MINING ONLINE SOCIAL DATA FOR DETECTING SOCIAL NETWORK MENTAL DISORDERS

NUKALA RAJA 1, P. VIJAYARAGHAVULU 2*

Research Scholar, 2. Asst.Professor
Dept of CSE,
Sri Annamacharya Institute of Technology and Science, Rajampet, Kadapa.

ABSTRACT An increasing number of social network mental disorders (SNMDs), such as Cyber-Relationship Addiction, Information Overload, and Net Compulsion, have been recently noted. Symptoms of these mental disorders are usually observed passively today, resulting in delayed clinical intervention. In this paper, we argue that mining online social behavior provides an opportunity to actively identify SNMDs at an early stage. It is challenging to detect SNMDs because the mental factors considered in standard diagnostic criteria (questionnaire) cannot be observed from online social activity logs. Our approach, new and innovative to the practice of SNMD detection, does not rely on self-revealing of those mental factors via questionnaires. Instead, we propose a machine learning framework, namely, Social Network Mental Disorder Detection (SNMDD), that exploits features extracted from social network data to accurately identify potential cases of SNMDs. We also exploit multi-source learning in SNMDD and propose a new SNMDbased Tensor Model (STM) to improve the performance. Our framework is evaluated via a user study with 3126 online social network users. We conduct a feature analysis, and also apply SNMDD on large-scale datasets and analyze the characteristics of the three SNMD types. The results show that SNMDD is promising for identifying online social network users with potential SNMDs.

Keywords Online social network, mental disorder detection, feature extraction, tensor factorization

INTRODUCTION

"As we expect more from technology, do we expect less from each other?" asked Sherry Turkle, the Abby Rockefeller Mauz´ professor of the Social Studies of Science and Technology in MIT.1 With the explosive growth in popularity of social networking and messaging apps, online social networks (OSNs) have become a part of many people's daily lives. While OSNs seemingly expand their users' capability in increasing social contacts, they may actually decrease the face-to-face interpersonal interactions in the real world. Studies show that some people's behavior is bolder on the OSNs because they can put a mask when communicating with others there, i.e., hide who they really are. However, do those OSN users still know how to connect with others when the masks are off? Lying between receiving positive attention from OSNs and face-to-face interactions may be a great gulf in the real life. Most research on social network mining focuses on discovering the treasure of knowledge behind the data for improving people's life. In contrast, much less attention has been drawn to remedy the problems incurred from various social network applications. Indeed, some social network mental disorders (SNMDs), such as Information Overload and Net Compulsion [1], have been recently noted.2 For example, studies suggest that 1 in 8 Americans suffers from problematic Internet use.3 Moreover, as reported by the BBC News, 25% of the population in Korea are estimated to suffer from SNMDs.4 Due to the epidemic scale of these phenomena, 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. Moreover, leading journals in mental health, such as American Journal of Psychiatry [2], have reported that the SNMDs may incur excessive use, depression, social withdrawal, and a range of negative repercussions.

The contributions of this paper are summarized as follows.

• Today online SNMDs are usually treated at a late stage. To address this issue, we propose an approach, new to the current practice of SNMD detection, by mining data logs of OSN users to actively identify potential SNMD cases early.

• We develop a machine learning framework for detecting SNMDs, namely Social Network Mental Disorder Detection (SNMDD). Moreover, we design and analyze many features from OSNs, such as disinhibition, parasociality, self-disclosure, etc., which serve as important factors or proxies for identifying SNMDs. The proposed framework can be deployed as a software program to provide an early alert for potential patients and their advisors.

• We study the multi-source learning problem for SNMD detection. By leveraging tensor algebra and considering the SNMD characteristics into the tensor model, we propose STM to better extract the latent factors from different sources, thus improving the accuracy. • We conduct a user study with 3126 users to evaluate the effectiveness of the proposed SNMDD framework. To the best of our knowledge, this is the first dataset crawled online for SNMD detection. Also, we apply SNMDD on large-scale real datasets and perform a social network analysis on the detected cases. The result reveals interesting insights on the network structures in SNMD types, which can be of interest to social scientists and psychologists.

