Predicting Mental Health Status Based on Web Usage Behavior

Tingshao Zhu, Ang Li, Yue Ning, and Zengda Guan

School of Information Science and Engineering,
Graduate University of Chinese Academy of Sciences,
Beijing 100190, China
tszhu@gucas.ac.cn
http://wsi.gucas.ac.cn

Abstract. To build a predicting model for mental health status based on Web Usage Behavior, we collect data from 571 first-year graduate students using our own Internet Usage Behavior Check-List (IUBCL) and Psychological Health Inventory (PHI). We build six logistic regression models, in which Web usage behavior features are as independent variables while mental health status as dependent ones. We find that the accuracy is about 72.9%−83.1%, which demonstrates it is applicable and feasible to identify each individual's mental health status by analyzing his/her Web usage behaviors.

Keywords: Mental Health, Logistic Regression, Web Usage Behavior.

1 Introduction

The faster the modern society develops, the more mental pressure people may undertake. Accordingly, the mental health problems have become a major concern in our society [20,5,17,21]. As a crucial premise of psychological intervention, identifying the target group with certain mental health problems accurately will improve the effectiveness and pertinence of the works later on [2]. At present, Internet has changed the pattern of communication among our human beings [18]. Therefore, the diagnosis procedure based on traditional way of mutual communication should be adjusted as well.

Specifically, because the techniques of internet could support communication beyond time and space boundary, individuals' physical existence is no longer the basis of all social activities. In other words, certain communication mediated by Internet is featured with the absence of physical body [26,6]. Such form of computer mediated communication (CMC) will filter much behavior information which could be observed or conveyed in a face-to-face way, such as expressions, tones, pauses and body languages. In that way, behaviors under the Internet environment have been simplified into digital information in the screen of computers of which the essence are virtual behaviors expressed by text, images and symbols [10,11].

N. Zhong et al. (Eds.): AMT 2011, LNCS 6890, pp. 186–194, 2011.

Now, many researchers have found that the variable of internet usage behavior is related to some kinds of mental health problems [9,13,23,19]. However, most relevant conclusions only draw some correlational results. They do not conduct sufficient discussion of mutual distinctiveness among certain internet usage behaviors mirrored by different kinds of mental health problems. Thus, such conclusions, which cannot improve the efficacy of internet usage behavior to distinguish mental health problems, have limited practical value.

Our research attempts to use the internet usage behavior data as independent variables and the different kinds of mental health problems as dependent variables. Based on this, we could build a series of regression models and then estimate the distinctiveness among all of them.

2 Methods

The research objects are the first grade graduate students in science and technology major in Graduate University of Chinese Academy of Sciences and the sample size is 571. The average age is 23.67 years (S.D.=1.37). In such research sample, the male population is 419, accounting for 73.4%. the female population is 152, accounting for 26.6%. The Han ethnic population is 542, accounting for 94.9%, and other ethnics' population is 29, accounting for 5.1%. The population from only-child family is 215, accounting for 37.7%, and population from not-only child family is 356, accounting for 62.3%. The descriptive statistical results about the subjects are the following:

Table 1. Descriptive statistics of subjects(n=571)

Variables	Population	Percentage
Hometown location		
City	200	35.0
Town	157	27.5
Rural area	214	37.5
family monthly income		
< 2000	201	35.2
$2001 - 4000$	220	38.5
$4001 - 6000$	98	17.2
$6001 - 8000$	30	5.3
$8001 - 10000$	10	1.8
> 10000	12	2.1

2.1 Measurement

Internet Usage Behavior Check-List(IUBCL): This questionnaire is self-designed and contains 52 items. The content lists common internet usage behavior including “General Internet behavior”, “Information Retrieval preference”, “Social network and Instant message preference”, “Web pages topics preference”, “Web pages affection preference”, “Web pages function preference”.

Psychological Health Inventory(PHI) [24]: This questionnaire contains 7 dimensions of mental health including “Somatic disorder”, “Depression”, “Anxiety”, “Psychopathic Deviate”, “Hypochondria”, “Unrealistic”, “Hypomania”. Besides, we add two validity scales: L(Lie) and F(Fake).

2.2 The Statistical Method

Our research uses SPSS13.0 to conduct the Logistic regression analysis towards the data we have collected.

