iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 Issue: III Month of publication: March 2019

DOI: http://doi.org/10.22214/ijraset.2019.3184

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue III, Mar 2019- Available at www.ijraset.com

Prediction of Opinion Behavior using Content based Sequential Opinion Influence Model

Karthik Elangovan[1], Sharmila R[2], Vijayadharshini B I[3], Sarika Yadav R[4]
[1] Department of Computer Science and Engineering, S.A Engineering College

***Abstract:** Micro-blogging and social networking sites play a vital role in identifying and extracting distinct opinions and emotions from public. Social media has become a viral medium for public to express and exchange their opinions through the user-generated text. The aim of this paper is to classify the tweets using Content based algorithm and Support Vector Machine (SVM). The tweets are also analyzed and classified in terms of polarity: Positive, Negative and Neutral. To keep track of user opinion behaviors and infer user opinion influence from the historical exchanged textual information, we develop a content-based sequential opinion influence framework with the help of two opinion sentiment prediction models with alternative prediction strategies are proposed.*
***Keywords:** Support vector machines (SVM), Opinion influence, Collaborative filtering, Social media, Twitter*

I INTRODUCTION

Classification of opinion is becoming very important as they play a vital role in decision making. Sentiments expressed via micro blogging sites such as twitter, Facebook, etc. helps in understanding the mindset of people. Several models have been proposed to investigate and extract distinct opinions from users where the accuracy is important. Text mining and data mining helps to identify patterns and establish relationships to solve problems in analyzing large datasets. Tools that are used for data mining allow commercial enterprises to draw conclusion for future trends. We propose a content-based sequential opinion influence framework to incorporate the content information with the historical information for sentiment prediction in opinion dynamics. Two models with different prediction strategies are proposed. The experiments conducted on the Twitter dataset demonstrate the effectiveness of the two proposed models. The prediction ability of the proposed model is further verified on the opinion word prediction task. Based on the learned influence, we explore the expression styles of users with different influence powers, which provide the valuable information for companies to manage their accounts and design marketing plans. The current work only considers the influence from neighbors. There are other useful information sources such as the Twitter trends. In our future work, we will consider to extend our proposed framework to capture the influence of the external information sources on user's opinion behaviors.