SOCIAL NETWORK MENTAL DISORDER DETECTION

In this paper, we aim to explore data mining techniques to detect three types of SNMDs [1]: 1) Cyber-Relationship (CR) Addiction, which includes addiction to social networking, checking and messaging to the point where social relationships to virtual and online friends become more important than real-life ones with friends and families; 2) Net Compulsion (NC), which includes compulsive online social gaming or gambling, often resulting in financial and job-related problems; and 3) Information Overload (IO), which includes addictive surfing of user status and news feeds, leading to lower work productivity and fewer social interactions with families and friends offline. 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 two-phase framework, called Social Network Mental Disorder Detection (SNMDD), as shown in Figure 1. 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 Feature Extraction We first focus on extracting discriminative and informative features for design of SNMDD. This task is nontrivial for the following three reasons. 1. Lack of mental features. Psychological studies have shown that many mental factors are related to SNMDs, e.g., low self-esteem [3], loneliness [15]. Thus, questionnaires are designed to reveal those factors for SNMD detection. Some parts of Psychology questionnaire for SNMDs are based on subjective comparison of mental states in online and offline status, which cannot be observed from OSN logs. For example: Q1. How often do you feel depressed, moody, or nervous when you are off-line, which goes away once you are back online? Q2. How often do you prefer the excitement of the Internet to intimacy with your partner? Consider Q1. The feel of depression and nervousness offline can not be observed online. To tackle this problem, we have to leverage the knowledge from Psychology, such as withdrawal or relapse patterns, and exploit some proxy features extracted from online social activity logs to approximate them. For Q2, the preference of excitement of the Internet to intimacy with users' partners is important questions for SNMD detection. As it is difficult to directly observe these 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. 2. 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 SNMDD. However, 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. How to distinguish them by extracting discriminative features is critical training and use of classifiers built upon Transductive SVM

Social Interaction Features We first extract a number of social interaction features to capture the user behavior on social media. Parasocial relationship (PR). Research shows that the mental factor of loneliness is one of the primary reasons why the users with SNMDs excessively access online social media [5]. As the loneliness of an OSN user is hard to measure, we exploit the parasocial relationship, an asymmetric interpersonal relationship between two people where one party cares more about the other but the other does not, to capture loneliness (as studies show that they are correlated [16]). The feature of parasocial relationship is represented as |aout|/|ain|, where |aout| and |ain| denote the number of actions a user takes to friends and the number of actions friends take to the user, respectively.8 As the ratio increases, the extent of parasocial relationship also grows.

Online and offline interaction ratio (ONOFF). As observed by mental health professionals, people who indulge themselves in OSNs tend to snub their friends in real life [7]. Therefore, the number of online interactions is inclined to significantly exceed their interactions offline. We extract the number of check-in logs with friends and the number of "going" events as an indicator of the number of offline activities to estimate the online (|aon|)/offline (|aoff |) interaction ratio. Although the number of offline events observed from online is smaller than the actual number, the ratio is relative and is a good indicator (as pointed out in [17]), because the frequent check-in records of a user imply that the user is active in offline activities, which is an indicator of non-SNMD. Social capital (SC). Two types of friendship ties are usually involved in the theory of social capital [18]: i) Bond strengthening (strong-tie), which represents the use of OSNs to strengthen the relationships; and ii) Information seeking (weak-tie), which corresponds to the use of social media to find valuable information. The first type usually creates more interactions in order to increase the social tightness and is related to Cyber-Relationship (CR) Addiction, while the second type concentrates more on finding and reading the information and is thus related to Information Overload (IO) [19]. Therefore, the ratio between the number of strong ties (nstrong) and weak ties (nweak) could be used for differentiating the CR and IO types. Moreover, since the number of strong ties is much smaller than that of weak ties [20], and the number of friends that CR users frequently interact with is less than that of IO users, we exploit the ratio between the number of friends the user interacts online (likes, comments, and posts) and the total number of the user's friends as proxy features to differentiate the CR and IO types. Social searching vs. browsing (SSB). 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 7The third challenge is addressed in Section 3.2. 8The actions include like, comment, and post in our work. social browsing (passively reading personal news feeds) [21]. 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 SMNDs perform social searching more frequently than non-SNMDs, it is not easy to distinguish these two behavior on social media. Let ni denote the total number of the i-th action for posts among friends. For example, if a user is the second one among her friends who click "likes" on a post, the n2 increases 1 for the user. As most social media provide friends' comments and "likes" in the form of news feeds to users, we consider the number of likes/comments on news feeds from friends as social browsing (∞ i=2 ni). On the other hand, if users take an initiative to search for someone's profile and like/comment on it, we consider this as a social searching (i.e., the number of likes/comments on others' news feeds that are not liked/commented by his friend before (n1)). Therefore, we use n1 ∞i=2 ni as a feature. The social searching features are related to CR because CR users tend to find social supports, whereas social browsing is more related to IO. Compared with social capital, SSB focuses on different behavior in reading news feeds, rather than the different types of friend ties.

