International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

Available at http://inpressco.com/category/ijcet

Research Article

Emotion recognition via real-time analysis of Twitter posts

Anjali Deshpande and Prof. Ratnamala Paswan

Department of Computer Engineering Pune Institute of Computer Technology Pune, India

Received 10 Nov 2020, Accepted 10 Dec 2020, Available online 01 Feb 2021, **Special Issue-8 (Feb 2021)**

Abstract

The analysis of social media posts is extremely challenging as it concerns the detection of user communities. As emotions play a pivotal role in human interaction, the capability to detect them via analysing social media posts has various applications such as detecting psychological disorders in individuals or quantitatively detecting the public mood of a community. Previous studies on emotion classification made use of lexicons and bag-of-words classifiers. However, the existing work of emotion recognition on Twitter was carried out with the help of deep learning techniques on static Twitter data by taking into account only the hashtags present. The proposed method tries to increase the overall accuracy of emotion recognition via machine learning algorithms on real-time streaming data fetched from Twitter. The overall aim is to accurately recognize the various emotions that a particular tweet expresses semantically.

Keywords: *Emotion recognition; text mining; machine learning; unison model; twitter*

Introduction

Emotions can be defined as mental states that are associated with certain chemical changes within the nervous system. Recent studies have been carried out on emotion detection on social media using opinion mining. However, emotion recognition faces certain challenges in the form of limited length of post or ambiguous expression of the user. The main focus of previous studies in this area was on using lexicons and machine learning methods for analysis. The performance of these methods depends upon quality of extracted features as well as emotion lexicon. Paul Ekman has defined six basic emotions viz. anger, disgust, fear, joy, sadness, surprise. An extension to it was given by Robert Plutchik with two additional emotions; those were trust and anticipation. Thus, a Profile of Mood States (POMS) was proposed which is a psychological instrument defining a six-dimensional emotional state representation. Our proposed system is a Profile of Mood States (POMS) defining twelve-dimensional emotional-state representation using 65 adjectives combining basic as well as well as supplementary emotions. The emotions representing the twelve dimensions are vigour, joy, depression, confusion, anger, trust, fatigue, surprise, fear, anticipation, sadness and disgust. Traditional previous studies mostly focused on the detection and classification of single emotion or a couple of emotions using one-vs-all classification models. Building a multiclass classification model is desirable as we can predict multiple classifications at the same time using a single model. This, in turn, allows for a more in-depth analysis.

The motivation behind this work is that we can timely detect the level of stress of a person based on the emotions identified. Also, long-time monitoring of latest real-time emotional data helps in detecting the public mood of a community regarding a particular topic. As previous systems only considered static data, deployment in real-time applications becomes extremely difficult. Hence, the analysis of real-time data is taken into consideration. The proposed model aims to develop a single model for predicting multiple emotion classifications at the same time simultaneously achieving good results. We take into consideration basic emotions as well as their synonyms for classification purposes. To perform analysis on the Twitter data streamed in real-time.

Literature Survey

N. Colneric and J. Demsar [1] made use of hashtags of Twitter posts to create three emotion-labelled data sets that correspond to various basic as well as certain supplementary emotions. They made use of CNN and compared its performance with traditional bag-of-words model. A single model for prediction of multiple emotion classification at the same time was proposed. Improved performance using the new hashtag emotion lexicon. However, it doesn't take into account the derivatives or synonyms of words while working on static data considering the hashtags. S. M. Mohammad and S. Kiritchenko [2] state that hashtags corresponding to emotions are good labels of emotions in tweets that can be obtained without human intervention. Also created a huge corpus of word–emotion associations which was the first of its kind. Mainly worked with six basic emotions categories. SVM with sequential minimal optimization was used for automatic detection of personality from tweets. B. Plank and D. Hovy [3] proposed that certain personality traits correlate with linguistic behaviour of individuals. They made use of social media as a resource for large-scale, open vocabulary personality detection. Logistic regression was used for analysis and features predictive of personality traits were found out. Thus, a lexicon of approximately 1 million English tweets associated with Myers-Briggs gender and personality type was created. However, the model only detected two personality distinctions with high reliability. X. Liu et al. [4] propose a Multi-Task DNN for representation learning, combining semantic classification and semantic information retrieval tasks. The proposed MT-DNN model maps arbitrary text queries and documents into a low dimensional latent space using semantic vector representations. It outperforms strong baselines across all web search and query classification tasks. O. Irsoy and C. Cardie [5] explored application of deep recurrent neural networks to the task of sentence-level opinion mining. RNNs outperformed previous traditional methods S. M. Mohammad et al. [6] analyse electoral tweets for sentiment, the emotion, the purpose or intent behind the tweet and the style of the tweet. Made use of SVM with 10 - fold cross-validation. Automatically classified tweets into emotional categories. Mostly handled the emotions concerned with only disgust or trust without consideration of past tweets. J. Schnebly and S. Sengupta [7] studied the hazardous bots infesting Twitter. Proposed a generalized model that can detect existing Twitter bot accounts with 90.25% accuracy with random forest. Y. Hegde and S. K. Padma [8] carried out the sentiment analysis of Kannada documents. Used random forest ensemble of classifiers to identify the sentiment. Improved the overall accuracy from 65% to 72% indicating the efficiency of the proposed model.

