International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 1 | ISSN : 2456-3307

A Survey on Sentiment Computing for the Opinions Based on the Twitter

Aishwarya Mankar[1], Harshada Patil[1],Chetan Arage[2], Mahesh Gaikwad[2]
[1]BE , Department of Computer Science and Engineering, Sanjay Ghodawat Institute Atigre, Kolhapur, Maharashtra, India
[2]Professor, Department of Computer Science and Engineering, Sanjay Ghodawat Institute Atigre, Kolhapur, Maharashtra, India

ABSTRACT

The sentiment computing of opinions is a significant component of the social media big data. The explosive increase of the social media data on the web has created and promoted the development of the social media big data mining area. It has attracted many researches, which could support many real-world applications, such as public opinion monitoring for governments and recommendation for websites. However, existing sentiment computing methods are mainly based on the standard emotion thesaurus or supervised methods, which are not scalable to the social media big data. Therefore, we propose an innovative method to do the sentiment computing for opinions. More specially, based on the social media data (i.e., words and emoticons) of a Tweets, a Word Emotion Association Network (WEAN) is built to jointly express its semantics and emotions, which lays the foundation for the opinion sentiment computation. Based on WEAN, a word emotion computation algorithm is proposed to obtain the initial word emotions which are further refined through the standard emotion thesaurus. With the word emotions in hand, we can compute every sentence's sentiment. Experimental results on real-world datasets demonstrate the excellent performance of the proposed method on the emotion computing for opinions.

Keywords: Text mining, sentiment computing, emotion classification, social media big data, opinions.

I. INTRODUCTION

This Sentiment computing for opinion based on twitter is a desktop-based web application, which is purposely used for the decision-making. This web application gives the exact emotion behind any opinion. A large amount of Web content is subjective and reflects people's opinions. It is mainly done through blog posts, online forum, social media, reviews on product. The amount of content generated by user is too vast for normal user to make decision. So there is need to automate this, sentiment analysis becomes a hot area in decision making. Sentiment analysis is a category of natural language processing for tracking the mood or review of public about a particular product or topic. Sentiment analysis uses the computational techniques to extract or classify sentiment from opinions and reviews. We provide a survey and a comparative analysis of existing techniques for sentiment analysis like Fuzzy logic, Hybrid approach, combining lexicon-based and learning-based methods. Nowadays, some researchers tend to compute the multidimensional emotions of text. According to commonly used emotions [9], the emotion is a six dimensional vector based on joy, love, surprise, fear, sad, anger.

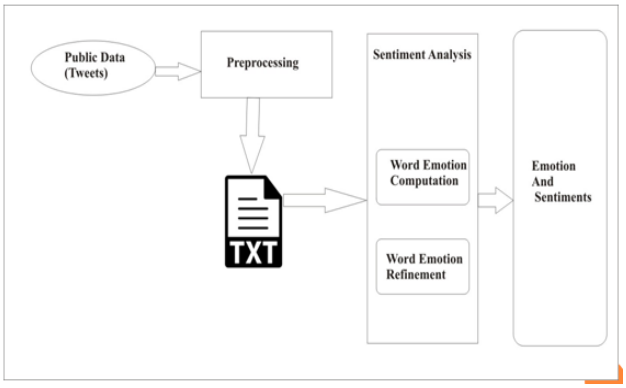

Figure 1. System Architecture

Our innovative idea is to do the sentiment computing for a news event by the help of its semantics. As in Fig. 1, the main part of the opinion sentiment-computing task is the word emotion computation, which can be splitted into two procedures: word emotion computation through word emotion association network and word emotion refinement through standard sentiment thesaurus.

II. LITERATURE SURVEY

A. Sentiment Analysis on Twitter Data

V. Sahayak, V. Shete and A. Pathan [1] proposed in 2015 about "Sentiment Analysis on Twitter Data" suggested the hybrid approach that classifies the tweets from twitter dataset in sentiment categories like positive , negative, and neutral. They developed the process of sentiment analysis of tweets, which contains three sections. First section is data extraction, which helps to extract opinion words from tweets. Second section does preprocessing of all the extracted words, which includes emoticons handling, filtration, tokenization, removal of stop words, n-gram construction. Third section classifies the sentiments using machine-learning classifiers.