Analysis of SNMD Types in Large Datasets In this analysis, we first apply the proposed SNMDD framework (with TSVM) on some large-scale OSN datasets, i.e., FB L and IG L, to classify their users. In Figs. 3(a) and 3(b), we analyze the detected SNMD cases among the friends of an SNMD user. In Fig. 3(a), the leftmost bar indicates that in FB L, among all CR users, about 45% of their friends are also CR users, which is greater than the percentage of other SNMD types. On the other hand, the 8th bar from the left in Fig. 3(a) indicates that in FB L, about 59% of NC users' friends are NA (non-SNMD users). Figs. 3(a) and 3(b) show that, in FB L and IG L, CR and IO users have similar friend types. This is because CR and IO cases, by their nature, are similar, i.e., they are both seeking social satisfaction (e.g., relationships and information) from the OSNs. Moreover, among different SNMD cases, CR and IO users are likely to be friends with other CR and IO users. For CR users, this phenomenon has been described as "loneliness propagates" [16]. Furthermore, Infomap community detection [41] is performed on FB L and IG L to derive the relationships between different types of SNMD users in their communities. Figs. 3(c) and 3(d) analyze the community structures of SNMD users with different SNMD scores, where each point represents the characteristic of a community. Specifically, each community in the dataset is represented by three different types of points, i.e., CR, NC, and IO. For example, each CR point is represented as score, ratio, where score is the average CR score in that community, and ratio indicates the proportion of CR users in the community. It is similar for each IO/NC point. As Figs. 3(c) and 3(d) show, for each SNMD type, when the average SNMD score is higher, it is likely to have more SNMD users in the community. Moreover, there are many communities with large IO scores in IG L that have IO ratios close to 1. This implies that the users with large IO scores in IG L are more inclined to form homogeneous groups. At the first glance, one may feel that NC users frequently appear in many communities, and there seems to be a large number of NC users, especially in FB L (i.e., Fig. 3(c)). However, after carefully examining these communities, we find that those communities (with large ratios of NC users) are usually very small (usually with the size around 5) because NC users are less-active. On the other hand, in IG L, when SNMD scores are larger, the ratios of IO users in communities are also larger. This is because IO users can view, like, or follow others in Instagram more easily (not necessary to be friends first). Fig. 3(e) compares the ratios of different types of

SNMD users identified in FB L and IG L. There are more CR users in IG L probably because CR users seek social supports online to compensate the loneliness in real life. We argue that the Instagram platform makes it easy to freely create social relationships with strangers. In contrast, it is not that easy to create new social relationships on Facebook since the friend requests need to be approved. Finally, Fig. 3(f) compares the average number of hops from each SNMD user to the nearest user with the same type of SNMDs. The leftmost bar shows that the average hop distance from each CR user to the closest CR user is 1.07 hop, indicating that CR and IO users are close to other same-type users, i.e., average hop distances are within 1.15, where Figs. 3(a) and 3(b) also report similar results.

CONCLUSION In this paper, we make an attempt to automatically identify potential online users with SNMDs. We propose an SNMDD framework that explores various features from data logs of an OSN and a new tensor technique for deriving latent features from multiple OSNs for SNMD detection. This work represents a collaborative effort between computer scientists and mental healthcare researchers to address emerging issues in SNMDs. As for the next step, we plan to study the features extracted from multimedia contents by techniques on NLP and computer vision. We also plan to further explore new issues from the perspective of a social network service provider, e.g., Facebook or Instagram, to improve the well-beings of OSN users without compromising user engagement.

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] J. Block. Issues of DSM-V: internet addiction. American Journal of Psychiatry, 2008.

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

[4] I.-H. Lin, C.-H. Ko, Y.-P. Chang, T.-L. Liu, P.-W. Wang, H.-C. Lin, M.-F. Huang,

Y.-C. Yeh, W.-J. Chou, and C.-F. Yen. The association between suicidality and Internet addiction and activities in Taiwanese adolescents. Compr. Psychiat., 2014.

[5] Y. Baek, Y. Bae, and H. Jang. Social and parasocial relationships on social network sites and their differential relationships with users' psychological well-being. Cyberpsychol. Behav. Soc. Netw., 2013.

[6] D. La Barbera, F. La Paglia, and R. Valsavoia. Social network and addiction. Cyberpsychol. Behav., 2009.

[7] K. Chak and L. Leung. Shyness and locus of control as predictors of internet addiction and internet use. Cyberpsychol. Behav., 2004.

[8] O. Turel and A. Serenko. Is mobile email addiction overlooked? CACM, 2010.

[9] E. Baumer, P. Adams, V. Khovanskaya, T. Liao, M. Smith, V. Sosik, and K. Williams. Limiting, leaving, and (re)lapsing: an exploration of Facebook non-use practices and experiences. CHI, 2013.

[10] K. Dinakar, B. Jones, C. Havasi, H. Lieberman, and R. Picard. Common sense reasoning for detection, prevention, and mitigation of cyberbullying. ACM TiiS, 2012