2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology

Predicting Depression of Social Media User on Different Observation Windows

Quan Hu
Institute of Psychology, Chinese Academy of Sciences (CAS)
School of Computer and Information Engineering, Henan University

Ang Li
Department of Psychology, Beijing Forestry University
Black Dog Institute, University of New South Wales

Fei Heng, Jianpeng Li
School of Information Science and Engineering
Lanzhou University

Tingshao Zhu*
Institute of Psychology, Institute of Computing Technology, CAS
Email: tszhu@psych.ac.cn

***Abstract*—Depression has become a public health concern around the world. Traditional methods for detecting depression rely on self-report techniques, which suffer from inefficient data collection and processing. This paper built both classification and regression models based on linguistic and behavioral features acquired from 10,102 social media users, and compared classification and prediction accuracy respectively among models built on different observation windows. Results showed that users' depression can be predicted via social media. The best result appears when we make prediction in advance for half a month with a 2-month length of observation time.**

***Index Terms*—Machine learning, Depression, Microblogging behavior, Classification, Prediction**

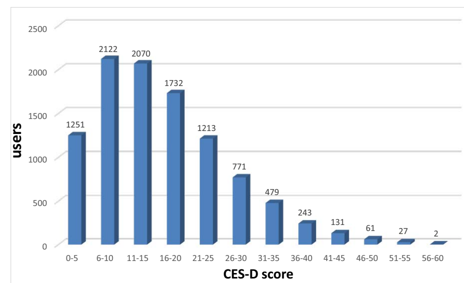

Fig. 1. Distribution of depression scores on CES-D questionnaire

I. Introduction

Mental health is an essential component of human health [1]. There are different types of mental disorders [2]. Among them, depression is recognized as one of the most burdensome diseases in the world [3] [4], and the leading cause of disability and suicidal behaviors in developed countries [3], particularly in China [5].

Efficient depression treatments work on the basis of early detection of depression. Currently, self-report methods (e.g. questionnaire survey, structured interview and clinical judgement) have been widely used to assess depression. However these methods suffer from inefficient data collection and processing, which leads to delayed intervention.

The rise of social media (e.g. Microblogging services) provides an opportunity to detect depression efficiently. On one hand, millions of users are motivated to express themselves online; on the other hand, social media data can be collected and processed in real time, suggesting the possibility of early detection of depression via social media.

Many studies have been done to explore the relationship between psychological characteristics and social media behaviors [7]. Recently, a few studies have begun to examine how to identify individual mental health status through social media [8]. However, there exist contradictions among previous studies [9], which inspires us to do further analysis in a much more comprehensive way.

Sina Weibo is one of the most popular Microblogging service providers in China. Since Sina Weibo data is rich and informative [6] [11] [14], in this paper, we propose to predict user's depression via Sina Weibo data. Specifically, we aim to build classification models for differentiating participants with high and low scores on self-report depression measurement (CES-D), and train regression models for predicting the CES-D score of any individual user.

Three steps were followed in this study: (1) Data collection, (2) Features extraction and selection, and (3) Models training and evaluation.

II. Data collection

We randomly selected a seed user A and crawled A's social network by breadth-first search. This study only focused on active users whose Weibo data might be rich enough for further analysis. We define active users in this paper as follows:

(1) published more than 532 Microblogs since registration.
(2) average number of daily posts ranged from 2.84 to 40.

After identifying active users and receiving their consent, we instructed them to complete an online depression questionnaire CES-D composed of twenty items [12], between August 1 and August 15 2013. After removing participants who provided invalid answers, we got 10,102 completed online questionnaires. The distribution of scores on CES-D is presented in Figure 1, which is consistent with results of other research in China [16].

Because levels of depression change over time [15], it is important to select an appropriate observation window for predicting depression accurately.

978-1-4673-9618-9/15 $31.00 © 2015 IEEE
DOI 10.1109/WI-IAT.2015.166

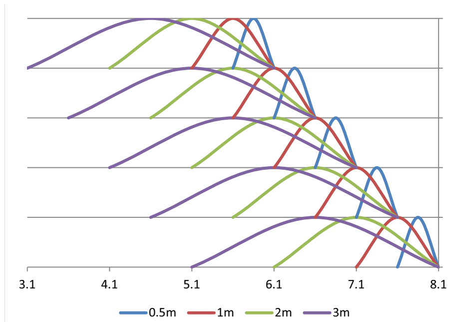

Fig. 2. Different lengths of observation time on each time node

TABLE I
DIFFERENT OBSERVATION WINDOWS

	3m	2m	1m	0.5m
O1	aug01-may01	aug01-jun01	aug01-jul01	aug01-jul15
O2	jul15-apr15	jul15-may15	jul15-jun15	jul15-jul01
O3	jul01-apr01	jul01-may01	jul01-jun01	jul01-jun15
O4	jun15-mar15	jun15-apr15	jun15-may15	jun15-jun01
O5	jun01-mar01	jun01-apr01	jun01-may01	jun01-may15

