International Journal of Advanced Science and Technology
Vol. 29, No. 6, (2020), pp. 6262 – 6270

A Comparative Study of Various Classifiers for Emotion Detection from Text

Akhilesh Kumar Singh[1], Krishn Kant[2], Darshika Srivastava[3], Dhruv Goel[4]
GLA University
Mathura,India
akhilesh.kumar@gla.ac.in[1] ,krishn.kant_cs16@gla.ac.in[2],
darshika.srivastava_cs16@gla.ac.in[3], dhruv.goel_da17@gla.ac.in[4]

Abstract

Emotion detection from text has become more popular in recent times. The reason behind its large scale use is its enormous potential applications in human-computer interaction marketing, psychology, political science, artificial intelligence, etc. Emotion is a way of expressing the thoughts of something that is one of the hardest work to find. A person's real emotion through his / her text. From the human face, we can analyze the present state of that person. Every human being will express their feeling/emotion implicitly or explicitly by their gestures, facial expressions, text, and speech. In this paper, we analyze text through many techniques introduced in the past, through which emotion can be detected. Here, we are working on four i.e. Linear Regression, SVM, Random Forest, Naive Byes methods using the TF-IDF technique to detect the emotions.

Keywords*: Naïve Byes, Tf-IDF, Random Forest, SVM, Linear regression, Logistic Regression.*

1. Introduction

In the field of computational linguistic when we are finding discrete emotion which is expressed in the form of text then this process is known as emotion detection. In the natural term, emotion analysis can be viewed as an expansion of sentiment analysis. For example, in marketing, emotion detection can be utilized to analyze consumer's feedbacks to products [25]. With the results of emotion detection systems that can also be used as input to other systems, like what [27] has done in profiling authors by analyzing the presence of emotions in their text. Therefore, understanding the emotions can be advantageous for any entity and organization such as commercial institutes, political campaigns, managing the response to a natural disaster. We are focused on evaluating people's feelings, which is a very challenging task. In recent times, Emotion detection from a text is too much popular field of research [16]. The various scenario where fearful people tend to have a pessimistic view about the future, whereas angry people tend to have a more optimistic view [28]. Likewise, fear generally is a passive emotion, whereas anger is more likely to lead to action [29]. It is mounted on a wired model. We are focused on evaluating people's feelings, which for these organizations is a very challenging task, application-based like a business, social well-being, etc. It is mounted on a wired model. Here we will use numerous techniques such as Linear regression, SVM, Random Forest using the TF-IDF method to achieve better results among them. To identify specific words in the text, we used some matching patterns that search the entire sentence to find the correct keyword. The words are like "disgust, sadness, happiness, anger, fear, and surprise".

This paper analyzes a person's text and says that person's emotional state on any website or blog. It is very important for many occasions as it is difficult to search the fact because the emotional word is not clearly expressed in the text. We must function on the level of the sentences to extract the emotions from the text. Through this, we can help the

ISSN: 2005-4238 IJAST
Copyright ⓒ 2020 SERSC

user or person to get known the emotion of the other person like in an organization. The main motivation of the project is to overcome the problem such as finding the emotion of user on particular things like a company made a product and launched and the company wants to know whether a product will survive in the market or not then our tool will detect the user mood using different techniques. So our tool finds the emotion on any blog, chats, Product reviews, YouTube comments using a different technique which will help to find better results. This paper aims to find the best method which will detect the emotional state of a person just by analyzing the texts which he/she is using in their daily lives or socially which will be very important to give them the right path and especially in the organization. This will provide a way to think and control over the wrong things like threats, reviews of the products, suicide attempts by analyzing their state of mind through their texts. We have organized this paper as follows: In Section 2 we have explained related work regarding textual emotional detection. In Section 3 we have specified proposed work. Using various approaches, we evaluate the performance which is presented in Section 4. Finally, we have concluded the paper and provide some future research work in Section 5.

