IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 13, 2020, accepted October 18, 2020, date of publication October 21, 2020, date of current version November 5, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3032850

The Method of How to Predict Weibo Users' Recovery Experience on the Weekend Based on Weibo Big Data

PENG WANG[1,2,3], YU SUN[1], WEIXUAN MENG[1,4], RUNSHENG PAN[1], JUN WANG[1], XIANGPING ZHAN[1], XIAOYUE LI[1], AND DENGHAO ZHANG[2,3]

[1]School of Psychology, Shandong Normal University, Jinan 250358, China
[2]Department of Psychology, Renmin University of China, Beijing 100872, China
[3]Laboratory of the Department of Psychology, Renmin University of China, Beijing 100872, China
[4]The School of Psychology and Cognitive Science, East China Normal University, Shanghai 200062, China

Corresponding author: Weixuan Meng (784453463@qq.com)

This work was supported by the Building World-Class Universities (Disciplines) of Renmin University of China under Grant RUCPSY0017.

ABSTRACT The prevailing "996" overtime phenomenon in China has raised extensive consideration and discussion towards the topic of work-life balance. Following this trend, this study focused on the topic of work recovery experience. Based on Lens Model, we aimed to construct prediction models of weekend recovery experience with individuals' social media footprints, which include their social media posts, behavioral information, and demographic information. We acquired Weibo data and Recovery Experience Questionnaire results from 493 participants and extracted Weibo data features for model training through two methods. As a result, two types of model were constructed: regression models which applied Ridge Regression, LASSO Regression, and Elastic Net; classification models which applied Gradient Boosting Decision Tree, Logistic Regression and Support Vector Machine. For the results of regression models, Pearson correlation coefficients between predicted values and self-reported scores ranged from 0.40 to 0.84; for classification models, F1-score ranged from 0.49 to 0.78. The results showed that individuals' recovery experience on weekends could be predicted by their social media footprints. What is more, the methodology proposed in this study could help organizations to evaluate large groups of employees' work recovery in real-time, which will have further implications for both theoretical and practical purposes.

INDEX TERMS Recovery experience, big data, Weibo, machine learning.

I. INTRODUCTION

In contemporary Chinese society, with the prevalence of the "996" overtime culture, working overtime has become a primary daily stressor for many people. The "996" refers to a situation that an employee has to work from 9 a.m. to 9 p.m., 6 days a week [1]. With "Work 996, Sick ICU" once became a hot topic on Weibo [2], employees' dissatisfaction and resistance towards the "996" work system began to erupt, which also reflected that job stress and burnout has become a serious and common problem in today's society. If these negative consequences of lacking rest are not recovered on time within the appropriate time cycle, they could further affect not only the physical and mental health of employees but also their performance in organizations.

The recovery from work refers to the process of an individual's functional system activated during work recovering and stabilizing at pre-stress levels, which is an essential factor affecting employees' happiness [3]–[5]. For most employees, weekends are valuable opportunities to replenish their physical and mental resources consumed during workdays. Previous research has shown that the level of recovery from work over weekends not only improves employees' Monday's work status but also leads to more positive emotional experiences in the following new week [6]. Therefore, it is essential for employees and organizations to assess recovery from work during weekends reasonably.

According to Sonnentag and Geurts, recovery from work can be assessed in terms of the recovery process and recovery outcomes, and both of them are related to recovery experience [7]. Thus, this study is concerned with the recovery experience, which is the psychological experience associated

The associate editor coordinating the review of this manuscript and approving it for publication was Muhammad Asif.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

with recovery activities during the non-work period [8]. While Sonnentag and Fritz [9] argued that those specific experiences that contribute to emotional repair were the basis of effective recovery processes, they also proposed four meaningful recovery experiences, including psychological detachment, relaxation, mastery experiences, and control.

In the field of recovery experience measurement, the traditional self-reported diary research method remains mainstream [9]–[15]. Participants are generally required to provide data on their feelings, moods, and activities at least once a day for several consecutive days [16]. However, in addition to the common flaws of traditional questionnaires, including lack of accuracy and representativeness, there are also limitations such as cumbersome procedures, difficulty in recruiting participants, and delay of data feedback. These defects make it challenging to conduct real-time assessments of large populations. However, Weibo (http://weibo.com/), a social media platform, which is similar in function to Twitter, offers an excellent opportunity for an alternative approach of psychometric measurements. For recovery experience, a processual psychological experience may be reflected by individuals' social media posts, it could be predicted by ~~the~~ large-scale Weibo text. Meanwhile, the most critical issues of timeliness, scale, and accuracy were addressed.

