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

A Sentiment Computing for the Opinions Based on the Twitter

Pooja Dhamanekar[1], Pooja Bindage[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 era of social networking has increased the amount of data generated by the user. People from all over the world share their opinions and thoughts on the micro-blogging sites on daily basis. As the use of internet such as websites, social networks, and blogs increases online portals reviews, opinions, recommendations, ratings, and feedbacks are also generated by users. Twitter is one of the most widely used micro-blogging site where people share their reviews in the form of tweets. This user can give their opinion on anything like books, people, hotels, products, research, events, etc. These sentiments become very useful for businesses, governments, and individuals. However, there are several challenges facing the sentiment analysis and evaluation process. These challenges become mountain in analyzing the accurate meaning of sentiments and measuring sentiment polarity. Therefore, we propose an innovative method to do the sentiment computing for opinions. Our method is based on the social media data 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.

Keywords: Sentiment computing, Emotion classification, Social media big data, Opinions, Text mining.

I. INTRODUCTION

This Sentiment computing for opinion based on twitter is a desktop-based web application. This web application gives the exact emotion behind any opinion. Sentiment analysis is the practice of applying natural language processing and text analysis techniques to identify and extract subjective information from text. Now a day's large quantity of data is available on internet, data mining is applied to collect knowledge from the data in many domains. Users express their opinions on day-to-day basis about various services or products using micro-blog posts, review sites etc. Among the 111 tweets, twitter is one of the most popular sites. It allow for valuable and well-timed statement of Semantic distance from a word as good or bad had information. Twitter supports brief explanation of ideas via short messages of tweets that are no longer than 280 characters.

The average number of tweets to report different opinions is at least 100,000. Each tweet is composed of different materials. The social media data is highly dynamic. The rich information hidden in the social media data is a perfect testing ground for the researchers in the big data area and a powerful tool for the corporations and governments to make specific decisions or global strategies. However, the vital issue faced is that the data available may be unstructured. Therefore, getting the sentiments from the huge unstructured data is very difficult and automatic classification of sentiment is at great demand.

Today, learning based hashing methods draw much attention because these methods exploit data distribution and even additional supervisory information. These methods employ statistical learning to directly learn hash functions from the data instead of using randomly generated projections. Here, we propose a method to do the sentiment computing for opinions. Proposed method 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. The methods proposed almost classify texts into six categories: joy, love, surprise, fear, sad, anger. With the word emotions in hand, we can compute every sentence's sentiment.

II. RELATED WORK

The study of people's point of view or emotions towards a product or an event is "Sentiment Analysis". Sentiment analysis helps to track the reputation of product or services in general. Sentiment classification can be at sentence level or document level. Document level classification needs to filter out the sentence that does not contain opinion words before classifying it into positive or negative. The method for classifying the phrases first extracts the opinionated text, then estimates the positions of these texts in the phrases and finally positive or negative value is assigned to the given phrase. As the social media developes, analyzing short texts sentiment becomes more and more popular. So the work on sentiment computation divided into two parts, sentiment computation (original) and short text sentiment computation.

1. Sentiment computation

At first, sentiment computation almost aims at dealing with long texts. Long text sentiment computation is a text classification problem indeed, so the existing methods for text classification can all be used in this problem.

Pang et al. [1] used machine-learning techniques to analyze the sentiment of film reviews, and they divided film reviews into two categories: positive sentiment and negative sentiment. Three machine-learning methods were taken in their experiments, and the result shows that using bag of words as features and using SVM as classifier has the best effect.

Kim et al. [2] computed sentence level sentiment with position feature and comment word feature.
Taboada et al. [3] computed each document's emotion score with the features of lexicon including emotion words and phrases about emotion orientation.

Yamamoto et al. [4] constructed emotion lexicon according to film review, which is based on ten dimensions of emotion. Emotion lexicon includes emoticons and emotion scores. Emoticon role was divided into conversion, addition, assuagement and emphasis and then decided one tweet sentiment score based on emoticon role.
Tang et al. [5] used deep learning to learn semantic representations of user and products for document level sentiment classification.

2. Short text sentiment computation

Recent years have witnessed the rapid development of social media platforms, such as Facebook, Twitter, Sin a micro blog and so on. Read et al. [6] generated positive Twitter corpus and negative Twitter corpus with appropriate emoticons.

Go et al. [7] built a machine learning classifier with features of Bigrams, unigrams and POS tags to divided Twitter corpus into two categories: positive and negative.

Linguistics and computational linguistics are both related to our study. Researches in linguistics are primarily concerned with syntax and semantics of comparisons, rather than computational identifying technology.

