Web Use Behaviors for Identifying Mental Health Status

Ang Li, Fan Zhang, and Tingshao Zhu

Institute of Psychology, University of Chinese Academy of Sciences
Chinese Academy of Sciences, China
tszhu@psych.ac.cn

Abstract. It is very important to identify mental health problems early and efficiently, but traditional method relies on face-to-face communication which suffers from the limitations in practice. This study aimed to propose an innovative method of detecting mental health problems via web use behaviors. 102 graduates were administrated by SCL-90 questionnaire to get their actual mental health status with 10 dimensions, and their web use behaviors were acquired from Internet access log recorded on the gateway. A computational model for predicting scores on each SCL-90 dimension was built based on web use behaviors. Results indicated that the value of Pearson Correlation Coefficient between predicted scores and actual scores on each dimension ranged from 0.49 to 0.65, and the value of Relative Absolute Error (RAE) ranged from 75% to 89%. It suggests that it is efficient and valid to identify mental health status through web use behaviors, which would improve the performance of mental health care services in the future.

Keywords: web use behaviors, mental health status, computational model, graduates.

1 Introduction

Mental health refers to "a state of well-being in which the individual realizes his or her own abilities, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to his or her community" [1]. Because of its prevalence and severity, mental health problem has been regarded as the focus of public interest. It is reported that the prevalence of any mental health disorder was 12.0% - 47.4% [2]; while, in China, the number has reached 17.5% [3]. Nowadays, mental health disorders are the leading disabling illnesses around the world [4]. The growth in mental health disorders has resulted in negative consequences accounting for 37% of all years lived with disability from disease [5]. Besides, mental health disorders also weaken individual subjective well-being [6], social adaptability [7] and even physical health condition [8].

The efficient delivery of mental health care services is based on early identification of individual mental health status. Mental health problem is a complex concept, including different types of mental health disorders (e.g., depression, anxiety and obsessive-compulsive) [4, 9]. To acquire mental health status, we might need to examine as

K. Imamura et al. (Eds.): BHI 2013, LNAI 8211, pp. 348–358, 2013.

many types of mental health disorders as possible, which is time-consuming and inapplicable. In psychology, self-report survey is the widely-used method of measuring psychological features, which collects respondent's ratings on items in questionnaire [10]. For mental health measurement, a few famous questionnaires (e.g., MMPI-2 and SCl-90) could examine different types of mental health disorders simultaneously [11-12]. However, because of its limitation in recruiting large and diverse sample and inefficient in collecting longitudinal and real-time data, self-report survey could not be suitable for monitoring mental health status in a large scale [13-14]. Besides, traditional psychological research has been accused of neglecting actual human behaviors, which reveals the nature of psychological testing [15-16]. Psychologists would like to use direct observation of behaviors for testing in psychological research [17].

We proposed to detect user's online mental health status by means of analyzing his/her web use behaviors. In addition, because mental health disorders have physiological basis [18-19], which could be insensitive to variation of social settings, the equivalence of online and offline mental health status should be recognized. It means that web user's online mental health status could be labeled as results of offline measurement. That is, in order to predict online mental health status through web use behaviors, we could build computational models predicting offline mental health status based on web use behaviors instead.

2 Related Work

For decades, Internet has become increasingly popular, which implies an opportunity for improving mental health measurement. By June 2012, the world Internet population has exceeded 7 billion (http://www.internetworldstats.com/stats.htm), and in China the number has reached 564 million until December 2012 (http://www1.cnnic.cn/IDR/ReportDownloads/201302/P020130312536825920279.pdf). In cyber world, interpersonal contact would be accomplished through computer-mediated communication rather than face-to-face communication. With the help of information technology, online behavioral residue (e.g., user's IP address, domain name, access time and URL) could be recorded as web log automatically and instantaneously [20-21]. It means that it is feasible for psychologists to acquire and trace actual web use behaviors from a large and diverse sample in real time.

In order to understand web user's behaviors on the Internet, psychologists are interested in the relationship between web use behaviors and psychological features. According to Brunswik's "Lens Model", personal environment contains information cue, which could indicate occupant's mental status [22]. By means of "behavioral residue", information cue would form in diverse social settings [23]. It suggests that, if a real-world behavior can manifest one's psychological features (e.g., mental health status), a virtual-world behavior can manifest one's psychological features as well.

