



EMOTION RECOGNITION OF NEWS DATA FEED USING SOCIAL MEDIA DATA

Mr.R.Subash¹, Mr.Kotha.Saideshwar²,
Mr.Prashanth.Mamillapalli³
Computer Science & Engineering
SRM Institute of Science and Technology
Kattankulathur, Chennai, India
subash.r@ktr.srmuniv.ac.in
ksaideshwar@gmail.com
lovelyprashanth23@gmail.com

July 6, 2018

Abstract

The Emotion recognition of the present day events of new data that is growing across the globe is significant increase in the analysis of social media data for computing which has approached to the market trends for analysis for researchers that which could solve some real world applications for the benefit to the society. For example government policy making to the public to take opinion of the people for better e-governance. In day-to-day life we have come across so many difficulties for the purpose of performing required tasks. But for that tasks we should depends upon some third party members which include people who write blogs, forums and some electronic-commerce company websites for analysis purposes. By invoking semantics all possible results for similar emotions are obtained, hence results are efficient when compared with keyword based search engine. Finally

user views, opinions and attitude towards objects, individuals, events and topics are classified in 6 categories namely fear, joy, sad, love, anger and surprise. But we live today world that social media data is getting more and more every day. This shows that the data is more data in the future. Today people are educated to know whether the purpose of work to be done which will be positive or negative. These people only gives opinion and feelings publically to make valuable decision making process. From these approaches, they are being extracted and analysed in proper manner to get accurate progress. Some accuracy of emotion finding algorithms has been proposed to improve the accuracy.

Keywords: Emotion computing, Big data, Word Emotion Association Network, Social Media.

1 INTRODUCTION

Irregular increase in population has ensued people personal views to express and exchange their interests and opinions on the new data that they come across some stages. The media platforms like facebook, google, tawk, and wechat have become the most important platforms for the people to express their opinions. But there is no system that is effective enough to classify peoples opinions and interests. At present News social media data are news events which have occurred in the today's world.

Due to the following reasons the demand for a new system has been increased. The first one is the volume that is the large number of the news events are been coming daily. The second is the variety in which the events formed are different from each other. The third is the at which the News events are coming. For example the every day data extracted from twitter is about more than 700TB now-a-days. In this way after the occurrence of the news event, many people discuss on these events on social media platforms. This would be helpful for the corporations and governments to find the valuable information through the news event analysis.

This type of system has the resistance capability to discover the emotion of a particular event that had happened. This may be used in many areas for successful implementation for projects.

For suppose there is lot of information available on the net about a news event so, there is greater possibility to monitor public opinions and concerns which helps the government to respond quickly. Many companies are interested to know about their product Feedback from the customers which will be helpful to correct their mistakes. This system helps to accurately determine the emotions of the customers which could help the corporations to evaluate the design and make decision on future design. The Existing system has few drawbacks which can be corrected in the proposed system. The existing system only focus on adjectives and adverbs for emotion classification. In existing system the semantics are ignored and there is no word net effect. The Emotion recognition mainly uses machine learning methods to extract, identify, the emotion of the news events. The Classification of the news events is done into two categories (positive and negative). Though there are efficient systems and algorithms for Emotion computation, it lacks. Emotion analysis increases the efficiency of the existing system.

2 LITERATURE SURVEY

The literature survey is done from different papers to get an idea of how this proposed system is implemented. Dual Sentiment Analysis that which provides for the emotion analysis of the data that which is collected divided into positive, negative and neutral. It is very effective for polarity classification and extending the DSA algorithm to DSA3.

Gurpreet Kaur, Opinion Mining and Sentiment Analysis paper provides Emotion analysis. Emotion detection that which is subject that being approached by many students in different levels of research work which is nothing but called Opinion mining. It can perform Emotion analysis quickly and in an effective manner. The drawback is Cannot handle more complex polarity shift patterns.

Md. Ansarul Haque, Emotion Analysis by using Fuzzy Logic works with truth tables that which are membership values that are 0 and 1. It is helpful for anyone in any way to meet up their interests or what they deserve. In this Analysis with more tweets

reviews is a complex process.

Muhsina KM, Product Rating using Emotion Analysis deals with Pos tagging method means nothing but that sentence is taken into consideration and then sentence is tokenized into words .That words are applied for Parts of Speech tagging and then it gets analysed.The drawback is main analysis is done only for mobile phones.

Nirmal Robert, Emotion Analysis depends upon the data that which is collected by the researcher to analysis purposes which includes the review of the people upon so many issues for the opinion mining.These data that which is collected should be more enough because analysis can be done on more data for accuracy purpose.In this way of process text summarization also becomes more importance because more data means indirectly taking more time and inefficient for the programming side.