2. Related Work

As we have seen many methods were used to provide emotion detection systems or tools or techniques which detect emotions, then they are grouped into three categories of algorithms. Based on Knowledge, machine learning, and based on the hybrid. In this paper, the classification scheme is the same as these are, but here our work is less or more related to these algorithms. We will explain it in this section one by one.

2.1. Sentiment Analysis

In natural language processing, Sentiment analysis is a well-established field. The computer is very important because it collects the relevant data or information from sites or blogs by analyzing the person's satisfaction. It tells about the data that is positive, negative, or neutral that's why it is sometimes said to be an opinion mining that tells about the nature or behavior of the person. It helps in various areas like for the business where they analyze their customer's satisfaction for the products, in public action, etc. This has been proven very beneficial in numerous applications such as marketing, advertising [18], question answering systems [19-21] summarization [22], as part of recommendation systems [23], or even improving information extraction [24], and many more. Sentiment Analysis [1] which is also called opinion mining, which focuses to retrieve information from blogs or document. A computer is very important because it collects the relevant data or information from sites or blogs by analyzing the person's satisfaction.

2.2. Emotion recognition on text-based approach using keyword

It is a traditional way as well as an easy way to find the emotional keyword, it generally depends on what they provided an input sentence. The evaluation shows that a word is close to happy or sad, in which they have shown the negativity of sentence and positivity, they also used a term Potency which shows high and low times of word and activity shows that currently it is the passive activity or can be active activity. To identify a particular word in the text, apply some matching patterns to check the whole sentence for finding a specific keyword. The words are "disgust, sadness, happiness, anger, fear, and surprise". This approach is shown in Figure:1

ISSN: 2005-4238 IJAST
Copyright ⓒ 2020 SERSC

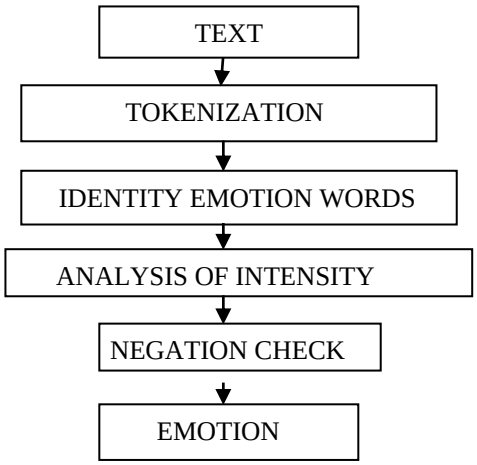

Figure 1. The text-based approach using keyword

2.3. Keyword-based method

Keywords based approach utilizes synonyms and Word Net antonyms to describe word sentiments based on a set of seed opinions. This method is implemented in a bootstrap that uses a small collection of specified seed opinion terms to find their synonyms and antonyms in WordNet to predict the adjective's semantic orientation. For Word Web, the adjectives are in an organization's bipolar cluster shape and have the same alignment as the synonyms. Since all the adjectives are connected, it creates a sequence and contributes to the emotion expressed by the word.

2.4. Vector Space Model

Many categorizations are used in the Vector Space Model. The co-occurrence frequency vector matrix is used for the dimensional representation of the data set. Words are defined by rows, and columns will reflect sentences, paragraphs, or documentation. The table, therefore, made a relationship. Using the Term Frequency and Inverse document frequency weighting scheme, VSM weighs these same frequencies. The Term Frequency-inverse document frequency finds the occurrence of each word in terms of its essentiality within the data sets of documents. The rating is split down into Term Frequency and frequency of the related papers. The tf means term frequency which is the occurrence of a term within a blog or comment. The formula to calculate Term Frequency:

Term Frequency = ml,b/ s*n

In the formula, m, l, b is a couple of times the phrase t, occurs in record n and s*n is the complete number of characters in document n.

