Computers in Human Behavior 33 (2014) 69–78

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Predicting self-monitoring skills using textual posts on Facebook

Qiwei He [a,*], Cees A.W. Glas [a], Michal Kosinski [b], David J. Stillwell [b], Bernard P. Veldkamp [a]

[a] Department of Research Methodology, Measurement, and Data Analysis, University of Twente, P.O. Box 217, GW-OMD, 7500AE Enschede, The Netherlands
[b] The Psychometrics Centre, Cambridge University, Psychology Department, Downing Street, Cambridge CB2 3EB, United Kingdom

ARTICLE INFO

Article history:

Keywords:
Facebook
Self-monitoring
Text mining
Item response theory
Data validation
Natural language processing

ABSTRACT

The popularity of the social networking site Facebook (FB) has grown unprecedented during the past five years. The research question investigated is whether posts on FB would also be applicable for the prediction of users' psychological traits such as self-monitoring (SM) skill that is supposed to be linked with users' expression behavior in the online environment. We present a model to evaluate the relationship between the posts and SM skills. The aim of this study is twofold: first, to evaluate the quality of responses to the Snyder's Self-Monitoring Questionnaire (1974) collected via the Internet; and secondly, to explore the textual features of the posts in different SM-level groups. The prediction of posts resulted in an approximate 60% accuracy compared with the classification made by Snyder's SM scale. The variable "family" was found the most significant predictor in structured textual analysis via Linguistic Inquiry and Word Count (LIWC). The emoticons and Internet slangs were extracted as the most robust classifiers in the unstructured textual analysis. We concluded that the textual posts on the FB Wall could partially predict the users' SM skills. Besides, we recommend that researchers always check the validity of Internet data using the methodology presented here to ensure the data is valid before being processed.

1. Introduction

As the second most visited website on the Internet (Alexa Internet Inc., 2011), Facebook (FB) attracts a global audience of over 606 million people for a daily use (Gonzalez, 2011). When joining the FB community, the platform requires users to compose an online "self" and allows them to share their emotions and problems by posts on the Wall, which are viewed by the users' self-selected and mediated audience. This composition is inherently an act of self-presentation, which is "the goal directed activity of controlling information of self in order to influence the impressions formed by audiences" (Schlenker, 2004).

Among all the multimedia formats of the posts, textual input is predominantly used for updating users' status on the FB. The "status updates" are short user-generated public messages that generally contain information about what the FB user is doing or thinking at that point of time, i.e., "what's on your mind?" (Ryan & Xenos, 2011). Such language is regarded as the most common way for people to translate their internal thoughts and emotions into a form that others (i.e., online audiences) can understand (Tausczik & Pennebaker, 2010).

Individuals' daily expressions can also be used to predict personality traits (e.g., Mairesse, Walker, Mehl, & Moore, 2007; Markovikj, Gievska, Kosinski, & Stillwell, 2013; Pennebaker & King, 1999). Among these traits, self monitoring (SM; Snyder, 1974) is identified as a special trait linked with users' expression behavior in the online environment (Ellison, Heino, & Gibbs, 2006; Toma, Hancock, & Ellison, 2008). On the virtual platform, the self-presentational affordances led by SM skills create "dialectical tensions between an accurate and an ideal self and between a truthful and a deceptive self" (Hall & Pennington, 2013). In the research regarding degree and type of online deception, SM skills showed promise in explaining variance in online misrepresentation (Hall, Park, Song, & Cody, 2010). Hall et al. (2010) suggested that SM was the strongest and most consistent predictor of strategic misrepresentation compared with the Big-Five personality traits (i.e., openness, conscientiousness, extraversion, agreeableness, and neuroticism; see more in Digman, 1990; John, 1990) and demographic variables (e.g., gender, age, and education). Therefore, given concerns on the specialty and importance of SM skills in the online environment, the current study focused on exploration of the relationship between the users' SM skills and their messages on FB.

Recent studies have explored the relationship between SM skills and the linguistic cues on FB. For instance, Hall et al. (2010) demonstrated that similar as in a face-to-face environment, the SM skill that is applied in various self-presentation tactics played an important role in controlling oneself to the online social appropriateness and reflecting individual's expression behavior. Hall and

* Corresponding author. Tel.: +31 53 4892829; fax: +31 53 4894239.
E-mail addresses: q.he@utwente.nl (Q. He), c.a.w.glas@utwente.nl (C.A.W. Glas), mk583@cam.ac.uk (M. Kosinski), ds617@cam.ac.uk (D.J. Stillwell), b.p.veldkamp@utwente.nl (B.P. Veldkamp).

0747-5632/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.chb.2013.12.026

Pennington (2013) found that the high self-monitors would be more extravert and used cues on FB that might promote extraversion on FB (i.e., to receive more "likes" by their FB friends). For instance, the high self-monitors were more likely to use profile pictures at a younger age and use shorthand abbreviations, such as OMG (oh my god) and BTW (by the way). In comparison, the low self-monitors were more likely to be honest on FB. They promoted a conscientious self to their FB public. For example, a higher proportion of family talks was found in this group in their FB status updates. Rosenberg and Egbert (2011) also suggested that the low self-monitors were less sensitive to social cues than the high self-monitors, and therefore were less skilled at assessing appropriate behaviors and self-presentation in various situations. Unlike high self-monitors who regulate their own words and behaviors perceived favorably by others, low self-monitors often chose actions and words in accordance with their dispositions in a social network.

However, the relationship between SM skills and users' online expression features is not the only important aspect of a social network behavior analysis; the predictability of the online posts for users' SM level is also of interest. That is, in addition to identifying the expression features in different SM levels, it is also important to assess whether these FB textual features can predict the level of users' SM skills. Little investigation in this area has been done in earlier studies.

The aim of this study is twofold. The first is to evaluate the quality of responses to the Snyder's Self-Monitoring (SM) Questionnaire collected via the Internet. This will be done using an item response theory (IRT) model. Besides giving an indication of the scalability and reliability of the responses, the model also provides estimates of the personal level of SM skills of the sampled FB users. Secondly, the textual features of the posts for different SM-level groups will be extracted using structured and unstructured textual analysis using a concurrent model of the measured SM skills and posts on the FB Wall.

1.1. Self-monitoring

The SM construct was introduced by Snyder (1974) as a trait that describes and explains individual differences in the self-control of expressive behavior for the sake of the demands and norms of an audience or context (von Davier & Rost, 1996). There are striking and important individual differences in the extent to which individuals can and do monitor their self-presentation, expressive behavior, and nonverbal affective display (Snyder, 1974). The SM-scale developed by Snyder (1974) was specifically designed to discriminate individual differences in concern for social appropriateness, sensitivity to the expression and self-presentation of others in social situations as cues to social appropriateness of self-expression. This instrument covers 25 self-report items like "I find it hard to imitate the behavior of other people" that are usually analyzed in a quantitative fashion, i.e., by summing the item responses after coding all items in the same direction. A median-split (sum score = 12) is generally applied to these sum scores in order to differentiate between two groups of people, the high self-monitors and the low self-monitors (von Davier & Rost, 1996).

The high self-monitors are characterized as persons who behave strategically to obtain desired outcomes by regulating public presentations. That is, when persons are made certain of their emotional reactions, they look to the behavior of others for cues to define their emotional states and model the emotional expressive behavior of others in the same situation who appear to be behaving appropriately (Schachter & Singer, 1962). For instance, such a person would be more likely to laugh at a comedy when watching it with amused peers than when watching it alone (Fuglestad & Snyder, 2009). In comparison, the low self-monitors present themselves in ways that reflect their authentic attitudes, values and beliefs. They express it as they feel it rather than monitored, controlled, and molded to fit the situation (Snyder, 1974).

