A Synoptic Survey of Social Network Mental Disorder Identification via Social Media Mining

Meghana M[1], Thippeswamy K[2]

[1]Fourth Sem M.Tech, Department of Studies in CS&E, VTU PG Center, Mysuru
[2]Professor & Chairman, Department of Studies in CS&E, VTU PG Center, Mysuru

---***---

Abstract - *The strongest weapon to conquer the knowledge in today's world - "Internet" , has unfortunately turned out to be one of our greatest obsessions in killing time and is affecting our daily activities and responsibilities with a massive desire to get rid of everything to be able to 'Netflix and relax' all the time. Though the 'Internet Addiction' is gaining attention in the mental health field and had been recently added to the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) as a disorder, it needs a lot of research and standardized diagnosis. Their detection at an early stage is extremely important because the clinical interventions only during the last stage will make things worse and critical. In this paper, we argue that the potential Social Network Mental Disorder (SNMD) users can be automatically identified and classified into various categories like Virtual Relationship Addiction, Obsessive Online Gambling and Information Glut using SNMD based tensor model, with the data sets collected from data logs of various Online Social Networks (OSNs). The proposed model stands out in the list as the users are not involved in revealing their habits to understand and diagnose the symptoms manually. We also exploit multi-source learning in SNMDI (Social Network Mental Disorder Identification) and propose a new SNMD-based Tensor Model (STM) to improve the accuracy. The results show that SNMDI is reliable for identifying online social network users with potential SNMDs.*

***Key Words*: Online social network, Social network mental disorder identification, feature extraction, data logs, tensor decomposition.**

1. INTRODUCTION

Internet Addiction has undoubtedly become the growing epidemic as the number of cases getting registered for the treatment of these mental disorders due to excessive Internet Usage every year is drastically increasing. As per the latest report, this addiction has got so much to do with depression, anxiety disorders, insomnia, isolation, mood swings, procrastination and many more. New terms such as Phubbing (Phone Snubbing) and Nomophobia (No Mobile Phone Phobia) have been created to describe those who cannot stop using mobile social networking apps. Conceptually, it's diagnosis is a compulsive-impulsive spectrum disorder that involves online and/or offline computer usage and consists of at least three subtypes: excessive gaming, sexual preoccupations, and e-mail/text messaging. All of the variants share the following four components: 1) excessive use, often associated with a loss of sense of time or a neglect of basic drives, 2) withdrawal, including feelings of anger, tension, and/or depression when the computer is inaccessible, 3) tolerance, including the need for better computer equipment, more software, or more hours of use, and 4) negative repercussions, including arguments, lying, poor achievement, social isolation, and fatigue.

These symptoms form important diagnostic criteria for SNMDs like Cyber-Relationship Addiction, Information Overload, Net Compulsion, Cyber-Sexual and Computer Addiction. The symptoms of these disorders were till now observed passively and hence the clinical intervention got delayed. Research shows that the early diagnosis of such mental disorders can greatly reduce the risk. Hence the practice of SNMD identification, that relies on self-revealing of those mental factors via questionnaires in Psychology is not adopted in our proposed model as the users might try to over smart the diagnosis by answering questions dishonestly. We propose a new innovative machine learning framework called Social Network Mental Disorder Identification (SNMDI) that detects potential SNMD users by designing and analyzing many important features for identifying SNMDs from OSNs, such as disinhibition, parasociality, self-disclosure, etc.

Furthermore, users may behave differently on different OSNs, resulting in inaccurate SNMD detection. When the data from different OSNs of a user are available, the accuracy of the SNMDI is expected to improve by effectively integrating information from multiple sources for model training. A naive solution that concatenates the features from different networks may suffer from the curse of dimensionality. Accordingly, we propose an SNMD-based Tensor Model (STM) to deal with this multi-source learning problem in SNMDI. Specifically, we formulate the task as a semi-supervised classification problem to detect three types of SNMDs and the new framework can be deployed to provide an early alert for potential patients.

