ISSN No:-2456-2165

Self Behavioral Analysis using Social Media Mining

Anu KC
Department of Computer Science and Engineering
SRM Institute of Science and Technology
Chennai, India

Jrollin Ninan Jayan
Department of Computer Science and Engineering
SRM Institute of Science and Technology
Chennai, India

Mira Kumar
Department of Computer Science and Engineering
SRM Institute of Science and Technology
Chennai, India

Ashutosh
Department of Computer Science and Engineering
SRM Institute of Science and Technology
Chennai, India

Abstract:- In recent times, we have seen an explosion in the growth and popularity of social networking, which resulted in its problematic usage. There is an increase in number of social network addiction. Symptoms of these addictions are observed passively today, resulting in affecting the users adversely. In this paper, we propose that mining online social behavior provides an opportunity to monitor the addictive usage of the user. It is challenging to detect behavior because the mental status cannot be directly observed from social activity logs. Here, we propose a deep learning framework that exploits features extracted from social network data to accurately self - analyze the behavior in online social network users. We perform a feature analysis, and also machine learning on large datasets and analyze the characteristics of the user. Sentiment Analysis is used to identify and study affective states and subjective information. The retrieved result is displayed to the user in the form of statistical graphs.

I. INTRODUCTION

The growth in popularity of social networking and messaging apps have made online social networks, a part of people's daily lives. Most of the research on social network mining focuses on exploring the knowledge behind the data for the improvement people's life. While Social Networks are expanding the users' capability in increasing social contacts, meanwhile they are actually decreasing the face-to-face interactions in the real world.

Due to the wide-scale usage of these phenomena, new terms such as Phubbing (Phone Snubbing) and Nomophobia[3] (No Mobile Phone Phobia) have been coined to describe those who cannot stop using social networking apps. Some social network disorders such as Net compulsion Information Overload have also been noted.

The recent studies have reported that the Social Network Disorder may result in excessive use, depression, loneliness, social withdrawal, and a range of other negative side effects. These symptoms are important components of diagnostic criteria for Social Network Disorders e.g., abnormal usage of social networking apps – usually associated with a loss of the sense of time or and withdrawal of feelings like anger, tension, and/or depression when the computer or apps are not available. Social Network Disorders are social-oriented and tend to happen to users who usually interact with others via online social media. Those users with Social Network Disorders usually lack offline interactions, and as a result, seek cyber-relationships to compensate. Today, identification of potential mental disorders often falls on the shoulders of teachers or parents passively. However, the patients usually do not actively seek medical or psychological services. Therefore, patients would only seek help when their conditions become very severe. However, a recent study shows a strong connection between suicidal attempt and Social Network Disorders which indicates that adolescents suffering from social network addictions are more suicidal prone than non-addictive users. The research also reveals that social network addiction may negatively impact emotional status, causing depressive mood, and compulsive behavior. Even more alarming is that the delay of early intervention may seriously damage individuals' social functioning. In short, it is desirable to have the ability to actively detect potential Social network Disorders in users on social media at an early stage. This task is very challenging. For example, the extent of loneliness and depression of users are not easily observable. Therefore, there is a need to develop new approaches for detecting the disorders amongst the affected users. We propose that mining the social network data of individuals as a complementary alternative to the conventional psychological approaches provides a better opportunity to actively identify those cases at an early stage. In this paper, we develop a deep learning framework for analyzing the behavior of the user. We extract tweets by sending a post request to the rest API, these extracted tweets are used for training. The extracted tweets are stored in the database(MongoDB). We use NLTK for sentimental Analysis and segregate the positive and negative tweets. We also use NLU(Natural Language Understanding) for intent classification. Using NLU we can train the system to classify the nature of the sentence(tweets or any comments). This will act as a double check while filtering out the tweets for training.The filtered data is trained using deep learning, and the result data,i.e the nature of tweets(happy, sad, angry, indifferent in the form of a graph).

II. RELATED WORKS

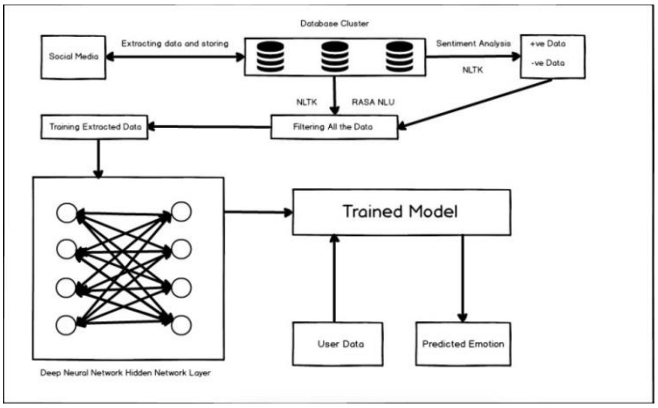
A. Existing System

The existing System uses a machine learning framework, which evaluates features extracted from social network data to accurately identify the severe cases of mental disorders. The framework was evaluated via a user study with 3126 internet users. Feature analysis was conducted, and framework was applied to large-scale datasets and analyze the characteristics of the disorders. The results manifest that framework was promising for identifying internet users with potential mental disorders.[3] The dataset was obtained by taking a survey, and it was used to deduce the mental state of a person. The framework used Naive Bayes algorithm to classify the characteristics of the mental disorders.

B. Proposed System

Here in the proposed system to, perform emotional analysis, we have used the techniques of the existing system to further modify it and implement it on social media. We have obtained the anonymous data by crawling facebook and twitter. All the mined data is stored in the database(MongoDB). We have used NLTK(Natural Language Tool Kit) for performing Sentimental Analysis on the dataset, to segregate the positive and negative statements. We also use NLU(Natural Language Understanding) for the intent classification of the dataset, this measure is used as a double check before filtering out the tweets with respect to the various emotions (angry, sad, happy, indifferent). Data obtained after sentimental analysis and NLU are compared with each other to filter the results. Finally, the obtained dataset is used for training for which we have used Deep Neural Network.

