Depression Intensity Estimation via Social Media Data Using a Deep Learning Framework

Mr. Gaurav Kawade
Mr. Yash Thakare
Mr. Divyaratna Bhusari
Students, Department of Computer Engineering
Prof. Shiv Shankar
Assistant Prof. Computer Department ,
Padmabhooshan Vasantdada Patil Institute of Technology, Pune

s

Abstract*-Stress and depression are the most widely perceived and incapacitating mental problems that affect society rightly. Automated health monitoring systems may be key and important for improving the framework of sadness and stress recognition using social networking. Sentiment analysis refers to the use of natural language processing and content mining approaches to plan sentiment or opinion recognition. Sensible computing is the examination and advancement of frameworks and gadgets that can understand, understand, process, and replicate human effects. Sentiment analysis and deep learning techniques can provide powerful algorithms and frameworks for targeted assessment and observation of psychological issues, and depression and stress in particular. The use of sentiment analysis and deep learning methods for depression and stress detection and monitoring is discussed. In addition, the basic scheme of an inclusive multimodal framework for investigating stress and depression, including appraisal investigation and emotion processing strategies, is studied. In particular, the paper explores the fundamental issues and approaches such frameworks comparatively.*

Index Terms - Deep Learning, Health, Stress and Depression, Sentiment Analysis, Social Media.

I. INTRODUCTION

Social media is the richest source of human-generated text input. Opinions, feedback and criticism given by Internet users reflect attitudes and feelings towards certain topics. We present a knowledge-based system, including an emotional health monitoring system to detect users with potential psychological disorders, especially depression and stress [1] [4]. Symptoms of this mental illness are usually passive. In this situation, the author argues that online social behavior extraction offers an opportunity to proactively identify mental illness at an early stage [5].

Depression and stress are among the most common and disabling mental disorders, and have a corresponding impact on society [5]. Currently, methods for screening and diagnosing depression and stress rely on self-reporting with informed assessment by health care practitioners. Provision of effective health monitoring systems and diagnostic aids can be crucial and important for improvement

Work of health professionals and lower healthcare costs. Sentiment and deep learning technologies can help meet these

objectives by providing effective tools and systems for objective evaluation. Such tools and systems do not aim to replace the psychologist or psychiatrist but can support their decisions.

Our approach is new and innovative to the practice of psychiatric diagnosis, so it does not rely on self-disclosure of psychological factors. Instead, a machine learning technique that detects mental illness in social networks uses features extracted from social network data to accurately identify potential cases of mental illness [5] [8].

II. LITERATURE SURVEY

Renata L. Rosa, Gisele M. Schwartz, Wilson V. Ruggiero, and Dem´ostenes Z. Rodr´ıguez - Online Social Sites (OSN) give important data on users feeling about various topics. Along these lines, applications, for example, checking and suggestion frameworks (RS) can gather and dissect this information. This paper exhibits a Knowledge-Based Recommendation System (KBRS), which incorporates an enthusiastic wellbeing observing framework to distinguish clients with potential mental unsettling influences, explicitly, depression and stress using CNN, BLSTM-RNN algorithms and the eSM2 opinion metric for disposition appraisal.

Guang Yang, Haibo He, Fellow, IEEE, and Qian Chen

- Estimation investigation on microblog posts has been examined inside and out, sentiment analysis of posts is as yet testing a result of the restricted logical data that they ordinarily contain. In microblog situations, emojis are much of the time utilized and they have clear passionate implications. They are significant enthusiastic signs for microblog nostalgic analysis. They address this issue by developing an enthusiastic space as a component portrayal framework and anticipating emojis and words into the passionate space dependent on the semantic composition using enhanced convolutional neural network algorithm.

M. Al-Qurishi, M. S. Hossain, M. Alrubaian, S. M. M. Rahman, and A. Alamri - In this paper, author propose a coordinated web-based social networking content investigation stage that use three level of features, i.e., user produced content, social graph associations, and user profile exercises, to dissect and identify atypical behaviors that go amiss altogether from the standard in huge scale social networking sites. A few sorts of investigations have been directed for a superior comprehension of the distinctive user practices in the discovery of exceptionally versatile vindictive users. This system used PCA algorithm for feature extraction and Profile-Based Collection Technique, Time-Based and Gradual Enhancement Technique for real time data collection.