In this study, participants were instructed to complete CES-D questionnaire between August 1 and August 15 2013. We defined several time nodes, that is node O1 (August 1), O2 (July 15), O3 (July 1), O4 (June 15), and O5 (June 1). Besides, we download users' Sina Weibo data with four different lengths of observation time (0.5m, 1m, 2m and 3m) on each time point.

Finally, we got 20 different observation windows (5 different nodes * 4 different lengths of observation time), and they are depicted in Table I and Figure 2.

III. Features Extraction and Selection

After we got Weibo data with 20 different observation windows, we extracted both of linguistic and behavioral features as predictive variables. The total number of features is 927 (88+39+800).

A. Linguistic features.

In this study, we used a Chinese text analysis software "Wen Xin"(http://ccpl.psych.ac.cn/textmind/) to process the text of Weibo posts and extract 88 designed linguistic features.

Specifically, we got 46 general features, such as the total number of words (*WC*), the average number of words used in each Microblog (*WPS*), the number of figures (*Numerals*) and different types of punctuations (*AllPct*), the number of pronouns (*we, you*). Besides, we got another 42 features reflecting psychological characteristics, such as the frequency of words reflecting positive emotion (*posemo*) (e.g. benediction), negative emotion (*negemo*) (e.g. worry), death (*death*) (e.g. die), work (*work*) (e.g. salary), health (*health*) (e.g. doctor), anger (*anger*) (e.g. detestable), sadness (*sad*) (e.g. dismay), anxiety (*anx*) (e.g. uneasy) and so on.

B. Behavioral features.

We got 39 static features, which keep stable over time:

(1) Profiles (4 features). It included features implying personal information, such as one's gender (*gender*), and location (*city_id*).

(2) Self-expression behaviors (18 features). It included features implying impression management online, such as the length of screen name (*screen_name_length*), and length of self-statement on personal page (*description*).

(3) Privacy settings (3 features). It included features implying personal privacy protection, such as filtering out messages delivered by strangers (*allow_all_act_msg*).

(4) Interpersonal behaviors (14 features). It included features implying social interaction online, such as the number of followers (*followers_count*).

We also got dynamic features, which change obviously over time. We defined dynamic features by following four steps.

(1) We defined and grouped initial 40 dynamic features into 4 categories, such as the Microblogs updates (12 features), @ mentions (3 features), use of apps (6 features) and recordable browsing behaviors (19 features).

(2) Defining "hour" and "day" as two dimensions to describe temporal properties of each initial dynamic feature. Therefore, for each observation window, we can get a two-dimensional matrix.

(3) Based on this two-dimensional matrix, we can export time series data. In this paper, we define 5 behavior series for each initial dynamic feature. Taking Microblogs updates as an example, we can export 5 behavior series, such as 1) the time when a user update the first post in each day; 2) the time when a user update most frequently in each day; 3) the total number of posts updated in each day; 4) the total number of posts updated in each hour across all days; 5) the total number of posts updated in all days. We totally got 200 behavior series (40*5).

(4) We further extracted dynamic features from behavior series. For each behavior series, we extracted four kinds of features: mean, variance, sum and weighted sum. We got a total of 800 (40*5*4) dynamic features.

C. Features Selection

After getting 10,102 scores on CES-D and extracting 927 features, we totally got 20 matrices (10,102*927) with different observation windows. Because some users may not be active within a specific period, for reducing the influence of noisy data, we removed users whose total number of words (WC) were less than 50 in the corresponding observation window for each matrix.

To select sensitive features for indicating depression, we did features selection. For each matrix, we calculated the correlation between features and scores on CES-D. Those features which were associated with scores on CES-D significantly would be picked out for training models. Then, we used Greedy Stepwise (GS) algorithm for features selection.