According to the theory of Lens Model [17], [18], the personal space context of individuals contains some clues that can reflect individual psychological characteristics (such as the furnishings of desk or bedroom), while the digital behavioral footprints of individuals could also reflect their psychological state to a certain extent. At the same time, a considerable number of studies [19], which were driven by advances in big data computing methods, had shown that individuals' psychological characteristics could be assessed and predicted from their digital footprints. These digital footprints would come from Facebook status updates [20]–[22], Twitter messages [23], Flickr online selfies [24], music preferences [25], and even bank account payment information [26]. In China, Weibo has become one of the most popular social media platforms. The abundant data information with accessibility on Weibo makes it an ideal social media platform for examining individual psychological characteristics. [27]. Many researchers have already conducted a series of psychometric studies using Weibo as a source of data [28]. As a result, several high-quality psychological researches were successfully achieved from the perspective of predicting individual personality [27], [29], accessing individuals' mental health status [30], [31], and predicting group social state of mind [32]–[36].

In the view of the importance of recovery experience in maintaining physical and mental health, improving individuals' well-being, job performance and organizational effectiveness [4]–[6], [37], and the limitations of traditional measurement methods used in previous studies of recovery experience, this study proposed a new approach to predict Weibo user's recovery experience on the weekend by applying Weibo text and behavior data with text mining technology. It is expected to provide a scientific means of measurement for subsequent researches in the field of recovery experience.

II. MATERIALS AND METHODS

A. ETHICS APPROVAL

This study was carried out under the recommendations of the World Medical Association's Declaration of Helsinki. The protocol was approved by the Ethics Committee of Shandong Normal University on 5 March 2019. All participants gave their informed consent for inclusion before they participated in this study.

B. PARTICIPANTS

We recruited a total of 1,221 participants through the sample service provided by www.wjx.cn, an online questionnaire survey platform. Considering the gap between students' recovery from academic and employees' recovery from work may affect the result, we excluded 256 self-identified student participants. Besides, we excluded 310 participants who did not respond seriously (e.g., the response time was too short or long; fail to provide valid Weibo ID account). Finally, the questionnaire responses of 493 participants from 162 cities in China were retained. Participants ranged in age from 18 to 58 years old ($M = 20.16$, $SD = 6.25$). The distribution of demographic variables of the sample is shown in Table 1.

C. MEASURES

1) Weibo USER DICTIONARY OF RECOVERY EXPERIENCE

a: VOCABULARY SOURCE

To proceed with the text mining analysis of recovery experience, we first set up a dictionary of recovery experience based on the theory of recovery experience [9], [10], representative Weibo posts, and Text Mind System (Chinese semantic analysis system) [38], [39]. Considering the particularity of Weibo language and ensuring the representativeness and timeliness of selected words, we obtained the original Weibo text of the top 8 bloggers in the year from 1 October 2018 to 1 October 2019 as one of the sources. Besides, the dictionary of Text Mind System was selected as another source. The Text Mind System is a word segment tool developed by Computational Cyber Psychology Lab (CCPL), a research team at the Institute of Psychology of the Chinese Academy of Sciences, which has been proved to be a robust tool in personality and mental health research. The dictionary of Text Mind System contains two parts. The first part is the simplified version of traditional Chinese Linguistic Inquiry and Word Count (SCLIWC) dictionary (converted by CCPL), which has 6,547 words, 263 wildcards, and 73 punctuation marks. The second part contains 5,627 new words added by CCPL based on Weibo data.

b: VOCABULARY SELECTION

After determining the source, group evaluation, and online questionnaire surveys were conducted to make sure that

TABLE 1. Distribution of demographic variables in the sample (N = 493).

Demographic Variables	Categories	Percent (%)
Gender	Male	36.11
	Female	63.89
Education background	Master or above	8.92
	Undergraduate	80.93
	Senior College	8.32
	Technical secondary school	0.41
	High school	1.22
	Middle school	0.20
Marital status	Single	25.76
	No single but unmarried	13.39
	Married without children	8.32
	Married with children	51.93
	Divorced without children	0.20
	Divorced with children	0.41
Professional	Government cadre or civil servant	8.11
	Enterprise manager (Including grassroots, senior and junior management)	48.28
	Professional Staff (e.g., doctor, lawyer, journalist, teacher)	24.75
	General worker (e.g., factory workers, manual workers)	5.68
	Service staff (e.g., salesperson, store clerk, waiter)	3.85
	Self-employed operator and contractor	3.85
	Free-agent	4.87
	Retirement	0.20
	Underemployed	0.61
Last month income (RMB)	More than 10,000	20.69
	5,001-10,000	48.48
	4,001-5,000	15.62
	3,001-4,000	9.13
	2,001-3,000	3.45
	1,001-2,000	1.83
	501-1,000	0.20
	Less than 500	0.61

selected words can reflect the recovery experience. The group evaluation was done by 9 college students from the school of psychology. As for the online survey, a total of 234 participants from 106 cities, including 102 males and 115 females with work experience, answered our survey, and 217 valid responses were obtained. Finally, consistent with Sonnentag and Fritz's theory [9], a work recovery dictionary with 245 words including 63 words in the psychological separation dimension, 83 words in the relaxation dimension, 62 words in the mastery dimension and 37 words in the control dimension was constructed. As a result, this work recovery dictionary covered all four dimensions of work experience.

