International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 04 | Apr 2020 www.irjet.net p-ISSN: 2395-0072

Stress detection methods in Social networks: A Comparative study

P.Iswarya[1], D.Maheswari[2], R.Rajeswari[3]

[1]*Assistant Professor, Dept. of. CT, Dr.N.G.P Arts and Science College, Coimbatore, TamilNadu*
[2]*Professor, Dept. of. CT, Dr.N.G.P Arts and Science College, Coimbatore, TamilNadu*
[3]*Assistant Professor, Dept. of. CT, Dr.N.G.P Arts and Science College, Coimbatore, TamilNadu*

Abstract - *Psychological stress is an important factor that affects our healthy life. Traditional stress detection methods use face-to-face interviews which is time consuming and laborious task. Due to rise in social media networks, instead of making interaction in real; most of the people spend their daily interaction with their family and friends through social media. The user emotions such as angry, sad, happy, joy and depressed conditions in social media can be identified through their weekly tweets. An identified emotion categorized in to positive and negative tweets, finding their stress state through continuous negative tweets and informing the user regarding the stress state to prevent them from suicide and also from other attacks. The sentence pattern labeling method in facebook contains abundant information for data analysis. Utilizing above information and features extracted from multiple modalities through convolution neural network model. The extracted features are fed into the several classifiers for tweet classification. Experimental results show that the Random Forest (RF) model provides higher accuracy rate than Support Vector Machine (SVM) and Probabilistic Neural Network (PNN) models.*

Key Words: **Convolution Neural Network, Probabilistic Neural Network, Random Forest, Support Vector Machine, Tweet.**

1. INTRODUCTION

The digital era that makes the relationship is being impacted by the digital revolution. Modern technology diverts time and attention from spouse, families and friends [1]. Due to rise in interaction of people in social media platforms, making it feasible to identify online social network data for stress detection. The paper implements various stress detection methods to find the user's psychological stress state, and able to identify the higher efficiency stress detection method. Many studies on social media based emotion analysis are at the tweet level, using text-based linguistic features and classic classification approaches. A system that make emotion analysis on the Chinese micro-blog platform called Weibo using Mood Lens [2] for classifying the emotion categories into four types, i.e., angry, disgusting, joyful, and sad.

An existing system studied the emotion propagation problem in social networks, and found that anger has a stronger correlation among different users than joy, indicating that negative emotions could spread more quickly and broadly in the network. As stress is mostly considered as a negative emotion, this conclusion can help us in combining the social influence of users for stress detection.

Traditional psychological stress detection in student environment is mainly based on behavior of student in the class. Those students may be inattentive, lack of confidence, alone and low self-esteem etc. However, finding those stressed students can possible by continuous monitoring of the faculty which is laborious task in mass student group and unpredictable. We first define a set of stress-related textual, visual, and social attributes from various aspects, a Convolution neural network used for feature extraction and test data label classification is done using Support Vector Machine (SVM), Random Forest (RF) and Probabilistic Neural Network (PNN) respectively that leverage a tweet content and social interaction information for stress detection. The rest of the paper organized as follows. Section 2 describes about the architecture of the system. The Feature extraction strategies explained in Section 3. The comparative result analysis and discussion for different classifiers detailed in Section 4. The conclusion is presented in Section 5.

2. SYSTEM ARCHITECTURE

The system architecture for stress detection methods is shown in Fig.1.

To collect user information from face book, get API access token from graph internet explorer and then access right will be provided to collect user information from facebook. In this paper our college students facebook information were collected. The raw data from social network are preprocessed using stop word removal and lemmatization techniques to generate tokens. From the data several features are extracted and fed in to the machine learning model to obtain various emotions i.e happy, joy, surprise, angry and sad. The first three are positive emotions, and remaining are negative. The students weekly tweets are analyzed, if negative tweets are in higher degree, their stress is detected and intimated to their faculties. The stressed student is monitored and counseled by the psychologist.