C. Architecture



D. *Modules*

➢ *Extracting and Storing data:*

We have used rest api to extract facebook and twitter data using hashtags of different categories using synonyms of Happy, Sad, Angry and Indifferent. Synonyms were obtained from the thesaurus. The mined data is stored in the database. We have used MongoDB as the database.

➢ Sentimental Analysis and Emotion Detection:

For Sentimental Analysis, we have used NLTK(Natural Language Tool Kit).The algorithm we have used is Naive Bayes Algorithm. It works on conditional probability. Conditional probability is the probability that something will happen, given that something else has already occurred. The conditional probability, can be used to calculate the probability of an event using its prior knowledge.

Below is the formula for calculating the conditional probability.

$$P(H \mid E) = \frac{P(E \mid H) * P(H)}{P(E)}$$

where

- P(E) is the probability of the evidence.
- P(H) is the probability of hypothesis H being true. This is known as the prior probability.
- P(H|E) is the probability of the hypothesis when the evidence is there.
- P(E|H) is the probability of the evidence when the hypothesis is true.

Along with Sentimental Analysis, we have used Natural Language Understanding(NLU), which is a subtopic of natural language processing in artificial intelligence that deals with machine reading comprehension. NLU is an intent classifier. NLU can be broken down into four parts:

- Distributional
- Frame-based
- Model-Theoretical
- Interactive Learning

We have used the Interactive Learning Technique. In this, we provide the System with the inputs, outputs and the system understands the given data and designs a model with respect to data provided. We use the designed model to predict the intent and classify the emotions.

E. Training and Output

After sentimental analysis, the data we got was used for training. For training purpose, we have used the Deep Neural Network.

Neural networks are a set of algorithms, modeled like the human brain, which is designed to recognize patterns. They interpret sensory data through a type of machine perception, labeling /clustering raw input. The patterns they recognize are numerical, contained in vectors, into which all real-world data, such as images, sound, text or time series, must be translated.

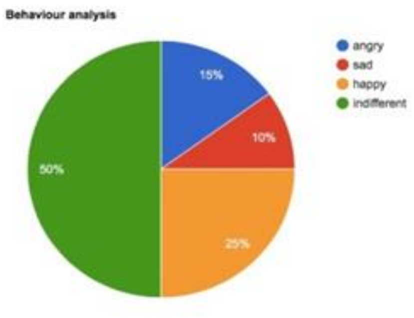
With the help of appropriate classifications, deep learning can establish correlations between different factors. This can be termed as a static prediction. When the same token is exposed to enough of the right data, deep learning is able to establish correlations between current events and future events. It can run regression analysis between the past and the future. We can say that the future event is like the label in a way. Deep learning doesn't necessarily care about time or the fact that something hasn't happened yet. Given a time series, deep learning may read a string of number and predict the number most likely to occur next. Eg:

Details of the datasets	Dataset Description
FB U	User profile, the friends of each user, the news feeds created by users with metadata (likes, comments, stickers, and geotag), the news feeds users like or comment (stickers also).
Tweet U	User profile, the followers/followees of each user, the media created by users with metadata (likes, comments, and geotags), and the contents users like or comment
FB A	Anonymous user ID that performs the action, anonymous user ID that receives the action, and timestamp of action creation
Tweet A	Anonymous media ID, the anonymous ID of the user who created the media, timestamp of media creation, set of tags assigned to the media, number of likes and number of comments received.

- Hardware issues
- Health issues

We have used the predictive algorithm in the Deep neural network. The data obtained is trained in order to predict the emotions of the user and the final behavioral analysis will be displayed in the form of a graph.

F. Table for Datasets



III. CONCLUSION AND FUTURE WORKS

In this paper, we make an attempt to identify the behavior of the user from his/her social media activity. We used Deep Learning framework that explores various features from data logs of users. This would help the users to self-analyze their behavior and would help them to use the social media in a positive way. As of now, only English comments are analyzed. In future, we can implement techniques to analyze multiple languages.

Techniques like real-time messages can also be integrated which would notify the user when there is an excess of inappropriate usage. This would help in reducing the instability amongst certain social media users.

REFERENCES

[1]. K. Young, M. Pistner, J. O'Mara, and J. Buchanan. Cyber-disorders: The mental health concern for the new millennium. *Cyberpsychol. Behav.*, 1999.

[2]. K.-L. Liu, W.-J. Li, and M. Guo. Emoticon smoothed language models for Twitter sentiment analysis. *AAAI*, 2012.

[3]. Hong-Han Shuai, Chih-Ya Shen, De-Nian Yang, Yi-Feng Lan, Wang-Chien Lee, Philip S. Yu, and Ming-Syan Chen: A Comprehensive Study of Social Network Mental Disorders Detection via Online Social Media Mining.IEEE,2017.

[4]. Witten and E. Frank. Data mining: practical machine learning tools and techniques with Java implementations. Morgan-Kaufmann, San Francisco, 2000.

[5]. M. Saar-Tsechansky and F. Provost. Handling missing values when applying classification models. *JMLR*, 2007.

[6]. P. Comon, X. Luciani, and A. L. D. Almeida. Tensor decompositions, alternating least squares and other tales. *Journal of Chemometrics*, 2009.

[7]. F. Chang, C.-Y. Guo, X.-R. Lin, and C.-J. Lu. Tree decomposition for large-scale SVM problems. *JLMR*, 2010.

[8]. C.-C. Chang and C.-J. Lin. LIBSVM: a library for support vector machines, 2001.