In computational scenario, machine learning and pattern match approach are two popular methods for identifying comparisons. Past experiments show machine learning has a better performance in comparative sentence identification. Jindal and Liu mined comparative information between products based on sequential rule mining with continuous Part-of-Speech sequence within the radius of three of each keyword. The sequential rules are then used as features of machine learning [8]. Park and Blake have investigated comparisons in scientific articles. They trained three different classifiers using the dependency syntactic features [9]. Compared with machine learning, pattern match is an unsupervised learning and pattern database is difficult to contain all of patterns. Song et al. manually constructed a Chinese pattern database and applied it to mine comparative sentences [10].

There are almost 111 micro-blogging sites today over the internet. These micro-blogs are actually social media that the people use to share their posts. Among the 111 micro blogs, twitter is one of the most popular sites. Twitter lets the people post tweets (message) of 148 characters in length Micro-blogging websites are social media that helps users to make short and frequent posts. As one tweet only consists of 148 characters, it makes the process of sentiment analysis easier.

Most of the methods classify the texts into two categories: positive and negative, which is too simple to deal with sentiment. Public sentiment is very complex that it should be classified into multidimensional sentiments.

III. PROPOSED WORK

The sentences that represent observations or attitude that is expressed as emotions are called as sentiments. The users post their tweets in twitter. These tweets are extracted in the form of unstructured data. The unstructured dataset is converted into structured form then extracts features from structured review. The features of the words are choosed and then sentiment classification technique is applied on extracted features to classify them into its sentiment polarity.

In our study we going to refer the research of Dandan Jiang, Xiangfeng Luo [11] which describe Word emotion computation through Word Emotion Association Network and word emotion refinement through standard sentiment thesaurus. Word emotions are the basis for the text emotion computation. In the following, we will successively introduce our method in detail.

Word emotion computation through word emotion association network:

People can publish views, opinions and attitudes for objects, individuals, events or topics on social media whenever and wherever. These views, opinions and attitudes contain users' emotion. We call it social media text emotion. Up to now, there are many researches on social media text emotion, and many multidimensional emotions have been reported from different views. We use multidimensional emotion classification, in which emotions are divided as love, joy, anger, sad, fear and surprise.

We propose a method to compute emotion of words in tweets. In a given tweets, a word in different stage may have different emotions, so we need to compute the word emotion at a specific time.

Algorithm to compute required results is described below.

Algorithm 1:

Input: Tweets from twitter
Output: Word emotions

Step 1: Create dictionary or thesaurus from the web corpus for each emotions**.**

Step 2: Get the twitter data as input using
getTwitter()String
SEARCH_TERM= "#book";
Twitter twitter = getTwitter();

Step 3: Pre-processing of twitter data.
3.1 Read the input file and clean it by using cleanText()
String cleanText(String book);
3.2 Tokenize the each statement using tokenizeText() and search in dictionary.
3.3 Create hashtable for each word specifying the count of emotions.

Step 4: Make use of POS tagger to find the tag i.e. adjective, verb, noun etc.

Step 5: Check the polarity of each statement for example
"very_strong_positive": count += 2;
"strong_positive": count += 1;
"positive": count += 0.5;
"weak_positive": count+=0.25;
"neutral": count+=0;
"weak_negative": count -= 0.25;
"negative": count - = 0.5;
"strong_negative": count-=1;
"very_strong_negative": count -= 2;

Step 6: Send the adjective to the sentiword and get the rating of each statement.

Step 7: Calculate the count of each emotions and display the result in the form of Pie chart.

The above results are generated for a file containing approx 500 tweets into it, which are directly fetched from twitter. But in some cases we want to calculate the result from only single tweet, then in order to achieve this scenario we need to follow given below algorithm:

Algorithm 2:

Step1: Store a single tweet in text file.
Step2: Create database containing dictionary words related to all emotions.
Step3: Now compare this input tweet with database in order to generate result.
Step4: Result will be displayed on screen as a output.

After WEAN is constructed, the next step is to compute word emotion. The different scale and intension of word circumstance can affect word emotion. The larger of the scale and the stronger of the intension, the more intense of the word emotion. Based on these, we propose two assumptions:

Quantity Assumption: In WEAN, if one word has connection link with other emotive words the more connection exist, the stronger the word emotion is. On the contrary, the less of the l connection links, the weaker of the emotion.