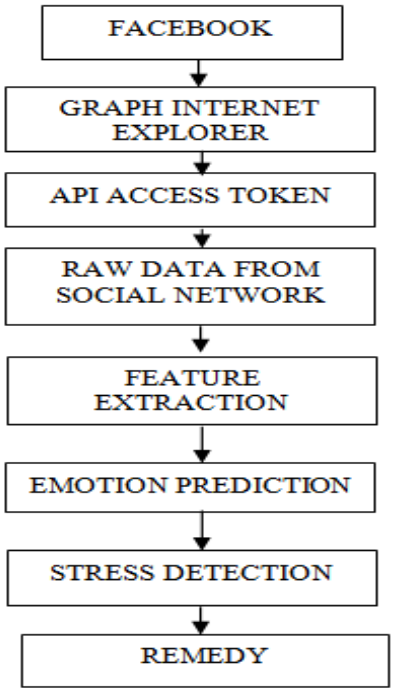

Fig-1: Overview of Stress detection in social media

3. FEATURE EXTRACTION

From the preprocessed data, set of attributes used for stress detection are tweet level attributes and user level attributes [2].The tweet level attributes consists of linguistic, visual and social interaction, where these gathered from single user tweet having feature vector length of 35. The user level attributes extracted from user weekly tweets of posting behavior and social interaction. The total vector length of 100 features need to be extracted for a user. The Cross Auto-Encoder (CAE) with Convolution Neural Network (CNN) used for analyzing the social network data. The CNN is found be effective in modeling multiple data sources such as text, images and audios. The user level attributes are derived from daily tweet attributes containing their posting behavior, social interaction and content attributes. The data from different modalities is used to train the CAE and which implements based on stochastic gradient descent method. The tweets from different modalities generated over a time series form a one dimensional series. Then 1-D data is processed using 1-D CNN model and it generates several attribute maps. The mean over time pooling technique is applied to several attribute maps to single attribute map.

4. STRESS DETECTION METHODS

The stress detection methods such as Support vector Machine, Probabilistic Neural Network and Random forest are modeled and incorporated with time variant tweet content attributes and social interactions used for learning and detecting user emotion states. The negative emotions such anger, sad, depressed is carried over a week it is considered as stressed state.

Support Vector Machine: Joachims [3] was the first person who introduced and used the Support vector machine classifier in solving categorization task. SVM works based on structural risk minimization principle, and its idea is to minimize the true error by finding a hypothesis h [4]. Suppose SVM is an example of two class problem, linear classifier assumed to have label y as +1 (positive) and -1 (negative) instance respectively [5]. The X denotes a vector with components x_i, where i=1,2…..n, and linear classifier of their dot product is defined as

$$W^T X = \sum_{i=1}^{n} W_i x_i, \text{ for all } x_i \in X. \text{Here } (x_i, y_i)_{i=1}^{n} \text{ every } x_i$$

vector associated with labeled with y_i. Linear classifier based on linear discriminant function is of the form

$$f(X) = w^T X + b$$

To classify a linear separable test document, SVM finds a hyper plane that has maximum Euclidean distance to nearer training samples. For non-separable training sets the amount of training error is measured using a slack variable **ξi** . Computing the hyper plane to solve an optimization problem is given in equation (3) and (4)

$$\text{minimize:} \quad V(\vec{w}, b, \vec{\xi}) = \frac{1}{2}\vec{w} \cdot \vec{w} + C \sum_{i=1}^{n} \xi_i$$

$$\text{subj. to:} \quad \forall_{i=1}^{n} : y_i[\vec{w} \cdot \vec{x}_i + b] \geq 1 - \xi_i$$

$$\forall_{i=1}^{n} : \xi_i > 0$$

Random Forest: Random forest, like its name implies, consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model's prediction. The fundamental concept behind random forest is a simple but powerful one — the wisdom of crowds [2]. A large number of relatively uncorrelated models (trees) operating as a committee will outperform any of the individual constituent models.
The low correlation between Just like how investments with low correlations (like stocks and bonds) come together to form a portfolio that is greater than the sum of its parts, uncorrelated models can produce ensemble predictions that are more accurate than any of the individual predictions. The reason for this wonderful effect is that the trees protect each other from their individual errors (as long as they don't constantly all err in the same direction). While some trees may be wrong, many other trees will be right, so as a group the trees are able to move in the correct direction.

Probabilistic Neural network: Probabilistic Neural Network (PNN)[6] is multilayer feed forward neural network which is similar to back propagation neural network but differs in their learning process. PNN uses supervised learning algorithm and it consists of 3 nodes; input layer, hidden layer and output layer. The input layer consists of N nodes, which represents feature vectors and this layer is fully interconnected with hidden layer. There are no weights in

hidden layer and each hidden node represents an example vector, with example acting as the weights to that hidden node. Hidden layer is not fully connected with output layer. The most important function of output layer is determination of class which is done through a winner takes all approach.

4. EXPERIMENTAL RESULTS AND DISCUSSION

Experimental results show that by exploiting the users' social interaction attributes, the different model detection performance is implemented and observed. The performance metrics used in evaluation of dataset is in terms of accuracy, precision and Recall. The experiments were conducted with only fine tune the network with facebook dataset in 5-fold, and achieved the highest accuracy of 79.75 percent which is shown in Table1.