1.2. Online assessment and Internet data

The past decade has witnessed a rapid expansion of the Internet. The Internet has not only significantly changed the way people conduct business, communicate, and live, but also influenced the practice of psychology as it related to testing and assessment (Naglieri et al., 2004). The Internet brings benefits of speed, costs, convenience and flexibility to the online assessment but introduce new problems such as testing security and data validity as well. A major limitation of previous researches regarding FB is that the data primarily assessed through self-report and few attempts have been made to evaluate the criterion validity of these measures. In other words, few studies has been conducted yet to evaluate whether self-reported data on FB are, in fact, related to actual data collected offline. A recent study was conducted by Junco (2013) regarding the criterion validity of measures of FB frequency by comparing self-reported time spent on the site and number of logins against actual usage as measured by computer monitoring software. Although there was a strong positive correlation between self-reported and actual time spent on FB in that study, a significant discrepancy was also shown between the two: Students spent an average of 26 min per day on FB, significantly lower than the average of 145 min per day obtained through self-report. Researches in other areas of human behavior have also shown that self-report measures through Internet tests could raise the risks of inaccuracy when compared to actual behaviors. For example, online self-reported measures of physical activity underestimated health risk biomarkers by as much as 50% when compared to accelerometer measurements (Celis-Morales et al., 2012), and self-reported TV watching time was underestimated by an average of 4.3 h per week when compared to data from a TV monitor (Otten, Littenberg, & Harvey-Berino, 2010).

Although some misrepresentation and outright deception unavoidably occurs on mediated platforms like FB, on the whole, users present themselves online in a manner that approximates their offline self (Gosling, Gaddis, & Vazire, 2007; Toma et al., 2008). However, as there is growing interest in researching on the Internet data and the psychosocial effect of FB use, it is important to come up with measurement methods that are both accurate and useful. The requirement of validity is of utmost importance for such tests on the Internet (Buchanan, 2002). How to validate the Internet data would be an interesting and essential topic to be addressed before stepping into data processing. In the present study, a validation method based on IRT (Lord, 1980; Rasch, 1960) was introduced to examine the authenticity of the Internet data.

1.3. Structured textual analysis – LIWC

Computer-based textual analysis is generally divided into two categories: structured and unstructured one. Structured textual analysis usually involves tight structures from existing software, such as Linguistic Inquiry and Word Count (LIWC; Pennebaker, Francis, & Booth, 2001). LIWC is a textual analysis software program that looks for words and counts words in psychology-relevant categories across multiple text files, for instance, essays, emails, blogs, novels and so on. It has two central features – the processing component and the dictionaries. During the processing, the program goes through each file word by word. Each word in a given text file is compared with the dictionary file. A dictionary refers to the collection of words that define a particular category such as "family", "positive emotion" and "work". There are 80 word categories (i.e., variables) in LIWC. These variables are divided into five dimensions: (a) Linguistic Process, in which variables as word count, word count per sentence, first person pronouns, verbs are output; (b) Psychological Process, in which variables related to positive emotions, negative emotions, family words, anxiety words, and etc. are calculated; (c) Personal Concerns, in which variables related to hobbies, work, life and etc. are output, (d) Spoken Categories, which focus on elements used in spoken language, and (e) Punctuations (Tausczik & Pennebaker, 2010). Each word or word stem defines one or more word categories or subdictionaries. For example, the word "cried" is part of five word categories: sadness, negative emotion, overall affect, verb, and past tense verb. Hence, if it is found in the target text, each of these five subdictionary scale scores will be incremented.

1.4. Unstructured textual analysis – text mining

Unstructured textual analysis focuses on extraction of patterns from loose structures. The development of information technology demonstrated breakthroughs in handling unstructured textual data during the past decade. A promising technique is text mining, which exploits information retrieval, information extraction, and corpus-based computational linguistics. This technique also plays a fundamental role in extracting correlation patterns between personality and variety of user's data captured from multiple sources (Markovikj et al., 2013).

Supervised text classification is a commonly used approach for textual categorization with text mining techniques. It generally involves two phases, a training phase and a prediction phase (Jurafsky & Martin, 2009). During training, the most discriminative keywords for determining the class labels are extracted. The input for the machine learning algorithm consists of a set of pre-specified keywords that may potentially be present in a document and labels classifying each document. The objective of the training phase is to "learn" the relationship between the keywords and the class labels. The prediction phase plays an important role in checking how well the trained classifier model performs on a new dataset. The test set should consist of data that were not used during training. In the testing procedure, the keywords extracted from the training are scanned in each new input. Thus, the words that were systematically recognized are fed into the "trained" classifier model, which predicts the most likely label for each new self-narrative. To ensure proper generalization capabilities for the text classification models, a cross-validation procedure is generally applied.

1.5. The present study

The objective of the present study is to investigate whether the textual features of posts on FB Wall can predict the users' SM ability. After a validity checking on the Internet data, both structured textual analysis by LIWC and unstructured textual analysis by text mining techniques were conducted in this study. The predictions from textual analyses were further compared with the binary classification i.e., high or low self-monitors defined by the SM-scale (Snyder, 1974). As stated earlier, the purpose of this study is twofold: first, to evaluate the quality of responses to the Snyder's SM Questionnaire (1974) collected via the Internet; and secondly, to explore the textual features of the posts in different SM-level groups.

2. Method

2.1. Dataset

A sample of 39,218 instances from Facebook (activity and demographic data) with approximately 1.8 million status updates used in the present study was provided by the MyPersonality project (http://mypersonality.org/wiki; Celli, Pianesi, Stillwell, & Kosinski, 2013). All the instances participated the Snyder SM test on FB at least once within the time period from January, 2009 to April, 2011. The respondents were highly motivated to answer honestly and carefully, as the only gratification they received for their participation was the feedback on their results (Kosinski, Stillwell, & Graepel, 2013). Among the respondents, 37,360 people took the test once, and 1858 people took the test at least twice. (We took the most recent responses into analysis for the duplicate cases.) Applying the median-split rule (threshold = 12) on the SM scale, 54.8% of the total are low self-monitors, while 45.2% are high self-monitors. The SM score followed a normal distribution with mean equaling to 11.97 and standard deviation equaling to 4.13.

Of the 39,218 instances, 2972 respondents had at least one textual post in status updates during the collection time period, that is, they gave both textual expressions and responses to the SM-scale. To simplify the study, we only focused on English-speaking people, which resulted in 2655 participants. Further, to concentrate on the investigation of predictability of posts on FB, we followed the approach of Argamon, Dhawle, Koppel, and Pennebaker (2005) by only including respondents with extremely high or low scores on SM-scale and excluding the middle scorers. The extreme groups were defined as SM scores were above the 75% (SM > 15) or below the 25% (SM < 9) (Snyder, 1974). The typical group of extreme SM skills might be professional stage actors as high self-monitors and psychiatric patients as low self-monitors, respectively. This approach was testified in the study of Mairesse et al. (2007) that a 2–3% increase in overall accuracy scores yielded compared to datasets that included the middle scorers. The two groups with extreme SM scores, consisting of 1128 respondents, were finally used in the current study for textual analysis. The sum of their posts on FB Wall was approximately 140 thousand. Of the 1128 respondents, 552 (48.9%) were extremely low self-monitors, while 576 (51.1%) were extremely high self-monitors. The majority of respondents were female (55.7%). The age of respondents ranged from 18 to 60 years, with a mean of 25.7 and standard deviation as 9.1.

