International Journal of Research in Advent Technology (IJRAT) Special Issue
E-ISSN: 2321-9637
Available online at www.ijrat.org
National Conference on "Role of Information Technology in Social Innovations"
26th & 27th February 2019

Mental Disorder Detection on Social Media using Machine Learning

Dr. Meenakshi A. Thalor, Monika S. Budhavant, Kajal R. Sonawane, Pradnya J. Adagale
[1]*Associate Professor, Computer Engineering', AISSMS Institute of Information Technology, Pune*
Email:thalor.meenakshi@gmail.com; monikabud6@gmail.com

Abstract- In today`s world Social networking craze is increasing day by day. Social media get popular among youngsters, senior citizens and almost all the peoples. But with the excessive use of social network communication, it affects the human life very badly. Recently, a growing number of mental disorders have been observed in social networks, such as the dependence on cybernetic relationships, information overload and network restriction etc. Currently, the symptoms of these mental disorders are observed, which causes several dangerous issues. In this article, we state that using social behaviour online offers the opportunity to actively identify the detection of mental disorders at very early stage. It is difficult to identify the disorder because the mental state cannot be observed directly in the registers of online social activities. Our new and innovative approach of identifying mental disorders is not based on the psychology questionnaires. Instead of that, we propose a framework of machine learning which exploits the characteristics extracted from social network data to accurately identify possible cases of disorder cases. We also proposing a new disorder based tensor model to improve accuracy. To increase the scalability of tensor model, we further improve efficiency with performance guarantees. Our framework is evaluated through a user study with no of users of the social network.

Index Terms- Social network, mental disorder detection, feature extraction, machine learning Classifier.

1. INTRODUCTION

Psychological mental disorder is becoming a dangerous threat to peoples health now a days. With the rapid pace of life, more and more people are feeling mentally disturb. It is not that much easy to detect users mental disorder in an early stage to protect user from getting depressed. With the ease of social networking, individuals users are sharing their day by day activities, facts, feelings, emotions, various information and interacting with friends via networking media stages, making it possible to use online social network data for mental disorder detection. In our system, we somewhat find that users disorder state is closely related to that of his/her friends on social media, and we employ a large-scale dataset from real-world social platforms to systematically study the correlation of users disorder states and social interactions. We first define a set of mental disorder-related textual, visual, and social attributes from various aspects. With the rapid growth of life, progressively and more individuals are feeling stressed as well as depressed. Though Mental disorder itself is non-clinical and common in our life, excessive usage can be harmful to peoples physical and mental health. Users social interactions on social networks contain useful cues for stress detection. Social psychological studies have made two interesting observations. The first is mood ;contagion that means a bad mood can be transferred from one person to another person during social interaction. The second one is Social Interaction that means people are known to social interaction of user. The advancement of social media networks like Twitter, Facebook and SinaWeibo2, an ever increasing number of people will share their every day events and moods, and interact with friends through the social networks on daily basis. We can classify this using machine learning framework. Due to leverage both Facebook post content attributes and social interactions to enhance mental disorder detection. After getting the output that is in the form of percentages as well as in disorder level, system can recommend the user to hospital for further treatment, we can show that nearby hospitals on map and system also recommended to take precautions to user for avoid stress.

2. LITERATURE REVIEW

1. F. Chang, C.Y. Guo, X.R. Lin, and C.J. Lu.: The support vetor machine is used to pattern classification, growing trends, at the same time speed up the time of training. for the dual optimization two methods are applied namely sequential minimal optimization and another is support vector machine. In this divide and conquer strategy is decomposes an input space. decision tree is used to decomposes the data space.

2. Chi Wang, Jie Tang, Jimeng Sun, Jiawei Han: Author proposed, the PFG(pair wise factor graph) model is used to formalization of the inuence between the two users in wide variety of large social network.

DFG(dynamic factor graph) is used for the time **information can show that influence maximization** problem. To overcome the problem in sum product algorithm where the message passing between

neighbouring nodes need to wait for all message to arrive and that does not **directly to reflect the influence**
between two user**'**s.