This section works in two steps:

1. Model construction. 2. Model usage to check accuracy of classification. When complexity of emoticons and opinions increases, it becomes difficult for this approach to give right answer. This can be a drawback. For example, "The product was awesome but the services were gruesome". In this case, this approach may get confused for the result of sentiment.

B. Sentiment Analysis using Fuzzy Logic

Md. A. Haque&T. Rahman [2] proposed in 2014 "Sentiment Analysis with the help of Fuzzy Logic" by ranking the review in terms of positive and negative is the ranking perspective and it is achieved using fuzzy logic. The need of sentiment analysis is based on the two sectors i.e. classification of documents according to the orientation of sentiments such as positive and negative, other sector is gathering information by identifying the subjective or objective (SO) polarity of the comment or post, identifying the positivity or negativity (PO) polarity of comment or post and by identifying the degree of PN-polarity in terms of good, better or best. The tool to determine the polarity of lexical (the sentence is converted into sequence of tokens) is SentiWordNet. This gives numerical score to token range from 0 to 1. By having the values of the post and the weights, the result can be computed by calculating the weighted and arithmetic mean, from that, percentages of the individual sentiment (subjective &objective) is declared and by using concept of normalization the results are present in better way.

C. Sentiment Analysis and Opinion Mining

Y. Sharma, V. Mangat and M. Kaur [3] proposed in 2015 about "Sentiment Analysis & Opinion mining" that suggested various approaches based on which the sentiments can be analysed. The different approaches for sentiment analysis include the feature selection method, which reduces the unconnected information. It enhances the classification accuracy, and decreases the running time of algorithm. The selection step is to remove the target, stop words, URL, & stemming. The two broad ways in which the sentiment analysis is done are Machine-learning approach and Lexicon based approach. Machine learning approach learns from the previously generated results whereas the Lexicon based approach is usually fixed and gives approximated results.

D. A Hybrid Approach for Twitter Sentiment Analysis

N. Mittal & B. Agarwal [4] proposed "A Hybrid Approach for Twitter Sentiment Analysis" which is a three stage hierarchical model for sentiment extraction, in first stage the emoticons are labeled, then tweets are assigned sentiments using pre-defined lists of words with polarity and finally based on subjectivity of lexicon, the proposed probability based method assign the weight to all the tokens. The lexicons are weighted using various approaches like SentiWordNet, proposed probability based method, SentiWordNet (SWN) then probability based method, probability based method then SentiWordNet or Hybrid approach. The accuracy measured for hybrid approach was comparative higher than all the approaches i.e. 72.563 %. Hybrid method uses both SWN and probability based method to calculate the polarity of the token. Hence proposed hybrid approach improves the sentiment classification accuracy.

E.Opinion Mining of Real Time Twitter Tweets

A.Shrivatava, S. Mayor and B. Pant [5], proposed "Opinion Mining of Real Twitter Tweets," In this proposed system, a tweet puller is developed which automatically fetch the public opinion on a topic and using SVM the opinions are classified into positive, negative and neutral. First, tweets are collected using twitter API then creating domain specific dictionary. Tweet puller does extracting all the tweets from Twitter when server is connected. Using classification tool to generate threshold frequency for each feature and generate a text file. Text file is input to LIBSVM tool, which is proposed to provide accurate rate for testing the classification.

F. Combining Lexicon-based and Learning-based Methods for Twitter Sentiment Analysis

L. Zhang, R. Ghosh, M. Dekhil, M. Hsu, and B. Liu [6], proposed "Combining Lexicon-based and Learning-based Methods for Twitter Sentiment Analysis". The system is a new entity-level sentiment analysis approach for Twitter, which is done using lexicon based method, the input preprocessed tweets are analysed and categorize into sentence type detection, coreference resolution, using opinion rule aggregate opinions are formed which is input to train sentiment classifier that is Learning-based method and finally the extract opinionated tweets are classified. Coreference resolution gives the closest entity. For example, "Amit went to Arijit Singh's concert then for a long drive, it was amazing". "It" can be resolved by considering the closest entity that is "long drive". This system gives high precision, recall, and F-score.

G. Mining Sentiments from Tweets

A.Bakliwal, PiyushArora, SenthilMadhappan, Nikhil Kapre, Mukesh Singh, VasudevaVarma [7] proposed "Mining Sentiments from Tweets". In this system, the method for sentiment analysis is used on Stanford dataset &Mejaj dataset with achieves 88% accuracy. The feature vector approach is used to form feature using unigrams, bigrams, hash tags (#), targets (@), emoticons, special letters, & semi-supervised SVM classifier. The feature is distinguishing into Twitter specific and NLP feature. This approach is very useful when the user has to extract maximum information out of small content. It includes Emotion and Punctuations handling, spell correction, stemming, stop word removal using unigram model, noun indication, and finally score of the tweet is based on all the factors.

