Depression: OSN Users Depression Detection and Auto Motivation System using Machine Learning Techniques

[1]V. SHANTHINI, [2]Ms. K. PRAVEENA

[1]MCA, [2]ASSISTANT PROFESSOR
A.R.J COLLEGE OF ENGINEERING AND TECHNOLOGY, MANNARGUDI.

Abstract-
Depression is viewed as the largest contributor to global disability and a major reason for suicide. Generally, clinical psychologists diagnose depressed people via face-to-face interviews following the clinical depression criteria. However, often patients tend to not consult doctors in their early stages of depression. Nowadays, people are increasingly using social media to express their moods. Sentiment Analysis (SA) is the examination of the polarity of emotions and opinions expressed in the text by using computational methods. Sentiment could be expressed implicitly or explicitly in the text. Numerous studies on mental depression have found that tweets posted by users with major depressive disorder could be utilized for depression detection. The potential of sentiment analysis for detecting depression through an analysis of social media messages has brought increasing attention to this field. In this project, we aim to predict depressed users as well as estimate their depression intensity via leveraging social media (Twitter) data, in order to aid in raising an alarm. we propose a lexicon-enhanced LSTM model. The model first uses sentiment lexicon as an extra information pre-training a word sentiment classifier and then get the sentiment embedding's of words including the words not in the lexicon. Combining the sentiment embedding and its word embedding can make word representation more accurate. Derived from a combination of feature extraction approaches using sentiment lexicons and textual contents, these features are able to provide impressive results in terms of depression detection.

I.INTRODUCTION

Depression is a mood disorder that causes a persistent feeling of sadness and loss of interest. Also called major depressive disorder or clinical depression, it affects how you feel, think and behave and can lead to a variety of emotional and physical problems. In today's world, many of us rely on social media platforms such as Facebook, Twitter, Snapchat, YouTube, and Instagram to find and connect with each other. Many people in today's world live with their smartphones as virtual companions. Traditional Machine Learning classifiers were used. However, as with most other research that attempts to predict depression, the analysis was dependent on self-reported cases, and to date, approaches aiming to identify individuals who are as yet unaware of their depression diagnosis remain rare. Early detection is critical for rapid intervention, which can potentially reduce the escalation of the disorder Moreover, in the earlier study, the authors did not perform an early detection evaluation. Sentiment Analysis is the process of determining whether a piece of writing is positive, negative or neutral. Sentiment Analysis is a powerful text analysis tool that automatically mines unstructured data (social media, emails, customer service tickets, and more) for opinion and emotion, and can be performed using machine learning and deep learning algorithms. Deep learning (DL) is considered an evolution of machine learning. It chains together algorithms that aim to simulate how the human brain works, otherwise known as an artificial neural network, and has enabled many practical applications of machine learning, including customer support automation and self-driving cars.

II.SYSTEM ANALYSIS

Social networks have been developed as a great point for its users to communicate with their interested friends and share their opinions, photos, and videos reflecting their moods, feelings and sentiments. This creates an opportunity to analyses social network data for user's feelings and sentiments to investigate their

moods and attitudes when they are communicating via these online tools. The evaluation follows a time-aware approach that rewards early detections and penalizes late detections. The use of linguistic markers as a tool in the analysis and diagnoses of depression has enormous potential. Depression can so quickly be seen in text, even without the use of complex models. Simply by collecting, cleaning, and processing available data, visual analysis alone can illuminate the difference between random Tweets and Tweets that have depressive characteristics. The potential of linguistic analysis in the arena of mental health cannot be overstated. By analysing a person's words, you have a clear and valuable window into his or her mental state. Even the simplest analysis of social media can provide us with unprecedented access to individuals' thoughts and feelings and lead to substantially greater understanding and treatment of mental health. The findings offer a potential new way to uncover and get help for those suffering from depression. Deep learning's power can also be seen with how it's being used in social media technology. Let's take Pinterest for example, which includes a visual search tool that lets you zoom in on a specific object in a "Pin" (or pinned image) and discover visually similar objects, colors, patterns and more. The company's engineering team used deep learning to teach their system how to recognize image features using a richly annotated data set of billions of Pins curated by Pinterest users. The features can then be used to compute a similarity score between any two images and identify the best matches.
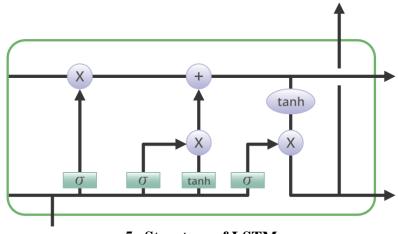
III.EXISTING SYSTEM