2.2. Validation of the Internet data

The purpose of data validation is to determine that the data are valid, sensible, reasonable and secure before they are processed. In the present study, the data of the SM-scale collected from FB were validated by using an IRT model. In psychological and educational measurement, instruments are developed that are used in a population of persons and item fit is used to evaluate to what extent an IRT model fits an instrument in a particular population (Glas & Dagohoy, 2007). Analyses were carried out using the public domain software MIRT (Glas, 2010). The SM-scale (Snyder, 1974) was developed based on a population of Stanford University undergraduates ($n = 533$) and it is necessary to use item fit to evaluate whether the IRT model fits the SM-scale for the FB users. If the model holds fit, it implies that the Internet data are as valid as the original data that were used for the instrument development; otherwise, the Internet data can be determined as invalid.

Responses of the whole sample 39,218 FB users to the 25 items of the SM-scale were used as input for the statistical analysis. A unidimensional two parameter logistic (2PL) model was used to estimate the individual's latent trait of SM. In this model, that is, the probability of a score in category "yes" ($X_{ni} = 1$) of item i is given by the item response function

$$P(X_{ni} = 1|\theta_n) = \frac{\exp[\alpha_i(\theta_n - \beta_i)]}{1 + \exp[\alpha_i(\theta_n - \beta_i)]}, \quad (1)$$

where θ_n is the latent SM level of person n, β_i is an item location parameter representing the difficulty level of each SM item, and α_i is an item discrimination parameter indicating the extent to which the item response is related to the latent scale. The item parameters in the IRT model were estimated by marginal maximum likelihood (MML; Bock & Aitkin, 1981).

We investigated item fit using Lagrange Multiplier (LM; Glas, 1999) tests. Given the size of the data set, the focus will not be on the significance probabilities of the test, but on the observed and expected response frequencies and the effect sizes on which the test is based. To compute the LM statistic, the sample of respondents is divided into subgroups labeled $g = 1, 2, \ldots, G$. We defined the subgroups as three total-score level groups (i.e., Level 1: total scores 0–9, Level 2: total scores 10–14, and Level 3: total scores 15–25) which were formed in such a way that the numbers of respondents in each group were approximately the same. The statistic is based on the difference between average observed scores on every item i in the subgroups, namely, $S_{ig} = \frac{1}{N_g}\sum_{n|g}^{N_g} X_{ni}$ (where the summation is over the N_g respondents in subgroup g), and their posterior expectations $E(S_{ig})$. The differences are squared and divided by their covariance matrix (for more details refer to Glas, 1998, 1999; Glas & Falcon, 2003). The LM statistic has an asymptotic chi-square distribution with $G-1$ degree of freedom. The statistics are accompanied by effect size $d_{ig} = \max_g |S_{ig} - E(S_{ig})|$ that shows the degree of model violation. Since the effect size d_{ig} is on a scale ranging from 0 to the maximum score m_i, effect size $d_{ig} < 0.10$ although somewhat arbitrary, is commonly suggested as an indicator of minor and acceptable model violation (He, Glas., & Veldkamp., 2014; van Groen, ten Klooster, Taal, van de Laar, & Glas, 2010). In the current study, an item was identified as misfit when the effect size was above the cutoff point $d_{ig} = 0.10$.

Besides the item fit analysis, person fit is also necessary to take into consideration when handling the Internet data, because specific persons may still produce patterns that are highly unlikely given the model, although the IRT model may generally fit the data. For instance, some persons may give random responses because they are unserious to take the test. Using person fit statistics, the fit of a score pattern can be determined under the null-hypothesis that the IRT model holds. To test the person fit, we used the LM test for the constancy of θ over response patterns for the 2PL model introduced by Glas and Dagohoy (2007). The LM person fit test is based on a split two subtests: say the first part of the test (1–11 item) and the second part (12–25 item) of the test. In addition, to show that the quality of the Internet data is appropriate, we also compared the distribution of SM scores of the Stanford undergraduates sample and the FB sample by applying non-parametric Wilcoxon Rank Sum Test (also called Mann–Whitney Test).

2.3. Textual analyses

The textual posts on users' FB Wall were analyzed via both structured and unstructured approaches by using LIWC and text mining techniques, respectively. In the present study, we focused on the two extreme SM groups, the low self-monitors (LSM, SM < 9) and the high self-monitors (HSM, SM > 15). A sample of 1128 respondents with approximately 140 thousand posts were included.

2.3.1. Structured textual analysis using LIWC

All the 80 variables in LIWC, including 26 variables in the dimension of Linguistic Process, 32 variables in the dimension of Psychological Process, 7 variables in the dimension of Personal Concerns, 3 variables in the dimension of Spoken Categories and 12 variables in the dimension of Punctuations, were input for the structured textual analysis. Two approaches, logistic regression and classification trees were used to classify the individuals into two categories – HSM and LSM based on the features of their textual input.

In the logistic regression, the dependent variable was defined in a binary category, i.e., 0 (i.e., LSM) and 1 (i.e., HSM). We input all the 80 variables as predictors. The logistic regression model is defined as

$$\mathrm{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = b_0 + b_1 x, \tag{2}$$

where $p = \pi(x)$ is the probability that the dependent variable equals 1, b_0 and b_1 are regression coefficients.

In the approach of classification trees, we used the classification and regression tree (CRT) growing method with maximum tree depth equaling to 100. CRT splits the data into segments that are as homogeneous as possible with respect to the dependent variables. A terminal node in which all cases have the same value for the dependent variable is a homogeneous and "pure" node (Kotsiantis, 2007). The minimum number of cases in each parent node and child node were set as 100 and 50, respectively. Pruning tree was used to avoid over-fitting (for more on pruning trees refer to Bruha, 2000; Elomaa, 1999). The minimum change in improvement of each depth was set at 0.0001. To ensure the proper generalization capabilities for the classification tree model, a 10-fold cross validation procedure was applied.

2.3.2. Unstructured textual analysis using text mining techniques

In the unstructured textual analysis, a preprocessing was conducted first to ensure the textual data following a standardized format for a further use. Unlike the normal preprocessing in text classification, we included the stop words (e.g., "I", "is", "the", and etc.) in the present study, because some literature mentioned that the inclusion of stop words could increase the classification accuracy in textual analysis of online blogs (e.g., Iacobelli, Gill, Nowson, & Oberlander, 2011). Further on, all the words were stemmed using Porter's stemming algorithm (Porter, 1980). We noticed that the Internet language was more casual, thus resulted in more spelling mistakes than the normal writings. For instance, the "wrong" spelling "soooooo big" was often used in FB posts to emphasize the degree of bigness, but such coined word was often difficult to be standardized in the stemming process. In the current study, we handled these typical Internet words by two steps: first, transforming them into the original status (e.g., "sooooo" was transformed into "so") and secondly stemming them by the Porter stemming algorithm. This approach avoided the mighty confusion in keywords extraction. For example, "soooo" (i.e., "so" with four o's) and "sooooooo" (i.e., "so" with seven o's) would be extracted as a unique stem "so" instead of two different ones. However, the writer's latent intention to emphasize the degree of bigness was lost in preprocessing.

We deployed a supervised text classification in the present study, that is, to divide the textual analysis into two phases, training and testing. 70% of the dataset were randomly selected into training data, while the remaining 30% of the dataset were used to test the trained model. During training, the most discriminative keywords to determine the SM-level were extracted by using chi-square feature selection model (Oakes, Gaizauskas, Fowkes, Jonsson, & Beaulieu, 2001). A recently developed machine learning algorithm, product score model (PSM; He, Veldkamp, & de Vries, 2012) was employed in conjunction with three representative models – unigrams, bigrams, and a combination of uni- and bigrams – to learn the patterns between the extracted keywords and the labels.