2.5. Word Based Approach

NLTK package was used, which is used for human language data. Assigned all the feelings Labels i.e. (fear, joy, sadness, anger, happiness). We will pinpoint from the text and it is also essential to assign offensive words to it to detect the negative Feelings and then we will get the information then delete the unwanted words in those paragraphs that are available in the textual data from which identification to made correctly. Word scanning is done to classify the words into different categories using the nltk.pos() function, after which the data frame sturdy and creations are made. In the heading phase,

ISSN: 2005-4238 IJAST
Copyright ⓒ 2020 SERSC

all labeled words will be merged and put in a separate file and linked to each other, and in the data frame a predetermined tabular sentiment will be generated to allow the system to learn the relationship between keywords and labels, the columns and rows form contain the feelings in the first column and the words connected to it in the second and third column and the last column contains the emotion mark in which that line or paragraph is inserted. The WordNet is created between the phrase of columns 2 and 3. And by doing so is passed the new lines whose emotion is required to be identified, and the outcome is obtained.

2.6. Emotion Lexicon

In this technique, there is a recording of the presence of emotional lexicon by depending on WordNet Affect's association of words and emotional categories. In this method, they measure the Term frequency-inverse document frequency to each category of emotions (e.g., happiness, love, passion, sorrow, disloyalty, etc.) focused on the event happening of words in the WordNet Affect classification associated with them. Danescu Eet al. [2] proposed a method in which he measured the politeness which is specified in the text. SentiStrength [14] which is an openly accessible tool for computing positive and negative sentiment score. De Smedt et al. [3] developed a tool that can measure uncertainty. This tool focused on the uncertainty in terms of grammatical emotional states and adverbs. In this, the author will tokenize texts using the Stanford NLP library. During the cleansing process, he has removed HTML tags, code fragments, and URLs that can lead to the introduction of noise in the training. Unlike Ortu et al. [9], which has conducted stemming or lamentation, due to which inflected forms can convey important sentimental information. In this paper, the author uses the two Emotxt data sets on the gold standard and the Jira. Their set of data contains 4000 comments that are posted by various software developers on Jira4. The Jira gold standard includes manually numbered sentences with the emotions of passion, happiness, anger, and sadness. They feature EmoTxt, an open-source toolkit for text-based emotion detection, taught and tested on two large set gold data sources mined from Stack Overflow and Jira. This author uses categorization models that are to be used to detect emotions. Apart from grouping, EmoTxt facilitates emotion perceptron training from personally annotated training data. The training strategy leverages a set of features separate from the theoretical model implemented for data labeling.

2.7. Naïve Bayes

This algorithm is taken from the Bayesian theorem and works on probability, which is done from Unconditional probability to conditional probability, it says based on condition there might be an event.

$P(W|Z) = (P(Z|W)\ P(W))/P(Z)$

And

$P(W|Z) =(P(Z|W)\ P(W))/(P(Z|W) + P(Z|W'))\ P(W|Z)$

This means that the probability of W occurs when Z already occurs. $P(Z|W)\ P(W)$ called joint probability. This means that P(W) occurs when the P(Z|W) when A occurs.

2.8. Linear Regression

It is the simplest of regression in which analysis of the relation between the emotional words is done. It analyses the various emotions in which it tells the relation between dependent words and the various independent words to easily recognize the behavior or the emotion of a person by just examining the words of emotions that the person shows.

ISSN: 2005-4238 IJAST
Copyright ⓒ 2020 SERSC

As it is linear therefore it just examines it linearly by comparing the person's emotional word to the various words in a database or dataset. Mathematically, it's equation is shown in equation (1).

$$k=m+nc \quad (1)$$

where k is the dependent word, c is the collection of non- dependent words, n is a slope, and m is an intercept.

2.8. Random Forest

It is random because it randomly creates a decision tree and collection of many decisions tree itself is a forest or can be said as a random forest. In this, the emotion is analyzed by making a decision tree and this collection of trees tells the more accurate result regarding the emotion of a person. If more number of the decision tree is accurate then the more satisfying result can be found. The creation of trees is done which is more uncorrelated to each through which a committee can predict the accurate result than that individual decision tree.