2) RECOVERY EXPERIENCE ON WEEKEND

The Chinese version of The Recovery Experience on Weekend Scale revised by the research group was adopted [9], [15], [40]. The scale has a total of 16 questions, including four dimensions of psychological detachment (4 items; e.g., "I did not think about work at all on this weekend."), relaxation (4 items, e.g., "I took time for leisure on this weekend"), mastery (4 items, e.g., "I learned new things on this weekend")and control (4 items, e.g., "I felt like I can decide for myself what to do on this weekend"). Participants rated their agreement with the items on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The higher the score is, the higher the level of recovery experience on the weekend is reflected. Cronbach's alpha reliability for this sample was 0.78. According to the definition of reliability standard 0.70, this scale has reached a high reliability level [41]. A confirmatory factor analysis (CFA) using Mplus 7.0 indicated that the scale has good structural validity (χ^2 = 242.88, df = 98, χ^2/df = 2.48, RMSEA = 0.05, CFI = 0.93, TLI = 0.92) [42].

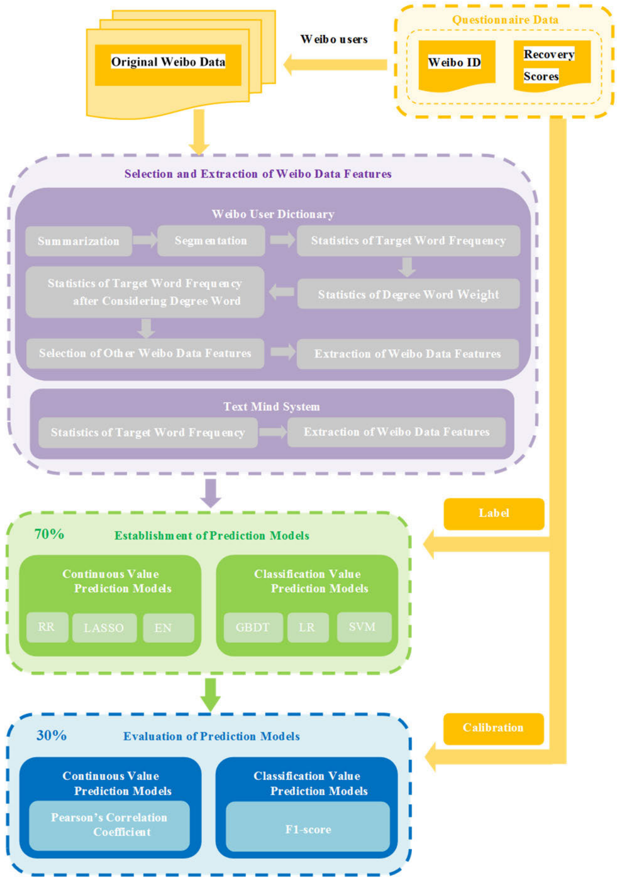

FIGURE 1. Procedures of this stud. *Note.* RR = Ridge Regression; LASSO = Least Absolute Shrinkage and Selection Operator Regression; EN = Elastic Net; GBDT = Gradient Boosting Decision Tree; LR = Logistic Regression; SVM = Support Vector Machine.

D. PROCEDURES

The whole procedures of this study are shown in Figure 1.

1) ACQUISITION OF QUESTIONNAIRE DATA

When issuing the questionnaire through www.wjx.cn, we asked the recruited participants to fill out the questionnaire only between 12:00 and 23:59 on Sunday, and the questionnaire system was closed at other times. We opened the system to collect questionnaire data on the day of 15 December 2019, 29 December 2019, 5 January 2020, and 19 January 2020, respectively.