Previous studies have demonstrated the relationship between web use behaviors and mental health status. Take depression for example, Morgan and Cotton found that, for undergraduates, increased hours spent on e-mail, instant messaging and chat room were associated with decreased depressive symptoms, while increased hours spent on online shopping, online gaming and scientific research were associated with increased depressive symptoms [24]. Ozcan and Buzlu found that, increased problematic Internet use was associated with increased depressive symptoms [25]. Ceyhan and Ceyhan found that, depressive status was an influencing factor towards problematic Internet use [26]. van den Eijnden et al found that, increased frequency of using instant messaging software and chatting online were associated with increased compulsive Internet use 6 months later [27]. Bessiere et al found that, compared with web users who used the Internet for information, entertainment and escape purposes, those who used the Internet for communicating with family and friends would have lower depression scores 6 months later [28]. Selfhout et al found that, increased Internet use for communication purposes (e.g., chatting online) was associated with decreased depression scores, while increased Internet use for non-communication purposes (e.g., random surfing) was associated with increased scores on both depression and social anxiety [29]. Peng and Liu found that, increased dependency on online gaming was associated with increased depression scores [30]. These studies imply that it is rational to detect one's mental health status based on his/her web use behaviors.

However, the existing studies seldom to discuss how to detect mental health status by using actual web use behaviors. They merely examine the correlations between mental health status and web use behaviors based on self-report technique. Recently, a few studies have begun to predict web user's personal features through actual web use behaviors. Gosling et al found that, web user's personality could be manifested on actual web use behaviors [31]. Kosinski et al found that, digital records of web use behaviors (Facebook Likes) could be used to predict personal features, including sexual orientation, ethnicity, religious and political views, personality traits, intelligence, happiness, use of addictive substances, parental separation, age, and gender [32]. So far, there were few studies which tend to predict mental health status through actual web use behaviors.

3 Method

3.1 Participants and Procedure

A total of 102 Chinese graduates agreed to take part in this study voluntarily (78 men and 24 women and 23.54 ± 0.91 years old on average).

The length of the experimental period was 10 weeks, from March, 2011 to May, 2011. In this study, in order to collect scores on mental health status, participants were required to complete a psychological testing for once two weeks (one participant should have five testing results). Their web use behaviors were obtained from Internet access log on gateway. Each result of psychological testing would be paired with one set of web use behaviors, and each set of web use behaviors would have a two-week length of recording period before corresponding testing time. Thus, this yielded 207 pairs of completed assessments.

3.2 Measures

Web Use Behaviors. Participants' web use behaviors were acquired from two sources. NetentSec Internet Control Gateway (NS-ICG 5.6) and NetentSecData Center (NS-DC 2.1) were both used to record IP-labeled detailed web use behaviors (e.g., time spent online, types of web request and classification of browsed URL). Such two equipments were developed by Beijing NetentSec Ltd. The Internet access log recorded on gateway from Network Information Centre only recorded IP-labeled web use behaviors (e.g., login and logout time) and corresponding ID (student number). Because of dynamic host configuration protocol (DHCP), in order to obtain identity mark on each record of web use behaviors, we combined data from two data sources based on clues of IP and surfing time and then labeled each record with ID.

Mental Health Status. 90-item Symptom Checklist-90 (SCL-90) [12] was administrated to measure participants' mental health status. SCL-90 is one of the most popular measures of mental health status [33-34]. The instrument assesses 10 dimensions, including total score, somatization, obsessive-compulsive, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation and psychoticism. The items consist of multiple symptoms defining each mental health dimension. Participants rated themselves on each item by a 5-point Likert-type scale (1 = not at all to 5 = extremely).

3.3 Behavioral Features Extraction

Richness in valid behavioral features would improve the performance of established models. Because of its flexibility, interactivity and complexity, extracting behavioral features from raw log data seems to be difficult. The ontology of web use behaviors is needed to guide the behavioral features extraction.

In order to describe web use behaviors completely and guarantee the behavioral features sensitive to the variance of web user's psychological features, we build the ontology of web use behaviors based on psychological theory. According to the parent-child relationships of psychological terms displayed in "Thesaurus of Psychological Index Terms (10th Edition)", we selected "behavior" as root concept to develop a concept tree of human behavior (see Fig. 1). This concept tree was regarded as a picture of human behavior in psychological perspective. Some concepts were deleted from concept tree, which were less likely to appear in cyberspace (e.g., eating behavior and animal ethology). Guided by this modified concept tree, we combined different web elements (e.g., behavioral agent, behavioral object, applied web service, behavioral trace and time stamp) as web use behaviors to represent these concepts and provided them with instances. Through this process, the ontology of web use behaviors would be built in the future.