3. Proposed Work

In this paper of emotion detection, we have analyzed the texts from blogs, comments to detect the emotion of a person to help the organization to provide them an effective way to measure the emotions by our implementation of tools which method is best to detect the emotions. So in our paper, we are making different methods to deal with this concern and to provide help. We have made four methods to detect the emotion i.e. Linear regression, SVM, Random Forest, Naive Byes using Term Frequency-inverse document frequency. Term Frequency-inverse document frequency a very common way which helps to categorize the text and to analyze them. The Term Frequency which tells or calculate the number of occurrence of any text or term in any blog or document to total terms in that blog or document. And the inverse document frequency is important because it tells the significance or importance of that term which is calculated by term frequency. Here we have used a formula to calculate the inverse documents frequency i.e. inverse document frequency = log (n/df), where n is several documents, and df is several documents that have that term. Term Frequency Inverse document frequency is an effective way to convert the textual information into a vector-space model.

4. Experimental Result

In this, we are attaining our goals by using various methods on the defined datasets which we have selected from the sites. We have given our analysis and tried to give our best results on the detection of emotion just by its texts. First of all, we have checked the phrases which are present in our dataset. The result is shown in Figure.2.

ISSN: 2005-4238 IJAST
Copyright ⓒ 2020 SERSC

	Emotions	Phrases
0	joy	On days when I feel close to my partner and ot...
1	fear	Every time I imagine that someone I love or I ...
2	anger	When I had been obviously unjustly treated and...
3	sadness	When I think about the short time that we live...
4	disgust	At a gathering I found myself involuntarily si...

Figure 2. Emotion and corresponding phrases

We have also created a word cloud of the dataset after various processing of the data which is shown in Figure.3.

Figure 3. Word cloud for labels of emotions

The accuracy of various classifiers using various machine learning algorithms is shown in the table: 1. We have used Linear regression, Logistic regression, Naïve Bayes, Random forest, and Support vector machine.

ISSN: 2005-4238 IJAST
Copyright ⓒ 2020 SERSC

Table 1. Accuracy of various classifiers

S.No	Classifiers	Accuracy
1	Linear Regression	0.537
2	Logistic Regression	0.53
3	Naïve Bayes	0.543
4	Random Forest	0.549
5	Support Vector Machine	0.558

In Figure.4 we have shown a graph that represents the accuracy of the algorithm. From the graph, it is clearly stated that the support vector machine will give the best accuracy among the various algorithm.