3 Results

Our study intends to find out if we can exactly locate and distinguish different kinds of mental health using Internet usage behavior. Therefore, we need to make an appropriate assessment of prediction accuracy of our regressive model. Logistic regressive analysis meets the demands of this study, and has more factual values and intuitiveness of understanding in our statistical results comparing with other methods.

3.1 Data Pre-processing

According to the scores of validity scales in PHI, we delete data of 9 subjects and retain other 562 copies as statistical samples. Then we use these samples to computes T score of 7 dimensions of mental health in PHI. Each specific dimension is sorted according to the score in descending order. By the means of “extreme grouping”, top 27% and bottom 27% subjects named “higher” and “lower” are extracted from each dimension. Uniform naming of every dimension score is classified into “higher” or “lower” (code in binary variables of 0/1).

3.2 Logistic Regression Analysis

Our research presumes the binary variables(high performance and low performance from the dimensions of the mental health problems) as the observation variables. Series of internet usage behavior variables and demographic variables are assigned as the prediction variables. Then we implement logistic regression analysis($\alpha = 0.05$). It turns out that distinction between high performance and low performance is not obvious because the score of high and low performance in “Depression” dimension overlaps a lot. Thus we build 6 regression models in the following table:

Because the current SPSS could not provide multiple collinearity diagnostics in logistic regression analysis, alternate method is using same observation and prediction variables to simulate linear regression model and promoting corresponding multiple collinearity diagnostics [28]. In such way, we assume the prediction variables including Tolerance(the tolerance index) and Variance Inflation Factor(VIF) in the following table:

Table 2. Statistic Results Using Regression Models in Mental Health Dimensions

Model	**Dimension**	χ^2	*df*	*p*	**A**	**P**	**CRR**
1	Somatic disorder	65.874	6	< .01	72.9%	64.6%	80.3%
2	Anxiety	101.062	10	< .01	78.4%	73.3%	82.6%
3	Psychopathic Deviate	105.594	12	< .01	79.0%	72.8%	84.1%
4	Hypochondria	94.420	10	< .01	77.9%	64.6%	88.6%
5	Unrealistic	122.106	14	< .01	83.1%	80.0%	85.6%
6	Hypomania	102.629	15	< .01	77.1%	72.4%	81.5%

Note: **A**: Accuracy; **P**: Precision; **CRR**: Correct Rejection Rate

Table 3. Assignments of indices

Model	**Tolerance**	**VIF**
Model 1	0.896 – 0.931	1.074 – 1.116
Model 2	0.810 – 0.964	1.037 – 0.234
Model 3	0.754 – 0.926	1.080 – 1.326
Model 4	0.739 – 0.874	1.144 – 0.352
Model 5	0.416 – 0.971	1.030 – 2.402
Model 6	0.752 – 0.952	1.051 – 1.329

According the corresponding standards, the above 6 logistic regression models do not have serious multiple collinearity problem. Finally, 6 logistic regression models are the following(where PSI is Preference of Social network and Instant message, PCA is Preference of Contents Affections , PFS is Preference of Functional Service, PIR is Preference of Information Retrieval, PCT is Preference of Contents Topics, GIB is general internet behavior, and "*" represents the dummy variable):

(1)logit (P|y=Somatic disorder)= 2.134 + 0.806×PSI (Participation in topic groups)+0.735×PSI (Publishing journals)−1.229×PSI (Visiting friends' personal pages)+0.447×PCA (Fear)−0.789×PFS (E-mail)+0.297×PFS (On-line shopping);

(2)logit(P|y=Anxiety)= 1.812 + 2.375×PIR (The means of information retrieval: Others)+1.160×PSI (Participation in topic groups)−1.213×PSI (Visiting friends' personal page)−2.512×PCT (Time of focusing on academic materials:> 3 hours)*+0.443×PCA (Fear)−0.876×PFS (Search engine)−0.389× PFS (Short Messaging Service)+0.853×PFS (Random surfing);

(3)logit (P|y=Psychopathic Deviate)= 3.769 − 2.030×GIB (Average lingering time on one page: 30 seconds-1 minute)*−2.911×GIB (Average lingering time on one page:> 5 minutes)*+2.202×PIR (Information Retrieval means: Others)+0.936×PSI (Participation in topic groups)−0.416×PCA(Heartwarming)-0.964×PFS(Downloading of softwares)+ 0.567×PFS(On-line shopping)-0.332× PFS(Alumi record)+ 0.690×PFS(Random surfing);

