Study of Depression Analysis using Machine Learning Techniques

Objectives and motivations: Detecting depression level using machine learning. The classification plays a major role in determining the kind of help a depressed person needs and also, the person with suicidal thoughts need to be identified and helped according to his condition.

I have collected dataset from various websites such as Twitter, Instagram, Facebook. First the collected data set is analyzed with the help of syntax and semantics analysis which gives the sense of depression stigma among posts posted by different age groups. In this process the syntax is analyzed for finding certain key words and relevance of those key words is mode with the help of semantic analysis which find the general emotion of the text also known as Emotion detecting system.

Now depression is a very serious mental illness that affects our mental state and increase our mental pressure. Depression weakens us physically and mentally and increases discomfort and anxiety. Depression is different from sadness. Some people believed that depression is the result of chemical imbalance but it does not capture the complications of the disease. Researchers suggest that depression does not occur because of chemical imbalance in brain. I think depression is also caused by family unrest and increasing distance from close people and various inconsistencies in society. According to world health organization (WHO) depression is the most common illness worldwide and a leading cause of mental disability. The 7.6% of people suffering people these diseases are from age group 12 which shows how early these disease can arise in an individual. Recently Bollywood superstar Sushant singh Rajput committed suicide due to depression and day by day depression is increasing in our society. The recent report on the global burden of disease and in it the depression among different genders with help of data obtained from such sources predicted that unipolar depressive episodes to be 1.9% in case of men and 3.2% in case of women and one year one-year prevalence have been estimated to be 5.81% in case of men and incase of women 9.54%. It is estimated that by the year 2020 if these trends continue then the burden of depression will rise to 5.7% of total burden of diseases.

Classification is the backbone of the analysis system which aims to produce the result on the basis of the particular features of the entities constituting the dataset. Data obtained after the process of filtering is inputted to the classifier algorithm. In depression analysis model we have different sets of inputs such as text, speech and images and according to different sets of these we use the different classifiers to properly classify them. A popular technique in classification is SVM.

I have collected dataset from various websites such as Twitter, Instagram, Facebook. I have not been able to collect data properly from these websites. This is the main flaws for this paper. The filtering mechanism constitutes the system for preprocessing of the input where the aim is to provide the data set to the learning algorithm without any redundancies. The aim is to get proper learning dataset so as to improve the accuracy of classification of the problems by the system. The detection systems run on classification algorithms which bring out a feature in the entity that is useful for characterizing the entity accordingly. When it is fed with wrong or redundant dataset then , the learning algorithm fails to comprehend the meaning to the dataset and that can lead to misclassification of the input. The varying nature of input dataset can prove challenging to these filtering algorithms where at a time an algorithm has to filter the discrepancies in data set input in text , speech or in image.

This paper provided a large dataset for identification of common traits among depressed people and identify them. This paper provides the survey about the use of machine learning techniques in the analysis of depression with their research issues.

A MULTITASK DEEP LEARNING APPROACH FOR USER DEPRESSION DETECTION ON SINA WEIBO

Objectives and motivations: The online social network provides researchers with an other perspective for detecting individuals suffering from depression. However existing studies of depression detection based on machine learning still leave relatively low classification performance suggestion that there is significant improvement potential for improvement in their feature engineering. With the rapid development of the online social network (OSN) such as Twitter and Facebook, people are more frequently using the OSN to express opinion and emotions. It provides researchers with a novel and effective way to detect the mood, communication, activity, and social behavior pattern of individuals.

According to a survey of the World Health Organization (WHO), more than 300 million people worldwide suffer from depression. Depression can cause great psychological pain, even suicidal tendencies. Moreover, evidence from a health action plan of WHO shows that people suffering from depression are much more likely to end their life prematurely that the general population.

In this paper I have collected large data set on Sina Weibo (a leading OSN with the largest number of active users in the Chinese community), namely Weibo user depression detection Dataset(WU3D). It includes more than 20000 normal users and more than 10000 depressed users, both of which are manually labeled and rechecked by professionals. I construct a Deep Neural network classification model. It implements a multi task learning strategy to process text-based word vector and statistical features simultaneously. Experimental results show that it achieves both the highest classification performance and the best robustness to unbalanced training sample.

The current methods for online depression detection mainly include two direction. (1) Traditional Machine Learning (TML) models for classification. (2) Using deep learning (DL) approaches to automatically extract features and constructing deep neural network models as classifiers.

Mining depression users based on TML mostly uses features , that is numeric vector that have been manually analyzed and extracted from users to represent the predicted object (a user, a tweet , a posted picture , etc). Modeling

approaches based on DL are mainly for jointly considering user social behavior and multimedia information such as text, pictures, videos, etc. Among them the modeling of the text information is the main research direction. Researchers have adopted NLP approaches to embed text into a high dimensional continuous vector to automatically mine word features.

However current approaches to online depression detection still face many unresolved challenges. Firstly many current studies are not user-oriented modeling. Those works usually aim to analyze and model the language style of the user. Through sentiment analysis and feature engineering of the tweet text, a classification model is developed to detect whether a specific tweet has a depressive tendency. However such results cannot be directly applied to user-level depression detection, or it may lead to an incorrect prediction. The size of thousand data samples being used. Moreover not enough studies of user depression detection have been proposed on weibo compare to Twitter and Facebook. To the best of our knowledge there is no published large Weibo user depression detection dataset available currently.