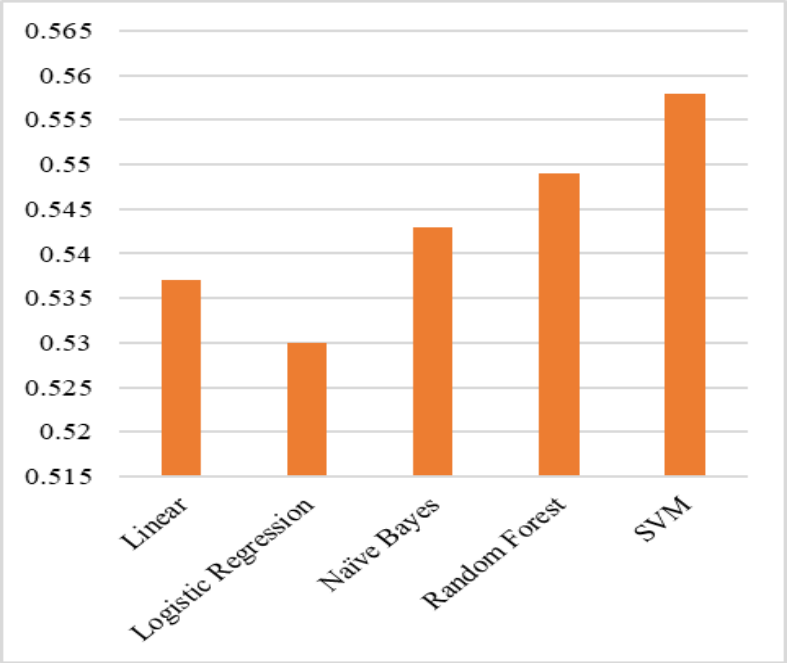

Figure 3. Accuracy of various classifier in Emotion Detection

5. Conclusion

In this paper, we have classified five methods in which the Term Frequency-Inverse document frequency method is used to find the accuracy of all the four algorithms i.e. Linear Regression, logistic regression, Naive Byes, Random Forest, SVM. They are applied to the dataset which we have selected from the various sources. After applying all the algorithms, we found that the accuracy is different for each algorithm and the best accuracy is 0.558 of SVM among them and the lowest accuracy is of Linear Regression. In this paper, we will examine that SVM is better and will be more useful to detect emotions. In the future, we can apply other algorithms like LSA and LDA to get a better result. We can also apply Deep learning to improve the result of the accuracy.

ISSN: 2005-4238 IJAST
Copyright ⓒ 2020 SERSC

Acknowledgments

We would like to thank our project guide Mr. Akhilesh Kumar Singh for guiding in finding the result. I also like to thank our collogues who have motivated me to do this project and write a paper on it.

References

[1] Basile, P., Basile, V., Nissim, M., Novielli, N., & Patti, V. (2018). Sentiment Analysis of Microblogging Data.
[2] Danescu-Niculescu-Mizil, C., Sudhof, M., Jurafsky, D., Leskovec, J., & Potts, C. (2013). A computational approach to politeness with application to social factors. arXiv preprint arXiv:1306.6078.
[3] T. De Smedt and W. Daelemans, "Pattern for Python," The Journal of Machine Learning Research, vol. 13, no. 1, pp.2063-2067, 2012.
[4] R.E. Fan, et al. "Liblinear: A library for large linear classification". J Mach Learn Res 9:1871–1874, 2008.
[5] D. Gachechiladze, F. Lanubile, N. Novielli, and A. Serebrenik. "Anger and its direction in collaborative software development". In Proceedings of the 39th International Conference on Software Engineering: New Ideas and Emerging Results Track. IEEE Press, 11-14, 2017.
[6] T. Helleputte. "LiblineaR: Linear Predictive Models Based on the LIBLINEAR C/C++ Library". R package version 1.94-2, 2015.
[7] T. Joachims. "Training linear SVMs in linear time". In: Proceedings of the 12th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining, ACM, New York, NY, USA, KDD '06, 217–226, 2006.
[8] M. Kuhn et al. "Caret: Classification and Regression Training". R package v. 6.0-70. https://CRAN.R-project.org/package=caret, 2016.
[9] M. Ortu, B. Adams, G. Destefanis, P. Tourani, M. Marchesi, and RTonelli."Are bullies more productive?: an empirical study of affectiveness vs. issue fixing time". In Proc. of MSR '15. IEEE Press,303-313, 2015.
[10] R. Plutchick. "Emotions: A general psychoevolutionary theory." In K.R.Scherer & P. Ekman (Eds) Approaches to emotion. Hillsdale, NJ;Lawrence Ealrbaum Associates, 1984.
[11] P. Shaver, J. Schwartz, D. Kirson, C. O'Connor. "Emotion knowledge: Further exploration of a prototype approach. Journal of Personality and Social Psychology 52(6):1061–1086, 1987.
[12] C. Strapparava and A. Valitutti. WordNet-Affect: an affective extension of WordNet. In Proc. of LREC, 1083–1086, 2004.
[13] C. Tantithamthavorn, S. McIntosh, A. E. Hassan, and K. Matsumoto."Automated parameter optimization of classification techniques for defect prediction models". In ICSE, pages 321–332. ACM, 2016.
[14] M. Thelwall, K. Buckley, G. Paltoglou. "Sentiment strength detection for the social web". J Am Soc Inf Sci Technol 63(1):163–173, 2012.
[15] Robert Jenke, Angelika Peer," Feature Extraction and Selection for Emotion Recognition from EEG" IEEE Transactions on Affective Computing, Vol. 5, NO.3, July-September 2014.
[16] https://www.microsoft.com/developerblog/2015/11/29/emotion-detection-and-recognition-from-text-using-deep-learning/
[17] [Jin et al.2007] Xin Jin, Ying Li, Teresa Mah, and Jie Tong. 2007. Sensitive webpage classification for content advertising. In Proceedings of the 1st International Workshop on Data Mining and Audience Intelligence for Advertising, ADKDD '07, pages 28–33, New York, NY, USA. ACM.
[18] [Qiu et al.2010] Guang Qiu, Xiaofei He, Feng Zhang, Yuan Shi, Jiajun Bu, and Chun Chen. 2010. DASA:Dissatisfaction-oriented Advertising based on Sentiment Analysis. Expert Systems with Applications,37(9):6182–6191.
[19] [Somasundaran et al.2007] Swapna Somasundaran, Theresa Wilson, Janyce Wiebe, and Veselin Stoyanov. 2007.Qa with attitude: Exploiting opinion type analysis for improving question answering in on-line discussions and the news. In ICWSM.
[20] [Stoyanov et al.2005] Veselin Stoyanov, Claire Cardie, and Janyce Wiebe. 2005. Multi-perspective question answering using the opqa corpus. In Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, pages 923–930. Association for Computational Linguistics.
[21] [Lita et al.2005] Lucian Vlad Lita, Andrew Hazen Schlaikjer,WeiChang Hong, and Eric Nyberg. 2005. Qualitative dimensions in question answering: Extending the definitional qa task. In PROCEEDINGS OF THE NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE, volume 20, page 1616.Menlo Park, CA; Cambridge,MA;London; AAAI Press; MIT Press; 1999.
[22] [Seki et al.2005] Yohei Seki, Koji Eguchi, Noriko Kando, and Masaki Aono. 2005. Multi-document summarization with subjectivity analysis at duc 2005. In Proceedings of the Document Understanding Conference (DUC).
[23] [Terveen et al.1997] Loren Terveen, Will Hill, Brian Amento, David McDonald, and Josh Creter. 1997. Phoaks: A system for sharing recommendations. Commun. ACM, 40(3):59–62, March.