3. Liqiang Nie, Member: Medical terminology scheme is used to bridging the vocabulary gap between health seeker and healthcare knowledge.novel scheme is used to code medical records with corpus related terminology.health tap example in which the question is asked by patient and same quetion is written by health seekers and the answer to this question is provided by the well trained experts with multiple

possible meanings.In future , the investigation on how **to flexibly organize** the unstructured medical content

into user need-aware technology by leaverging the recommended medical terminology. 7. Quan Guo,

4. Huijie lin ,Jia ff Jia,Quan Guo:Author states that By combining a Deep sparse neural network to incorporate ffi di erent feature from cross media microblog data,the framework is quite feasible and e cient for dtress detection.The Proposed model consist of 3 level framework to formulate the problem and propose a middle level representation according to psycological and art theories which can narrow the gap between low level cross media features and high level stress semantics.

5. Jenifer Golbeck: Openness conscientiousness, extroversion, agreeableness , and neuroticism based on this attributes author able to develop a model that can predict **personality on each of the five personality** factors within between 11 percentage and 18 percentage of the actual values.

6. Sepandar D. Kamvar: In this paper,proposed model **is worked as,firstly a** URL server collects URLs of pages to be crowled. these pages contains post, pages of social network.Then crowler sends fetched page to feeling indexes which extracts sentences which contains the feeling or emotion along with date and

location.After processing of indexes the extracted data **is send to the we feel fine database.API** server **translates specified URL to SQL queries and then**
return the SQL result to browser as XML.Query cache used for faster processing.The frontend is java applet which translates user actions into API queries and send them to API server and translate the returned result in interactive visualization.

7. Yuan Zhang, Jie Tang, Jimeng Sun, Yiran Chen, and Jinghai Rao: Moodcast deal with the users emotional status according to the past experiences which incorporates users dynamic status information

(eg. locations, activities and attributes) and social **influence from users friends into unified model. It uses**
the dynamic continuous factor graph model. Moodcast can accurately predict emotion status of more than 62 percent of users and 8+ percent improvement than baseline methods.

8. Andrey Bogomolov, Bruno Lepri, Michela Ferron, Fabio Pianesi, and Alex Pentland: In this paper, An alternative approach providing evidence that daily stress is recognized from phone calls, sms and similarly based on the weather conditions. The family background or status is also responsible for the stress.Stress is condition known as the physical , psychological and behavioral conditions . Two classes are used stress**ffi**vs non-stressed.The random forest algorithm is used for the **tree classification which**
found very e cient. The stress information is distributed by the participants felting seven items scaled with 1=**"** not stressed**"** 4=**"**neutral**"** 7=**"**extremely stress**"** in this experiment 117 subjects are used.This multifunctional statistical model which is person independent obtain the accuracy score of 72.28 percentage for the two classes daily stress recognization problem.

9. Maxim stankevich, Vadim Isakov: The main idea of the task is to classify users into 2 groups: risk case of

depression and non risk case. In this paper we consider **the classification task of Reddit users by processing**
their text messages in order to detect depression.

10. Sharath Chandra Guntuku, David B Yadens:In todays world, the rates of diagnosing mental illness have over increased. Symtoms associated with mental illness are observable on twitter,facebook,web forums.For feature extraction, Linear regression and then SVM are used.Author consider 4 approaches to pre dict depression i.e a) prediction based on Survey responses b) prediction based on self declared mental health statusc) prediction based on forum membership d) prediction based on annotated posts.

3. PROPOSED SYSTEM ARCHITECTURE

A system architecture is the conceptual as well as effective model that defines and describes the structure, behaviour , and more detail perspectives or views of the system. The following architecture shows the general flow of our system.