II RELATED WORK

Fiernad Napitu [1] has proposed that Churn rate analysis has numerous techniques that make use of the following customer activities such as payment behavior, usage, complaint data and tenure. This paper deals with the churn rate prediction with the help of churn analysis techniques which are in need to acknowledge a number of other important factors. The value of the churn rate of the product is correlated with the collection of opinions about a product. Convolution neural network (CNN) architecture helps in classifying positive, negative, and neural sentiments. Granger causality analysis helps in cross validating the accuracy of churn rate prediction. The network's parameter is trained with the 5-step process to achieve accuracy as the resulting Mean Absolute Percentage Error (MAPE) is 1.47%. This paper as mentioned makes use of CNN for classification and recurrent neural network (RNN) in order to increase the performance. Sanchita Kadambari, Kalpana Jaswal, Praveen Kumar, Seema Rawat [2] has proposed that Twitter has emerged as a popular micro-blogging platform in expressing one's opinion and ideas. It is because on twitter millions of users are tweeting each and every second which eventually generates a continuous influx of real time data. Business and Commercial organizations are tapping the data posted on twitter.com in order to extract the general opinion or sentiment expressed on a topic. The applications of this paper are numerous and wide ranged. The real time automated sentiment analysis process will extract the trends and helps in generating values. The Dictionary based and Corpus based approaches are used to understand the granularity of twitter tweets at word level. This paper makes use of Machine learning and Natural language processing algorithms to understand the sentiments in an automated manner. Brian Heredia, Joseph D. Prusa and Taghi M. Khoshgoftaar [3] have proposed that Tweets from twitter application are popularly used to express opinions, especially when the topic of choice is polarizing, as it is in politics. The creation of dataset consisting of about 3 million tweets on twitter starting from 22nd September to 8th November about Hillary Clinton and Donald trump is utilized. This paper deals with two approaches in analyzing the voter's opinions. The Convolution neural network (CNN) architecture is trained on Sentiment140 dataset and the Tweet election dataset. The location and the need to weigh tweets based on their population are necessary to be concerned about. The election prediction is said to be successful when the poll results which are extracted before 13 days of election appears to be approximately close. The Bots detection algorithm must be used as the 2016 election was flooded with bots. Considering the twitter application as a polling resource for the first time has worked efficiently and turned out to be successful. Pragya Tripathi, Santosh Kr Vishwakarma, Ajay Lala[4] have proposed that Data mining techniques are used for the purpose of sentiment classification analysis on opinions shared by the people on twitter.com. This paper focuses on collecting the tweets from twitter API which is normally present in natural language and applying text mining techniques such as Tokenization, Filtering stop words Stemming, etc. This is required for converting them into a useful form and for building classifiers to predict and categorize a single tweet into positive, negative, and neutral. Rapidminer is the tool that helps in building classifier and testing dataset. The two classifiers used are Nave bayes and K-NN where the latter gives more accurate prediction. The maximum size of testing dataset should be considered. The Text mining is performed on rapidminer to build classifiers for testing. This paper as mentioned helps in categorizing emotions and additionally can be applied to other Indian languages as well (Tamil, Hindi, Marathi, Gujarati, etc.).Yoad Lewenberg, Yoram Bachrach, Svitlana Volkova[5] have proposed the technique that is focused on this paper depends on the concept of training Machine Learning models to categorize the opinions expressed via tweets, considering the Ekman's six high level emotions. Raters used are sourced from Amazon's mechanical Turk in order to investigate many twitter profiles and to extract the user's interested areas like economics, politics, sports, news, movies, etc. It is designed in such a way to analyze self-disclosure processes both in online and offline environment. Several models are presented to make use of emotions of twitter user by analyzing their tweets to predict their degree of interest on a particular topic. This paper can help tracking the issues that is discussed between two people and in understanding about how people come forward to disclose information about them. The accuracy and causal inference requires improvement. Rincy Jose, Varghese S Chooralil [6] has proposed a Lexicon based sentiment analysis method is adopted to exploit the sense definitions, as the semantic indicators of sentiment. This paper deals with the novel based approach for accurate classification of sentimental twitter messages with inclusion of lexical resources or sentiment lexicons such as SentiWordNet and WordNet along with the Word Sense Disambiguation (WSD). The sentiment lexicons are used for finding political statements from real time twitter messages. Word sense disambiguation and Negation handling in sentiment analysis are used for improving the accuracy. The negation handling is used as a pre-processing step in order to achieve high accuracy. This paper says that negation handling gives 1% of improvement while WSD gives 2.6% of improvement in the classification accuracy. The sentiment classifiers are used only along with the Word Sense Disambiguation. Wafa Alorainyi, Pete Burnapi, Han Liui, Amir Javed, Matthew L. Williams [7] have proposed the problem of the unreliable Human annotation in detecting the hate speech on social media or web is addressed in this paper. The implementation of emotion analysis is done on three sources of datasets like Suspended, Active and Neutral accounts. Two random Forest classifiers are made use on twitter messages of both suspended and active accounts. The result on two classifiers shows that the tweets from suspended accounts have outperformed the tweets from active accounts by 16% of overall F-score. It is expected to analyze and extract more subtle harassment produced by suspended accounts. This paper results in an increased rate of 16% in detecting the hateful tweets. Here the data annotation is used which is more efficient than human annotation because the emotion extracted from web states that the tweets from suspended accounts express more anger, fear, disgust and sadness emotions than the case of active accounts but the human annotators assume both as hateful ones only. Namita Mittal, Divya Sharma, Manju Lata Joshi [8] have proposed a Image sentiment concept like Adjective Noun Pairs (ANP) is used to automatically discover the tags on web or social media images for detecting the emotions on image. The main aspect to be concerned is identifying emotions of the unlabeled images. This paper helps in overcoming it by using the deep learning techniques for sentiment analysis as the deep learning models can learn the image behavior effectively. This paper deals with some other noteworthy models of deep learning such as Deep learning neural network (DNN), Convolution Neural Network (CNN), Region-based CNN and Fast R-CNN. Among these models, CNN is found to be efficient and the training time of the model is also reduced. Many other models can be designed to get even better optimum results by making use of other datasets such as flicker, twitter testing and SentiBank datasets. S.Geetha, Vishnu Kumar Kaliappan [9] has proposed the usage of emoticons and text messages helps in expressing one's state of mind. The Novel Future Prediction Architecture based on Efficient Classification (FPAEC) is designed by making use of several classification algorithms such as Fisher's Linear Discriminant Classifier (FLDC), Support Vector Machine (SVM), and BIRCH clustering algorithms. Initially, the distinct classification algorithm's efficiency is evaluated. Finally, the performance of the text analysis is improved by using efficient classification algorithm. This paper uses FPAEC which is used for word context in association with the emotions and punctuations to predict the piece of different texts, instead of using the particular word and emotion alone. It should be