TABLE II
THE MEAN VALUE OF EACH MATRIX

	O1	O2	O3	O4	O5
3m	16.11	16.09	16.09	16.06	16.05
2m	16.11	16.09	16.09	16.06	16.06
1m	16.08	16.10	16.06	16.04	16.04
0.5m	16.09	16.05	16.09	16.02	15.98

TABLE III
THE STANDARD DEVIATION OF EACH MATRIX

	O1	O2	O3	O4	O5
3m	9.87	9.86	9.85	9.83	9.83
2m	9.85	9.85	9.86	9.83	9.84
1m	9.85	9.82	9.82	9.82	9.85
0.5m	9.85	9.84	9.81	9.84	9.81

TABLE IV
THE PERFORMANCE OF REGRESSION MODELS ON DIFFERENT OBSERVATION WINDOWS

	O1	O2	O3	O4	O5
3m	0.3128	0.3219	0.3142	0.2837	0.3041
2m	0.3553	0.3894	0.3479	0.3154	0.3209
1m	0.3317	0.3512	0.3618	0.3453	0.3145
0.5m	0.2538	0.3023	0.3072	0.2792	0.2963

TABLE V
THE PERFORMANCE OF CLASSIFICATION MODELS ON DIFFERENT OBSERVATION WINDOWS

	O1	O2	O3	O4	O5
3m	74.68%	73.93%	70.46%	71.49%	68.97%
2m	78.71%	81.84%	75.46%	77.72%	74.37%
1m	72.41%	79.66%	76.21%	78.33%	73.49%
0.5m	63.30%	78.07%	73.41%	66.76%	74.73%

D. Feature Analysis

For features selected by correlations, in each matrix, we sorted their absolute values of correlation coefficients in descending order, and selected the top ten. By this way, we can get 200 (20*10) features for 20 matrices (duplicated features existed). We then calculated the frequency of appearance of each feature.

For features selected by GS, we run the same process. Those features which appeared repeatedly in selected matrix can be recognised as features with higher weights.

IV. MODELS TRAINING AND EVALUATION

We built classification models using Logistic regression method [13] (see Table V). For training classification models, we calculated the mean value and standard deviation of depression scores in each matrix (Table II and Table III), and averaged values of twenty matrices. Then, we got one mean value (16.068) and one standard deviation (9.839). Participants can be divided into two groups (high-scoring and low-scoring) by (mean value $\pm$ standard deviation). The threshold of high-scoring group is 26 (16.07+9.84), and the threshold of low-scoring group is 6 (16.07-9.84). Besides, we built regression models using linear regression method (see Table IV). We run 10-fold cross-validation while training and testing classification and regression models.

After getting 20 classification models and 20 regression models, we compared the performance of models across different observation windows. In this paper, we used Precision (P) to evaluate the performance of classification models, and used Pearson's Correlation Coefficient (CC) to evaluate the performance of regression models.

V. DISCUSSION

For the regression models, the correlation coefficient between predicted scores and questionnaire scores ranged from 0.25 to 0.39. For the classification models, the accuracy ranged from 63% to 82%. It suggests that both of behavioral and linguistic features can be used to identify users' levels of depression accurately [10].

Besides, we run further analysis from two aspects: longitudinal and transverse.

For the longitudinal analysis, we compared the CC (Figure 3) and P (Figure 4) across different observation windows on the same time point. With an increasing observation window, both CC and P showed an inverted U-shaped function. The optimal observation window was either one-month or two-month.

For the transverse analysis, we compared the CC (Figure 5) and P (Figure 6) across different time points on the same observation window. From O1 to O5, both CC and P showed an inverted U-shaped function. The optimal time point was either O2 (half a month before the administration of online questionnaire) or O3 (one month before the administration of online questionnaire).

In overall, the optimal observation window is that if we want to predict individual current depression, we should use two-month observation window on the time point half a month in advance. In other words, if we collect Weibo data with a two-month length of observation time based on the time node half

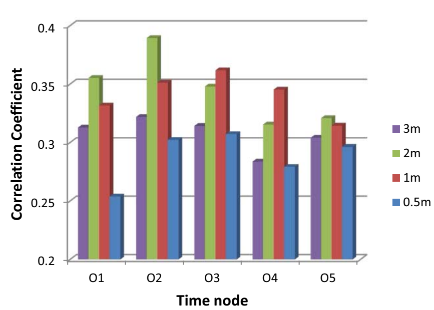

Fig. 3. Correlation coefficients on different observation windows

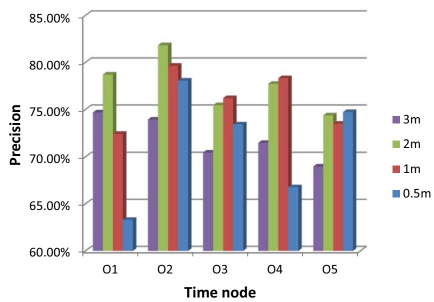

Fig. 4. Classification precision on different observation windows

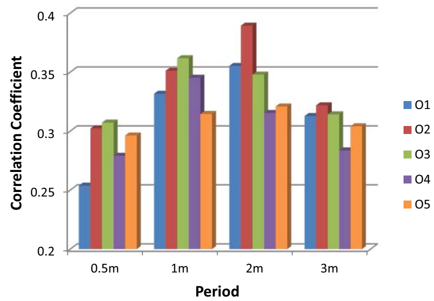

Fig. 5. Correlation coefficients in different periods

a month in advance, we can predict the current depression efficiently. It suggests that using social media behaviors to predict depression may have an effect of hysteresis.