The PSM is an alternative machine learning algorithm, which features in assigning two weights for each keyword (in binary classification) – the probability of the word w occurs in the two separate corpora, U_w and V_w – to indicate to how much of a degree the word can represent the two classes. The weights are calculated by

$$\begin{cases} U_w = (n_w + a)/len(C_1) \\ V_w = (m_w + a)/len(C_2) \end{cases}, \quad (3)$$

where n_w and m_w are the word occurrences in HSM Corpus (C_1) and LSM Corpus (C_2), respectively. Note that a smoothing constant a (we use $a = 0.5$ in this study) is added to the word occurrence in Formula (3) to account for words that do not occur in the training set, but might occur in new texts. (For more on smoothing rules, see Jurafsky & Martin, 2009; Manning & Schütze, 1999.) The name *product score* comes from a product operation to compute scores for each class, i.e., S_1 and S_2, for each input text based on the term weights. That is,

$$\begin{cases} S_1 = P(C_1) \cdot \prod_{w=1}^{k} U_w = P(C_1) \cdot \prod_{w=1}^{k} [(n_w + a)/len(C_1)] \\ S_2 = P(C_2) \cdot \prod_{w=1}^{k} V_w = P(C_2) \cdot \prod_{w=1}^{k} [(m_w + a)/len(C_2)] \end{cases}, \quad (4)$$

where $P(C)$ is the prior probability for each category given the total corpora. The classification rule is defined as:

$$\text{choose} \begin{cases} C = 1 \text{ if } \log(S_1/S_2) > b \\ C = 2 \text{ else} \end{cases}, \quad (5)$$

where b is a constant and was defined equal to 0 in the current study (for more on PSM, see He & Veldkamp, 2012; He et al., 2012).

To avoid mismatches caused by randomness, unclassification rules are also taken into account. Based on the chi-square selection algorithm, the keywords are labeled as positive indicators or negative indicators. We defined a text as "unclassified" when either one of the following conditions was met: (a) no keywords are found in the text; (b) only one keyword is found in the text; (c) only two keywords are found in the text, and one is labeled as a positive (i.e., HSM) indicator while the other as a negative (i.e., LSM) indicator.

In the current study, 1000 keywords, including 500 HSM indicators and 500 LSM indicators, were extracted as robust classifiers and used for text classification. To generalize the results from unstructured textual analysis, a 5-fold cross validation was also applied.

2.4. Analytic strategy

In the present study, we defined the label made by the SM-scale as "standard" and the label predicted via textual analysis as "test", respectively. The performances of the structured and unstructured textual analysis were compared on five metrics; accuracy, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). Accuracy, the main metric used in classification, is the percentage of correctly defined individuals. Sensitivity and specificity are the proportion of actual positives (HSM) and actual negatives (LSM) that are correctly identified, respectively. The predictive values, PPV and NPV, are estimators of the confidence in predicting correct classification; that is, the higher predictive values are, the more reliable the prediction is.

3. Results

3.1. Validation of the Internet data

Table 1 shows the item parameters that were calibrated by the MML on a sample of 39,218 instances from FB. As shown in Table 1, the discrimination parameters varied in the interval [0.256, 1.543], with a mean value around 0.69 (S.D. = 0.37). The difficulty parameters were included in the range [−0.808, 1.243], with a mean of 0.09 (S.D. = 0.64).

The validity of Internet data was investigated between the observations and expectations predicted by the IRT models using the LM statistics. The observed total score is the sum score of the responses on all items except the item targets. Table 2 reports the outcomes of analysis of model fit. The columns Obs and Exp give the observed and expected scores under the model, respectively. The last column (Dif) gives the effect sizes d_{ig}. Note that the highest effect size was 0.05, which is well below the criterion of 0.10. Further, in the person fit analysis, the detection of inconstancy of θ-estimates identified 365 individuals. Thus, the detection rate was 0.009, which is far below the significance probability of 0.05. In addition, the reliability of the SM score predictions of FB users was 0.732, which is well acceptable. Finally, in the comparison of distribution between the two samples of Stanford undergraduates and the FB users, the Wilcoxon Rank Sum Test resulted in a p-value of 0.25. Therefore, the null hypothesis that there is no systematic difference between the two independent populations was not rejected. As shown in Fig. 1, the SM scores of the FB users (solid line) follow a normal distribution, with a mean of 11.97 and standard deviation of 4.13. The SM scores of the Stanford undergraduates (dot line) also follow a normal distribution, though a bit condensed than the FB curve, with a mean of 12.41 and standard deviation of 3.48.

The overall conclusion is that the IRT model fitted the Internet data very well, and the hypothesis that a latent scale pertained to the FB users was not rejected. Thus, the Internet data used in this study was valid enough to be further processed.

3.2. Textual analysis

In the structured textual analysis approach, the logistic regression using LIWC showed a model fit in the Hosmer and Lemeshow test of goodness-of-fit ($\chi^2 = 4.507$, $df = 8$, $p = 0.809$). This test assesses whether or not the observed event rates match expected event rates in subgroups of the model population. Table 3 lists the logistic regression coefficients of the top twenty LIWC predictors based on the significance of the score statistic. It was found that only the first three predictors were significant in parameter estimation ($p < 0.05$), including two variables, i.e., "family" ($b = -0.496$) and "discrepancy" ($b = -0.311$), in the dimension of Psychological Process and one variable, i.e., "leisure" ($b = 0.198$), in the dimension of Personal Concerns.

Fig. 2 presents the classification tree model based on the 80 LIWC variables. The tree resulted in four depths, nine nodes and five terminal nodes. The four robust classifiers were extracted in a decreasing order: "anger", "family", "preps" and "word per sentence". Note that the improvement in each depth was very marginal and the highest improvement was produced by "anger" as 0.012, implying that the predictors were not very powerful to make the decision. This might also be the reason of the shortness of the tree. Based on a 10-fold cross validation, the risk of misclassification was estimated as 0.38 with standard error of 0.014.

Further, we also investigated the correlations between each variable in the LIWC and the SM scores. Using a 95% confidence interval, thirteen variables were found significantly correlated with the SM scores, including eight variables had positive correlations and five variables had negative correlations, though the values were not high (see Table 4). It was interesting to find that the words related to assent had the highest positive correlation with the SM scores, which implied that the higher SM skill a person has, the more often he/she may use assent words like "ok", "yes", "agree" in the posts. We also noticed that the lower SM skill a person had, the more likely he/she might use words related to the third episode and family terms (home, sister, brother, etc.) to update

Table 1
Item parameters of 25-item Self-monitoring scale (Snyder, 1974).