2. RELATED WORK

Internet Addiction is like any other compulsive behaviour such as drug and alcohol addictions, but this is online-related. It's also termed as Internet Compulsivity that dominates

addict's life. There cannot be a single behaviour pattern defining Internet addiction. National surveys revealed that over 70% of Internet addicts also suffered from other addictions, mainly drugs, alcohol, tobacco and sex. Trends also show that the majority of Internet addicts suffer from emotional problems such as depression, mood disorders, social unrest and anxiety disorders, and will use the fantasy world of the Internet to psychologically escape unpleasant feelings or stressful situations.

There are many research studies in Psychology and Psychiatry stating the important factors, co-relations and effects of Internet Addiction Disorder. At first, it began with mental health practitioners' reporting increased caseloads of clients whose primary complaint involved Internet and hence surveyed therapists who treated clients suffering from cyber related problems. It was specifically, discovered that young people with narcissistic tendencies are particularly vulnerable to addiction with OSNs. Further examinations reflected the risk factors related to Internet addiction and the association of sleep quality and suicide attempt of Internet addicts [1],[2]. By using an NLP-based approach, the linguistic and content-based characteristics from online social network are extracted to detect Borderline Personality Disorder and Bipolar Disorder patients [3]. The topical and linguistic features derived from online social networks helped in analysing patients with depression patterns [4]. Further research analysed the emotion and linguistic styles of social media data for Major Depressive Disorder (MDD)[5].

There were cross-sectional studies to examine the associations of suicidal thoughts and attempt with Internet addiction and Internet activities in a large representative adolescent population where students aged 12–18 years were selected using a stratified random sampling and were asked to complete the questionnaires. The questions were used to inquire as to analyze the participants' suicidal thoughts and attempt in the past one month. The kinds of Internet activities that the adolescents participated in were also noted. The associations of suicidal thoughts and attempt with Internet addiction and Internet activities were examined using logistic regression analysis to control for the effects of demographic characteristics, depression, family support and self-esteem. Online gaming, online searching for information, and online studying were associated with an increased risk of suicidal thoughts. While online gaming, chatting, watching movies, shopping, and gambling were associated with an increased risk of suicidal attempt, watching online news was associated with a reduced risk of suicidal attempt.

The results of this study clearly indicated that adolescents with Internet addiction have higher risks of suicidal thoughts and attempt than those without. However, most previous research focuses on individual behaviours and their generated textual contents but does not carefully examine the structure of social networks and potential Psychological features. In addition, the developed schemes

are not designed to handle the sparse data of multiple OSNs. But most of these features cannot be directly observed in OSNs and hence the new system proposes to automatically identify SNMD patients at the initial stage according to their OSN data with a novel tensor model that efficiently integrate heterogeneous data from different OSNs and proposes a new multi-source machine learning approach, i.e., STM, to extract proxy features in Psychology for different diseases that require careful examination of the OSN topologies, such as Virtual Relationship Addiction and Obsessive Online Gambling.

The new framework is based on the support vector machine, which has been widely used to analyse OSNs in many areas. Further, we present a new tensor model that not only adapts the domain knowledge but also estimates the missing data and avoids noise to properly handle data from multiple sources. Many research works focused on approximating the probability of mortality in ICU by modelling the probability of mortality as a latent state that evolves over time and proposed a hierarchical learning method for event detection and forecasting by first extracting the characteristics from different data sources and then learning through a geographical multi-level model [6],[7].

However, the SNMD data from different OSNs may be incomplete due to the heterogeneity. For example, the user profiles may be empty due to the privacy problem, different functions on different OSNs (e.g., game, check-in, event), etc. We propose a new novel tensor-based approach to address the problems of using heterogeneous data and incorporate domain knowledge in SNMD identification. Because many users are inclined to use different OSNs, and it is expected that the data records of these OSNs can provide enriched and complementary information about the behavior of the user. Therefore, our goal is to explore multiple data sources (i.e., OSNs) in SNMDI, to obtain a more complete picture of user behavior and effectively deal with the problem of data scarcity. To exploit the learning of multiple sources in SNMDI, a simple way is to directly concatenate the characteristics of each person derived from different OSNs as a huge vector. Therefore, we explore tensor techniques that have been increasingly used to model multiple data sources because a tensor can naturally represent data from multiple sources. Our goal is to use the decomposition of the tensor to extract common latent factors from different sources and objects.

