www.ijecs.in
International Journal Of Engineering And Computer Science ISSN: 2319-7242
Volume 5 Issue 10 Oct. 2016, Page No. 18655-18657

Sentiment Analysis Through Text Mining-A Review

Anupriya Pandey[1], Rakhi Verma[1], Deepika Bariha[1], Varsha Dewangan[2]
1) Department of Computer Science and Engineering,
New Government Engineering College, Raipur, Chhattisgarh, India
anu.20nov1995@gmail.com,
rakhiverma1008@gmail.com,
deepika.bariha1996@gmail.com
2) Assistant Professor, Department of Computer Science and Engineering,
New Government Engineering College, Raipur, Chhattisgarh, India
varsh61@gmail.com

Abstract: *Emotions are an important part of human life. It creates direct effect when the gestures of people are visible to the audience, but it is unclear when only texts are used rather than used in combination with gestures and facial expression. As the wide use of internet and mobile, text data has become the main source of communication in several social networking sites. One way of making such textual based interaction more interactive is to make a system that can recognize the sentimental purpose present behind the text. Now a days several researches are ongoing to make computer and human interaction as effective as it can detect the sentimental state of a person. This paper focuses on studying different work done in the area of sentiment analysis through text mining and provides a review that covers the best method and techniques used based on the previous experimental result.*

Keywords: Sentiment analysis, Precision, Recall, F-score, Textual data analysis, Emotional classes.

1. Introduction

Emotion means energy in motion. This energy can be transmitting in the form of text, audio or video. Among all communication through text is of more important because of small storage and easily transmit through a network. Now a day in there is a large amount of textual data available on the web and it's really interesting to extract the sentiment behind the text for different goals like [1-4].

- **Opinion Analysis:** Opinion analysis is used to take the product reviews and classify them into positive and negative reviews, by this way, it helps companies to examine the product based on product reviews.
- **Human Computer Interaction System:** To make human interaction with computers more effective, sentiment recognition techniques are used to recognize the emotions of users and make the system feel like a human.
- **Speech generation from Text:** To make speech generation from text more accurate and effective it's needed to extract the emotions behind the text as it is illustrated in Fig. 1 [1-6].

On the whole, emotions can be categorized into two: Basic and Complex. Basic emotions classes are joy, sadness, fear, disgust, surprise as defined by Ekman. Complex emotions classes are combinations of Basic emotions. Although emotions don't have clear boundaries, Plutchik classifies the emotions in the form of a wheel as it is illustrated in Fig 1 [4-8]. There has a scientific conclusion that at least five different emotional classes (fear, disgust, anger, happiness, sadness) are different in the logic of triggering different combinations of brain parts. To detect these different classes in text a technique called text mining is used. Sorensen defines text mining as a knowledge discovery technique that provides computational intelligence [4,7,8].

Fig.1: Plutchik's Proposed Wheel of Emotions [8]

It consists different fields such as text exploration, feature extraction, processing natural language and mined data classification based on the logical and nontrivial pattern from the large dataset. Unique challenges exist when we try to apply text mining to social media data, as unstructured data is going to be processed [3-8].

2. Related Work

The efficiency of different models used for sentiment analysis is based on three matrices. They are Recall, Precision, and F-Score. Precision is a measure of accuracy about a class has been predicted. The recall is a capability to select an instance of a particular class. F-Score is the harmonic mean of precision and recall matrices. Based on above three

DOI: 10.18535/ijecs/v5i10.52

Work	Emotion Models	Accuracy (%)
Sentiment Analysis Using Artificial Neural Network	KABNN	58.6
Sentiment Analysis Customized Decision Tree Algorithms	Decision Tree	80.2
Sentiment Analysis in Text for Reading in 3-D Facial Expression	SVM	70.55
Sentiment Analysis in Text Mining	Naive Bayes Classifier	89.38

Table 1: Emotion Models for sentiment analysis

It is an approach for text classification in which we use tokens of file, by using MAP (Maximum a Posterior) decision rule, we define the following classifier-

2.1 Decision Tree

From the document, relevant data is extracted by language preprocessing. This model works on the analysis of a set of training data and if then else condition. As such tree built by applying conditions can easily be converted into classification rules the overview of classification as it is shown in Fig. 2 [10]. In some cases data set contains objects that have completely different behaviour, these data objects are outliers. For analysis of these data separate mechanism called outlier analysis is used.