H. Sentiment analysis using multidimensional emotion classification

Dandan Jiang, XiangfengLuo∗, Member, IEEE, JunyuXuan, ZhengXu[8] proposed "Sentiment Computing for the News Event Based on the Social Media Big Data". This method consists two procedures: word emotion computation through word sentiment association network and word emotion refinement through standard sentiment thesaurus. For the word emotion computation through word sentiment association network, a Word Emotion Association Network (WEAN) has been built to jointly express its semantics and emotions. Based on WEAN, a word emotion

computation algorithm has been proposed to obtain the initial word emotions. After WEAN is constructed, we propose two assumptions: Quantity assumption and intensity assumption. Furthermore, a word emotion refinement algorithm has been proposed to improve the accuracy by in corporating the common prior knowledge: standard emotion thesaurus. After computing word emotion, we can classify tweets into six dimensional emotions i.e. Joy, love, surprise, fear, sad, anger.

III. CONCLUSIONS

Many of the organizations are putting their efforts in finding the best system for sentiment analysis. Some of the algorithms give good results but still many more limitations in these algorithms. As the twitter users are increasing day by day and the posts shared by the people are short messages (tweets) it can be very useful to analyze its data set. There are many techniques developed to do sentiment analysis but the problems are still not solved. The traditional way is very complex and time consuming but the recent approaches i.e. multidimensional emotions classification, which mentioned in this survey paper are quite simpler and efficient. More researches are done using SVM (support vector machines) classifier, and improvising its efficiency by introducing new rules and solving parsing problem. Fuzzy logic helps sentiment analysis provide efficient results as it is based on reasoning on the approximate values. Sentiment analysis when used with fuzzy logic helps to take decisions effectively but sometimes it may differ from the real time values. Future work may combine many different types of techniques to overcome individual's limitations, benefit from each other's merit, and measure the performance of classification technique.

IV. REFERENCES

[1]. Varsha Sahayak, Vijaya Shete, Apashabi Pathan, "Sentiment Analysis on Twitter Data", International Journal of Innovative Research in Advanced Engineering (IJIRAE), Issue 1, Volume 2, January 2015, pp. 178-183.

[2]. Geetanjali S. Potdar, Prof R. N. Phursule, "A Survey Paper on Twitter Opinion Mining", International Journal of Science and Research (IJSR) ISSN (Online): 23197064, Volume 4 Issue 1, January 2015, pp. 19-21.

[3]. Yakshi Sharma, Veenu Mangat, And Mandeep Kaur, "Sentiment Analysis & Opinion Mining", Proceedings Of 21St Irf International Conference, 8th March 2015, Pune, India, Isbn: 978-93-82702-75-7, Pp. 35-38.

[4]. Namita Mittal, Basant Agarwal, Saurabh Agarwal, Shubham Agarwal, Pramod Gupta, "A Hybrid Approach for Twitter Sentiment Analysis," Proceedings of ICON2013: 10th International Conference on Natural Language Processing, Noida, India, 2013, pp: 116-120.

[5]. A. Shrivatava, S. Mayor and B. Pant, "Opinion Mining of Real Twitter Tweets," International Journal of Computer Applications, Volume 100-No. 19, August 2014.

[6]. L. Zhang, R. Ghosh, M. Dekhil, M. Hsu, and B. Liu, "Combining Lexicon-based and Learning-based Methods for Twitter Sentiment Analysis, "Technical report, HP lLaboratories, 2011.

[7]. Akshat Bakliwal, Piyush Arora, Senthil Madhappan, Nikhil Kapre, Mukesh Singh, VasudevaVarma, "Mining Sentiments from Tweets", 3rd Workshop on Sentiment and Subjectivity Analysis (WASSA), Report No: IIIT/TR/2012/-1,July 2012.

[8]. Dandan Jiang, XiangfengLuo∗, Member, IEEE, JunyuXuan, ZhengXu "Sentiment Computing for the News Event Based on the Social Media Big Data"

[9]. W. McDougall, An introduction to social psychology. Courier Corporation, 2003.