3 PROPOSED SYSTEM

In the proposed system, Emotion computing for a news event is done using semantics. By invoking semantics more possible words with similar emotions are obtained. Emotion recognition aims at dealing with social media News feed. The data is been collected from various social media platforms like twitter . Then the data is been classified into two categories that is positive ,negative and neutral. The proposed system mainly consists of two procedures. One is finding the emotion of the sentence based on the word itself. The other one is by finding the semantics of the sentence. The Existing system has few drawbacks which can be corrected in the proposed system

Fig.1 describes the entire process of our proposed system. Here in the first step of the process the data collection is been done from various social media platforms. Then the next step of the process is to categorize then into different categories. Then the next phase of the process is data modeling .There are two types of modeling text-based modeling and network-based modeling. In text-based modeling the data is been classified based on its emotions like sad, anger, love, surprise. In network-based modeling the classification is been done based on type of network. The last step of the process

is Machine learning. In this there are three process that are used. They are data mining, optimization and validation.

3.1 MODULES

Social media news feed prepared to receive positive feedback and negative feedback. Latest technology has given people a chance to express their opinions and views in public platforms. In order to respond immediately to the people, it is important to take immediate action and act forward. Both negative and positive feedback arise in social media. The form of people views will be in form of sentiments and reviews. To get involved in social media then you need to be

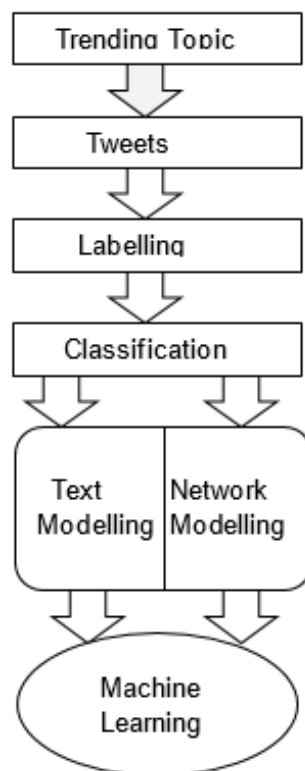


Figure 1: Architecture Diagram

1. Pre-Processing

Data preprocessing is done to remove the incomplete noisy and inconsistent data. Data must be pre-processed before using in feature selection task. In pre-processing following are some tasks: Removing URLs, Special characters, Numbers, Punctuations etc. Removing Stop words Removal of Retweets (in case of twitter dataset), Stemming, Tokenization.

2. Feature Selection

Feature selection from pre-processed text is the difficult task in sentiment analysis. It is the process of preparing the models which to make in a convenient way of understanding to interpret for the users. This also includes shorter training datasets for times and curse of dimensionality can be neglected. By this way we can achieve reduction of variance. In this we can create new feature extraction from original features in which written in texts and array. POS tagging is used in feature selection techniques.

3. Sentiment Word Identification

Sentiment word identification is basically nothing but the words in the sentence are taken into consideration for the process of identifying that the particular word comes under which word emotion which leads to the robustness. These emotions may be positive, negative and neutral.

4. Sentiment Polarity Identification

The basic task in Sentiment Polarity is nothing but calculation of polarity in the particular sentence or word or paragraph given. The polarity is in three category i.e. Positive, Negative and Neutral. Polarity identification is done by using different lexicons which help to calculate sentiment score, sentiment strength etc.

5. Sentiment Classification

Sentiment classification of news review dataset and product review dataset is done using supervised machine learning approaches like naive Bayes, SVM. Accuracy is depends on

which dataset is used for which classification methods. In the case of Supervised machine learning approaches Training dataset is used to train the classification model which then help to classify the test data.

Algorithm

Naive Bayes Classifier

The fundamental of Naive Bayes algorithm has been classified into two types of features

1. independent
2. equal

Bayes' Theorem

According to the Naive Bayes Classifier before knowing about this we should know about Bayes Theorem of probability to solve the problem of analysis of dependencies of occurring in the dataset. To the probabilistic model is nothing but the conditional probability model. If it takes the values of the instance to which classification is to be done. If classification done only upon the decision rule to pick the hypothesis that is most probable. The calculation is made by assumptions that are equi-probable classes or calculating estimation probability from training dataset available. These type of the feature models are called Event model.

Gaussian Naive Bayes classifier

We use this when the data is in form of continuous. Assume, that in a training dataset, we have point called z with regard to the first segment the data by the class and then compute the mean and variance of z in every class of reference. In this training dataset we have one observation value. By applying the formula we get the probability distribution of occurrence of the particular event.