3. PRELIMINARIES

3.1 Social network mental disorder Identification

In this paper, we aim to analyse data mining techniques to identify three types of SNMDs [8]:

1) Virtual Relationship Addiction, which includes the addiction to social networking, checking and messaging to an extent where virtual and online friends become more

important than real-life relationships with family and friends.

2) Obsessive Online Gambling, which includes compulsive online social gaming or gambling, often leading to financial and job-related problems.

3) Information Glut, addresses how the information technology revolution would shape the world, and how the large amount of data available on the Internet would make it more difficult to sift through and separate fact from fiction. Accordingly, we formulate the detection of SNMD cases as a classification problem. We detect each type of SNMDs with a binary SVM. In this study, we propose a dual-phase framework, called Social Network Mental Disorder Identification (SNMDI). The first phase extracts various discriminative features of users, while the second phase presents a new SNMD-based tensor model to derive latent factors for training and use of classifiers built upon Transductive SVM (TSVM) [9]. Two major challenges in the design of SNMDI are:

i) We are not able to directly extract mental factors like those extracted via questionnaires in Psychology and hence need new features to learn the classification models.
ii) We aim to exploit user data logs from multiple OSNs and thus need new techniques to integrate multi-source data based on SNMD characteristics.

3.1.1. Feature Extraction

We first focus on extracting discerning and factual features for design of SNMDI. This task is nontrivial for the following three causes.

a. Lack of mental features. Psychological studies have shown that many mental factors are related to SNMDs, e.g., low self-esteem [10], loneliness [11]. Thus, questionnaires are designed to reveal those factors for SNMD detection. Some parts of Psychology questionnaire for SNMDs are based on the subjective comparison of mental states in online and offline status, which cannot be observed from OSN logs. As it is difficult to directly observe all the factors from data collected from OSNs, psychiatrists are not able to directly assess the mental states of OSN users under the context of online SNMD detection.

b. Heavy users vs. addictive users. To detect SNMDs, an intuitive idea is to simply extract the usage (time) of a user as a feature for training SNMDI. But, this feature is not sufficient because i) the status of a user may be shown as "online" if she does not log out or close the social network applications on mobile phones, and ii) heavy users and addictive users all stay online for a long period, but heavy users do not show symptoms of anxiety or depression when they are not using social apps. To distinguish them by extracting discriminative features is critical.

c. Multi-source learning with the SNMD characteristics. As we intend to exploit user data from different OSNs in SNMDI, extracting complementary features to draw a full portrait of users while considering the SNMD characteristics into the tensor model is a challenging problem.

To address the these challenges, we consider a number of factors to understand the mental states of users, e.g., self-esteem and loneliness. The goal is to distinguish users with SNMDs from normal users. Two types of features are extracted to capture the social interaction behavior and personal profile of a user. It is worth noting that each individual feature cannot precisely classify all cases, as research shows that exceptions may occur. Therefore, it is necessary to exploit multiple features to effectively remove exceptions.

3.2 Effective features as proxies to capture the mental states of users

A fundamental problem in text data mining is to extract meaningful structure from document streams that arrive continuously over time. Newsfeeds, messages exchanged, posts shared on an individual's wall are all the natural examples of such streams, each characterized by topics that appear, grow in intensity for a period of time, and then fade away. The published literature in a particular research field can also be seen to exhibit similar phenomena over a much longer time scale. Underlying much of the text mining work in this area is the following intuitive premise that the appearance of a topic in a document stream is signaled by a "burst of activity," with certain features rising sharply in frequency as the topic emerges. The human appetitive system is in charge of the addictive behavior. A recent study has shown that social searching (actively reading news feeds from friends' walls) creates more pleasure than social browsing (passively reading personal news feeds) [12].

This finding indicates that goal-directed activities of social searching are more likely to activate the appetitive system of a person as drug rewards do, and it is more related to SNMDs because the appetitive system is responsible for finding things in the environment that promote species survival(i.e., food, sexual mates) and thus is inclined to form addictive behavior after several rewards. While users with SNMDs perform social searching more frequently than non-SNMDs, it is not easy to distinguish these two behavior on social media. This example is just one such kind of a feature that could be used to analyse a user's social interaction and personal features. The new system will have many more similar features that are exploited to understand the mental status and habits of a SNMD user that considers online/offline interaction ratios, the temporal behavior, his self-obsessive characteristics hinting the possibility of SNMD.