Item	Question in NCS-R	Item parameters	
		α (SE)	β (SE)
1	I find it hard to imitate the behavior of other people. (F)	1.159 (0.020)	−0.532 (0.014)
2	My behavior is usually an expression of my true inner feelings, attitudes, and beliefs. (F)	0.271 (0.016)	1.238 (0.013)
3	At parties and social gatherings, I do not attempt to do or say things that others will like. (F)	0.651 (0.012)	−0.542 (0.012)
4	I can only argue for ideas which I already believe. (F)	0.544 (0.014)	0.352 (0.011)
5	I can make impromptu speeches even on topics about which I have almost no information. (T)	0.959 (0.017)	0.298 (0.013)
6	I guess I put on a show to impress or entertain people. (T)	1.137 (0.020)	0.587 (0.014)
7	When I am uncertain how to act in a social situation, I look to the behavior of others for cues. (T)	0.256 (0.014)	−0.578 (0.011)
8	I would probably make a good actor. (T)	1.543 (0.024)	−0.166 (0.015)
9	I rarely seek the advice of my friends to choose movies, books, or music. (F)	0.649 (0.015)	1.243 (0.013)
10	I sometimes appear to others to be experiencing deeper emotions than I actually am. (T)	0.257 (0.013)	0.288 (0.011)
11	I laugh more when I watch a comedy with others than when alone. (T)	0.300 (0.013)	0.151 (0.011)
12	In groups of people, I am rarely the center of attention. (F)	1.060 (0.019)	0.365 (0.013)
13	In different situations and with different people, I often act like very different persons. (T)	0.476 (0.015)	−0.034 (0.012)
14	I am not particularly good at making other people like me. (F)	0.812 (0.017)	−0.753 (0.013)
15	Even if I am not enjoying myself, I often pretend to be having a good time. (T)	0.318 (0.013)	−0.034 (0.011)
16	I'm not always the person I appear to be. (T)	0.405 (0.015)	−0.728 (0.012)
17	I would not change my opinions (or the way I do things) in order to please someone else or win their favor. (F)	0.384 (0.015)	1.057 (0.013)
18	I have considered being an entertainer. (T)	1.254 (0.020)	0.251 (0.014)
19	In order to get along and be liked, I tend to be what people expect me to be rather than anything else. (T)	0.343 (0.016)	1.164 (0.013)
20	I have never been good at games like charades or improvisational acting. (F)	1.203 (0.020)	−0.571 (0.014)
21	I have trouble changing my behavior to suit different people and different situations. (F)	0.908 (0.018)	−0.808 (0.013)
22	At a party, I let others keep the jokes and stories going. (F)	0.730 (0.016)	0.678 (0.013)
23	I feel a bit awkward in company and do not show up quite as well as I should. (F)	0.631 (0.015)	−0.050 (0.012)
24	I can look anyone in the eye and tell a lie with a straight face (if for a right end). (T)	0.663 (0.015)	−0.398 (0.012)
25	I may deceive people by being friendly when I really dislike them. (T)	0.440 (0.014)	−0.189 (0.011)

Note. The item parameters were estimated from 2PL model. α indicates the item discrimination parameter, β indicates the item difficulty parameter.

Table 2
Model fit in score level groups for Facebook users (n = 39,218).

Item	Level 1		Level 2		Level 3		Dif.
	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.	
1	0.40	0.40	0.62	0.62	0.81	0.80	0.00
2	0.15	0.19	0.22	0.22	0.30	0.27	0.02
3	0.48	0.49	0.64	0.63	0.76	0.75	0.01
4	0.31	0.31	0.41	0.41	0.53	0.53	0.00
5	0.28	0.26	0.43	0.43	0.60	0.62	0.01
6	0.17	0.19	0.35	0.37	0.62	0.58	0.03
7	0.55	0.59	0.64	0.64	0.73	0.69	0.02
8	0.31	0.28	0.54	0.54	0.75	0.77	0.02
9	0.21	0.15	0.25	0.23	0.26	0.34	0.05
10	0.38	0.37	0.43	0.43	0.48	0.48	0.00
11	0.37	0.40	0.47	0.46	0.55	0.53	0.02
12	0.27	0.24	0.42	0.42	0.58	0.62	0.02
13	0.36	0.41	0.50	0.51	0.67	0.61	0.04
14	0.53	0.51	0.67	0.68	0.79	0.80	0.01
15	0.40	0.44	0.51	0.51	0.62	0.58	0.03
16	0.55	0.59	0.68	0.68	0.79	0.75	0.03
17	0.17	0.20	0.25	0.26	0.36	0.33	0.02
18	0.27	0.24	0.45	0.45	0.63	0.67	0.02
19	0.14	0.19	0.24	0.24	0.34	0.30	0.03
20	0.44	0.41	0.62	0.63	0.78	0.81	0.03
21	0.48	0.51	0.70	0.69	0.84	0.82	0.02
22	0.26	0.22	0.34	0.34	0.45	0.49	0.03
23	0.43	0.38	0.51	0.51	0.59	0.64	0.03
24	0.45	0.46	0.59	0.60	0.74	0.72	0.01
25	0.41	0.45	0.54	0.55	0.69	0.64	0.03

Note. The columns labeled Obs and Exp give the observed and expected scores under the model, respectively. The observed total score is the sum score of the responses on all items. Dif gives the absolute value of effect size averaged across the three score levels. Level 1: total scores 0–9, Level 2: total scores 10–14, Level 3: total scores 15–25. Degree of freedom equals to 2.

the status on FB. This result kept consistent with the findings in the study of Hall and Pennington (2013).

In the unstructured textual analysis, 1000 keywords, consisting of 500 keywords for HSM and 500 keywords for LSM were used for text classification with text mining techniques. Table 5 presents the top 20 keywords (10 for HSM and 10 for LSM) extracted from the FB posts. The fourth column shows the chi-square score for each keyword. The last two columns give the number of occurrences of each keyword in the LSM and HSM corpora, respectively. It was noticed that among the top twenty keywords, eight were emoticons and four were Internet slangs. The robustness of emoticons and Internet slangs in prediction of SM skills aroused our special interests. An emoticon is a communicative pictorial representation of a facial expression to send the feelings of the user, for instance, ":)" indicates a happy face and ":(" indicates a sad face. The Internet slangs are expressions that coined and popularized by the Internet users to save time on keystrokes, for instance, "wow" is generally used to express astonishment or admiration, and "ugg" often indicates ugly. As shown in Table 5, the "happy faces", e.g., ":)", "=)", ";)", and "^_^", were found the most significant classifiers for low self-monitors whereas the "sad or puzzling face", e.g., ":(" and "++", were the robust classifiers in the high self-monitors. In addition, we also found that the Internet slangs (e.g., "wow", "ugg", "lol", "omg") were used more often by the group of high self-monitors. This was similar as the findings in the study of Hall and Pennington (2013) where the high self-monitors were reported more likely to use the shorthand abbreviation.

Table 6 exhibits the performance metrics compared between structured and unstructured textual analysis. The logistic regression using variables from LIWC yielded the highest accuracy (0.629) among all the models. It also resulted in the highest sensitivity, specificity and PPV. In general, the structured textual analysis approach performed better than the unstructured one. However, giving concerns on classifying FB posts solely based on the keywords, the over 50% accuracy rate is acceptable. The PSM in conjunction with the unigrams performed the best in the unstructured textual analysis. Although the PSM with a combination of unigrams and bigrams resulted in the highest NPV (0.678) among all, it was compensated by the lowest PPV (0.328). The bigram was not as powerful as it was shown in the study of Iacobelli and his group (2011) where the bigrams