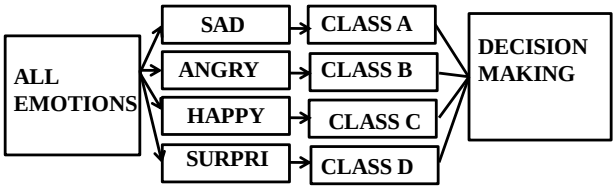

Fig.2: Overview of classification [10]

2.2 Space Vector Model (SVM)

It is an algebraic model for representing text document as in the form of vectors of identification documents and queries:

$D_j = (w_{1,j}, w_{2,j}, \ldots\ldots\ldots\ldots, w_{n,j})$

$Q = (w_{1,q}, w_{2,q}, \ldots\ldots\ldots\ldots, w_{n,q})$

Each dimension corresponds to separate term. If the term occurs in the document its vale in the vector is non-zero. Here terms represent the word, present in the domain of classes. Vector operation can be used to compare documents with queries.
In this model, long documents are poorly interpreted substring might result in a false positive match and the order
This means it should classify a new file/document. It must matrices different emotions models and their accuracy are given in Table 1 [1-9]

of terms appear is lost [6,10,11]. From the above 4 models, Naive Bayes classifier gives the best accuracy as 89.38%. The system builds by using it works on sentence level classification whether, it contain positive, negative or neutral emotions in each sentence.

2.3 Naive Bayes Classifier

$Cmap = \arg_{c \in C} \max(p(c/d)) = \arg_{c\,C} \max(p(c)\pi_{1\le k\le n}\, p\,(t_k/c))$
Where:
t_k– token (term/words) of document/file
C-set of classes used in the method
P(c/d)-conditional probability of class C given file d
P (c)-the prior probability of class C.
$P(t_k/c)$-conditional probability of token of given class

calculate the probability of product of each word of the file given a likelihood/particular class, multiplied by the prior class (particular class) after above calculation of all classes of set C, it was find the highest probability. If the highest probability value is in the decimal no. then it is an accurate value, if we will lead to float point no. it is not able to fit in the memory of it will be rounded to zero or analysis is useless. We will maximize the sum of the logarithms to avoid maximizing the probability of product.

$Cmap = \arg_{c \in C} \max (\log p(c) + \sum_{1\le k\le n} \log p\,(t_k/c))$

In place of selecting the highest probability class, we select the highest log score. The map remains the same if the logarithmic function is monotonic.
If any word does not appear in the particular class then its conditional probability become zero. If we use above decision method (product of probability) the product becomes zero. If we use the sum of logarithm methods log (0) is undefined. To avoid this problem we will use Laplace smoothing by adding 1 to each count [9, 12-14].

$$P\,(t/c) = \frac{Tct+1}{\sum(Tct'+1)} = \frac{Tct+1}{\sum(Tct')+B'}$$

Where B'= no. of terms contained in the vocabulary V.
The experimental based twitter data is shown in Fig.3.

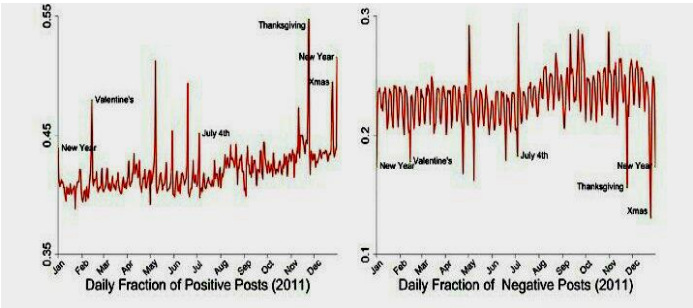

Fig 3: Experimental result on twitter data

3. Proposed Algorithm

This algorithm is based on Naive Bays classifiers- [13]

TrainNB (C,D)

V←Extract Vocabulary (D)
N←CountDocs (D)
For each c∈C
Do N_c←CountDocsInClass (D,C)
Prior[c]←N_c/N
textc←ConcatenateTextofAllDocsInClass (D,C)
for each t∈V
do T_{ct}←CountTokensOfTerm($text_c$,t)
for each t∈ V
do
CondProb[t][c]← $Tct + 1/\sum(Tct' + 1)$
return V,Prior,CondProb.