Proposed Methodology

Profile of Mood States or POMS is a psychological instrument for analysing a person's emotional state. It defines 65 adjectives that are rated on a five-point scale by the particular person in consideration. Each of these adjectives individually contribute to one of the six emotional categories in the case of basic emotions. All of these ratings are combined to form a six-dimensional emotion-state representation. This method can be extended to form the proposed twelve dimensional emotion-state representation. Advantages of Proposed System:

1) Increases human-computer interactions
2) Low-cost
3) Fast emotion recognition system
4) Scalable

A. Proposed Methodology

The proposed approach in Fig.1 shows a generalized outline of the various steps that need to be carried out for building a multiclass classifier for emotion recognition. Firstly, the raw data from the social media site (i.e. Twitter) is obtained in realtime via the Twitter API. Then, certain pre-processing is done to convert the obtained data into suitable format. Following it are the feature selection and extraction steps whereby the discriminating features for emotion recognition are identified. A classification model is built using some emotion-labelled twitter training data owing to which the incoming new tweets are classified based on the emotion they depict.

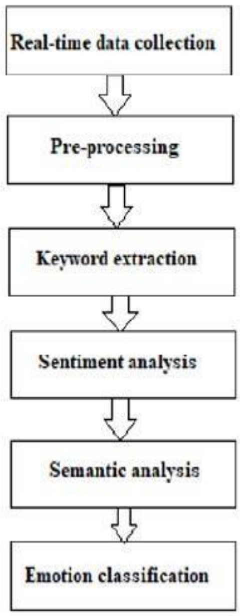

Fig. 1. Proposed Methodology

B. Architecture

The Fig.2 shows the proposed system architecture.

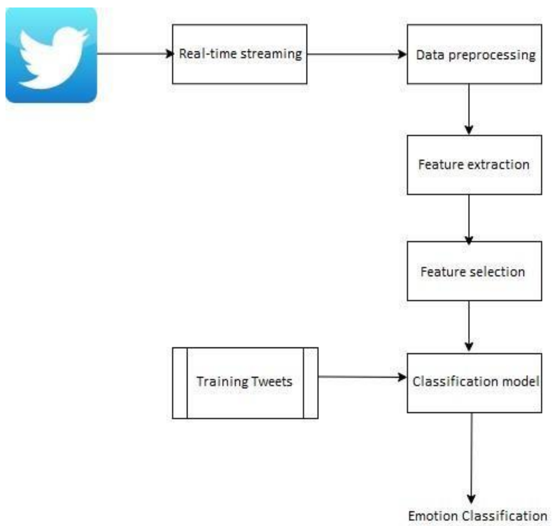

Fig. 2. System Architecture

After successful implementation of the formerly mentioned steps, a classification model is built using suitable algorithm. This model can henceforth classify any new incoming tweets with respect to the emotions expressed.

C. Algorithms

a. Latent Dirichlet Allocation (LDA) Algorithm:

LDA describes how the documents in a dataset were created using a generative version. Consider that a dataset is a collection of D documents which, in turn, can be considered as a collection of words. Hence, this model indicates how each document acquires its words. Initially, let's assume that K topic distributions for our dataset, meaning K multinomials containing V elements each, where V is the number of terms in the corpus. Let βi represent the multinomial for the i^th topic. The size of βi is $V : |\beta i| = V$. Given these distributions, the LDA generative process is as follows:

Steps:

1. For each document:
   (a) Randomly choose a distribution over topics (a multinomial of length K)
   (b) for each word in the document:
   (c) Probabilistically draw one of the K topics from the distribution over topics obtained in (a), say topic βj (ii) Probabilistically draw one of the V words from βj

b. Random Forest

Step 1: Let the number of training cases be N, and the number of variables in the classifier be M.

Step 2: The number m of input variables to be used to determine the decision at a node of the tree; m should be much less than M.

Step 3: Choose a training set for this tree by choosing n times with replacement from all N available training cases (i.e. take a bootstrap sample). Use the rest of the cases to estimate the error of the tree, by predicting their classes.

Step 4: For each node of the tree, randomly choose m variables on which to base the decision at that node. Calculate the best split based on these m variables in the training set. Step 5: Each tree is fully grown and not pruned (as may be done in constructing a normal tree classifier). For prediction a new sample is pushed down the tree. It is assigned the label of the training sample in the terminal node it ends up in. This procedure is iterated over all trees in the ensemble, and the average vote of all trees is reported as random forest prediction.