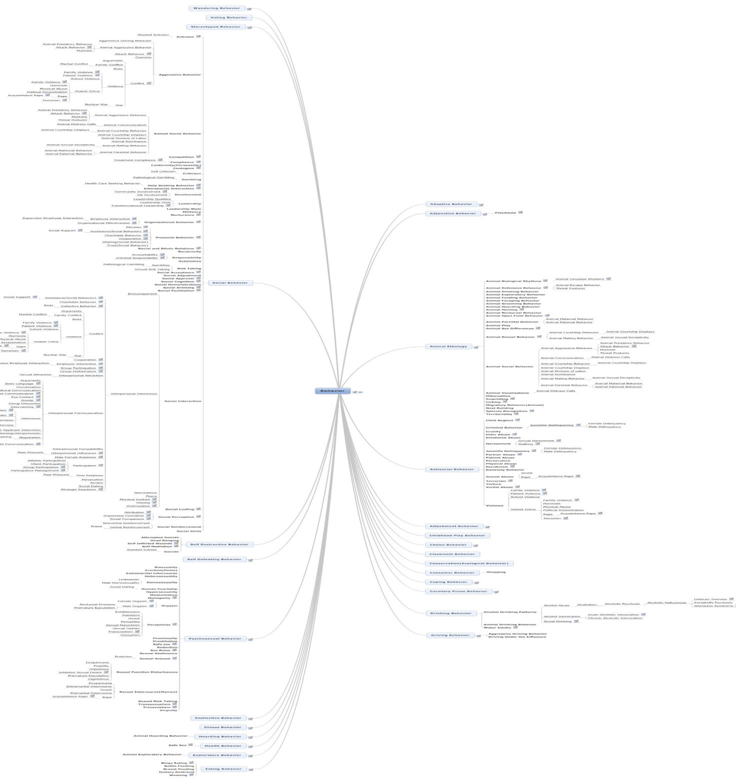

Fig. 1. Concept Tree of Human Behavior

In this study, according to established concept tree and recording contents of NS-ICG 5.6 and NS-DC 2.1, we extracted 131 behavioral features. Categories of extracted behavioral features were shown in Table 1, and types of web services involved in this study were shown in Table 2. Because most of extracted behavioral features could be analyzed daily and represented as time series data (e.g., daily time spent online), we used both average value and standard deviation to describe any time series.

3.4 Models Training

To predict each mental health dimension, web use behaviors were regarded as predictors to fit a pace regression model. Compared with ordinary least squares (OLS) regression, pace regression improves by "evaluating the effect of each variable" and "using a clustering analysis to improve the statistical basis for estimating their contribution to the overall regression" [35]. Previous studies have reached little consensus about how user's mental health status associates with web use behaviors. Because of the absence of theoretical or empirical evidence for assigning priority, we put predictors into pace regression model by using stepwise method [36]. 5-fold cross-validation was used to improve on the process of models training. Besides, both Pearson Correlation Coefficient, which estimated the correlation between predicted scores and actual scores on each mental health dimension, and Relative Absolute Error (RAE) were used to evaluate the performance of models.

Table 1. Description of Extracted Behavioral Features

Categories
Time and frequency of surfing, and preferred time period for surfing
Amount and percentage of using each type of web request (e.g., TELNET and FTP)
Amount, percentage and strategy of browsing each type of web page (e.g., social communication and entertainment)
Amount and percentage of sending and receiving messages, and number of friends and registered accounts on each kind of instant messaging software (e.g., QQ and MSN)
Amount and percentage of using each kind of search engine (e.g., Baidu and Google)
Amount and percentage of searching each type of content (e.g., books and pictures), and length of query words
Amount and percentage of using each kind of SNS (e.g., RenRen and Kaixin001)
Amount and percentage of using each kind of micro-blogging site (e.g., Sina and Tencent weibo)

4 Results

We trained 10 pace-regression models for predicting scores on each mental health dimension: total score, somatization, obsessive-compulsive, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation and psychoticism. For each model, the number of selected predictors was displayed in Fig. 2.

Results (see Fig. 3) showed that, from the perspective of Pearson Correlation Coefficient, all 10 correlation coefficients of pace regression models were higher than the criterion of 0.40 ($r = 0.49$ to 0.65), implying moderate correlations (mean $r = 0.58$). Among them, correlation coefficients of pace regression models for predicting total scores ($r = 0.62$), anxiety ($r = 0.61$), depression ($r = 0.65$) and obsessive-compulsive ($r = 0.62$) were higher than the criterion of 0.60, implying strong correlations. From the perspective of RAE, all 10 RAE of pace regression models were lower than the criterion of 100% (RAE = 75% to 89%, mean RAE = 81.82%), implying acceptable modeling error.