(4)logit (P|y=Hypochondria)= 1.994 + 1.954×Times of browsing unhealthy web sites (2 times)*+1.255×PSI (Participation in topic groups)+0.879×PCA

(Angry and Violence)−0.549×PCA (Pity)−0.596×PFS (Search engine)-0.481× PFS(Alumi record)+ 0.390× PFS(Random surfing);

(5)logit (P|y=Unrealistic)= 4.181 + 1.422×GIB (Average daily time of using social networks:0.5 – 1 hours)*+1.370×GIB (Average daily time of using social networks:1−2 hours)* −1.409×PSI (Visiting friends' personal page)−4.452×GIB (Percentage of initiative internet behavior's time:1%−20%)* −4.667×GIB (Percentage of initiative internet behavior's time:41% – 60%)*−4.627×GIB (Percentage of initiative internet behavior's time:61% – 80%)*−6.761×GIB (Percentage of initiative internet behavior's time:81% – 100%)*+1.230×PCA (Angry and Violence)−0.517×PFS (Browsing news on-line)−0.834×PFS (Search engine)+0.956×PFS (Random surfing);

(6)logit (P|y=Hypomania)= 12.741 – 0.484Age−1.813×PSI(The number of friends in instant message softwares:50 – 100)* −1.919×PSI(The number of friends in instant message softwares:100-200)*+1.944PSI(The average daily number of contacting web friends:5 – 10)*+1.829×PCT(Browsing news on-line:2 – 3 hours)*−0.590×PFS(E-mail)+0.646×PFS(Electronic magazines)-0.392×PFS(Entertainment of multiple media)+0.452×PFS(Random surfing);

4 Discussion

We find out that internet usage behavior did have some relations with the 6 dimensions of mental health through the 6 regression models that we have built. Specifically, we have some conclusions in 6 fields in the following:

(1)Somatic disorder

It usually means disorder or discomfort in physics, and people with somatic disorder always explain their psychological problems as physical problems to obtain others' empathy. In clinical manifestations, it is displayed in the trend of hypochondria. Because some means like participation in topic groups, publishing journals and on-line shopping may provide convenient confiding way of communication and many talking objects, this could also explain why the activities like communications with some special person or transferring limited information in a period of time are not accepted. Besides, some people may have fear or worry about physical diseases, and thus they may prefer browsing horrific web pages to attribute their bad mood experience outside and relieve the discomfort in mind.

(2) Anxiety

It usually represents the nervous, anxious and repeated thinking status with lacking of confidence. In clinical manifestations, it displayed in anxiety. Anxiety in mind makes people hard to focus on one single thing in a long time(for example, focusing on academic materials, visiting friends' pages, using search engines and using short message service). Therefore, their Internet usage behaviors are displayed in random both in content and form(like using uncommon method of information retrieval or random surfing). Besides, strong anxiety may also push them to pay attention to the horrific web pages or participate in topic groups to get relief and consolation in mind.

(3)Psychopathic Deviate

It usually means the skin-deep communications and low tolerance to frustrations displayed with personality deviation in clinical manifestations. In reality, unhealthy relationships may mirror in the internet behaviors(like do not care of friends' pages), while the communications that do not involved in building stable and deep relations will not get influenced. On the other side, low tolerance to frustrations may make them to avoid some sequential behaviors, and such behavior depends on the status of internet connections(like downloading of softwares and lingering on the web pages in lots of time) to be completed successfully. Besides, psychopathic deviate has some features like antisocial and counter-moral. Thus, people with psychopathic deviate usually reject the web pages with heart-warming affection. While "participation in topic groups","random surfing" and "uncommon method of information retrieval" may help them to get information belong to non-mainstream culture.

(4)Hypochondria

It usually means sensitive, argumentative and of independent tendency with hypochondria in clinical manifestations. Because of the sensitivity, people may be more defensive and fluctuate emotionally more easily(for example, pay attention to web pages with angry affection and avoiding web pages with pity affection). In the meantime, people may show doubt about the web sites that do not have verification(like do not prefer searching internet information), or people may turn to be curious about the forbidden behaviors(like browsing unhealthy information). Besides, independent tendency may lead to lacking of social networks' activities(including do not care about friends' pages or random surfing). They are argumentative and thus they may show favor of participation in topic groups.