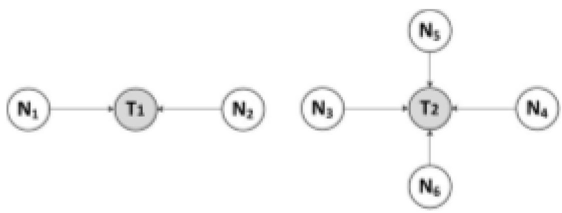

Figure 1. Quantity Assumption example

In Figure 1, each node denotes a word that appears in emotion E, such as love. N1...N6 are nodes we already know their emotion values, and suppose they have the same value. T1 and T2 are the words that we need to compute their emotion. T1 has two words linked with it, while T2 has four. According to quantity assumption, T2 has stronger emotion in love than T1.

• **Intensity Assumption:** Words can affect other nearby words through the links. The words who link with word k_n can have different intensity of emotion of the k-th dimention. The stronger of the word, the more emotion it transmits to other words. Hence, if one word links to many words whose emotion of the k-th dimention are strong, the emotion of the k-th dimention of word will be strong too.

Figure 2. Intensity Assumption example

In Figure 2, N7...N10 and T3,T4 are nodes that appears in emotion E. N7...N10 are nodes that already known emotion values, while N7 and N8 have stronger emotion in E than N9 and N10, Where N7 and N8 are smaller than N9 and N10. T3 and T4 all have two links linked to them, and we want to compute the emotion value of T3 and T4. According to intensity assumption, T4 has stronger emotion than T3 in emotion E.

After implementing algorithm 1 we come to following conclusion, which indicate pie chart as output, as shown in given Figuew 3.

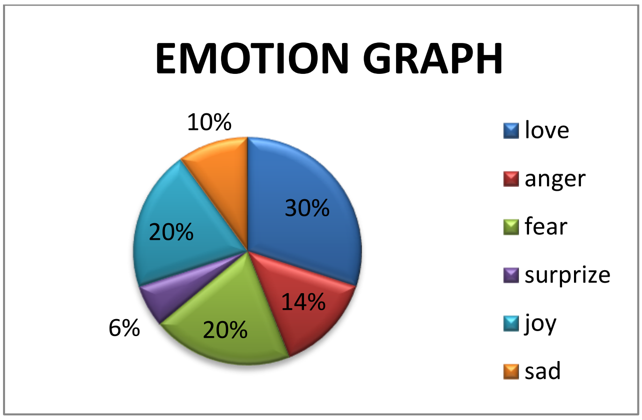

Figure 3. Output of a #books

We can draw the conclusion that our method is effective in emotion computation for the tweets.

IV. CONCLUSION AND FUTURE WORK

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 do sentiment analysis. There are many techniques developed to do sentiment analysis but the problem is still not solved. The traditional way is very complex and time consuming but the recent approaches mentioned in this paper are quite simpler and efficient.

In this paper, we have developed an innovative method to do the sentiment computing for the news events based on the social media big data. The proposed 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 incorporating 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.

In the future, we are interested in work with emoticons as well as other social media data rather than twitter for sentiment computation.

V. REFERENCES

[1]. 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. 79–86.

[2]. 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. 483–490.

[3]. M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," Computational linguistics, vol. 37, no. 2, pp. 267–307, 2011

[4]. D. G. A.-K. D. Matthias Steinbauer, Dr Maria Indrawan-Santiago, Y. Yamamoto, T. Kumamoto, and A. Nadamoto, "Multidimensional sentiment calculation method for twitter based on emoticons," International Journal of Pervasive Computing and Communications, vol. 11, no. 2, pp. 212–232, 2015.

[5]. D. Tang, B. Qin, and T. Liu, "Learning semantic representations of users and products for document level sentiment classification," in Meeting of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing, 2015.

[6]. J. Read, "Using emoticons to reduce dependency in machine learning techniques for sentiment classification," in Proceedings of the ACL student research workshop. Association for Computational Linguistics, 2005, pp. 43–48.

[7]. A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," CS224N Project Report, Stanford, vol. 1, p. 12, 2009.

[8]. N. Jindal and B. Liu, "Identifying Comparative Sentences in Text Documents", In Proceedings of SIGIR'06, (2006) , pp.244-251.

[9]. C.C. Chang, C.J.Lin, "Libsvm: a library for support vector machines", Trans IntellSystTechnol, vol. 2, no. 3, (2011), pp.27

[10]. R. Song, H. F. Lin, and F. Chang, "Chinese Comparative Sentences Identification and Comparative Relations Extraction", Journal of Chinese Information Processing, vol. 23, no. 2, (2009), pp.102-107.

[11]. Dandan Jiang, Xiangfeng Luo∗, Member, IEEE, Junyu Xuan, Zheng Xu, "Sentiment Computing for the News Event Based on the Social Media Big Data" DOI 10.1109/ACCESS.2016.2607218,