Social Network Stress Analysis Using Word Embedding Technique

E. Venukhaa[1], M.Varkees[2],

[1]PG Student, Department of CSE, DMI Engineering College, Tamilnadu, , India
[2]Assistant Professor, Department of Computer Science & Engineering,, DMI Engineering College, Tamilnadu, India

---***---

Abstract - *Stress is known to be the biggest threat to people's well-being, and anticipating its appearance before depression within serious problems is of significant importance. Proposed method for predicting the stress status of social networks from their annual, yearly, weekly, and regular data, these data may provide input to the proposed approach, taking into account the opinions of the users on the sentiment analysis of different online social network data. Using the word embedding method, translating words into vectors of real numbers, using the socialized Word2Vec technique, this implies user stress status and outcomes are classified to use decision tree text classification, to the system predicts various feedback states. Eventually the application gives an overview of the level of stress.*

Keywords: Online Social Network, Stress Status, Trained dataset Decision tree text classification, Word embedding.

1. INTRODUCTION

Over the last few years, online social networks have become increasingly popular as a way of sharing various types of content created by users, such as posting personal status updates, uploading photos and sharing current geographic locations. As well, users can communicate with other users by commenting on their posts and setting up conversations. Those interactions, users can express their feelings and thoughts, and report their everyday activities, creating a wealth of useful information about their social behaviors. The advancement of social networks such as twitter, facebook and Instagram, YouTube, a growing number of people will share their daily events and moods and interact with friends through social networks. This inspired us to propose the framework of word embedding technique and classification of decision tree text to predict the stress status of comment. The word embedding concepts then used to evaluate the overall outcome of the optimistic, negative, and neutral comments within the social networks.

1.1 Overview of Stress

Because stress is a very serious problem that is that day by day, a lot of people suffer from this problem. It seems to be a big issue and that is why it motivates us to focus on it. Older diagnoses of depressed patient were performed on the basis of questionnaires and their behavior was reported by family or friends. But the finding wasn't so precise and qualitative. Contrary to that, social media is a powerful tool in posting our comments to predict an individual's stress levels. Many people use social media platforms such as Face Book, Twitter, YouTube, LinkedIn, Instagram and others, we share thoughts, emotions, and feelings, feelings of guilt, worthlessness, human impotence and selfish existence etc. All they post is connected to their everyday activities & happenings. Social media helps to know the thought, mood, behaviors & socialization of the person. Thus we can assess the stress status of that particular individual by analyzing the social media data & applying algorithms on it. And diagnosing the person can help until he / she become more affected. User data is collected from social media using developer platform of social networks. Web application is an integral part of this paper.

The social media posts are considered in this paper, social network sites is a data source and screening tool for classifying users according to content generated by the user. The goal of the paper is to create a web application using real-world social media data as an input and to forecast performance as different stress status. Using sentiment analysis is one of the key emerging technologies in the effort to help people navigate the enormous amount of online content generated by users. Then word embedding approach used to map words to real-number vectors. These propose to use decision tree algorithm in this application. In this framework used the NLP libraries to access the Socialized Word2Vec algorithm in user comment.

2. LITERATURE SURVEY

Andrey Bogomolov, Alex Pentland, Michela Ferron, B. Lepri, Fabio Pianesi.[1] the study of "Daily Stress Recognition from mobile phone data, weather conditions and individual characteristics", That day-to-day stress can be interpreted reliably in terms of behavioral measurements, obtained from the customer's cell phone activity what's more, from additional markers, for example, the climatic conditions. In work environments where stress has become a serious problem affecting productivity, resulting in occupational problems and causing illnesses in the health. Our system could be extended and used to detect stress related conflicts and stress contagion early, and to support balanced workloads.

A. H. Alpaslan, Kadriye Avcl, Nusret Soylu, H. L. Guzel.[2] "The Association of Problematic Internet Uses, Suicide Probability, Alexithymia and Loneliness among Turkish Medical Students", the results of this study suggested a substantial correlation between Internet addiction and increased suicide risk among adolescents. Suicide prevention programs should pay more attention to internet dependent

teenagers. There should be stronger oversight of web content and more information on understanding web content. The goals of this cross-sectional research were to investigate suicidal ideation connections in a broad representative Taiwanese teenage population and attempt with internet addiction and internet activities.

Chenhoa Tan, Lee L., Tang J., Jiang L., Zhou M. [3] "User Level Sentiment Analysis Integrating Social Networks", a user-level sentiment analysis demonstration can be significantly improved by integrating social network connection information. Such links may suit interest, for instance when a Twitter user wants to pay attention to other status updates, or homophile. The main motivation behind our methodology is that users who are somehow related may have more probability of having similar opinions; thus, relationship knowledge can support what they can derive from their utterances about the views of a user.

