence is determined using different techniques, as internet has a large repository of natural language. To understand the sentiments of the people products reviews are necessary, therefore a summary of positive and negative reviews are need to be generated [11]. Author's main focus was on the review of algorithms and techniques used to extract feature wise summary of the product and analyzed them to form an authentic review. In future, focus will be on higher level natural language processing tasks. Best techniques or tools should be used for more accurate result in which only those keywords are used which are in dataset and rest of the words are eliminated by the system.

Pragya Juneja, et.al, (2017) discussed in this research novel approach to predict the Delhi Corporation Elections results. As today social networking sites like Twitter, Facebook, Instagram etc. plays a major role in sentiment analysis [12]. Basically sentiment analysis is used to identify the opinion and emotional states of the people in order to extract positive and negative views. Twitter analysis has to be done firstly, by extracting the twitter posts twitted by user and analyzes the present scenario, for example it was shown that BJP is more successful political party based on people opinion. Therefore, author concluded that sentiment analysis can be utilized for any purpose based on the tweets they collected like marketing, finance, media, and entertainment and many more.

Mika V. Mäntylä, et.al (2018) Proposed an approach that utilizes both the text mining and qualitative coding in order to provide the analysis on the large amount of gathered information by the author. With the advent in the technology, sentiment analysis is considered as the fastest growing technique in the area of research such as computer science [13]. Hence, author in this paper analyze various paper and concluded that area of the sentiment analysis is very vast as it cover stock markets, elections, disasters, medicine, software engineering and cyber bullying and many more.

Pulkit Garg, et.al, (2017) as surveyed that social media has becoming a medium for online sharing by the increase of more number of people coming online. In this paper, we study post- terror attack tweets by extracting it from twitter. The flow data posted on twitter is used to study factors like last retweet, number of retweets and number of favorites. Maximum number of retweets indicates maximum reach [14]. It creates widespread reaction on the social media. Governments are concentrating on digitalizing the whole nation. Due to increase in number of people, huge data is generated. Author discussed the Uri Terror Attacks that show more negative tweets tend to survive as compare to positive tweets, although their amount is low. It will lead to public unrest if people start targeting a community and provide negative information. Misleading information, the trends of retweets and number of favorites are the future scope to study its flow and survival.

Table 1: Table of Comparison

Authors Names	Year	Description	Outcomes
Mondher Bouazizi, et.al	2017	A novel technique was proposed to classify the texts that are gathered from Twitter in detailed manner. Amongst multiple sentiment classes, the text is classified through this technique.	On the basis of experimental results achieved it is seen that the multi-class classification is achieved with 60.2% of accuracy.
Ankit Kumar Soni	2017	In this paper, Naïve Bayes and Maximum Entropy classifiers are combined to generate one algorithm.	It is seen through the results achieved that the proposed technique has provided better results in comparison to other existing approaches.

Aldo Hernández, et.al	2016	A sentiment analysis technique was proposed that can help in predicting any kinds of future attacks that can possible arise within the web applications. From two sets of users, the tweets are gathered on daily basis which are used by this method.	Various experiments are conducted and it is seen that the proposed technique can help in providing an estimate of number of actual attacks and negative sentiments provided by the users in the form of tweets.
Venkata Sasank Pagolu, et.al	2016	The sentiment analysis and supervised machine learning principles were used in combined manner to analyze the sentiments of users on twitter platform.	On the basis of various experiential results achieved at the end, it is seen that the proposed technique provides better evaluation results in comparison to existing techniques.
Chhaya Chauhan, et.al,	2017	To understand the sentiments of the people products reviews are necessary, therefore a summary of positive and negative reviews are need to be generated.	Author's main focus was on the review of algorithms and techniques used to extract feature wise summary of the product and analyzed them to form an authentic review.
Pragya Juneja, et.al	2017	A novel approach was proposed to predict the Delhi Corporation Elections results. As today social networking sites like Twitter, Facebook, Instagram etc. plays a major role in sentiment analysis.	Author concluded that sentiment analysis can be utilized for any purpose based on the tweets they collected like marketing, finance, media, and entertainment and many more.
Mika V. Mäntylä, et.al	2018	An approach was proposed that utilized both the text mining and qualitative coding in order to provide the analysis on the large amount of gathered information by the author.	Author analyzed various paper and concluded that area of the sentiment analysis is very vast as it cover stock markets, elections, disasters, medicine, software engineering and cyber bullying and many more.
Pulkit Garg, et.al	2017	In this paper, we study post- terror attack tweets by extracting it from twitter. The flow data posted on twitter is used to study factors like last retweet, number of retweets and number of favorites.	Misleading information, the trends of retweets and number of favorites are the future scope to study its flow and survival.