Table 2. Description of Selected Web Services

Types	Instances	URL
Instant Messaging	QQ	http://im.qq.com/
	Fetion	http://download.feixin.10086.cn/pc/
	MSN	http://cn.msn.com/
Searching Engine	Baidu	http://www.baidu.com/
	Google	http://www.google.com.hk/
	Bing	http://cn.bing.com/
	Amazon	http://www.amazon.cn/
	Dianping	http://www.dianping.com/
	Ganji	http://bj.ganji.com/
	Aibang	http://www.aibang.com/
	Anjuke	http://beijing.anjuke.com/
	58	http://bj.58.com/
	JD	http://www.jd.com/
	Taobao	http://www.taobao.com/
	Iask	http://iask.sina.com.cn/
	Verycd	http://www.verycd.com/
	Sina	http://www.sina.com.cn/
	Sogou	http://www.sogou.com/
	Soso	http://www.soso.com/
	Soufun	http://www.soufun.com/
	Xunlei	http://www.kankan.com/
	Yahoo	http://cn.yahoo.com/
	Youdao	http://www.youdao.com/
SNS	Renren	http://renren.com/
	Kaixin001	http://www.kaixin001.com/
	Douban	http://www.douban.com/
	Pengyou	http://www.pengyou.com/
Micro-Blogging	Sina Weibo	http://weibo.com/
	Tencent Weibo	http://t.qq.com/

5 Discussion

In this paper, we trained a computational model to identify web user's mental health status by analyzing his/her web use behaviors. Results indicated that the web-behavior-based measurement had satisfying metric properties. According to the criterion of convergent validation in psychological research, test scores should be strongly correlated with other measures of the same psychological feature [37]. In Fig. 3, the correlation coefficients between predicted scores and actual scores on all 10

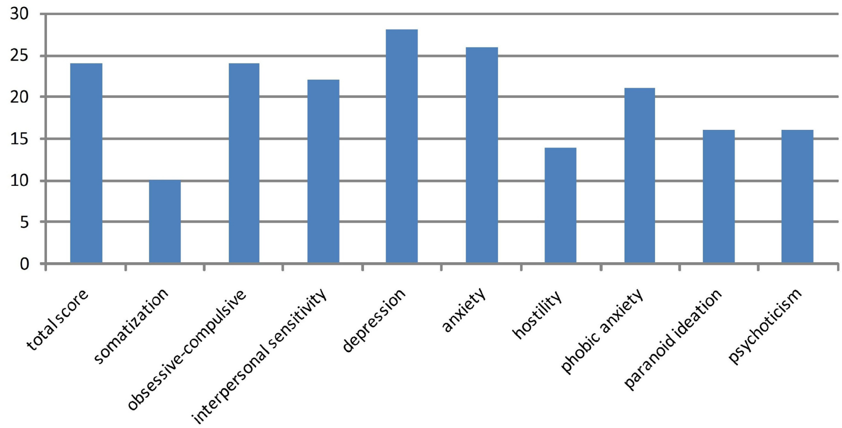

Fig. 2. Number of Predictors in Pace Regression Models Predicting Mental Health Status