B. D. Phung, Saha, Nguyen, and V. Svetha, [4] "A framework for classifying online mental health-related groups with an interest in depression", mental dispersion frequently occurs in mixtures, for example a patient with a nervousness problem may also generate gloom. This psychological wellness condition which attends gives consideration to our work in characterizing online networks with an enthusiasm for sorrow. We also expressed the psycho-etymological topics and features in. The creations are to add to our plan, use them. Following a vehicle Learning Technique, we have identified a joint presentation. Collection of these attributes together with existing features associated with the Online Psychological Health Network. From these functions, we have devised a collaborative modeling system to identify co-occurring online communities related to mental health.

Chi. Wang, Tang Jie, Han J. [5] "Dynamic Social Influence Analysis by Time Dependent Factor Graphs",(Feb 2011) A Pair Wise Factor Graph (PFG) model to formalize the problem in a probabilistic model and to extend it by integrating time information resulting in the Dynamic Factor Graph (DFG) style, the proposed approach will effectively discover dynamic social influences. In future work, parallelization of our algorithm can be done to further scale it up. Propose a Pair Wise Graph Factor (PFG) model to demonstrate the social impact of social systems. The purpose of a productive calculation is to take in the model and make induction. We also suggest a Dynamic Factor Graph (DFG) model to fuse the knowledge about the time. Trials are carried out on three distinct classifications of knowledge sets which show that the proposed methodologies can induce the dynamic social impact skillfully. The findings are related to the issue of impact boost, which aims to locate a small subset of hubs within an informal entity that could optimize impact spread. Trials show that the suggested solution will promote application

3. SYSTEM ANALYSIS & ALGORITHM

3.1 Existing System

Thanks to its extensive use and applications a lot of work has been done in this area. This section addresses some of the methods put in place to accomplish the same goal. These works are mainly distinguished from the Stress Detection Systems algorithms. Existing works have shown that it is feasible to use social media for healthcare, and particularly stress detection. Users don't always explicitly communicate their stressed states in the Social Network message. While no interruption of the publication itself is revealed, we can learn from the engaging follow-up comments made by the user and his buddies, that the user is really stressed at work. Users with high psychological stress may experience low social networking behavior.

3.1.1. Disadvantages of Existing System

Following are the disadvantages of existing system:

> Text data are not preprocessed in existing systems.
> Algorithms which are not suitable for large datasets in existing system
> Performance in prediction of stress status is low
> word grading (positive, negative, neutral) is not properly organized.

3.2 Proposed System

The suggested system covers server administrator, user of Online Social Network. The administrator deals with each of the information in this system. The User signs to become a part of social organizations. Creating a practical application to predict the stress status of users with social media data in online social networks, and creating an engaging user interface to present stress states for users and friends, with a collection of user feedback based on the input to the proposed system the data is being pre-processed for the analysis of sentiment. By using the word embedding technique to turn words into real number vectors, using Word2Vec socialized technique, those proposed user stress status are classified using the text classification of the decision tree. The administrator provides an overall analysis of the stress status of comments.

3.2.1. Advantages of Proposed System

- Word2vec technique works for both a small number of datasets and a large number of datasets.
- Textual data is pre-processed using method such as word embedding; word classification is correctly organized because it uses data from sentiment analysis.
- Easy-to-use GUI

3.3 Algorithm

Input: Social media data with N user

Output: Stress Status Prediction

Step1: User posts on social media

Step 2: User friends can comment on a post

Step 3: System collects social media comments

Step 4: Analyze data sets comments

Step 5: Word embedding concepts were used to create data sets. I.e., word count Collected and categorized as negative, positive, and neutral

Step6: Separation of every word from the sentence and testing with negative, positive and neutral data

A= [x y z]

x=[x1,... xn]

y= [y1, y2..........., yn]

z= [z1, z2..................., Zn]

If An in x:

Probability (positive words)

Else If An in y:

Probability (negative words)

Else:

Probability (neutral words)

If (positive words>=negative words and positive

words>>=neutral words):

Return positive

Otherwise

IF (negative words)