2) ACQUISITION OF ORIGINAL WEIBO DATA

In our questionnaire, we made it optional for participants to fill in their Weibo ID. Then, according to their Weibo ID, we obtained the user's Weibo data, such as Weibo text, post time, and post channel, through the crawler on Python software (Python Software Foundation, Beaverton, OR, USA). The date on which participants answered the questionnaire was taken as time node, and the Weibo data of 1 week (T1), 2 weeks (T2), 3 weeks (T3), and 4 weeks (T4) before the time node were used for analysis. For example, when a group of participants answered the questionnaire on the date of 15 December 2019, their Weibo data from 9 December 2019 to 15 December 2019 would be collected as T1 data. In this way, we obtained participants' Weibo data in the four weeks from 18 November 2019 to 15 December 2019.

3) SELECTION AND EXTRACTION OF WEIBO DATA FEATURES

After obtaining the participants' original Weibo data, it is necessary to clean and sort the data for further analysis. The following content will use the Weibo data from one participant as an example, and specifically introduce the feature selection and extraction steps.

a: THE MINING OF Weibo TEXTS THROUGH Weibo USER DICTIONARY

Step1: Segmentation of Weibo Data

According to Wang *et al.* [43], it is the basis of further target word frequency counts to split Weibo text by punctuation. This study adopted the method proposed by Hau *et al.* [42], and use ",", ".", ";", "?", "!" to split Weibo text into fragments. The Weibo text posted by participant A in the week from 9 December 2019 to 15 December 2019 could be divided into F_1, F_2, ……, and F_n (a total of n Weibo fragments).

Step2: Statistics of Target Word Frequency

F_1, F_2, ……, and F_n of Weibo fragments were matched to the corresponding files of each target word in the work recovery dictionary through fuzzy matching [43]. For example, the sentence "今天工作忒累了，下了班一定要好好休息一下。" (I am so tired from work today; I must have a good rest after work.) includes two fragments. Where the F_1 includes the target word "工作" (work), the F_1 was matched to the corresponding file of the "工作"(work), and the target word "工作"(work) is recorded to appear once in the F_1 (i.e., count of "工作" is 1). If the same target word appears more than one time in the same Weibo fragment, the corresponding count is recorded as its frequency. It is important to note that a Weibo fragment may contain multiple target words. In the above example, since two target words "下了班" (off duty) and "休息" (rest) appeared in F_2, the F_2 would be saved in the corresponding files of the two target words "下班"(off duty) and "休息"(rest). Then the count of "下班"(off duty) and "休息"(rest) were both 1.

Step3: Weights of Degree Adverb

Since the appearance of the degree adverb may increase or decrease the emotional tendency expressed by the target word, we adopted the Chinese Degree Words used by Wang *et al.* [43] and Han *et al.* [44] in this study. In the example sentence of "今天工作忒累了，下了班一定要好好休息一下"(I am so tired from work today, so I must have a good rest after work), the adverb "忒" (too) appeared in F1 once. Because the weight of the word "忒" (too) in Chinese Degree Words is 2, the degree weight of F1 is recorded as 2.

Step4: Statistics of Target Word Frequency after Considering Degree Adverb Weight

In this step, the target word frequency Count of fuzzy matching of fragment F_1 was multiplied by the weight of degree word. As a result, the target word frequency S_1 was calculated as ($S_1 = \text{Count}_1 * \text{Weight}_1$). Based on the Step1, the frequency of each target word is the sum of the target word frequency S_x after considering degree adverbs. Assume that, in the T1 Weibo dataset of participant A, four Weibo text fragments F_1, F_4, F_6, and F_8 were matched under the target words "工作" (work), the frequency of "工作" (work) in dataset T1 is the sum of the target word frequency of F_1, F_4, F_6, and F_8 fragments, which is the value of $S_1 + S_4 + S_6 + S_8$. Finally, the calculated target word frequency was used as one of the features in the following model training procedure.

Step5: Selection of Other Weibo Data Features

In addition to the features of Weibo text, we also selected the following five features from the existing Weibo data: Total number of posted Weibo during the corresponding time period, the total number of reposted Weibo during the corresponding time period, the channel of Weibo posting, the number of @ in Weibo during the corresponding time period, and the number of "# #" (refers to topics) in Weibo during the corresponding time period. Also, we collected the demographic information of the participants from questionnaires. Considering the participants' gender, age, region, marriage status, educational background, and professional and other information could also be available in Weibo data, we chose these demographic features as the independent variables.

Step6: Dimension reduction of Weibo Data Features

After the previous six steps, we preliminarily selected 260 Weibo data features in total, including 245 target words, 5 other Weibo data features, and 10 demographic features. We used the principal component analysis (PCA) method to reduce the dimension of the independent variables to extract the main features from the vast features of Weibo's big data. Before establishing the prediction models, we also carried out data preprocessing such as discrete data vectorization, feature normalization, and others.

b: THE MINING OF Weibo TEXTS THROUGH TEXT MIND SYSTEM

In addition, to further validate the effectiveness of Weibo user dictionary in predictive model construction, we used an alternative text mining method, the Text Mind System. This text mining approach has been widely used in big data psychology research [18], [27]–[30], [32]–[35], [38], [39].