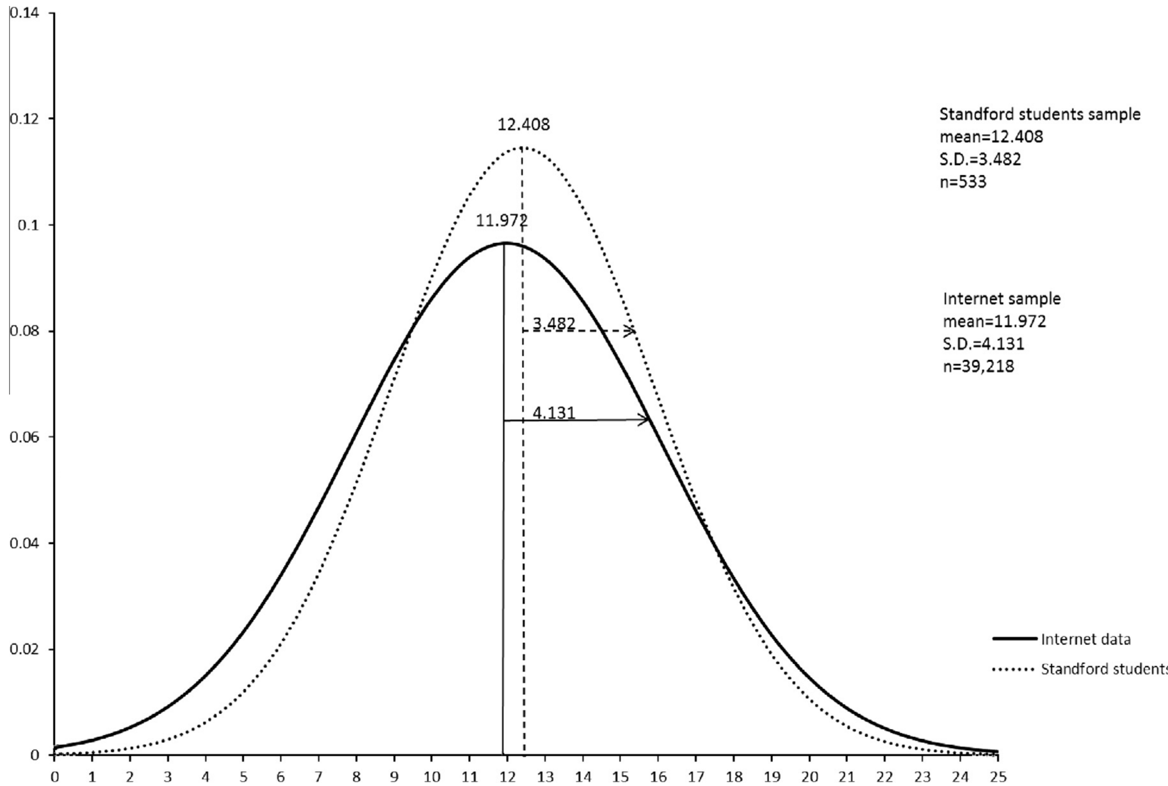

Fig. 1. SM score distribution of two samples: Stanford undergraduates ($n = 533$) and Facebook users on the Internet ($n = 39,218$).

Table 3
Logistic regression coefficients of LIWC predictors (top 20 predictors based on the significance of the score statistic).

	Predictor	b_1	p-Value	95% Confidence interval	
				Low	High
1	Family	−0.496	0.031	−1.105	−0.150
2	Leisure	0.198	0.046	0.020	0.460
3	Discrepancy	−0.311	0.049	−0.719	−0.003
4	Quantifiers	0.244	0.063	−0.009	0.565
5	Adverb	0.161	0.089	−0.032	0.397
6	Humans	−0.299	0.094	−0.770	0.010
7	Sad	−0.428	0.112	−1.033	0.190
8	Verb	−0.080	0.117	−0.216	0.037
9	Space	0.312	0.129	−0.103	0.868
10	Exclamation mark[a]	0.128	0.138	−0.039	0.350
11	Auxiliary verbs	0.154	0.161	−0.097	0.398
12	Function words	−0.135	0.172	−0.390	0.074
13	Religion	−0.156	0.172	−0.450	0.056
14	Period[a]	0.112	0.189	−0.067	0.327
15	Achievement	0.162	0.191	−0.154	0.435
16	Question mark[a]	0.111	0.196	−0.071	0.332
17	Parenthesis[a]	0.232	0.201	−0.173	0.673
18	Quote mark[a]	0.111	0.206	−0.067	0.336
19	All punctuations[a]	−0.102	0.221	−0.314	0.078
20	Assent	0.164	0.227	−0.102	0.486

Note. b_1 indicates the logistic regression coefficient for each predictor.
[a] Predictors in the dimension of punctuations in LIWC.

were found as the robust classifiers to represent the bloggers' personality. The probable reason might be that the blogs and FB posts are substantially different in textual format, though both of them are collected online. For instance, the blogs are generally longer texts and describe a relatively complete story while the posts are comparatively short and express thinking in mind in a relatively casual way.

4. Discussion

The present study evaluated the quality of responses to the Snyder's SM Questionnaire (1974) collected via the Internet, and explored the textual features of the posts in different SM-level groups and extracted patterns between FB users' SM skills and their posts on the FB Wall. The main contributions were made from the following three aspects: First, a method based on an IRT model was introduced to check the validity of the Internet data, which is of utmost importance for the online assessments. Secondly, by using both structured and unstructured textual analysis, we demonstrated that the textual posts on the FB Wall could partially predict the users' SM skills. The variable of "family" was found the most significant predictor in the approach of structured textual analysis. The accuracy of classification of the SM groups was above 60% when using LIWC and above 50% with the text mining algorithm PSM in conjunction with unigrams. Thirdly, the emoticons and Internet slangs were extracted as the most robust classifiers that played important roles in predicting the FB users' SM level.

It was found that the text classification of FB posts via an unstructured approach did not perform as well as the past researches that focused on story-based documents. He et al. (2012) used the PSM to analyze the patients' self-narratives to detect the posttraumatic stress disorder (PTSD) patients. The accuracy of text classification in that study was fairly high as 0.82. In another study of He and Veldkamp (2012), the PSM was applied to analyze undergraduates' life stories in order to understand their personality adaption. The computerized text classification resulted in over 70% accuracy compared with the human-raters' results. The reasons of relatively low classification accuracy in the current study might be addressed from three aspects. First, the contents of posts on FB Wall were much more diversified than the story-based documents. The FB users can express anything in their mind on the Wall whereas the respondents in the PTSD psychiatric

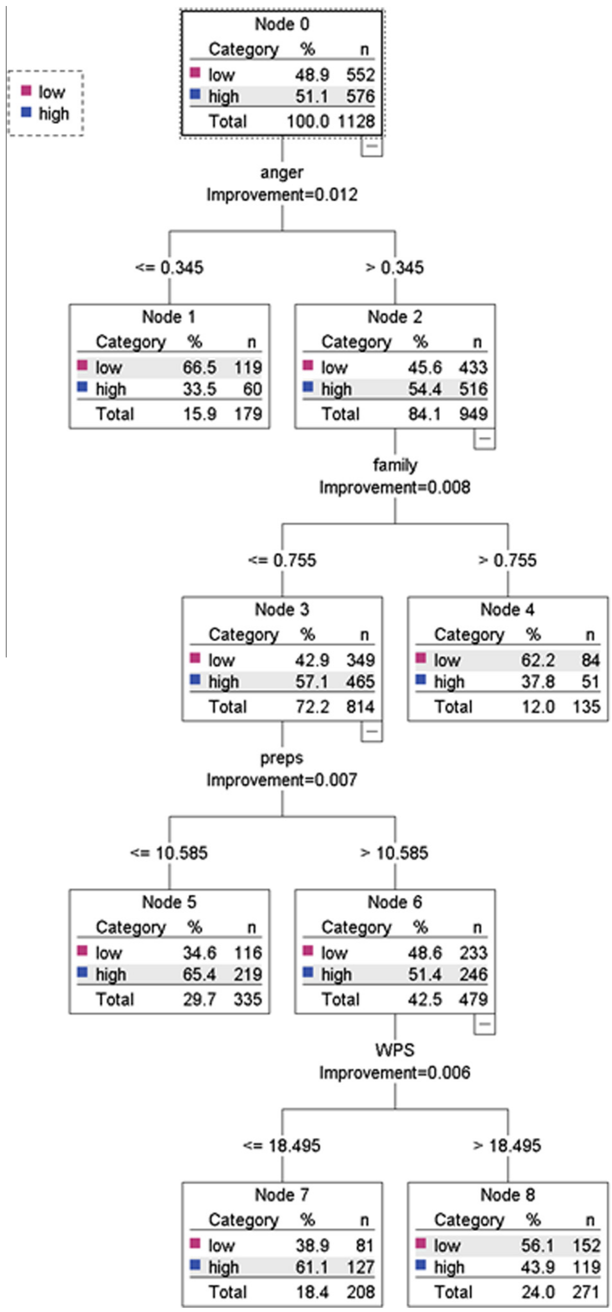

Fig. 2. Classification tree model on the 80 LIWC variables.
Note. Low and high represent low self-monitors and high self-monitors, respectively.