Huijie Lin, Jia Jia, Jiezhon Qiu, Yongfeng Zhang, Lexing Xie, Jie Tang, Ling Feng, and Tat-Seng Chua - In this paper, author find that users stress state is firmly identified with that of his friends in web-based social networking, and utilize a huge scale dataset from certifiable social stages to efficiently think about the connection of user's pressure states and social associations. author initially characterize a lot of stress related text, visual, and social traits from different perspectives, and afterward propose a novel half breed model - a factor diagram model joined with Convolutional Neural Network to use tweet substance and social collaboration data for stress discovery.

Budhaditya Saha, Thin Nguyen, Dinh Phung, Svetha Venkatesh - Psychological instability deeply affects people, families, and by expansion, society all in all. Informal communities enable people with mental issue to speak with others sufferers by means of online networks, giving a precious asset to thinks about on textual indications of mental medical issues. This paper discover quiet with a stress issue may likewise depression using Sequential minimal optimization (SMO) algorithm.

Chun-Hao Chang, Elvis Saravia, Yi-Shin Chen - In this paper, target building prescient models that influence language and standards of conduct, utilized especially in online life, to decide if a user is experiencing two instances of mental issue. These prescient models are made conceivable by utilizing a novel information assortment process, authored as Subconscious Crowdsourcing, which gathers a quicker and progressively solid dataset of patients. Our tests recommend that extricating explicit language examples and social connection highlights from solid patient datasets can enormously add to advance examination and identification of mental issue. This paper used Linguistic Inquiry and Word Count (LIWC) and Latent Dirichlet Allocation (LDA) algorithms.

Andrey Bogomolov, Bruno Lepri, Michela Ferron, Fabio Pianesi, Alex (Sandy) Pentland- In this paper, propose an alternative approach providing evidence that daily stress can be reliably recognized based on behavioral metrics, derived from the user's mobile phone activity and from additional indicators, such as the weather conditions (data pertaining to transitory properties of the environment) and the personality traits (data concerning permanent dispositions of individuals). Our multifactorial statistical model, which is person-independent, obtains the accuracy score of 72.28% for a 2-class daily stress recognition problem. Moreover, author identify and discuss the indicators which have strong predictive power. This paper used TFIDF and LIWC algorithms for feature Extraction and Random forest for classification.

Bimal Viswanath† Alan Mislove Meeyoung Cha Krishna P. Gummadi –In this paper, study the development of action between users in the Facebook social network to catch this idea. Likewise find that connections in the action organize will in general travel every which way quickly after some time, and the quality of ties displays a general diminishing pattern of movement as the informal community interface ages. For instance, just 30% of Facebook user sets connect reliably starting with one month then onto the next. Curiously, and locate that despite the fact that the connections of the movement organize change quickly after some time, many chart theoretic properties of the action arrange stay unaltered.

III. PROPOSED METHODOLOGY

In the proposed systemic approach, we formulate the task as a classification problem to detect four types of psychological disorders in social networks using sentiment analysis and the deep learning framework:

i tension

ii. Depression

iii positive comments

iv negative comments

An innovative solution to monitor and detect potential users with emotional disorders, according to the classification of sentences with sad or stressful content.

A. Architecture

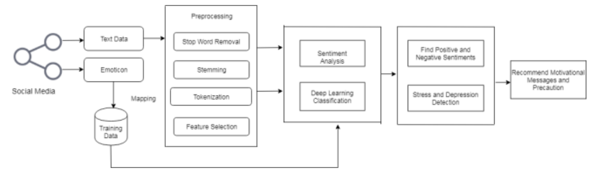

Fig. 1. Proposed System Architecture

B. Algorithm

1. Naive Base Steps:

• Given a training dataset D which consists of documents of different classes named class A and class B

• Calculate the prior probability of class A = number of objects of class A / total number of objects

Calculate the prior probability of class B = number of objects of class B / total number of objects

• Find NI, total number of frequencies of each class Na = total number of frequencies of class A Nb = total number of frequencies of class

• Find the conditional probability of occurrence of a keyword given the class:

P (Value 1/Class A) =count/ni (A) P (Value 1/Class B) =count/ni (B) P (Value 2/Class A) =count/ni (A)

P (value 2/Class B) =count/ni (B)

. . . .