(5)Unrealistic

It means isolation or shrinking back from the reality with disordered mind, unusual and eccentric experience. In clinical manifestation, it is displayed in deviation from reality. In the internet environment, this feature may turn into passive and negative progress of interaction with internet to avoid getting environmental information(like random surfing, low level active interaction with internet, less average daily time of internet usage, pay attention to news' browsing, not care about usage of search engine). Preference of pages with angry or violent affection may also reflect one's disordered status from aside.

(6)Hypomania

It usually means being enthusiastic about communication, too much vigor, active mind, high self-evaluation,and low control of behavior. Young people's vitality may promote this status. Although being enthusiastic about social communication(like contacting $5-10$ friends on-line daily), some people may have some shortages in the self-evaluation and behavior control, and their social communication effect is not ideal(like less friends in the instant message software). Besides, people with more vigor may pay a lot attention to news and magazine that are instant and timely information and care less about information that is not instant like E-mail. In the meantime, they may avoid those activities that need to think, and consider "random surfing" as a way of killing time.

5 Conclusions

Firstly, based on the 6 logistic regression models and the corresponding accuracies(72.9%-83.1%), we can find that there is a particular prediction variable in each model at least(like “publishing journals” in model 1). Using different combination of regression coefficient in the model, we can locate more precisely in some specifical category of mental health and distinguish it from other problems.

Secondly, according to the category that the prediction variables belong to logistic regression models, we can see that “preference of browsing content with affection” accounts for an important part. The former researches mostly focus on other content like “general internet usage behavior”, “preference of contents' topics” [1,16]. Therefore, in further research, we expect to pay more attention to the study of “preference of content with affection” which is valuable, and methods of content analysis will be necessary in related work.

Thirdly, some other researches and our experiment all use the questionnaire method and some evidences show that internet psychology research based on questionnaire may have deviation [27,25,7]. We conclude a few shortages including: (1) Because some researches may involve topics like socially deviant behavior [22,14,15], subjects may be influenced by the social desirability in the collection of internet usage behavior data. (2) The quality of internet usage behavior data may be restricted by the cognitive ability of subjects. When subjects need to answer the questions by memories(like “time of surfing the internet”, “times of surfing the internet”, and “frequency of surfing the internet”), the faculty of memory may affect the authenticity of the results. (3)Restricted by the patterns of questions in the questionnaires, the data of internet usage behavior that could be researched is limited. Generally speaking, closed questions just reveal the tendency and degree information of internet usage behaviors(like frequency), while some other information like navigation strategy and interface applications cannot be studied further [4]. The discrete options are hardly to make sure the inflection point of behavior to appear among them, and this may cause the decrease of accuracy and discrimination. Therefore, the form of questionnaire cannot record the internet usage behavior features effectively, while the technique of computer science like capture of internet usage behavior data and analysis of data can solve this problem.

In addition, the visiting behavior data could be recorded by the server as the web log [8], thus the web log could provide more reliable internet usage behavior data [3]. Web logs contain abundant information like users' IP address, domain, visiting time, URL, requesting status, the bit data returned or total data count, visited pages and browser's toolbar that used. Besides, the record that aimed at the visitors' actual operations data not only promise the outside effectiveness to get optimization, and it also helps to avoid the deviation that caused by subjective records in the process of data collection [12]. On the other side, the variables of internet usage behavior which get much attention from psychology field is grounded on the web log data to get operational definition, like “general internet usage behavior”, “preference of information retrieval”, “preference of social networks and instant message”, “preference of contents' topics”, "preference of content with affection", "preference of functional service" in our research. Because the web log provides a instant, direct and automatic reflection of internet usage behavior, we can try to do some research to match some psychological features with time varying(like emotions) in future.

Acknowledgments. The authors gratefully acknowledges the generous support from Research-Education Joint Project, Chinese Academy of Sciences, the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry.

