A LITERATURE REVIEW ON ANALYSIS OF SENTIMENTS USING TWITTER DATA

Priya Vinod *, S.Sheeja

Abstract

Analysis of sentiments is also known as Opinion mining. The foremost purpose of applying opinion mining is to detect the behaviour, mood and attitude of the user. As social sites have taken an important place in people's lives, sentiment analysis is used to study the reactions of users in different areas such as politics, economy, finance and sociology. As the number of views or reviews, about various things, being shared by people on microblogging sites has increased, performing Sentiment Analysis of the same has been popularized in order to find out whether the engagement of the same is positive, negative or neutral and assess their product or new launches accordingly. In this paper I'm trying to review the work done for the same by data mining from Twitter by using different popular techniques.

Keywords - Sentiment Analysis, Twitter Data, Data Mining.

I INTRODUCTION

Social media has been the centre of the entire buzz even during the pandemic situation. That being said, Twitter being a player on its own status and being different from its cousins like Facebook and Instagram is gaining popularity easily. If we look at the statistics available, Twitter currently has 353 million monthly active users in 2021, which was just 54 million in 2010 This creates a gold mine of data available globally, granting access to opinions of people which can be leveraged for further studies related to socio-economic and even psychological aspects that affect markets and countries globally. But Twitter data that is available is unstructured and hence in need of a methodology to acquire the required

Department of Computer Science,
Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India
*Corresponding Author

aspects from the data available. An extensive set of filtering parameters will need to be set during the extraction of tweets from Twitter so that they match any specific criteria.

Following which Sentiment Analysis will be performed on the data to understand the sentiment or opinion of the people regarding a particular topic, for eg: how a newly released movie has fared with the public can be analysed by opinion mining. Sentiment Analysis or Opinion mining is to infer sentiment from text.

II SENTIMENT ANALYSIS

A. Why Twitter Data?

Earlier magazines and newspapers were the publishing medium for opinions. But, with a jump in technology, people have become more expressive on social media about their feelings. Hence making Twitter Data valuable to researchers as it can be used to further study behaviours of consumers, or for even far more important topics like political campaigns and even acceptance evaluating the acceptance of a policy by the public. Twitter information is open, personal, and in-depth. You can extract tons of data from a user by analyzing a number of their tweets and trends. You'll conjointly see however individuals are talking concerning specific topics exploiting keywords and business names. Tweets are unique, because they are shorter of lengths (280 characters). We can determine the sentiment of a tweet whether is it positive, negative, or neutral by doing sentiment analysis.

B. Why Sentiment Analysis

[1]Twitter sentiment analysis (TSA) may be an analysis space that aims to investigate the users' sentiment or opinions toward entities-such as topics, events, people, issues, services, products, and organizations-and their attributes

supported by the content of their tweets. Sentiment mining could be a work that takes benefit of the information processing and data extraction (IE) perspectives to research an in-depth variety of archives so as to collect the feelings of comments exposed by various authors.

Using Sentiment analysis we can make decisions of whether the written or spoken language is of a particular nature i.e. whether the sentiment is positive, negative or neutral. The tools available for analysis at present are able to handle huge volumes of client criticism dependably and exactly. We are able to not just deal with the contents under investigation, but in doing so, Sentiment Analysis also throws light on customer opinions on various topics, along with procuring things, arrangement of services or even presenting promotions.

So, Sentiment Analysis or opinion mining can be a technique of data mining which can overcome the difficulties by segregating and breaking up, in a methodical way, theinformation on the web without involving much delays.[9] On further examination, the advertisers were able to measure and gauge the emotions of shoppers' and their states of mind, even though the volume and structure of knowledge available was difficult. The passion of this study to utilize sentiment analysis to promote research is double.

III LITERATURE REVIEW

Already many studies have been conducted in the area of Sentiment Analysis or Opinion Mining. I would like to point out that the scope of research on the same is even more as data is readily available from various sources and the right kind of analyses done can result in an enhanced study report for the field concerned, it can be advertising or social issues or studying human behaviour further.

There are multiple ways of doing Sentiment Analysis. The first starting with the different levels of Sentiment Analysis and then moving on to the methods mentioned in the fig 2.
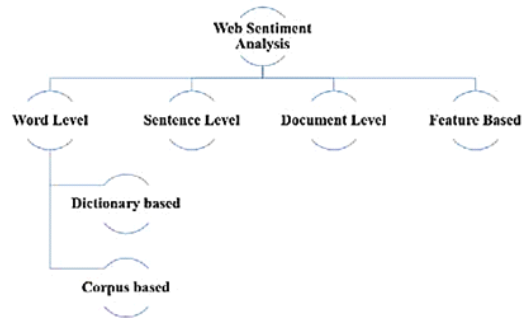
Fig 1. Levels of Sentiment Analysis

[2] Sentence level analysis gives us the opinion, sentence by sentence. Every sentence is analyzed in order to read the sentiment which can be positive, negative or neutral. Not providing an opinion is termed as a neutral sentiment. This type of analysis provides us with a subjective level of sentiments, whether the sentence is good or bad.