Table-1: Comparison of Efficiency and Effectiveness using Different models (%)

Method	Accuracy	Recall	Precision	CPU Time
SVM	74.58	82.39	74.16	~2 min
RF	79.75	87.99	79.24	80 sec
PNN	76.18	84.63	76.43	~5 min

The dataset contains 350 users; number tweets recorded are 4,560 over 12 weeks.Figure.2 represents comparison of accuracy % for different models. Random forest Model shows 5-6% improvement than other two models. Figure.3. shows the sample screenshot for user stress level detection using pie chart. Through observation of results out of 350 students 72 are in stressed state. One reason for this modest result is that it is hardest to model time varying tweets and more number of nodes. To extract more meaningful visual attribute features content based image retrieval techniques are used to improve the accuracy. Further to improve the performance using map-reduce and graph modeling techniques can be implemented.

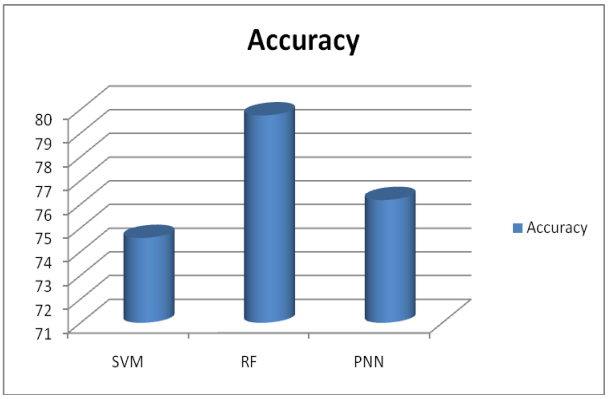

Fig-2: Experimental Analysis of different methods in terms Accuracy (%)

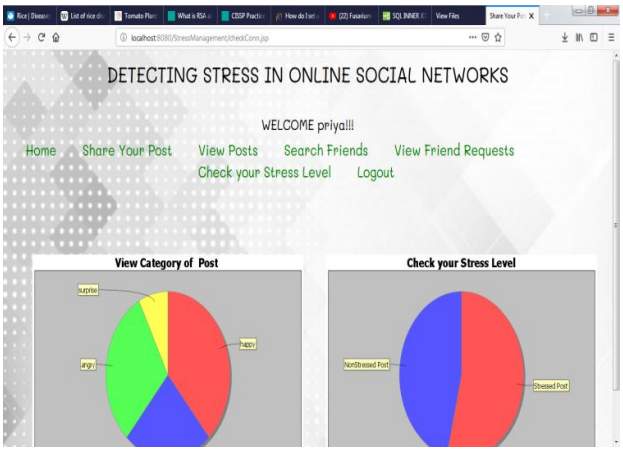

Fig-3: Each user tweet post category and their stress level

5. CONCLUSION

The paper presented a comparative analysis for detecting users' psychological stress states from users' weekly social media data, leveraging tweets' content as well as users' social interactions. Employing real-world social media data as the basis, we studied the correlation between user' psychological stress states and their social interaction behaviors. To fully leverage both content and social interaction information of users' tweets, we find a efficient model which combines the CNN with a SVM, RF and PNN. In this work, we also discovered several intriguing phenomena of stress. The RF model outperforms 5-6% higher than other two classification methods. The paper presented a comparative analysis of various classification methods for user stress detection. The data from multiple modalities such text, image and audio are collected from students facebook database. From raw data features are extracted using cross auto encoder with convolution neural network. Employing students facebook the social media data as the basis, we studied the correlation between user' psychological stress

states and their social interaction behaviors. To fully leverage both content and social interaction information of users' tweets, we find a efficient model which combines the CNN with a SVM, RF and PNN. In this paper, several emotions of the user are predicted and more negative emotions reveals pshyic stress to that student. The RF model outperforms 5-6% higher than other two classification methods.

REFERENCES

[1] Thilagavathi. P. et.al," Detecting Stress Based on Social Interactions in Social Network", International Journal of Engineering Research & Technology (IJERT), Vol.6,Issue.8,Pp.1-52018.

[2] Huijie Lin et.al.," Detecting Stress Based on Social Interactions in Social Networks", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 29, NO. 9, Pp.1820-1832 SEPTEMBER 2017.

[3] Jaohims, T.1998. Text Categorization with Support Vector Machines: Learning with Many Relevant Features, In Proceedings of the 10th European Conference on Machine Learning, London, and Pp.137-142.

[4] Lojo, D., Losada, D.E. and Barreiro, A. CIE-9-MC Code Classification with KNN and SVM, In Proceedings of Third International Work-Conference on the Interplay between Natural and Artificial Computation, IWINAC, Spain, Pp.499-508.

[5] Vapnik.V.N.1995. The Nature of Statistical Learning Theory, Springer.

[6] P.Iswarya and Dr.V.Radha, "Comparative Analysis of Feature Extraction techniques for Tamil Speech recognition", Proc. of International Conference on Emerging Research in Computing, Information, Communication and Applications, (ERCICA 2013), Bangalore**, *Elsevier***, Pp. 243-248.