screening are asked to focus on description of the traumatic events and related symptoms. Secondly, the Internet textual posts may have more loose linguistic structures. For instance, on the FB, it is more likely to see a sentence without a subject or use too many punctuations to express the emotions, e.g., "?!!!!!". Thirdly, the posts on FB are generally written in a more casual way, e.g., using slangs, shorthand abbreviations and emoticons and might have more spelling mistakes and coined words, e.g., "soooooo big". These wording variations bring new challenges in the unstructured textual analysis.

To link the extracted keywords with the psychology-oriented predictors in the LIWC is helpful to locate those verbal features into psychological dimensions. We mapped the 1000 keywords into the 80 categories in LIWC. The top five LIWC predictors that were most frequently matched by the keywords were "affect" (affective processes), "posemo" (positive emotion), "relativ" (relativity), "social" (social process) and "cogmech" (cognitive process). All these five predictors located in the second dimension of LIWC, i.e., Psychological Processes, which implying that the keywords extracted from the text mining approach were mainly the words with psychological attributes. Moreover, we also noted that the

Table 4
Correlation between LIWC predictors and SM scores (significance $p < 0.05$).

Predictor	Correlation with SM scores
Assent	0.085
Question mark[a]	0.065
Exclusive	0.064
Adverb	0.051
Feel	0.046
Cause	0.044
Body	0.044
Family	−0.079
Religion	−0.068
They (3rd pers plural)	−0.060
Inhibition	−0.051
Dash mark[a]	−0.041

[a] Predictors in the dimension of Punctuations in LIWC.

Table 5
Top 20 keywords extracted by chi-square feature selection model in unstructured textual analysis.

Label	Rank	Keyword	Chi-square score	Number of occurrences	
				LSM corpus	HSM corpus
LSM	1	=)	93.60	102	6
	2	god	83.28	137	27
	3	grace	76.04	88	7
	4	:-)	74.94	134	30
	5	;)	66.48	79	7
	6	work	65.74	293	142
	7	"–	63.47	69	4
	8	bless	53.88	212	96
	9	(^-^)	49.96	46	0
	10	repost	48.67	172	73
HSM	1	":	232.27	13	291
	2	wow	123.97	158	460
	3	!!!	88.48	1783	2583
	4	:(	86.76	41	193
	5	wit	79.73	50	202
	6	ugg	63.44	1	72
	7	fuck	55.86	426	727
	8	++	52.40	1	60
	9	lol	47.70	40	140
	10	omg	46.88	160	332

Note. LSM and HSM indicate low self-monitors and high self-monitors, respectively.

Table 6
Performance metrics compared between structured and unstructured textual analysis.

	Accuracy	Sensitivity	Specificity	PPV	NPV
LIWC					
Logistic regression	0.629	0.629	0.628	0.663	0.592
Classification Tree	0.621	0.637	0.607	0.601	0.643
PSM					
Unigrams	0.537	0.541	0.542	0.642	0.437
Bigrams	0.521	0.601	0.487	0.558	0.530
Uni + Bigrams	0.499	0.510	0.496	0.328	0.678

Note. The categories determined by the SM-scale are used as true standard in the classification. The structured and unstructured textual analysis are conducted by using the software LIWC and machine learning algorithm PSM, respectively. PPV and NPV represent the positive predictive value and negative predictive value, respectively.

keywords extracted from the group of HSM appeared more often in the predictors "negemo" (negative emotion) and "percep" (perceptual processes). The group of LSM had obviously more keywords mapped in the predictors of "funct" (function words), "verb", "present" (present tense), "home" and "relig" (religion). These findings demonstrated that the predictors in LIWC and keywords extracted from the text mining could be mutual supplements. To obtain the benefits from both methods, it might be interesting to put the 80 LIWC variables and the extracted keywords altogether in a pool of predictors for an entire pattern exploration in the future study. Further, as mentioned above, the emoticons and Internet slangs were extracted as the robust classifiers to distinguish the low and high SM groups. However, in the current version of LIWC, it is hard to map them into corresponding categories. With the increasing research interests in textual posts on social communication networks, like FB and Twitter, nowadays, it would be recommended to extend the dictionaries in the LIWC to a larger scope which could include the attributes of special Internet-related languages, such as emoticons and Internet slangs.

Besides the positive results, there are some limitations that also merit discussion. First, the current study only focused on the exploration of FB posts and SM skills, without taking other psychological traits of users into account. It would be interesting to extend the research scope in future research, for instance, to investigate whether the FB posts can predict other personality traits, besides SM skills, by using the textual prediction model presented in this study. Secondly, the sample used in the current study was selective on the users who used the FB MyPersonality application. Neither Internet users who do not have FB accounts nor FB users who do not use the MyPersonality application were included in this study. Thus, the result might have bias due to this sampling issue. A third limitation is that the FB posts used in this study were general counts, which were not sufficiently specific to the users' daily posts or comments to friends' status. To more clearly explore the features of different textual messages, i.e., "status updates" and comments to others, greater depth of analysis is needed. Different labels of textual sources (e.g., self-posts, comments to others' status, comments to photos) are recommended to be used in future studies.

In summary, the current study explored the relationship between posts on FB and SM skills and offered clear support for the claim that the textual posts on FB Wall could partially predict the users' SM skills. Both of the LIWC and text mining techniques were proved promising in handling the Internet textual posts. It is important to find that the typical networking language, emoticons and Internet slangs, are robust predictors to classify high and low self-monitors. Current findings also reconfirm the conclusions from the previous studies by Hall and Pennington (2013) that the expressions related to family topics were found more likely used by low self-monitors. In addition, this study emphasized the importance of checking the validity of Internet data and introduced a method to investigate it. We recommend that the researchers always perform a data validation study using the methodology presented here to ensure the Internet data is valid before being processed.

References

Alexa Internet Inc. (2011). *Alexna top 500 global sites*. <http://www.alexa.com/topsites> (03.02.11).

Argamon, S., Dhawle, S., Koppel, M., & Pennebaker, J. W. (2005). Lexical predictors of personality type. In *Proceedings of the 2005 joint annual meeting of the interface and the classification Society of North America*, St. Louis, Missouri, USA.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of an EM-algorithm. *Psychometrika, 46*(4), 443–459.

Bruha, I. (2000). From machine learning to knowledge discovery: Survey of preprocessing and postprocessing. *Intelligent Data Analysis, 4*, 363–374.

Buchanan, T. (2002). Online assessment: Desirable or dangerous? *Professional Psychology: Research and Practice, 33*(2), 148–154.