. . . .

. . . .

P (value n/Class B) =count/ni (B)

• Avoid zero frequency problems by applying uniform distribution

• Classify Document C based on the probability p(C/W)

a. Find P (A/W) =P (A)*P (value 1/Class A)* P (value 2/Class A). P(value n /Class A)

b. Find P (B/W) =P (B)*P (value 1/Class B)* P (value 2/Class B). P(value n /Class B)

• Assign document to class that has higher probability.

2. Recurrent Neural Network

Recurrent Neural Network(RNN) are a type of Neural Network where the output from previous step are fed as input to the current step. In traditional neural networks, all the inputs and outputs are independent of each other, but in cases like when it is required to predict the next word of a sentence, the previous words are required and hence there is a need to remember the previous words. Thus RNN came into existence, which solved this issue with the help of a Hidden Layer. The main and most important feature of RNN is Hidden state, which remembers some information about a sequence.

Steps:

Suppose there is a deeper network with one input layer, three hidden layers and one output layer. Then like other neural networks, each hidden layer will have its own set of weights and biases, let's say, for hidden layer 1 the weights and biases are (w1, b1), (w2, b2) for second hidden layer and (w3, b3) for third hidden layer. This means that each of these layers are independent of each other, i.e. they do not memorize the previous outputs.

• A single time step of the input is provided to the network.

• Then calculate its current state using set of current input and the previous state.

• The current ht becomes ht-1 for the next time step.

• One can go as many time steps according to the problem and join the information from all the previous states.

• Once all the time steps are completed the final current state is used to calculate the output.

• The output is then compared to the actual output i.e the target output and the error is generated.

• The error is then back-propagated to the network to update the weights and hence the network (RNN) is trained.

C. Mathematical Model

The mathematical model for Stress and Depression Monitoring System is as-

S = {I, F, O}

Where,

I = Set of inputs

The input consists of set of Words/Emoticon/Microblog. It uses Twitter and Facebook dataset.

F = Set of functions

F = {F 1, F 2, F 3 FN }

F1: Data Collection

Dataset =< T, F >

Where,

T- Twitter F-Facebook

F2: Sentimeter

Sentimeter-Br2 is a word-dictionary with its respective sentiment intensity (positive or negative words), considering n-grams, verbal tenses and adverbs. The

= sentiment intensity value of an S-sentence, using the Sentimeter-Br2

Sentimeter r2(S) =(SU+SB+ST)/(K+P+Q-R)

SU stands for unigram sentiment score, ST stands for trigram sentiment score, SB stands for bigram sentiment score, k is related to sentence tension, k = 1 if there is a verb in the previous clause of the sentence; and k = 0 if the sentence is in a different tense or the sentence has no verb, then p is the total number of unigrams in the F-sentence, including words without sensory intensity (words), q is the total number of bigrams. , and r is the total number of trigrams.

F3: **Enhanced emotional dimension**
eSM(S) = CentimeterBr2(s) ∗ C ∗ exp(a1 ∗ A1 ∗)
where
C stands for volatility scale; a1 ... represents a binary factor related to the age range, if one is equal to one, the other is zero; A1 ... The weighting factor of each age range, considering four ranges; g1 and g2 are binary factors related to gender; M and F are weight factors for sex, male or female, respectively; e1 and e2 represent dichotomous factors related to the level of education (higher education or not); G and nG are weighting factors for the level of education, higher education, or no education, respectively.

F3: **Sentiment analysis**

data = <w, N>
where
W - word
N - Naive Bayes
Classification Data = <w, rnn>

Where,
W – Words
rnn – Recurrent Neural Network
O=Find Disorder (i.e. Positive comments, Negative comments,Stressed user, Depressed user)

IV. RESULTS AND DISCUSSION

A. Dataset

The dataset used to classify stress, depression, and nonstressand non-depression expressions, in the training phase, was built using sentences written by Users on an OSN. In total, 30,000 labeled Facebook messages were used.