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue III, Mar 2019- Available at www.ijraset.com

designed in such a way that the neural tweets could be studied with datasets that are enhanced with analogue domain. The automatic interpretation of several emotions in other applications needs to be developed based on public moods. Rebeen Ali Hamad, Saeed M. Alqahtani, Mercedes Torres [10] have proposed the beneficial and accurate indicators of public emotion and sentiment are social media and micro-blogging sites. This paper deals with the classification of tweets to their respective classes by making use of cross validation and partitioning of data techniques across the cities using the supervised Machine Learning algorithms. The better and reasonable accuracy rates are provided by classifiers such as K-NN, Naive Bayes, SVM and ZeroR. However, K-NN has provided the best accuracy rate of 96.58% and 99.54% for ipad and iPhone emotion datasets. But when it comes to the cities, the K-NN achieves 100% for all polarity datasets. The attributes that needs to be concerned are handling implicit attributes of products and dealing with negation opinion expressions. Yujiao Li, Hasan Fleyeh [11] has proposed the capability of computers to learn without being explicitly programmed is known as machine learning. In this technology, customer opinion regarding the opening of new IKEA stores is analyzed with respect to the tweet frequency in four cities which is used as a data source for implementing sentiment analysis. This paper focuses on the lexicon based approach for English tweets and machine learning methods for computing sentiment polarity for Swedish tweets. Here in this paper, the analysis procedure comprises of sentiment prediction of individual tweets, correlation of words with IKEA and positive and negative words which are frequently used. Sentiment estimation for every tweet is done for every tweet with their individual scores i.e., the probability of positivity. Since tweets are restricted to be less than or equal to 140 characters, emoticons are be used for assuming the entire message. The main advantage of this paper is the sentiment estimation that IKEA event has been welcomed by residents due to high frequency of tweets related to IKEA are positive, and negative tweets were less. This paper tells us that positive opinions can improve reputation of government and negative opinion because direct loss of investment.

III PROPOSED WORK

Content-based sequential opinion influence framework is proposed to incorporate the content information with the historical information for sentiment prediction in opinion dynamics. Two models with different prediction strategies are proposed. The experiments conducted on the Twitter dataset demonstrate the effectiveness of the two proposed models. The prediction ability of the proposed model is further verified on the opinion word prediction task. Based on the learned influence, we explore the expression styles of users with different influence powers, which provide the valuable information for companies to manage their accounts and design marketing plans.
In this project, we are having following modules,

A. Twitter API

User needs to register first by giving his/her own information. While registering user should give their exact current location. If he/she is giving wrong location means, he is not supposed to register and login. And that user will be considered as a blocked user. So user needs to give only the current location. After registration user will login with username and password. Then he/she can see their profile and can view all user tweets. Admin needs to login with username and password. If both match, he/she will be considered as a valid person. After login, admin can view all blocked user who gave wrong location while registration. Admin can able to see all users profile and tweets.

B. Post Tweet

In this module registered user can post tweets in twitter. If the user tries to post any tweet which contains bad words means, it will not get posted in the twitter account. So the algorithm will restrict the user not to post bad words. The general tweets can be posted in application and as well as in twitter.

C. Search Query

Here in this module, user can search for any query in the application. The query has been processed and extracted live tweets from the real time twitter. The Keywords related 100 tweets are extracted from the live twitter.

D. Preprocessing

In this step all the tweets are extracted from twitter are processed and the noise data are removed.

1) Stop words Removal: A dictionary based approach is been utilized to remove stop words from tweets. A generic stop word list containing 75 stop words created using hybrid approach is used. The algorithm is implemented as below given steps. The target text is tokenized and individual words are stored in array. A single stop word is read from stop word list. The stop word is compared to target text in form of array using sequential search technique. If it matches, the word in array is removed, and the comparison is continued till length of array. After removal of stop word completely, another stop word is read from stop word list and again algorithm runs continuously until all the stop words are compared. Resultant text devoid of stop words is displayed, also required statistics like stop word removed, no. of stop words removed from target text, total count of words in target text, count of words in resultant text, individual stop word count found in target text is displayed.