The document-level analysis is performed over the whole document. The results obtained are that whether the document gives us a positive sentiment or negative sentiment. So, a document that contains comparative learning cannot be considered for document-level analysis.

Document and sentence level analysis does not provide us with any information about the likes and dislikes of the peoples. The feature-based level gives us direct information about the attitude of the people through their opinion. This level is also called Entity/Aspect based level.

A text classification was another issue that needed to be resolved in sentiment analysis. Machine learning approach and Lexicon based approach are the 2 broader categorizations of these classification approaches.

[3]The text is classified into different classes by the machine learning-based approach using many classification techniques. The two broader categorizations of these machine learning techniques are (1) unsupervised learning:

Here there is no category involved and the targets are not provided by them at all. So, clustering is considered to be an important factor here. (2) Supervised learning: The labelled dataset is used to develop this method. When the classification approach needs to be designed, the labels are provided to the model.

To get significant outputs when going through decision making, these labeled datasets are trained. The success of both these learning techniques lies in the fact that the determination and extraction of particular sets of features enable us to detect sentiments.

BM Bandgar [10] developed the indigenous Windows based user friendly application in Java to extract process and classify the real time social network tweet using unstructured models. [6]The meaningful real time tweets are obtained and the same is used for sentimental analysis. The processed meaningful tweets are classified into three different opinion mining classes positive, negative and neutral by using unstructured algorithms such as EEC, IPC and SWNC model. The SWNC Model gave better results over the EEC and IPC model. Their results are compared using the confusion matrix, precision and accuracy parameters. The results are also visualized using pie graph.
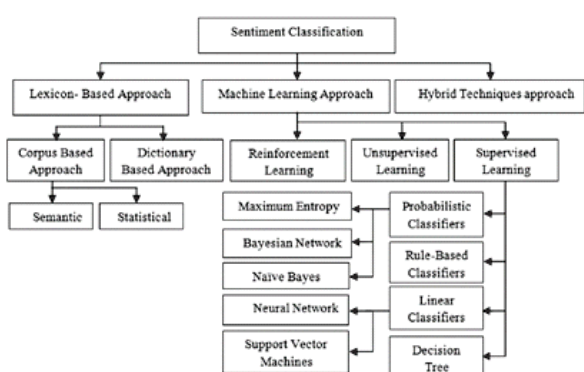

Fig 2. Classification of Sentiment Analysis Techniques

A. Lexicon Based Approach

The first technique that may be used for Sentiment Analysis is the lexicon-based technique. It uses a lexicon that consists of terms with various sentiment scores for every term. The term is related to one word, phrase or idiom. The sentiment is outlined and supported by the presence or absence of terms within the lexicon. The lexicon-based approach includes the corpus-based approach and dictionary-based approach.

1) Dictionary Based Approach

[7]The main plan behind the dictionary-based approach is to use lexical databases with opinion words to extract sentiment from the document. Supported by a group of seed sentiment words (e.g. good or bad) with their polarities are collected by hand. In the starting, this first set doesn't get to be massive, thirty opinion words is enough. The ensuing step is to use the polar words to complement a group by trying to look up several synonyms and antonyms in a computer database. Samples of such databases square measure WordNet , HowNet , SentiWordNet SenticNet, MPQA etc. The look-up procedure is repetitious. At every iteration, the formula takes an updated set of words (expanded set) and will search once more till there'll be no new words to incorporate. In the end, a group of sentiment words will be reviewed with the aim of deleting errors.[5]This methodshowed good results with an average accuracy of 84%. So, the current method can be effectively used for prediction of adjective semantic orientations and to find sentence polarity.

2) Corpus Based

The corpus-based approach may be applied in 2 cases. The primary case is the identification of opinion words and their polarities within the domain corpus employing a given set of opinion words. The second case is for building a brand-new lexicon among the actual domain from another lexicon employing a domain corpus. The findings recommend that

though the opinion words measure is domain-dependent it will so happen that the identical word can have an opposite orientation betting on the context.

In the field of machine learning, methods of classification to classify unlabeled information were developed that uses completely different methods. Although the classifiers would need a bit of coaching information. [8]Samples of machine learning classifiers are the Naive Thomas Bayes, most Entropy and Support Vector Machine. Further classified as supervised-machine learning ways as they need coaching information. It's vital to say that coaching a classifier effectively gives way to easier future predictions.