Conclusion

The sentiment analysis is the technique which is applied to analyze sentiment. The sentiment analysis techniques has various phases which are data collection, data cleaning, and classification. In this paper, various sentiment analysis techniques are review and analyzed in terms of certain parameters. In future, the SVM based classification method will be further improved for the sentiment analysis

References

- [1] Feng S., Wang D., Yu G., Yang C., Yang N., "Sentiment Clustering: A Novel Method to Explore in the Blogosphere" *Advances in Data and Web Management*, vol. 5446, pp. 332-344 , 2009.
- [2] Andreevskaia A. , Bergler S. , "Mining WordNet For a Fuzzy Sentiment: Sentiment Tag Extraction From WordNet Glosses", *Proceedings EACL-06, the 11rd Conference of the European Chapter of the Association for Computational Linguistics*, pp. 209–216, 2006.
- [3] Fei, Geli, et al. "A Dictionary-Based Approach to Identifying Aspects Implied by Adjectives for Opinion Mining." *24th International Conference on Computational Linguistics*, 2012.
- [4] M. Edkrantz, S. Truve, and A. Said Alan, "Predicting Vulnerability Exploits in the Wild," *Cyber Security and Cloud Computing (CSCloud), Proc. 2015 IEEE 2nd International Conferen* pp. 513-514, 2015.

- [5] O. Owoputi, C. Dyer , K. Gimpel, N. Schneider, and N.A. Smith, “Improved part-of-speech tagging for online conversational text with word clusters” Proc. Conference of the North American Chapter of the Association for Computational Linguistics, 2013.
- [6] K. Gimpel, N. Schneider, B. O'Connor, D. Das, D. Mills, J. Eisenstein, M. Heilman, D. Yogatama, J. Flanigan, and N. A. Smith, “Part-of-speech tagging for Twitter: annotation, features, and experiments”, Proc. 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers , vol. 2, pp. 42-47, 2011.
- [7] Mondher Bouazizi and Tomoaki Ohtsuki, “A Pattern-Based Approach for Multi-Class Sentiment Analysis in Twitter”, 2017, IEEE.
- [8] Ankit Kumar Soni, “Multi-Lingual Sentiment Analysis of twitter data by using classification algorithms”, 2017, IEEE.
- [9] Aldo Hernández, Víctor Sanchez, Gabriel Sánchez, Héctor Pérez, “Security Attack Prediction Based on User Sentiment Analysis of Twitter Data”, 2016, IEEE.
- [10] Venkata Sasank Pagolu, Kamal Nayan Reddy Challa, Ganapati Panda, Babita Majhi, “Sentiment Analysis of Twitter Data for Predicting Stock Market Movements”, 2016, IEEE.
- [11] Chhaya Chauhan, Smriti Sehgal, “Sentiment analysis on product Reviews” International Conference on Computing, Communication and Automation, vol. 10, pp. 1-8, 2017.
- [12] Pragya Juneja, Uma Ojha “Casting Online Votes:To Predict Offline Results Using Sentiment Analysis by machine learning Classifiers” vol. 5, pp. 1-6, 2017.
- [13] Mika V. Mäntylä, Daniel Graziotin, Miikka Kuutila, “The Evolution of Sentiment Analysis - A Review of Research Topics, Venues, and Top Cited Papers”, Volume 27, February 2018, Pages 16-32, ISSN, 2018.
- [14] Pulkit Garg, HimanshuGarg, VirenderRanga, “Sentiment Analysis of the Uri Terror Attack Using Twitter” Department of Computer Engineering National Institute of Technology Kurukshetra , vol. 8, pp. 1-4, 2017.