4. CONCLUSION

In this paper, we attempt to automatically identify potential online users with SNMD. We propose a new tensor method for deriving potential features from multiple OSNs for SNMD detection and SNMDI framework to search various characteristics from OSN data logs. This research represents the collaboration between computer scientists and mental health researchers to address new problems in SNMD. As a next step, we are planning to study features extracted from multimedia content by NLP and computer vision methods. We are also planning to further investigate new issues from the perspective of social network service providers such as Facebook and Instagram and to improve the well-being of OSN users without compromising user commitment.

ACKNOWLEDGEMENT

REFERENCES

[1] D. Li, X. Li, L. Zhao, Y. Zhou, W. Sun, and Y. Wang. Linking multiple risk exposure profiles with adolescent Internet addiction: insights from the person-centered approach. Computers in Human Behavior, 2017.

[2] K. Kim, H. Lee, J. P. Hong, M. J. Cho, M. Fava, D. Mischoulon, D. J. Kim, and H. J. Jeon. Poor sleep quality and suicide attempt among adults with internet addiction: a nationwide community sample of Korea. PLOS ONE, 2017.

[3] C.-H Chang, E. Saravia, and Y.-S. Chen. Subconscious crowdsourcing: a feasible data collection mechanism for mental disorder detection on social media, ASONAM, 2016.

[4] B. Saha, T. Nguyen, D. Phung, and S. Venkatesh. A framework for classifying online mental health-related communities with an interest in depression. IEEE Journal of Biomedical and Health Informatics, 2016.

[5] M. Choudhury, M. Gamon, S. Counts, and E. Horvitz. Predicting depression via social media. ICWSM, 2013.

[6] K. Caballero and R. Akella. Dynamically modeling patients health state from electronic medical records: a time series approach. KDD, 2016.

[7] L. Zhao and J. Ye and F. Chen and C.-T. Lu and N. Ramakrishnan. Hierarchical Incomplete multi-source feature learning for Spatiotemporal Event Forecasting. KDD, 2016.

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

[9] R. Collobert, F. Sinz, J. Weston, and L. Bottou. Large scale Transductive svms. JMLR, 2006.

[10] K. Young. Internet addiction: the emergence of a new clinical disorder, Cyberpsychol. Behav., 1998.

[11] L. Leung. Net-generation attributes and seductive properties of the internet as predictors of online activities and internet addiction. Cyberpsychol. Behav. Soc. Netw., 2004.

[12] K. Wise, S. Alhabash, and H. Park. Emotional responses during social information seeking on Facebook. Cyberpsychol. Behav. Soc. Netw., 2010.

[13] Hong-Han Shuai, Chih-Ya Shen, De-Nian Yang, Senior Member, IEEE, Yi-Feng Lan, Wang-Chien Lee, Philip S. Yu, Fellow, IEEE and Ming-Syan Chen, Fellow, IEEE, "A Comprehensive Study on Social Network Mental Disorders Detection via Online Social Media Mining", 2018.

BIOGRAPHIES

Meghana M is currently pursuing her M.Tech degree in department of studies in CSE at Visvesvaraya Technological University, PG Center, Mysuru and has completed B.E in CSE branch at Vidyavardhaka College of Engineering, Mysuru, Karnataka in the year 2016. She has an industry experience of more than a year in Mindtree, Bengaluru as Software Engineer in Quality Assurance Domain. Her M.Tech project area is Data Mining. This paper is a survey paper of her M.Tech project.

Dr. Thippeswamy K, currently working as Professor and HOD , Dept. of Computer Science & Engg. VTU PG Centre, Mysuru received his Ph.D. degree from the Department of CS&E, Jawaharlal Nehru Technological University, Ananthapur, Andra Pradesh in the year 2012, M.E degree

in CS&E from University Visvesvaraya College of Engineering (UVCE), Bangalore in 2004 and B.E. in CS&E from University B.D.T College of Engineering (UBDTCE), Davangere in the year 1998 .His major areas of research are Data Mining & Knowledge Discovery, Big data, Information Retrieval, and Cloud Computing. He is a Life member of India Society for Technical Education (LMISTE), Computer Society of India (CSI) & International Association of Engineers (IAENG).