4. SYSTEM DESIGN & IMPLEMENTATION

Architectural design depicts the data structure and application components expected to develop a computer based system. It's contains of the architectural style the system takes on the component structure and estate. Software Development Life Cycle is the most creative and challenging stage of developing a system. The first step in design is to determine how and in what form the output shall be generated.
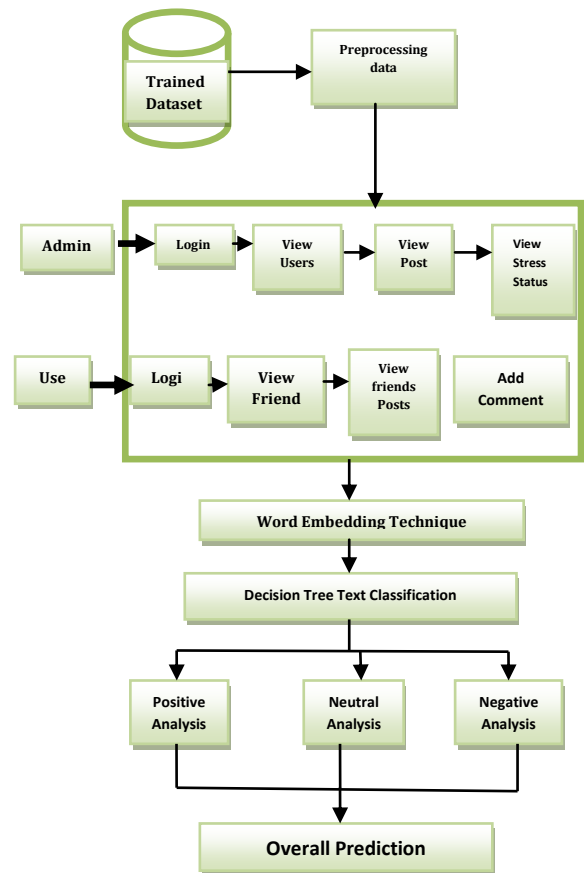

Fig 1: System Architecture

4.1 IMPLEMENTATION

4.1.1 MODULE

- Trained Dataset
- Preprocessing Data
- Admin
- User
- Word Embedding technique
- Decision Tree Text Classification
- Stress status prediction

4.1.1.2 Trained Dataset

The trained data set includes the various sentiment values of the user's social media comments, and different user ids.

4.1.1.3 Preprocessing Data

Converting the raw data into a clean data set, all the posted user information is stored in the database for analysis. Dataset of sentiment analysis used in this framework. The comments of the multi-class analysis performed over it are extracted from a sentiment analysis.

4.1.1.4 Admin

Admin will log into the dashboard with the right user _ id password. Upon successful login, some operations such as: View user, Display all request and answer mates, View message, View Stress status can be performed. The admin considers all end users and authorizes and applies the comment label for the stressed status as positive, negative, neutral. Using the graph the Admin offers an overall stress state review.

4.1.1.5 User

There are n numbers of users present in this module. User must register before performing any operations. When user accounts are made their information will be stored in the database. After successful registration, login using authorized user name and password, some operations, such as viewing your profile, searching friends and requesting friends, will be performed once login is successful. See all your friends, add a comment, post and show your friend's post and see your entire friend's comments by adding their feelings or comments

4.1.1.6 Word Embedding Technique

You have two primary ways to use word embedding: use pre-trained models and use your data and the Word2Vec technique to train custom models. Using the two natural language processing library for word embedding technique in this application which is Spacy library for good output purposes another is gensim for training access to the Word2Vec algorithms, and also allows for embedding by trained word.

4.1.1.7 Decision Tree Text Classification

One of the most important tasks of Natural Language Processing is interpretation of texts. It is the method of classifying strings of text or documents into different categories, based on the string contents. Use Scikit-Learn library to train a text classification model for machine learning.

Following are the steps that allow a decision tree text classification model to be built in this application

- Importing Libraries
- Importing the dataset
- Text Preprocessing
- Splitting data
- Converting text to numbers
- Training and testing sets
- Training decision tree text classification model and predicting sentiment

4.1.1.8 Results: Stress Status Prediction

The proposed system analyzes the users' positive, negative, and neutral status, and makes the estimation of various users' overall stress designation.