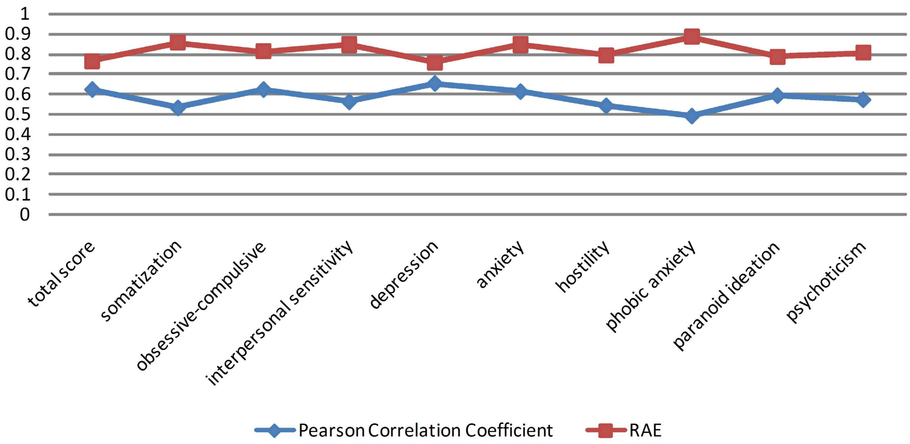

Fig. 3. Performance of Pace Regression Models Predicting Mental Health Status

dimen-sions ranged from 0.49 to 0.65, which were equal to both classical psychological measurement (r = 0.39 to 0.68) and empirical result of similar study for predicting personality (r = 0.43) [32], [38]. In addition, the value of RAE ranged from 75% to 89%, which suggested that labeling web user's mental health status with predicted scores would have less error than the average of test scores. Thus, from the perspective of both convergent validity and modeling error, the performance of measuring mental health status based on web use behaviors was ideal.

Our work can inspire future studies to detect and diagnosis mental health disorders by means of analyzing user's web use behaviors. In this paper, it is impossible to ensure that 131 behavioral features can describe web use behaviors completely. In the future, the ontology of web use behaviors needs to be further improved for searching more valid behavioral features, to build more accurate models on a large diverse population. With the help of established models, it is expected that user's mental health status can be measured momentarily and ecologically.

Acknowledgements. The authors gratefully acknowledge the generous support from National High-tech R&D Program of China (2013AA01A606), NSFC (61070115), Key Research Program of CAS (KJZD-EW-L04), Institute of Psychology (113000C037), Strategic Priority Research Program (XDA06030800) and 100-Talent Project (Y2CX093006) from Chinese Academy of Sciences.

References

1. Herrman, H., Saxena, S., Moodie, R., et al.: Promoting Mental Health-Concepts, Emerging Evidence, Practice. World Health Organization, Geneva (2005)
2. Kessler, R.C., Angermeyer, M., Anthony, J.C., et al.: Lifetime Prevalence and Age-of-Onset Distributions of Mental Disorders in the World Health Organization's World Mental Health Survey Initiative. World Psychiatry 6(3), 168–176 (2007)
3. Phillips, M.R., Zhang, J., Shi, Q., et al.: Prevalence, Treatment, and Associated Disability of Mental Disorders in Four Provinces in China during 2001-2005: An Epidemiological Survey. The Lancet 373(9680), 2041–2053 (2009)
4. World Health Organization: Mental Health: New Understanding, New Hope. The Journal of the American Medical Association 286(19), 2391 (2001)
5. Lopez, A.D., Mathers, C.D., Ezzati, M., Jamison, D.T., Murray, C.J.L.: Global Burden of Disease and Risk Factors. Oxford University Press/World Bank, New York (2006)
6. Diener, E., Suh, E.M., Lucas, R.E., Smith, H.L.: Subjective Well-Being: Three Decades of Progress. Psychological Bulletin 125(2), 276–302 (1999)
7. Andrews, G.: Reducing the Burden of Depression. Canadian Journal of Psychiatry 53(7), 420–427 (2008)
8. Phelan, M., Stradins, L., Morrison, S.: Physical Health of People with Severe Mental Illness. British Medical Journal 322(7284), 443–444 (2001)
9. Kitchener, B.A., Jorm, A.F.: Mental Health First Aide Manual. Centre for Mental Health Research, Canberra (2002)
10. Domino, G., Domino, M.L.: Psychological Testing: An Introduction. Cambridge University Press, New York (2006)
11. Butcher, J.N., Williams, C.L.: Essentials of MMPI-2 and MMPI-A Interpretation, 2nd edn. University of Minnesota Press, Minneapolis (2000)
12. Derogatis, L.R., Rickels, K., Rock, A.F.: The SCL-90 and the MMPI: A Step in the Validation of a New Self-Report Scale. The British Journal of Psychiatry 128, 280–289 (1976)
13. Buchanan, T., Smith, J.L.: Using the Internet for Psychological Research: Personality Testing on the World Wide Web. British Journal of Psychology 90, 125–144 (1999)
14. Carlbring, P., Brunt, S., Bohman, S., Austin, D., Richards, J., Ost, L., et al.: Internet vs. Paper and Pencil Administration of Questionnaires Commonly Used in Panic/Agoraphobia Research. Computers in Human Behavior 23, 1421–1434 (2007)
15. Anastasi, A., Urbina, S.: Psychological Testing. Prentice Hall, Upper Saddle River (1997)
16. Furr, R.M.: Personality Psychology as a Truly Behavioral Science. European Journal of Personality 23, 369–401 (2009)