Celis-Morales, C. A., Perez-Bravo, F., Ibañez, L., Salas, C., Bailey, M. E. S., & Gill, J. M. R. (2012). Objective vs. self-reported physical activity and sedentary time: Effects of measurement method on relationships with risk biomarkers. *PLoS One, 7*(5), e36345. http://dx.doi.org/10.1371/journal.pone.0036345.

Celli, F., Pianesi, F., Stillwell, D. S., & Kosinski, M. (2013). Workshop on computational personality recognition. In *Proceedings of the 7th international AAAI conference on Weblogs and Social Media*, Boston, MA, USA.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology, 41*, 417–440.

Ellison, N. B., Heino, R., & Gibbs, J. (2006). Managing impressions online: Self-presentation processes in the online dating environment. *Journal of Computer-Mediated Communication, 11*(2), 415–441.

Elomaa, T. (1999). The biases of decision tree pruning strategies. *Advances in Intelligent Data Analysis Proceedings, 1642*, 63–74.

Fuglestad, P. T., & Snyder, M. (2009). Self-monitoring. In M. R. Leary & R. H. Hoyle (Eds.), *Handbook of individual differences* (pp. 574–591). New York, NY: Guilford.

Glas, C. A. W. (1998). Detection of differential item functioning using Lagrange multiplier tests. *Statistica Sinica, 8*(3), 647–667.

Glas, C. A. W. (1999). Modification indices for the 2-PL and the nominal response model. *Psychometrika, 64*(3), 273–294.

Glas, C. A. W., & Dagohoy, A. V. (2007). A person fit test for IRT models for polytomous items. *Psychometrika, 72*(2), 159–180.

Glas, C. A. W., & Falcon, J. C. S. (2003). A comparison of item-fit statistics for the three-parameter logistic model. *Applied Psychological Measurement, 27*(2), 87–106.

Glas, C. A. W. (2010). *MIRT. Public domain software program*. <http://www.utwente.nl/gw/omd/Medewerkers/temp_test/mirt_package.zip>, and <http://www.utwente.nl/gw/omd/Medewerkers/temp_test/mirt-manual.pdf>.

Gonzalez, N. (2011). *Facebook marketing statistics, demographics, reports and news*. <http://www.checkfacebook.com> (Accessed 03.02.11).

Gosling, S., Gaddis, S., & Vazire, S. (2007). Personality impressions based on Facebook profiles. In B. Berendt (Ed.), *Authorial analysis. Symposium conducted at the international conference on Weblogs and Social Media*, Boulder, CO, USA.

Hall, J. A., Park, N., Song, H., & Cody, M. J. (2010). Strategic misrepresentation in online dating: The effects of gender, self-monitoring, and personality traits. *Journal of Social and Personal Relationships, 27*, 117–135.

Hall, J. A., & Pennington, N. (2013). Self-monitoring, honesty, and cue use on Facebook: The relationship with user extraversion and conscientiousness. *Computers in Human Behavior, 29*, 1556–1564.

He, Q., Glas, C. A. W., & Veldkamp, B. P. (2014). Assessing impact of differential symptom functioning on Posttraumatic Stress Disorder (PTSD) diagnosis. *International Journal of Methods in Psychiatric Research*. http://dx.doi.org/10.1002/mpr.1417.

He, Q., Veldkamp, B. P., & de Vries, T. (2012). Screening for posttraumatic stress disorder using verbal features in self-narratives: A text mining approach. *Psychiatry Research, 198*(3), 441–447.

He, Q., & Veldkamp, B. P. (2012). Classifying unstructured textual data using the Product Score Model: An alternative text mining algorithm. In T. J. H. M. Eggen & B. P. Veldkamp (Eds.), *Psychometrics in practice at RCEC* (pp. 47–62). Enschede, the Netherlands: RCEC.

Iacobelli, F., Gill, A. J., Nowson, S., & Oberlander, J. (2011). Large scale personality classification of bloggers. In S. D'Mello, A. Graesser, B. Schuller, & J.-C. Martin (Eds.), *Proceddings of affective computing and intelligent interaction* (ACII 2011, Part II) (pp. 568–577), Lecture Notes in Computer Science, 6975. New York, NY: Springer.

John, O. P. (1990). The "Big Five" factor taxonomy: Dimensions of personality in the natural language and in questionnaires. In L. A. Pervin (Ed.), *Handbook of personality theory and research* (pp. 66–100). New York: Guilford Press.

Junco, R. (2013). Comparing actual and self-reported measures of Facebook use. *Computers in Human Behavior, 29*, 626–631.

Jurafsky, D., & Martin, J. H. (2009). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Upper Saddle River, NJ: Pearson Prentice Hall.

Kosinski, M., Stillwell, D. J., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences of the United States of America (PNAS), 110*(15), 5802–5805.

Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. *Informatica, 31*, 249–268.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, CA: Erlbaum.

Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research, 30*(1), 457–500.

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.

Markovikj, D., Gievska, S., Kosinski, M., & Stillwell, D. S. (2013). Mining Facebook data for predictive personality modeling. In *Proceedings of the 7th international AAAI conference on Weblogs and Social Media (ICWSM 2013)*, Boston, MA, USA.

Naglieri, J. A., Drasgow, F., Schmit, M., Handler, L., Prifitera, A., Margolis, A., et al. (2004). Psychological testing on the Internet: New problems, old issues. *American Psychologist, 59*(3), 150–162.

Oakes, M., Gaizauskas, R., Fowkes, H., Jonsson, W. A. V., & Beaulieu, M. (2001). A method based on chi-square test for document classification. In D. H. Kraft, W. B. Croft, D. J. Harper, & J. Zobel (Eds.), *Proceedings of the 24th annual international ACM SIGIR conference on research and development in information retrieval* (pp. 440–441). New York, NY: ACM.

Otten, J. J., Littenberg, B., & Harvey-Berino, J. R. (2010). Relationship between self report and an objective measure of television-viewing time in adults. *Obesity, 18*(6), 1273–1275.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic inquiry and word count*. Mahwah, NJ: Erlbaum.

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology, 77*, 1296–1312.

Porter, M. F. (1980). An algorithm for suffix stripping. *Program: Electronic Library and Information Systems, 14*(3), 130–137.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Danish Institute for Educational Research.

Rosenberg, J., & Egbert, N. (2011). Online impression management: Personality traits and concerns for secondary goals as predictors of self-presentation tactics on Facebook. *Journal of Computer-Mediated Communication, 17*, 1–18.

Ryan, T., & Xenos, S. (2011). Who uses Facebook? An investigation into the relationship between the Big Five, shyness, narcissism, loneliness, and Facebook usage. *Computers in Human Behavior, 27*, 1658–1664.

Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinant of emotional state. *Psychological Review, 69*, 379–399.

Schlenker, B. (2004). Self-presentation. In M. R. Levy & J. P. Tangney (Eds.), *Handbook of self and identity* (pp. 492–518). New York, NY: Guilford.

Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology, 30*(4), 526–537.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psycology, 29*(1), 24–54.

Toma, C., Hancock, J., & Ellison, N. (2008). Separating fact from fiction: An examination of deceptive self-presentation in online dating profiles. *Personality and Social Psychology Bulletin, 34*, 1023–1036.

van Groen, M. M., ten Klooster, P. M., Taal, E., van de Laar, M. A. F. J., & Glas, C. A. W. (2010). Application of the health assessment questionnaire disability index to various rheumatic diseases. *Quality of Life Research, 19*(9), 1255–1263.

von Davier, M., & Rost, J. (1996). Self monitoring – A class variable? In J. Rost & R. Langeheine (Eds.), *Applications of latent trait and latent class models in the social sciences* (pp. 296–305). Münster, Germany: Waxmann.