The Text Mind System includes SCLIWC (the simplified version of traditional Chinese Linguistic Inquiry and Word Count) dictionary, which has 102 features, and software program. Through the Text Mind System, we could decompose each sentence into several phrases. For example, "我今天好高兴" (I'm so happy today) could be broken down into "我" (I), 今天(today), "好" (so) and "高兴" (happy). Then we could classify these words into different features according to their meanings, and finally calculate the percentage of the frequency of words in each feature in the total frequency of words, which can be taken as the linguistic features of users' Weibo texts. We still used the PCA method to extract the main features from the vast features of Weibo's big data.

4) ESTABLISHMENT OF PREDICTION MODELS

In this part, we established regression prediction models and classification prediction models, respectively, to predict Weibo users' recovery experience on the weekend. The regression prediction models could be used to estimate the recovery experience scores of employees on the weekend, and the classification prediction models can identify individuals with high or low recovery experience.

Based on a series of researches that used social media footprints to predict personality [27], subjective well-being [30] and mental health [18], [31], this study chose Ridge Regression, Least Absolute Shrinkage and Selection Operator (LASSO) regression, and Elastic Net algorithms to establish regression prediction models. As for classification models, Gradient Boosting Decision Tree (GBDT), Logistic Regression and Support Vector Machine (SVM) algorithm were chosen. The reason why we chose these algorithms is that they show steady performance in previous practical research. Ridge regression is often used to process high dimensional data, and its prediction effect is more stable than that of multiple regression and stepwise regression. LASSO regression is also used to deal with the selection of variables in higher dimensions. Compared with ridge regression, it can better exclude the influence of irrelevant information and artificial penalty parameters on the prediction model. However, the elastic net, which is more widely used, combines the advantages of ridge regression and LASSO regression [45], [46]. As for classification algorithms, numbers of big data studies have shown that these three algorithms perform well in building classification prediction models [27], [31], [47], [48]

For regression models, we used the participants' score on the Chinese version of The Recovery Experience on Weekend Scale as a label directly; for classification models, considering the sample size of this study and reducing data losses, we took the mean score as the critical value and labeled participants with high or low on weekend recovery experience. Participants with above-average scores were assigned to the high group, otherwise to the low group.

In the model training step, the ratio of the training set to the verification set was set to be 7:3. The training set was selected randomly without putting back. After the preprocessing steps, we trained models with six machine learning algorithms. During the training of regression prediction models, the optimal hyperparameters of ridge regression, LASSO regression, and elastic net were chosen by grid research. As for the training of classification prediction models, the training rate of GBDT algorithm was 0.1 as the default value, and the number of iterations was set as 20 by grid research; the learning rate of logistic regression was set to be 1 as default; the SVM algorithm applied Radial Basis Function (RBF), and the penalty parameter C was set as 1.0.

5) EVALUATION OF PREDICTION MODELS

We used the data from the validation set to evaluate the performance of our models. Pearson's Correlation Coefficient (r) between the predicted values and the self-reported scores of participants' recovery experience was used as an evaluation index for regression models. As for classification models, Precision, Recall, F1-score, and other indicators are often used as evaluation indexes in previous research [49]. Among them, F1-score, which is the harmonic mean of Precision and Recall, was selected to evaluate classification prediction models in this study. A higher F1-score indicates better performance of the prediction model.

$$\mathrm{F1-score} = \frac{2 \times \mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{1}$$

Finally, we compared the effectiveness of the prediction models under different time periods and different machine learning algorithms.

TABLE 2. The descriptive statistics of participants' recovery experience on weekend.

Group	Number	*Mean*	*SD*[1]	*Max*[2]	*Min*[3]
General	493	59.39	7.60	80.00	31.00
Male	175	59.72	7.59	77.00	31.00
Female	315	59.17	7.61	80.00	43.00

Note. [1]SD = standard deviation. 2Max represents the maximum value. 3Min represents the minimum value.

III. RESULTS

A. DESCRIPTIVE STATISTICS

The results of participants' recovery experience on the weekend were shown in Table 2. An independent sample t-test was conducted to compare males and females on the scores of recovery experience on the weekend. The results indicated that there was no significant difference between males and females ($t = 0.73$, $df = 491$, and $p = 0.46$).

TABLE 3. The correlation coefficient between the predicted value and the self-reported score.