Fig 2: Admin Login Page

Fig 3: Admin Login Page
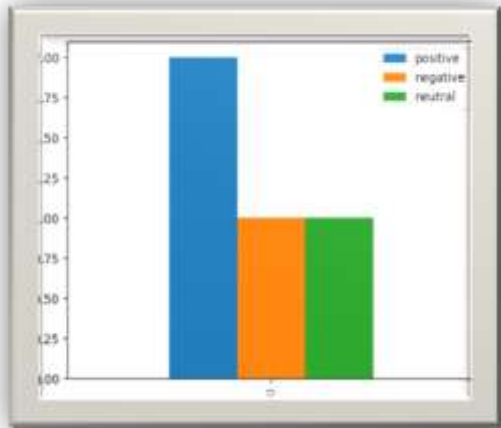

Fig 4: New User Registration Page
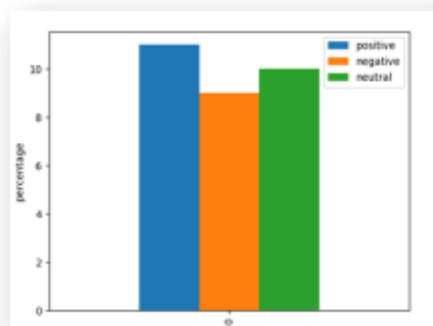
Fig 5: User Comment Stress

Fig 6: Overview of User Comment Stress Level

5 CONCLUSION & FUTURE ENHANCEMEN

The proposed framework for predicting the stress status of the social network from their annual, monthly, weekly, and regular data, In this context, in the sentiment analysis of various online social network data the comments of the users are considered and given these data as inputs to the proposed method. The comments are analyzed using the word embedding technique for the process of predicting stress status. To classify the comments in the stress state prediction process using the decision tree text classification, the framework predicts different categories of stress status comments. This work ensures that anyone suffering from stress and depression can ensure their privacy and at the same time get the best care possible.

There are several prospective paths worth exploring in the future, based on the present work. Additionally, social structure plays a vital role in defining an individual's stress level that is a useful reference for future related studies. The project's future work is to develop a system that will act as a survey method not only by predicting the stress but also by being able to analyze people mind. In the future we can use

picture as it is shared by a user for their prediction of stress. Using these photos we measure the degree of user tension. Using the method we will classify mental disorders in social networks.

REFERENCES

[1] A. Bogomolov, A. Pentland, Michela Ferron, B. Lepri, Fabio Pianesi "Daily Stress Recognition from mobile phone data, weather conditions & Individual Traits", in ACM International Conference on Multimedia pg. no 477-486, March 2014

[2] A.H. Alpaslan, Kadriye Avcl, N. Soylu, H. Guzel."The Association Between Problematic Internet Uses, Suicide Probability, Alexithymia & Loneliness Among Turkish medical Students," Journal of Psychiatry, 13-106, pages 1-8, Nov 2009

[3] B. Saha, Nguyen, D. Phung, and V. Svetha, "A framework for classifying Online Mental health-related communities with an interest in Depression", IEEE Journal Of Biomedical and Health Informatics, pages 2168-2194, March 2016

[4] Chenhoa Tan, L. Lee, J. Tang, L. Jiang, M. Zhou,"User level Sentiment Analysis Incorporating Social Networks," Journal of research gate DOI: 10.1145/ 20204 10.1145/2020408.2020614, September 2011

[5] Chi. Wang, Jie Tang, J. Han."Dynamic Social Influence Analysis through Time Dependent factor graphs," IEEE Trans on Knowledge & Data Engineering, Vol 6, No.7 pg. 123-156, Feb 2011

[6] H. Lin, J. Jia, Lexing. Xie, J. Tang, "Detecting Stress Based on Social Interactions In Social Networks", IEEE Transaction on Knowledge and Data Engineering, Vol.29, No.9, pges 126-154, September 2017

[7] Jennifer, C. Robles, M. Edmondson, Karen, "Predicting Personality from Twitter", In Passat/social com 2011, Privacy, Security, Risk and Trust, pages 149– 156, Feb 2011

[8] John T. Cacioppo, James H. Fowler, Nicholas A. Christakis, "Alone in the Crowd The Structure and Spread of Loneliness in a Large Social Network", in Journal of Personality and Social Psychology Vol. 97, No. 6, 977-991, Nov 2009

[9] M.D. Choudhury, Michael, Scoft Counts, E. Horvitz, "Predicting Depression via Social Media", Seventh International AAAI Conference on Web logs and Social Media, pages 128-134 April 2013

[10] Y. Zhang, J. Tang, Y. Chen, J. Rao."Mood cast: Emotion Prediction via Dynamic Continuous factor graph model,"

IEEE International Conference on Data Mining, Pages 10-16, March 2016

AUTHORS

Venukhaa E is currently pursuing her M.E degree in department of studies in CSE at DMI Engineering College, Aralvoimozhi, Tamilnadu, India.

Varkees M is currently working as Assistant Professor, Department of Computer Science & Engineering at DMI Engineering College, Aralvoimozhi, Tamilnadu, India.