ISSN: 2005-4238 IJAST
Copyright ⓒ 2020 SERSC

[24] Riloff et al.2005] Ellen Riloff, Janyce Wiebe, and William Phillips. 2005. Exploiting subjectivity classification to improve information extraction. Proceedings of the 20th national conference on Artificial intelligence,20(3):1106–1111.
[25] [Gupta et al.2013] Narendra Gupta, Mazin Gilbert, and Giuseppe Di Fabbrizio. 2013. Emotion detection in email customer care. Computational Intelligence, 29(3):489–505.
[26] [Voeffray2011] S Voeffray. 2011. Emotion-sensitive human-computer interaction (hci): State of the art-seminar paper. Emotion Recognition pages 1–4.
[27] [Rangel and Rosso2016] Francisco Rangel and Paolo Rosso. 2016. On the impact of emotions on author profiling.Information processing & management, 52(1):73–92.
[28] [Lerner and Keltner2000] Jennifer S Lerner and Dacher Keltner. 2000. Beyond valence: Toward a model of emotion-specific influences on judgment and choice. Cognition & emotion, 14(4):473–493.
[29] [Miller et al.2009] Daniel A Miller, Tracey Cronin, Amber L Garcia, and Nyla R Branscombe. 2009. The relative impact of anger and efficacy on collective action is affected by feelings of fear. Group Processes & Intergroup Relations, 12(4):445–462.

ISSN: 2005-4238 IJAST
Copyright ⓒ 2020 SERSC