Time Period	Machine Learning Algorithms		
	Ridge Regression	LASSO Regression	Elastic Net
T1	**0.66****	**0.52****	0.50**
T2	0.50**	0.40**	0.57**
T3	0.56**	0.51**	0.61**
T4	0.42**	0.45**	**0.70****

Note. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

TABLE 4. The correlation coefficient between the predicted value and the self-reported score (text mind system).

Time Period	Machine Learning Algorithms		
	Ridge Regression	LASSO Regression	Elastic Net
T1	**0.61****	0.61**	**0.84****
T2	0.58**	0.52**	0.71**
T3	0.56**	**0.77****	0.71**
T4	0.45**	0.67**	0.74**

Note. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

B. REGRESSION PREDICTION MODEL

After using the PCA method to reduce the dimension of the features, we further evaluated the correlations between the number of extracted principal components and the performance of models. The specific result was shown as following:

We examined the accuracy of ridge regression prediction models under 1-50 principal components. The results showed that the best prediction performance, $r = 0.66$ (hyperparameter $= 0.6$), was achieved with 16 principal components extracted from the T1 dataset.

We also evaluated the accuracy of the LASSO regression models under 1-50 principal components. The results indicated that the best prediction performance, $r = 0.52$ (hyperparameter $= 0.3$) was achieved with 46 principal components from the T1 dataset.

As for elastic net regression, the best prediction performance, $r = 0.70$ (hyperparameter $= 0.7$), was achieved with 23 principal components from the T4 dataset.

We further compared the effects of different Weibo data periods and different machine learning algorithms on model performance after determining the number of principal components that provided the best performance for each algorithm.

As was shown in Table 3, for regression models, the correlation coefficients between the predicted value and the self-reported score of recovery experience on the weekend ranged from 0.40 to 0.70, and all of them reached significance level of 0.01. In previous social media Big Data + Psychology studies, the range of correlations between users' subjective psychological variables and their social media footprints (such as Weibo text features) are usually 0.30-0.40 [50]. Thus, results are generally believed to be acceptable and significant when the correlation coefficient reaches 0.40 or higher [32].

From the perspective of machine learning algorithms, when building regression models with the Weibo data in the T1 period, ridge regression had the highest accuracy ($r = 0.66, p < 0.01$). The accuracies of the elastic net and ridge regression were close, slightly inferior to the LASSO regression algorithm ($r_{LASSO} = 0.52$, $r_{elastic} = 0.50$, $ps < 0.01$). The accuracy of elastic net regression was the highest when using T2 dataset ($r = 0.57$, $p < 0.01$), followed by ridge regression ($r = 0.50$, $p < 0.01$), and LASSO regression was the lowest ($r = 0.40$, $p < 0.01$). As for the T3 dataset, the accuracy of elastic net regression was still the highest ($r = 0.61, p < 0.01$). The accuracy of ridge regression was second-best ($r = 0.56$, $p < 0.01$), and LASSO regression was the lowest ($r = 0.51, p < 0.01$). When applying T4 dataset, the accuracy of elastic net algorithm was the highest ($r = 0.70, p < 0.01$), followed by LASSO regression ($r = 0.45, p < 0.01$), and the result of ridge regression was the lowest ($r = 0.42, p < 0.01$). From this point of view, the performance of the elastic net algorithm was the best among the three regression algorithms.

From the perspective of the data time period, when LASSO regression was used, the models established with fewer data were more accurate; when elastic net regression was used, the models constructed with more data were more accurate.

In summary, using the data in the T4 period and elastic net regression could build the best prediction model of Weibo users' recovery experience on the weekend. Based on previous studies of using social media footprints to predict individuals' psychological states [29], [31], [32], the result of $r = 0.70$ indicated that the regression procedure we deployed can be used to predict Weibo users' recovery experience on the weekend well.

We also adopted the Text Mind system, a common Chinese text mining method, to verify the predictive effect of regression models (see Table 4). The correlation coefficients ranged from 0.45 to 0.84, which indicated that the models with Text Mind System performed better than the Weibo user dictionary generally. What is more, the elastic net algorithm performed best in the T1 period ($r = 0.84$). This result suggests that when it comes to continuous value prediction, we could consider adopting the Text Mind System to mine the text. However, the performance of using the Weibo user dictionary for text mining and then constructing regression models is also acceptable.

C. CLASSIFICATION PREDICTION MODEL

In order to maximize the performance of classification prediction models, the same steps as described for regression models were applied: we traversed the F1-scores of the classification prediction models under the condition of 1-50 principal components for 3 classification algorithms. For GBDT algorithm, the best result (F1-score $= 0.70$) was achieved with 31 principal components extracted from T4 dataset; for logistic regression, the best result (F1-score $=$ 0.73) was achieved with 40 principal components extracted from T1 dataset; and for SVM algorithm, the best result (F1-score $= 0.73$) was achieved with 37 principal components extracted from T1 dataset.