D. Mathematical Model

Language Models:

Language models compute the probability of occurrence of a number of words in a particular sequence. The probability of a sequence of T words $\{w_1,...,w_T\}$ is denoted as $P(w_1,...,w_T)$. Since the number of words coming before a word, w_i, varies depending on its location in the input document, $P(w_1,...,w_T)$ is usually conditioned on a window of n previous words rather than all previous words:

$$P(w_1,...,w_T) \prod_{i=1}^{i=T} P = (wi|w1, \dots , wi - 1)$$

$$\approx \prod_{i=1}^{i=T} P(wi|wi - (n - 1), \dots , wi - 1) \qquad (1)$$

Equations (2) and (3) show this relationship for bigram and trigram models as follows:

$$P(w2|w1) = \frac{count(w1,w2)}{count(w1)} \qquad (2)$$

$$P(w3|w1, w2) = \frac{count(w1,w2,w3)}{count(w1,w2)} \qquad (3)$$

Results and Discussions

The following fig. (3) gives a comparative analysis of the proposed random forest model and Profile of Mood States (POMS) of the twelve-dimensional emotion classification on a sample of Twitter data. The following graph is plotted by using Weka Tool. It can be seen that our proposed model shows good performance as compared to the traditional POMS model in most of the cases.

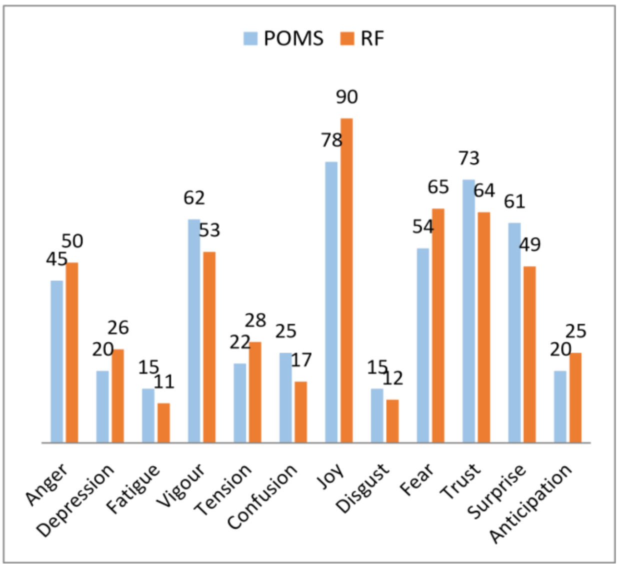

Fig. 3. POMS vs Random forest for 12-dimensional emotion categories

Based on the analysis above, the accuracy of POMS models to approx. 81.6% whereas that of proposed random forest approach models to 85.9%.

Conclusion

This paper proposes a model for emotion recognition on realtime streaming Twitter data that manoeuvres machine learning algorithm and a twelve-dimensional mood state representation combining Ekman's and Plutchik's emotions categories. It classifies the emotions with the help of multiclass classification and semantic analysis. The overall aim is to recognize and classify the emotions taken in to consideration as accurately as possible. The results are quite promising as the proposed model generally shows good performance as compared to POMS model in most of the cases.

References

[1]. N. Colneric and J. Demsar, -Emotion recognition on twitter: comparative study and training a unison model, IEEE Transactions on Affective Computing, 2018.

[2]. S. M. Mohammad and S. Kiritchenko, -Using hashtags to capture fine emotion categories from tweets, Computational Intelligence, vol. 31, no. 2, pp. 301–326, 2015.

[3]. B. Plank and D. Hovy, -Personality traits on twitter —or— how to get 1,500 personality tests in a week, Proc. of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, 2015, pp. 92–98.

[4]. X. Liu, J. Gao, X. He, L. Deng, K. Duh, and Y.Y. Wang, -Representation learning using multi-task deep neural networks for semantic classification and information retrieval, Proc. of the 2015 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 912–921, 2015.

[5]. O. Irsoy and C. Cardie, -Opinion mining with deep recurrent neural networks, Proc. of the Conf. on Empirical Methods in Natural Language Processing, ACL, 2014, pp. 720–728.

[6]. S. M. Mohammad, X. Zhu, S. Kiritchenko, and J. Martin, -Sentiment, emotion, purpose, and style in electoral tweets, Information Processing and Management, vol. 51, no. 4, pp. 480–499, 2015.

[7]. J. Schnebly and S. Sengupta, -Random forest twitter bot classifier, 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2019, pp. 0506-0512.

[8]. Y. Hegde and S. K. Padma, -Sentiment analysis using random forest ensemble for mobile product reviews in kannada, 2017 IEEE 7th International Advance Computing Conference (IACC), Hyderabad, 2017, pp. 777-782.

[9]. Huijie Lin, Jia Jia, Jiezhong Qiu, Yongfeng Zhang, Guangyao Shen, Lexing Xie, Jie Tang, Ling Feng and Tat-Seng Chua, -Detecting stress based on social interactions in social networks, IEEE Transaction On Data Knowledge And Engineering, Vol.12, No.02 June 2018

[10]. Andrey Bogomolov, Bruno Lepri, Michela Ferron, Fabio Pianesi and Alex (Sandy) Pentland, -Daily stress recognition from mobile phone data, weather conditions and individual traits, Proceedings of the 22nd ACM international conference on Multimedia,2014, pp. 477-486