2) Stemming Technique: After removing the unwanted words from the tweet, stemming technique is processed. Stemming is the process of reducing inflected (or sometimes derived) words to their word stem, base or root form generally a written word form. The stem need not be identical to the morphological root of the word; it is usually sufficient that related words map to the same stem, even if this stem is not in itself a valid root.

E. Classification

After stemming process, all the tweet terms containing the keyword are classified into positive, negative and neutral tweets. Collaborative filtering algorithm is used for classification. Here we are having good words and bad words datasets. By comparing with this, we can classify the tweets into positive, negative and neutral tweets.

F. Graph

Finally the graph is generated between positive, negative and neutral tweets. It is based upon the count of the positive, negative and neutral tweets. count of the positive, negative and neutral tweets.

IV CONCLUSION

The experiments conducted on the Twitter dataset demonstrate the effectiveness of the two proposed models. The prediction ability of the proposed model is further verified on the opinion word prediction task. Based on the learned influence, we explore the expression styles of users with different influence powers, which provide the valuable information for companies to manage their accounts and design marketing plans.

REFERENCE PAPERS

[1] F. Napitu, M. A. Bijaksana, A. Trisetyarso and Y. Heryadi, Twitter opinion mining predicts broadband internet's customer churn rate, 2017 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom), Phuket, 2017, pp. 141-146.Doi:10.1109/CYBERNETICSCOM.2017.8311699

[2] S. Kadambari, K. Jaswal, P. Kumar and S. Rawat, Using Twitter for Tapping Public Minds, Predict Trends and Generate Value, 2015 Fifth International Conference on Advanced Computing & Communication Technologies, Haryana, 2015, pp. 586-589.Doi: 10.1109/ACCT.2015.99

[3] B. Heredia, J. Prusa and T. Khoshgoftaar, Exploring the Effectiveness of Twitter at Polling the United States 2016 Presidential Election, 2017 IEEE 3rd International Conference on Collaboration and Internet Computing (CIC), San Jose, CA, 2017, pp. 283-290.Doi: 10.1109/CIC.2017.00045

[4] P. Tripathi, S. K. Vishwakarma and A. Lala, Sentiment Analysis of English Tweets Using Rapid Miner, 2015 International Conference on Computational Intelligence and Communication Networks (CICN), Jabalpur, 2015, pp. 668-672.Doi: 10.1109/CICN.2015.137

[5] Y. Lewenberg, Y. Bachrach and S. Volkova, Using emotions to predict user interest areas in online social networks, 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Paris, 2015,pp.1-10.Doi: 10.1109/DSAA.2015.7344887

[6] R. Jose and V. S. Chooralil, Prediction of election result by enhanced sentiment analysis on Twitter data using Word Sense Disambiguation, 2015 International Conference on Control Communication & Computing India (ICCC), Trivandrum, 2015,pp.638-641. Doi: 10.1109/ICCC.2015.7432974

[7] W. Alorainy, P. Burnap, H. Liu, A. Javed and M. L. Williams, Suspended Accounts: A Source of Tweets with Disgust and Anger Emotions for Augmenting Hate Speech Data Sample, 2018 International Conference on Machine Learning and Cybernetics (ICMLC), Chengdu, 2018, pp. 581-586.Doi: 10.1109/ICMLC.2018.8527001

[8] N. Mittal, D. Sharma and M. L. Joshi, Image Sentiment Analysis Using Deep Learning, 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI), Santiago, Chile, 2018, pp. 684-687.Doi: 10.1109/WI.2018.00-11

[9] S. Geetha and V. K. Kaliappan, Tweet Analysis Based on Distinct Opinion of Social Media Users', 2018 International Conference on Soft-computing and Network Security (ICSNS), Coimbatore, 2018, pp. 1-6.Doi: 10.1109/ICSNS.2018.8573619

[10] R.A.Hamad, S.M.Alqahtani and M.T.Torres, Emotion and polarity prediction from Twitter, 2017 Computing Conference, London, 2017, pp. 297-306. Doi:10.1109/SAI.2017.8252118

[11] Y. Li and H Fleyeh, Twitter Sentiment Analysis of New IKEA Stores Using Machine Learning, 2018 International Conference on Computer and Applications(ICCA) , Beirut, 2018, pp.411. Doi: 10.1109/COMAPP.2018.84677