TABLE 5. F1-scores of classification prediction models.

Time Period	Machine Learning Algorithms		
	GBDT[1]	Logistic Regression	SVM[2]
T1	0.67	**0.73**	**0.73**
T2	0.65	0.66	0.67
T3	0.66	0.66	0.65
T4	**0.70**	0.62	0.65

Note. 1 GBDT= Gradient Boosting Decision Tree. 2 SVM= Supported Vector Machine.

TABLE 6. F1-scores of classification prediction models (text mind system).

Time Period	Machine Learning Algorithms		
	GBDT[1]	Logistic Regression	SVM[2]
T1	**0.67**	**0.61**	**0.78**
T2	0.57	0.55	0.60
T3	0.62	0.49	0.50
T4	0.60	0.52	0.56

Note. 1 GBDT= Gradient Boosting Decision Tree. 2 SVM= Supported Vector Machine.

The F1-scores of classification models trained with different time period datasets and algorithms were presented as Table 5. Specifically, when the GDBT algorithm was used to construct classification prediction models, the maximum value of the F1-score was 0.70 in T4, and the minimum value was 0.65 in T2. When the logistic regression algorithm was used to build classification prediction models, the maximum value of the F1-score was 0.73 in T1, and the minimum value was 0.62 in T4. When the classification prediction models were established by using the SVM algorithm, the maximum value of F1-score 0.73 appeared in the T1 period, and the minimum value 0.65 appeared in both T3 and T4 periods.

To sum up, F1-scores ranged from 0.62 to 0.73, which indicated that the models we trained had excellent prediction results. When using the GBDT algorithm to build classification prediction models, it needed 4 weeks of Weibo data to obtain the best prediction effect. However, when logistic regression and SVM algorithms were used to construct classification prediction models, a functional classification effect could be obtained by only using Weibo data in the T1 period.

Table 6 shows the results of building classification models using Text Mind System. F1-scores ranged from 0.49 to 0.78, which indicated that the models with Text Mind System performed less well than the Weibo user dictionary generally. However, the SVM algorithm performed best in the T1 period (F1-score = 0.78). This result indicates that the classification models constructed using the mining approach of Weibo user dictionary well and were acceptable.

IV. DISCUSSION

Based on the digital footprints of Weibo users (text information and behavior information), we proposed an alternative approach to evaluating individuals' weekend recovery experience with the means of machine learning, which can make up for limitations of the traditional self-reported researches.

In order to analyze the features of Weibo text, we constructed a 245-word Weibo user dictionary of recovery experience based on the theoretical dimension of recovery experience. Besides, we selected 5 behavioral features and 10 demographic features of Weibo, together with 245 features of Weibo text as raw features. After determining the raw features of Weibo data, we then extracted features by PCA. Then, three regression algorithms, including ridge regression, LASSO regression, and elastic net, were used to build regression prediction models, while three classification algorithms including GBDT, logistic regression and SVM were used to establish classification prediction models. Finally, models' performances were evaluated, and the effect of datasets from different time periods was compared.

A. THEORETICAL IMPLICATIONS

Firstly, an alternative measurement approach of weekend work recovery experience was proposed in this study, which further enriched the research results in related fields. At present, there is much research done on the topic of work recovery, such as studies on the predictive variables and outcome variables of recovery from work [9]–[13], [15], [40]. However, the traditional approach of recovery experience measurement still needs further improvement because of the drawbacks we discussed previously. Filling this gap, the prediction models developed in this study could use large-scale and real-time Weibo data to predict recovery experience, which solves the drawbacks such as lack of timeliness and limitations of sample scale. With the new approach we proposed, researchers now can measure and study recovery experience from a different perspective, in some more flexible ways.

What is more, the validity of self-report material has become one of the threatens for the validity of psychological and organizational behavior studies. Currently, multitrait-multimethod (MTMM) is one of the solutions to this threaten [51]. However, MTMM relies on reports from others besides participants themselves, which will increase the cost of research and the difficulty of implementation. With the method proposed in this study, researchers could assess individuals' work recovery experience from a new perspective, with a more convenient approach.