There are few existing studies that have applied SVM, KNN, Decision Tree and Ensemble Learning. Some studies use single set features, such as bag of words (BOW), N-grams, LIWC or LDA to identify depression in their posts. Sentiment analysis could be a field dedicated to extracting subjective emotions and feelings from text. One common use of Sentiment Analysis is to work out if a text expresses negative or positive feelings. Written reviews area unit nice datasets for doing Sentiment Analysis as a result of they usually go with a score which will be wont to train a rule. Support Vector Machines also known as support vector networks. It is a non-probabilistic linear binary classifier that analyses data for classification or anomaly detection. It builds a hyper plane into high dimensional feature space and finds a hyper plane that isolates the data into two classes with the biggest separation to the closest training data purpose of any class. Decision tree is a simple and all around used classification based systematic approach that makes the hierarchical tree from the training dataset. The state of decision tree is to divide the data hierarchically that have different characteristics. For instance, of text documents classification, roots are commonly identified in terms and internal individual nodes may be sub-divided to its children in view of the yes or no of a term in the document. Ensemble methods use multiple learning algorithms of decision tree for better predictive performance. **Adaptive Boosting (AdaBoost)** is an ensemble technique that can combine many weak classifiers into one strong classifier. It is widely used for binary class classification problems. MLP is a special case of the artificial neural network often used for modeling complex relationships between the input and output layers. Due to its multiple layers and non-linear activation, it can distinguish the data that is not only non-linearly separable
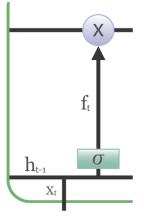
IV.PROPOSED SYSTEM

Propose a lexicon-enhanced LSTM model (LE-LSTM) to integrate sentiment lexicon into LSTM to capture more sentiment information of words. First, we use sentiment lexicon as the extra information to pre-train a word sentiment classifier. And then each word can get its sentiment embedding including the words not in sentiment lexicon. During the main training process, we concatenate the word embedding and its sentiment embedding as the input of LSTM and fine-tune the word sentiment classifier network. Long Short-Term Memory is a kind of recurrent neural network. In RNN output from the last step is fed as input in the current step. LSTM was designed by Hochreiter & Schmidhuber. It tackled the problem of long-term dependencies of RNN in which the RNN cannot predict the word stored in the long-term memory but can give more accurate predictions from the recent information. As the gap length increases RNN does not give an efficient performance. LSTM can by default retain the information for a long period of time. It is used for processing, predicting, and classifying on the basis of time-series data.

5a.Structure of LSTM

The information that is no longer useful in the cell state is removed with the forget gate. Two inputs x_t (input at the particular time) and h_t-1 (previous cell output) are fed to the gate and multiplied with weight matrices followed by the addition of bias. The resultant is passed through an activation function which gives a binary output. If for a particular cell state the output is 0, the piece of information is forgotten and for output 1, the information is retained for future use. Information is retained by the cells and the memory manipulations are done by the gates.

5b.Forget gate

To merge the lexical features obtained from datasets into the LSTM Model, we first perform a linear transformation to the lexical features in order to preserve the original sentiment distribution and have compatible dimensions for further computations. Later, the attention vector learned as in the baseline is applied to the transformed lexical features. In the end, all information is added together to perform the final prediction.

5c. Lexicon Integration into LSTM Model

Build a social networking service is an online platform which people use to build social networks or social relationships with other people who share similar personal or career interests, activities, backgrounds or

real-life connections. Social networking services vary in format and the number of features. The classification model has been exposed as a REST API which was consumed by a Web application built using Python's Flask framework. The main features include an Admin dashboard for visualization of Depression activities, an option to search tweets, and automatic generation and emailing of reports of Depression activity. In this module we developed the API for Depression analytics on chat or post user data. It focuses on keywords and analyzes chat or post according to a two-pole scale positive and negative. Datasets contain unnecessary data in raw form that can be unstructured or semi-structured. Such unnecessary data increases training time of the model and might degrades its performance. Pre-processing plays a vital role in improving the efficiency of DL models and saving computational resources. Text pre-processing boosts the prediction accuracy of the model. The preprocessing step is essential in Depression detection. It consists of both cleaning of texts (e.g., removal of stop words and punctuation marks), as well as spam content removal. In the proposed model, it has been applied to remove and clean unwanted noise in text detection. For example, stop words, special characters, and repeated words were removed. Then, the stemming for the remaining words to their original roots has been applied as a result of this preprocessing, and the dataset containing clean tweets is produced for the proposed model to be run and predicted. In this project, we executed various data preprocessing steps such as tokenization, spelling correction, stop words removing, punctuation removing, digit removing, removing a non-Bangla character, removing Emoticons, word normalization, and lemmatization and data splitting.