17. Baumeister, R.F., Vohs, K.D., Funder, D.C.: Psychology as the Science of Self-Reports and Finger Movements: Whatever Happened to Actual Behavior. Perspectives on Psychological Science 2, 396–403 (2007)
18. Rampello, L., Nicoletti, F., Nicoletti, F.: Dopamine and Depression: Therapeutic Implications. CNS Drugs 13, 35–45 (2000)
19. Goddard, A.W., Mason, G.F., Almai, A., et al.: Reductions in Occipital Cortex GABA Levels in Panic Disorder Detected with 1h-Magnetic Resonance Spectroscopy. Archives of General Psychiatry 58, 556–561 (2001)
20. Eirinaki, M., Vazirgiannis, M.: Web Mining for Web Personalization. ACM Transactions on Internet Technology 3, 1–27 (2003)
21. Schegga, R., Steiner, T., Labbena, G., Murphy, J.: Using Log File Analysis and Website Assessment to Improve Hospitality Websites. Information and Communication Technologies in Tourism 17, 566–576 (2005)
22. Brunswik, E.: Perception and the Representative Design of Psychological Experiments. University of California Press, Berkeley (1956)
23. Gosling, S.D., Ko, S.J., Mannarelli, T., Morriss, M.E.: A Room with a Cue: Personality Judgments based on Offices and Bedrooms. Journal of Personality and Social Psychology 82, 379–398 (2002)
24. Morgan, C., Cotton, S.R.: The Relationship between Internet Activities and Depressive Symptoms in a Sample of College Freshmen. Cyberpsychology & Behavior 6(2), 133–142 (2003)
25. Ozcan, N.K., Buzlu, S.: Internet Use and Its Relation with the Psychosocial Situation for a Sample of University Students. Cyberpsychology & Behavior 10(6), 767–772 (2007)
26. Ceyhan, A.A., Ceyhan, E.: Loneliness, Depression, and Computer Self-Efficacy as Predictors of Problematic Internet Use. Cyberpsychology & Behavior 11(6), 699–701 (2008)
27. van den Eijnden, R.J.J., Meerkerk, G.J., Vermulst, A.A., et al.: Online Communication, Compulsive Internet Use, and Psychosocial Well-Being among Adolescents: A Longitudinal Study. Developmental Psychology 44(3), 655–665 (2008)
28. Bessiere, K., Kiesler, S., Kraut, R., Boneva, B.S.: Effects of Internet Use and Social Resources on Changes in Depression. Information, Communication & Society 11(1), 47–70 (2008)
29. Selfhout, M.H.W., Branje, S.J.T., Delsing, M., et al.: Different Types of Internet Use, Depression, and Social Anxiety: The Role of Perceived Friendship Quality. Journal of Adolescence 32, 819–833 (2009)
30. Peng, W., Liu, M.: Online Gaming Dependency: A Preliminary Study in China. Cyberpsychology & Behavior 13(3), 329–333 (2010)
31. Gosling, S.D., Augustine, A.A., Vazire, S., et al.: Manifestations of Personality in Online Social Networks: Self-Reported Facebook-Related Behaviors and Observable Profile Information. Cyberpsychology, Behavior, and Social Networking 14(9), 483–488 (2011)
32. Kosinski, M., Stillwell, D., Graepel, T.: Private Traits and Attributes Are Predictable from Digital Records of Human Behavior. Proceedings of the National Academy of Sciences 110(15), 5802–5805 (2013)
33. Holi, M.M., Sammallahti, R.R., Aalberg, V.A.: A Finnish Validation Study of the SCL-90. Acta Psychiatrica Scandinavica 97(1), 42–46 (1998)
34. Holi, M.M., Marttunen, M., Aalberg, V.: Comparison of the GHQ-36, the GHQ-12 and the SCL-90 as Psychiatric Screening Instruments in the Finnish Population. Nordic Journal of Psychiatry 57(3), 233–238 (2003)

35. Wang, Y., Witten, L.H.: Pace Regression. Technical report, Department of Computer Science, University of Waikato (1999)
36. Leland, W.: Tests of Significance in Stepwise Regression. Psychological Bulletin 86(1), 168–174 (1979)
37. Campbell, D.T., Duncan, W.F.: Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix. Psychological Bulletin 56(2), 81–105 (1959)
38. Craig, R.J.: Assessing Personality and Psychopathology with Interviews. In: Handbook of Psychology: Assessment Psychology, vol. 10, pp. 102–137. John Wiley & Sons, Inc., Hoboken (2003)