References

1. Amichai-Hamburger, Y., Fine, A., Goldstein, A.: The impact of internet interactivity and need for closure on consumer preference. Computers in Human Behavior 20, 103–117 (2004)
2. Bennett, P.: Abnormal & Clinical Psychology. The McGraw-Hill Companies, Inc., New York (2003)
3. Burton, M.C., Walther, J.B.: The value of web log data in use-based design and testing. Journal of Computer Mediated Communication 6(3) (2001)
4. Catledge, L.D., Pitkow, J.E.: Characterizing browsing strategies in the world-wide web. Computer Networks and ISDN Systems 27, 1065–1073 (1995)
5. Demyttenaere, K., Bruffaerts, R., Posada-Villa, J., Gasquet, I., Kovess, V.: Prevalence, severity, and unmet need for treatment of mental disorders in the world health organization world mental health surveys. The Journal of the American Medical Association 291, 2581–2590 (2004)
6. Ding, D.: The interpersonal communication in cyberspace:a theoretical and demonstrative research. PhD thesis, Nanjing Normal University (2003)
7. Egger, O., Rauterberg, M.: Internet behavior and addiction. Technical report, Swiss Federal Institute of Technology (ETH) Zurich (1996)
8. Eirinaki, M., Vazirgiannis, M.: Web mining for web personalization. ACM Transactions on Internet Technology 3(1), 1–27 (2003)
9. Fu, Z.: The relationship between adolescents' internet service and pathological internet use. Master's thesis, Jilin University (2007)
10. Huang, S.: Social activities in cyberspace: A research on youth's Internet behavior. People's Publishing House, Bejing (2008)
11. Kiesler, S., Siegel, J., McGuire, T.W.: Social psychological aspects of computer-mediated communication. American Psychologist 39(10), 1123–1134 (1984)
12. Krishnamurthy, S.: Contemporary research in e-marketing. Idea Group Publishing, USA (2005)
13. Lei, L., Yang, Y., Liu, M.: The relationship between adolescents' neuroticism, internet service preference and internet addiction. Acta Psychologica Sinica 38(3), 375–381 (2006)
14. Leung, L.: Loneliness, self-disclosure, and icq use. CyberPsychology& Behavior 5(3), 241–251 (2002)
15. Li, X.: A study of the relationship between the preference of internet contents and personality characteristics of college students. Psychological Science 27(3), 559–562 (2004)

16. Lu, H.-Y., Palmgreen, P.C., Zimmerman, R.S., Lane, D.R., Alexander, A.L.J.: Personality traits as predictors of intentions to seek online information about stds and hiv/aids among junior and senior college students in taiwan. CyberPsychology& Behavior 9(5), 577–583 (2006)
17. Luo, M.: The Present Status of Mental Health Service Needs in Chinese Youth and Teenage. PhD thesis, Southwest university (2010)
18. Manasian, D.: Digital dilemmas: a survey of the internet society. Economist 25, 1–26 (2003)
19. Mazalin, D., Moore, S.: Internet use, identity development and social anxiety among young adults. Behaviour Change 21(2), 90–102 (2004)
20. Meltzer, H., Gatward, R., Goodman, R., Ford, T.: Mental health of children and adolescents in great britain. International Review of Psychiatry 15(1-2), 185–187 (2003)
21. Qiu, P., Yang, Y., Wu, F., Cao, X., Zhao, S., Ma, X.: The advancement and enlightment of researches about mental health for domestic and overseas floating population. Chinese Mental Health Journal 24(1), 64–68 (2010)
22. Rogers, M.K., Seigfried, K., Tidke, K.: Self-reported computer criminal behavior: A psychological analysis. Digital Investigation 3S, S116–S120 (2006)
23. Scealy, M., Phillips, J.G., Stevenson, R.: Shyness and anxiety as predictors of patterns of internet usage. Cyberpsychology & Behavior 5(6), 507–515 (2002)
24. Song, W., Mo, W.: Weizhen Song and Wenbin Mo. The compilation of psychological health inventory(phi). Psychological Science 2, 36–40 (1992)
25. Teoa, T.S.H., Limb, V.K.G., Laia, R.Y.C.: Intrinsic and extrinsic motivation in internet usage. Omegan. Mgmt. Sci. 27, 25–37 (1999)
26. Xiao, C.: Research methodology in cyber-psychology. Psychological Science 27(3), 726–728 (2004)
27. Young, K.S.: Internet addiction: the emergence of a new clinical disorder. Cyberpsychology and Behavior 3, 237–244 (1998)
28. Zhang, W.: Advanced textbook for SPSS statistical analysis. Higher Education Press, Beijing (2009)