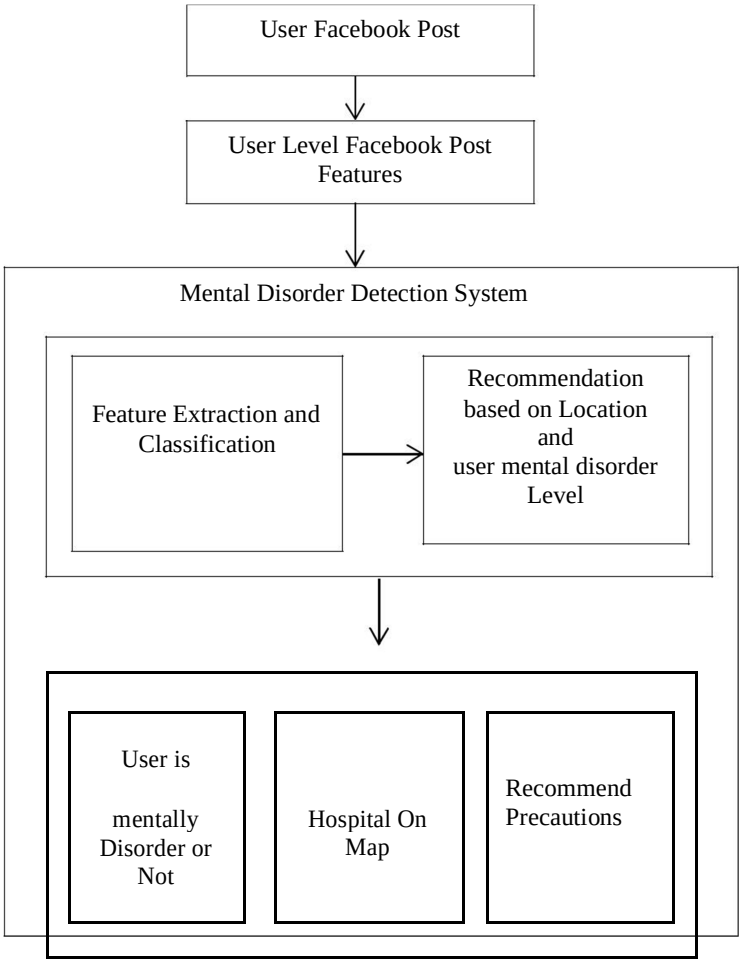

Fig. 1 Proposed System Architecture

In the proposed system architecture, we are try to detect whether the user is mentally disturb or not due to interaction with social network. In a social media network contain Facebook, twitter and many more applications. With the ease of Facebook, users are interacting with other people to share information. User can post different status base on mood on a Facebook. Facebook-level attributes describes the User level post behavior attributes as analyzed from a user's monthly Facebook postings, time of particular post, type of that post; social interaction attributes extracted from a user's social interactions with other people. In particular basis, the attributes of social interaction can further be broken or categorized into:

(i) the content attributes of social interaction are extracted from the content of users' social interactions with friends or with other peoples like feelings and

emotions; and (ii) structure attributes of social interaction extracted from the structures of users' social interactions with other friends. So based on this user input post, we can fetch the user level Facebook post features on that input of Facebook posts. For the features extraction step word and stemming algorithm are used and for classification the Naïve Bayes algorithm is used which gives result in percentage based on that mental status of user can be determined.

In the above architecture there are two modules namely Feature Extraction and Classification module and Recommendation module.

3.1. Feature Extraction and Classification:
3.1.1. Feature Extraction:-

In machine learning, pattern recognition as well as in image processing, feature extraction starts from an initial set of measured data and builds derived values or features intended to be informative and non-redundant, facilitating the subsequent learning and generalization steps, and in some cases leading to better human interpretations. Feature extraction is nothing but a dimensionality reduction process, where an initial set of raw variables is reduced to more manageable groups for processing, while still accurately and completely describing the original data set. For feature extraction the two popular algorithms are used that is stemming and step word.

- Stemming and step word :-

A computer program or subroutine that stems word called as a stemming program or stemmer.