B. Naïve Bayes Classifier

[9] Considerable numbers of features are utilized in the feature vector through the Naïve Bayes classifier. Since these features are independent equally, analyzing them exclusively is significant. The mathematical representation of the conditional probability for Naïve Bayes is given as:

$$P(X|yj) = \prod_{i=1}^{m} P(x_i|y_i)$$

Eq. 1

A feature vector denoted by "X" is included, which is defined by $X=\{x_1,x_2,....x_m\}$. The class label is represented by yj. The classification of various forms of independent features like positive and negative keywords, emoticons and emotional keywords is completed expeditiously by exploiting the Naïve Bayes method. The relationships amongst features are not considered within the Naïve Thomas Bayes classifier. Thus, the relationships that exist among emotional keywords, negation words and speech tag are not utilised in it.

The Naive Bayes Classifier is usedin ccategorizing texts into different classes.

C. Support Vector Machine Classifier

The Support Vector Machine (SVM) executes itself beautifully in sentiment analysis. [10] The SVM checks the information, differentiates the limited choices and makes use of the constituents for calculation, this is done in the input space. The critical information is then presented as two schemes of vectors, each of size m. Each datum (which is represented as a vector) is organized into a class at this point. Following which, the machine will then identify the partition between the two classes that are much further away from any place that were in the samples used in training. The separation defines the classification edge, broadening the edge will lessen the contradictory choices. The SVM has demonstrated to execute much more competently than the Naïve Bayes in multiple text categorizing problems.

In the papers I have studied, using both techniques have resulted in impressive results. So, I would not sway on choosing any particular method to mine data from Twitter.

IV PROBLEM FORMULATION

The behaviour of users is analyzed through sentiment analysis techniques. There are different social networking websites on which sentiment analysis can be applied to get insightful data about the different opinions of users regarding multiple topics. The pattern matching algorithm is applied in sentiment analysis to extract the features of input data. Further, the classification techniques are applied in order to detect any sarcasm. [11]Along with neural networks, the pattern-matching was applied in this proposed technique. For the classification of features, the Naïve Bayes classifier was applied in multiple papers. And there were two major issues highlighted in this research. Using the Naïve Bayes classifier alone would not allow us to evaluate sarcastic and ironic statements as a whole. And classifying sentiments as only positive, negative or neutral sentiments could also be another issue in terms of knowing a wider audience who are somewhat agreeing, but doesn't fully agree or the opposite

category. So, more inclusion of opinions may be required in many fields of study and research.

V CONCLUSION

Multiple techniques have been employed by researches in order to mine data from Twitter for the purpose of analysis of sentiments. The Naive Bayes and SVM have been able to achieve an accuracy of approximately 80%.

Alterations in the demonstration of Sentiment Analysis algorithms where many characteristics are considered can be an interesting area for the study in future. Combining multiple features has shown to lead in improvement of performances in most of the cases and showed substandard performances in others. Exploring why the instabilities occur in the performance is a compelling direction for future work. Another path for future study will be to find the issue of data sparsity by using together the ensemble and hybrid approaches. The reasonhere is to measure the robustness of various Twitter sentiment approaches to the data sparsity. Another space of study can be the use of active learning techniques to sight Twitter sentiments and to extend the boldness of decision-makers.

REFERENCES

[1] Abdullah Alsaeed and Mohammad Zubair Khan. A Study on Sentiment Analysis Techniques of Twitter Data, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 2, 2019.

[2] Priyanka Tyagi and Dr. R.C. Tripathi. A Review towards the Sentiment Analysis Techniques for the Analysis of Twitter Data, 2nd international conference on advanced computing and software engineering (ICACSE-2019).

[3] Vikas Malik and Amit Kumar. Sentiment Analysis of Twitter Data Using Naive Bayes Algorithm. International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-169 Volume: 6 Issue: 4 (April 2018)

[4] Pak, A., &Paroubek, P. (2010, May). Twitter as a corpus for sentiment analysis and opinion mining. In LREc (Vol. 10, No. 2010).

[5] Scott, J. (2011). Social network analysis: developments, advances, and prospects. Social network analysis and mining, 1(1), 21-26.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[6] Huma Parveen, Shikha Pandey. Sentiment analysis on Twitter Data-set using Naive Bayes algorithm. 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT).

[7] V. Kharde and P. Sonawane, "Sentiment analysis of twitter data: A survey of techniques," arXiv preprint arXiv:1601.06971, 2016.

[8] Boguslavsky, I. (2017). Semantic Descriptions for a Text Understanding System. In Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference "Dialogue"(2017) (pp.14-28).

[9] El Rahman, S. A., AlOtaibi, F. A., &AlShehri, W. A. (2019). Sentiment Analysis of Twitter Data. 2019 International Conference on Computer and Information Sciences (ICCIS).

[10] BM Bandgar, S Sheeja, "Analysis of real time social tweets for opinion mining", International Journal of Applied Engineering Research 11 (2), 1404-1407

[11] Shweta Rana, Archana Singh. Comparative analysis of sentiment orientation using SVM and Naive Bayes techniques. 2016 2nd International Conference on Next Generation Computing Technologies (NGCT).