ApplyNB (C,V,Prior, CondProb,D)

W←Extract Token FromDoc(V,D)
For each c∈C
Do score[c]← log prior[c]
for each t∈W
do score[c]+=log CondProb[t][c]
return arg $max_c \in_C$ score[c]

4. Conclusion and Future Enhancement

This report consists analysis of different models used for sentiment analysis with their accuracy rate. As decision tree algorithm and space vector model has several limitations and also because of highest accuracy rate Naive Bayes classifier based algorithm is proposed for best results.

Sentiment analysis has many uses in real world applications but only a limited research is done in this field. Resources available for sentiment analysis are works on English language only. Recent problems required resources that can handle not only English, Hindi but a mixture of different languages and the abbreviations i.e.msg, r u?, thanq, ttyl, sd, tc etc. used in daily conversation. A System requires, that can convert the whole conversation in a structured format without losing the emotions or sentiment behind the text. We can make a search engine that can provide tailored results based on user emotions about a product. We can include features to chat application to compare user's emotions with their friends and calculate statistics on their mood over time.

By successfully analyzing the sentiment behind the text, we can make a machine to work based on user's mood.

References

[1] S. Aman, S. Szpakowicz, "Identifying Expressions of Emotions in Text", pp. 196-205, 2007.

[2] P. Ekman, "An Argument for Basic Emotions", pp. 169-200, 1992.

[3] Diksha Sahni, Gaurav Aggarwal, "Recognizing Emotions and Sentiments in Text: A Survey", pp. 201-205, 2015.

[4] Sentiment Analysis: [Online], Available at https://en.wikipedia.org/wiki/Sentiment_analysis

[5] Text Mining: [Online], Available at https://en.wikipedia.org/wiki/Text_mining

[6] G. Vinodhini, R.M. Chandrasekaran,"Sentiment Analysis and Opinion Mining: A Survey" In International Journal of Advanced Research in Computer Science and Software Engineering, 2012.

[7] Guilan Hu, Xiaochun Cai, "Research on Future Optimization Method in Personal Information Mining",WASE International Conference on Information Engineering, 2009.

[8] Sanjeev Dhawan, Kulvinder Singh, Deepika Sehrawat, " Emotion Techniques in Social Networking Sites" pp. 1145-1153, 2014.

[9] Yong-soo seol, Dong-joo kim" Emotion Detection from Text Using Knowledge based ANN", ITC-CSCC 2008.

[10] Jaskaran Kaur, Sheveta Vashish, "Analysis of Different Clustering Techniques for Detecting Human Emotions Variation Through Data Mining", pp.27-36, 2013.

[11] Vector Space Model, https://en.wikipedia.org/wiki/Vector_space_model

[12] Shiv naresh sivhare, Sarita Kethawat,"Emotion Detection in Text".

[13] Christos Troussas, Maria Virvou "Sentiment Analysis of Facebook Statuses Using Naïve Bayes Classifier for Language Learning", 2012.

[14] Ricardo A.Calix, Bin Chen, "Emotion Recognition in Text for 3-D Facial Expression Rendering", In Proceedings of the IEEE Congress 2010.

[12] Tejasvini Patil, Sachin Patil "Automatic Generation of Emotions for Social Networking Websites Using Text Mining" in proceeding of IEEE-2012.

[13] Lily Dey, Nadia Afrozand, Rudra Pratap Deb Nath, " Emotion Extraction from Real Time Chat Messenger" in Proceedings of 3rd International Conference on Informatics, Electronics and Vision 2014.

[14] Sriram, Sivraman, Xiaobu Yuan " An Enhanced Approach for Classifying Emotions Using Customized Decision Tree Algorithm", 2012.

Author Profile

Rakhi Verma, B.E. Student in the Department of Computer Science and Engineering, New Government Engineering College, Raipur, Chhattisgarh, India, rakhiverma1008@gmail.com.