B. Limitation of Traditional Machine Learning Algorithm

- Machine Learning Algorithms Require Massive Stores of

  Training Data
- Labeling Training Data is a hard Process
- Machine learning algorithms are deployed, there will likely be more instances in which potential bias finds its way into algorithms and data sets.

C. Classification Accuracy

The test was done with a personal computer with the configuration: Intel (R) Core (TM) i5-2120 CPU @ 3.30GHz, 8GB memory, Windows 10, MySQL backend database and jdk 1.9. This application is code designed for web application in Eclipse Oxygen IDE and executed on Tomcat 9.5 server.
Overall accuracy of naive bay and recurrent neural network classification methods. Therefore, this work provides better classification results.
Calculation formula:
TP: True positive (that is, the number is correctly predicted)
FP: False positive (that is, the number is incorrectly predicted) TN: True negative (the number of false positives is correctly predicted)
FN is a false negative (false prediction of the number of unsolicited cases);
Based on these parameters, we can calculate four

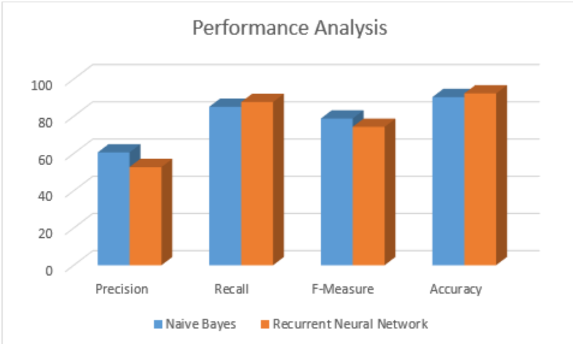

Fig. 2. Performace Analysis Graph

	Naive Bayes	Recurrent Neural Network
Precision	60.6	52.70
Recall	85.1	87.64
F-Measure	78.8	74.31
Accuracy	90.29	92.26

V. CONCLUSION

In this proposed system, automatically identifying potential online users with depression and stress poses a risk to public health. In this way, users suffering from depression can be identified and helped before taking drastic measures that can have long-term effects. Using data from real-world social networks as a basis, learn about psychological disorders of users and their social interaction habits, and recommend user health tips to be sent by mail for user interaction. Worked on social media. In future work, it is used in various services such as customer complaint system and user help framework to recognize unexpected changes in customer sentiment.

REFERENCES

[1] Renata L. Rosa, Gisele M. Schwarz, Wilson V. Ruggiero, and Dem´ostenes Z. Rodr´ıguez, "A knowledge-based recommendation system incorporating sentiment analysis and deep learning," IEEE 2019.

[2] Guang Yangang, Haibo He, Fellow, IEEE, and Qian Chen "Advanced Emotional-Semantic Neural Networks" IEEE 2019.

[3] M. Al-Kurishi, M. S. Hussain, M. Alrubaian, S. M. M. Rahman and

[4] A. Alamri, "User behavior analysis for malicious activity detection in large-scale social networks," IEEE Transactions on Industrial Informatics, vol. 14, no. 2, pp. 799–813, February 2018.

[5] H. Lin, J. Jia, J. Qiu, Y. Zhangang, G. Shen, L. Xie, J. Tang, L. Feng, and T. S. Chua, "Identifying stressors based on social interactions in social networks," IEEE Transactions on on Engineering Knowledge and Information, vol. 29, No. 9, 1820–1833, September 2017.

[6] Budhaditya Saha, Inche Nguyen, Dinh Phung, Sweta Venkates "A framework for classifying online mental health communities of interest in depression" IEEE 2016.

[7] Chun-Hao Chang, Elvis Saravia, Yi-Shin Chen "The Uneducated Crowd: A Potential Data Collection Mechanism for Detecting Mental Disorders in Social Media" 2016 IEEE/ACM

[8] Andrey Bogomolov, Bruno Lepri, Michel Ferron, Fabio Pianesi, Alex (Sandy) Pentland".

[9] 2015 IEEE Conference on Climate and Special Features

[10] Bimal Viswanath † Alan Mislove † ‡ Meeyoung Cha † Krishna P. Gummadi, "On User Interaction on Facebook," ACM 2011