3.1.2. Classification:-

Classification is a approach of supervised learning in which the computer system program learns from the input data given to the system and then uses this learning approach to classify the new observation. The dataset maybe bi-class as well as multi-class. some of the classification examples are like speech recognition, biometric identification, recognition of handwriting, document classification and so on. There are various classification algorithms are there such as linear classifiers like Logistic regression and Naïve Bayes classifier, Support Vector Machines(SVM), Decision Tree, Boosted Trees, Random Forest, Neural Networks. One of them is Naïve Bayes which is easy to implement at initial stage and gives estimated results. The explanation of Naïve Bayes algorithm is as follows

- Naïve Bayes :-

Naïve Bayes is an classification technique or approach based on Bayes theorem with an assumption of independence among the predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. Even if these features depend on each other or upon the existence of the other features, all of these properties independently contribute to the probability. Naive Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods.

3.1.3. Recommendation:-
A Recommendation engine which is also called as Recommender system is nothing but a software that analyzes the set of available data to make suggestions for something. Recommendation engine are very common among social media, e-commerce, and content based websites.

For the recommendation purpose, there are many algorithms used such as k-Means, fuzzy k-Means, matrix factorization, collaborative filtering etc. The best algorithm for recommendation is KNN (K-nearest Neighbour) which is easy to implement and gives

accurate result. KNN algorithm is explained as follows.

- KNN Algorithm:-

The k-nearest neighbors (KNN) algorithm is a very

simple, and supervised machine learning algorithm which is used to solve both classification as well as regression problems. A supervised machine learning algorithm that uses the labelled data to label the unlabeled data. K nearest neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure. KNN specially used in statistical estimation and pattern recognition purpose.

4. CONCLUSION

In this paper, automatically identify potential online users with SMMDs. Psychological Mental Disorder is threatening people's health. It is non-trivial to detect Mental Disorder timely for proactive care. Therefore we presented a framework for detecting users psychological Mental Disorder states from users monthly social media data, leveraging Facebook post content as well as users' social interactions. Employing real-world social media data as the basis, we studied the correlation between users' psychological Mental Disorder states and their social interaction behaviors we recommended the user for health consultant or doctor. We show the hospitals for further treatment on a graph which locate shortest path from current location user to that hospital. We recommended the user for health precaution send on mail for user interaction purpose.

REFERENCES

[1] F. Chang, C.-Y. Guo, X.-R. Lin, and C.-J. Lu.,"Tree decomposition for largescale SVM problems." JLMR, 2010.
[2] Chi Wang, Jie Tang, Jimeng Sun, Jiawei Han, "Dynamic social Inuence Analysis through time dependant factor graph," IBM TJ Waston Reaserch center, USA.
[3] Liqiang Nie, Member ,IEEE , Yi-Liang Zhao, "Bridging the vocabulary gap between health Seekers and Healthcare Knowledge," August 2013.
[4] Huijie lin , Jia Jia, Quan Guo, "Psychological stress detection from cross media microblog data using deep sparse neural network" Jenifer Golbeck, "Cristina Robles Predicting personality from twitter," 2011.
[5] Sepandar D. Kamvar, "We feel _ne and searching the emotional web". In Proceedings of WSDM, pages 117126, 2011.
[6] Yuan Zhang, Jie Tang, Jimeng Sun, Yiran Chen, and Jinghai Rao."Moodcast: Emotion prediction via dynamic continuous factor graph model." 2013 IEEE 13th International Conference on Data Mining, pages 11931198, 2010.
[7] Andrey Bogomolov, Bruno Lepri, Michela Ferron, Fabio Pianesi, and Alex Pentland. "Daily stress recognition from mobile phone data, weather conditions and individual traits". In ACM International Conference on Multimedia," pages 477486,2014.
[8] Maxim stankevich, Vadim Isakov,"Feature Engineering for Depression detection in Social Media",2017.
[9] Sharath Chandra Guntuku, David B Yadens," Detecting depression and mental illness on social media: an integrative review," sciences 2017,18.
[10]