VI.CONCLUSION

The detection of spam emails can be evaluated by different performance measures. Confusion Matrix is being used to visualize the detection of the emails for models Several measurements are used for performance evaluation of classifiers like accuracy, precision, recall, and f-score. The results and comparisons of different classifiers after data training and testing are presented in this section. We gathered 49799 tweets from the online resource 'kaggle' and translated them into English using the python library Googletrans, which uses the Google Translate Ajax API. 42797 tweets were used to train various ML and DL models. One seven thousand tweets were used for testing in order to quantify accuracy and assessment metrics. As explained about evaluation measures in chapter 9, we have evaluated accuracy, precision, recall, and f-measures that are evaluation measures measured using LR, XGBM and Naive Bayes, LSTM-CNN and BiLSTM. Finally, using various graphs, a comparison of models is presented below. The findings in Table 4 show that the deep learning algorithm (BiLSTM) is a stronger method for detecting Depression tweet classification, with high accuracy of 98.4%.The algorithms are designed to analyze the tweet for emotion detection as well as for detection of suicidal thoughts among people on social media. We validated the performance of our method by conducting extensive experiments on a standard dataset and outperformed the other alternatives for polarity estimation.

REFERENCES:
1. B. A. Primack, A. Shensa, J. E. Sidani, C. G. Escobar-Viera, and M. J. Fine, ``Temporal associations between social media use and depression,'' Amer. J. Preventive Med., vol. 60, no. 2, pp. 179-188, Feb. 2021.
2. S. Maxim, N. Ignatiev, and I. Smirnov, ``Predicting depression with social media images,'' in Proc. 9th Int. Conf. Pattern Recognit. Appl. Methods, vol. 2, 2020, pp. 128-138.
3. J. A. Diaz-Garcia, M. D. Ruiz, and M. J. Martin-Bautista, ``Non-query based pattern mining and sentiment analysis for massive microblogging online texts,'' IEEE Access, vol. 8, pp. 78166-78182, 2020.
4. G. Xu, Y. Meng, X. Qiu, Z. Yu, and X. Wu, ``Sentiment analysis of comment texts based on BiLSTM,'' *IEEE Access*, vol. 7, pp. 51522_51532, 2019.
5. Y. Li and H. B. Dong, ``Text emotion analysis based on CNN and BiLSTM network feature fusion,'' Comput. Appl, vol. 38, no. 11, pp. 29-34, 2018.
6. A.Yadav and D. K.Vishwakarma, ``Sentiment analysis using deep learning architectures: A review,'' Artif. Intell. Rev., vol. 53, no. 6, pp. 4335-4385, Aug. 2020, doi: 10.1007/s10462-019-09794-5.

7. W. Meng, Y. Wei, P. Liu, Z. Zhu, and H. Yin, ``Aspect based sentiment analysis with feature enhanced attention CNN-BiLSTM," IEEE Access, vol. 7, pp. 167240-167249, 2019.
8. F. Hao, G. Pang, Y. Wu, Z. Pi, L. Xia, and G. Min, "Providing appropriate social support to prevention of depression for highly anxious sufferers," IEEE Trans. Comput. Social Syst., vol. 6, no. 5, pp. 879–887, Oct. 2019.
9. G. Shen et al., "Depression detection via harvesting social media: A multimodal dictionary learning solution," in Proc. 27th Int. Joint Conf. Artif. Intell., Aug. 2017, pp. 3838–3844.
10. T. Shen et al., "Cross-domain depression detection via harvesting social media," in Proc. 27th Int. Joint Conf. Artif. Intell., Jul. 2018, pp. 1611–1617.
11. R. Xu and Q. Zhang, "Understanding online health groups for depression: Social network and linguistic perspectives," J. Med. Internet Res., vol. 18, no. 3, p. e63, Mar. 2016.