VI. Conclusion

This paper aims to establish an innovative method for detecting depression efficiently in a large population. That is to build computational models based on linguistic and behavioral features acquired from social media data. Besides, we also compared the performance of models on different observation windows. The results indicate that:

(1) It is feasible to predict individual user's depression via social media data.

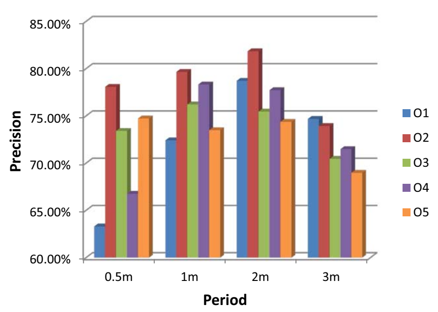

Fig. 6. Classification precision in different periods

(2) There is an effect of hysteresis for predicting depression through social media data.

In future, we plan to improve the current job based on a longer observation period and diverse participants.

VII. Acknowledgments

The authors gratefully acknowledges the generous support from National High-tech R&D Program of China (2013AA01A606), National Basic Research Program of China (2014CB744600), Key Research Program of Chinese Academy of Sciences (CAS) (KJZD-EWL04), and CAS Strategic Priority Research Program (XDA06030800).

References

[1] H. Herrman, S. Saxena, and R. Moodie, "Promoting mental health: concepts, emerging evidence, practice: a report of the World Health Organization, Department of Mental Health and Substance Abuse in collaboration with the Victorian Health Promotion Foundation and the University of Melbourne," World Health Organization, 2005, pp. 288.

[2] B. A. Kitchener, A. F. Jorm, and D. C. Kelly, "Mental Health First Aid Manual," ORYGEN Research Centre, University of Melbourne, 2002.

[3] C. D. Mathers, and D. Loncar, "Projections of global mortality and burden of disease from 2002 to 2030," PLoS medicine, 3(11), e442.

[4] Centers for Disease Control and prevention (CDC), "Behavioral Risk Factor Surveillance System Survey Data," Atlanta, Georgia: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2008, 2011, 2012.

[5] People's Daily, "The ghost of depression beside us," August 17, 2012.

[6] Sina first quarter financial report, 2013.

[7] Y. A. Hamburger, and E. Ben-Artzi, "The relationship between extraversion and neuroticism and the different uses of the Internet," Computers in Human Behavior, 16(4): 441-449, 2000.

[8] K. J. Mitchell, C. Sabina, D. Finkelhor, and M. Wells, "Index of problematic online experiences: item characteristics and correlation with negative symptomatology," CyberPsychology & Behavior, 12(6): 707-711, 2009.

[9] W. Peng, and M. Liu, "Online gaming dependency: a preliminary study in China," CyberPsychology, Behavior, and Social Networking, 13(3): 329-333, 2010.

[10] F. Zhang, T. Zhu, A. Li, Y. Li, and X. Xu, "A survey of web behavior and mental health," In Pervasive Computing and Applications (ICPCA), 2011 6th International Conference on (pp. 189-195), IEEE.

[11] L. Li, A. Li, B. Hao, Z. Guan, and T. Zhu, "Predicting Active Users' Personality Based on Micro-Blogging Behaviors," PloS one, 9(1), e84997, 2014.

[12] L. S. Radloff, "The CES-D scale a self-report depression scale for research in the general population," Applied psychological measurement, 1(3): 385-401, 1977.

[13] C. M. Bishop, "Pattern recognition and machine learning," New York: springer, p:205, 2006.

[14] W. Guan, H. Gao, and M. Yang, et al, "Analyzing user behavior of the micro-blogging website Sina Weibo during hot social events," Physica A: Statistical Mechanics and its Applications, 395: 340-351, 2014.

[15] P. A. Thoits, "Self, identity, stress, and mental health," Handbook of the sociology of mental health. Springer Netherlands, pp: 357-377, 2013.

[16] C. Zhiyan, Y. Xiaodong, and L. Xinying, "Psychometric Features of CES-D in Chinese Adolescents," Chinese Journal of Clinical Psychology, Vol 17 (4), 443-448, 2009.