4 IMPLEMENTATION

Implementation of the project is done by first selecting the topic of the present day new feed that is from either twitter or websites through web scraping method or any extracting method available and taking apart of the performance. After collecting datasets of

$$P(Z/A)=P(A/Z)*P(Z)/P(A)$$

$$P(A/Z_1,Z_2,...Z_n)=P(Z_1/A)*P(Z_2/A)...P(Z_n/A) / P(Z_1)*p(Z_2)...P(Z_n)$$

Figure 2: Naive Assumption

data, then we should form some training dataset for analysis which is our main aim of the project.

Collected data set is divided into like comma separated value tables to separate which is like tweet or review or comment. After segregating the data into csv format then the pre-processing module starts its action in this particular module five types of extracting will be done.

First process is removal of stopwords, numbers, punctuations, special characters from datasets for analysis is done in a convenient way. Second process is removal of stopwords that are present in datasets.

Third process is removing of duplicate data that is repeated. Fourth process is stemming which means that data is converted into semantics for the analysis.

Fifth process is tokenization which is converting data into pieces for better execution of the process.

Feature selection is the process of collecting the data after pre-processing is done successfully and applying the parts of speech tagging so that the data segregated which be converted into particular feature which it belongs.

After feature selection is done perfectly then Sentiment Word Identification module is executed to see that data is extracted before module into sentiment of the each and every word with help of Natural Language Tool Kit. This will help to divide every single word into positive or negative or neutral.

Sentiment Polarity Identification is a next module when a sentence or a document is given for sentiment analysis. Polarity is classified into three types positive, negative, neutral. Polarity identification is done by using different lexicons which help to calculate sentiment score, sentiment strength etc.

Lastly after polarity identification our Navie Bayes Algorithm is applied through the dataset is collected after all modules executed data, algorithm classify the particular each and every sentence into sentiment classification into three categories positive, negative, neutral.

Finally, analysis of the result is very important to make valuable decision to the particular company. Analysis is the best business intelligence.

5 CONCLUSION

In this proposed work, we used Twitter API using Microsoft Visual Studio tool which is open source. Tweets from twitter has been collected and gives to pre-processing task in that tool. Microsoft Visual Studio open source tool is used in text mining and also to crawl streaming data from social media like twitter and Facebook etc. Movie reviews data also pre-processed in R tool for sentiment analysis and opinion mining. There are different supervised and unsupervised approaches and different lexicons, dictionaries and corpus based methods which are very helpful in Sentiment Analysis. Different dataset are available for movie review, product review, Epinions dataset etc. In this method sentiment score has been calculated and counted number of positive, negative and neutral tweets for given Hashtag and can predict the public opinion of particular event. As per above analysis of different #Hashtags tweets for sentiment analysis, individual and industry can find the public opinion behind that event.

6 ACKNOWLEDGEMENT

It gives us immense pleasure to express our profound gratitude to our Department of Computer Science and Engineering, SRM Institute of science and technology, Kattankulathur, for encouraging and giving us support and good environment to gain valuable experiences.

References

- [1] H. Binali, V. Potdar and C. Wu, "A state of the art opinion mining and its application domains", 2009 IEEE International Conference on Industrial Technology, 2009
- [2] Sciencedirect.com, "On newspaper headlines as relevance optimizers", 2015. [Online]. Available:<http://www.sciencedirect.com/science/article/pii/S0378216602001340>
- [3] V. Singh, R. Piryani, A. Uddin and P. Waila, "Sentiment analysis of movie reviews: A new feature based heuristic for aspect-level sentiment classification", 2013 International Multi-Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4s), 2013. [4]
- [4] M. Karamibekr and A. Ghorbani, "Sentiment Analysis of Social Issues", 2012 International Conference on Social Informatics, 2012. [5] T. Thet, J. Na, C. Khoo
- [5] S. Zhang, J. Liang, R. He, and Z. Sun, "Code consistent hashing based on information-theoretic criterion," Big Data, IEEE Transactions on, vol. 1, no. 3, pp. 8494, Sept 2015.
- [6] J. Xuan, X. Luo, G. Zhang, J. Lu, and Z. Xu, "Uncertainty analysis for the keyword system of web events," Systems, Man, and Cybernetics: Systems, IEEE Transactions on, vol. PP, no. 99, pp. 11, 2015.
- [7] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. Association for Computational Linguistics, 2002, pp. 7986.
- [8] H. Cui, V. Mittal, and M. Datar, "Comparative experiments on sentiment classification for online product reviews," in AAAI, vol. 6, 2006, pp. 12651270.

- [9] S.-M. Kim and E. Hovy, “Automatic identification of pro and con reasons in online reviews,” in Proceedings of the COLING/ACL on Main conference poster sessions. Association for Computational Linguistics, 2006, pp. 483490.
- [10] W. McDougall, An introduction to social psychology. Courier Corporation, 2003.