Secondly, this study expanded the method of big data research to the field of work recovery. Currently, the field of big data research is booming, and the big data approach is becoming increasingly popular in psychological research, which in turn promotes the improvement of machine learning algorithms to adapt to psychological research. For psychological researchers, what is more important is to propose questions and hypotheses, select the feature system, build models, and analyze results in a theory-driven way, rather than conduct relevant research driven by data entirely. Following this principle, when we built the work recovery dictionary, the process was based on the theory of recovery experience dimension [9]. As a result, the work recovery dictionary was a relatively scientific and comprehensive collection of Weibo words. With this theory-driven dictionary, we have the confidence to say that the selection of Weibo data features became more scientific and accurate, to better link the features of Weibo data with the Weibo user's recovery experience, and then to build more effective prediction models scientifically. Such a move provided some research ideas for the future research of big data psychology. What is more, the work recovery dictionary developed in this research can be used for further big data research in related fields.

B. PRACTICAL IMPLICATIONS

Firstly, according to the research of Fritz *et al.* [6], the recovery experience of employees on weekends directly affects their work performance in the new week. As a rare opportunity to supplement the physical and mental resources consumed during work, how to reasonably and effectively evaluate this recovery status is essential. Therefore, the prediction models of Weibo user's recovery experience on the weekend constructed in this study are of considerable significance to both the individual and the organizational level.

Furthermore, under the situation of the rapid development and broad application of the Internet, massive social media big data provides an excellent opportunity for the development of traditional psychological research methods. When dealing with large-scale data, the critical issues of timeliness and sample scale in psychological research could be solved by using machine learning strategies, which are more efficient and convenient than manual operations. Recently, Li *et al.* have used the preconstructed prediction models to make large-scale and real-time predictions of the mental health state of Wuhan residents during the COVID-19 epidemic [35]. To some extent, the prediction models constructed in this study could assist organizations in making real-time measurements on the recovery experience of large-scale employees, which has specific practical implications for enterprises and other organizations.

C. LIMITATIONS AND FUTURE PROSPECTS

Firstly, the sample size of this study can be further increased. Future research can focus on the expansion of sample size and the enlargement of the data feature system. Among the vast amount of Weibo data, there must be more features that can reflect users' psychological tendencies. In the future, researchers need to focus on extracting features from massive Weibo data to build more reliable psychological prediction models and test the performance of the prediction models among a more massive range of Weibo users.

Secondly, restricted by the terms of Weibo platform, we failed to obtain the data of Weibo users' Like behavior, which has been verified many times in other social media (such as Facebook) studies and found to be able to effectively predict users' psychological characteristics such as personality traits [20], [22]. The number and category of Like are ''behavioral residues'', which could well represent the psychological inclination and attitude of social media users. It is expected that future research could take this as a breakthrough to expand the scope of recovery from work research using Weibo big data.

Thirdly, we adopted the method based on a semantic dictionary to extract the features of Weibo text, but the mining of Weibo text is not limited to this method. For example, Schwartz *et al.* used Differential Language Analysis (DLA) to conduct open word analysis to extract Facebook text vocabulary and topic features and obtained a better prediction effect [52]. In future research, we could try to use various methods in the step of feature extraction, and compare their similarities and differences. On the premise of a larger data scale, we also could consider a deep learning algorithm for more accurate results.

Last but not least, participants' subjective self-reported questionnaire scores were used as labels in model training. It is undeniable that there are some limitations in using self-reported data as a criterion to verify models. In future studies, the method of combining self-reported with other evaluations (such as colleagues or family members reported) could be considered, or objective characteristics such as work performance can be used as a criterion to verify prediction models, to evaluate the recovery experience on weekend of Weibo users from a more comprehensive and objective perspective.

V. CONCLUSION

This study mainly draws the following conclusions: (1) It was practical and feasible to predict Weibo users' recovery experience on the weekend through the features of social media footprints. (2) The best-performed regression model was the one established by combining one weeks of Weibo data before the target weekend with elastic net regression and Text Mind System. (3) The best-performed classification prediction model was the one using one week of Weibo data before the target weekend with SVM algorithm and Text Mind System.

ACKNOWLEDGMENT

(Peng Wang, Runsheng Pan, Jun Wang, Xiangping Zhan, Xiaoyue Li, and Denghao Zhang are co-first authors.)

REFERENCES

[1] J. Tang and M. Fang, ''How Internet businesses are responding to the 996 phenomena,'' *Manager' J.*, no. 8, pp. 90–91, Aug. 2019. [Online]. Available: http://www.cnki.com.cn/Article/CJFDTotal-GLZJ201908035.htm

[2] C. Kong, *Do Not Let the '996' Overtime Culture Overwhelms the Sound of the National Law*. Yunnan Economic Daily, Apr. 2019, p. A01.

[3] S. A. Geurts and S. Sonnentag, ''Recovery as an explanatory mechanism in the relation between acute stress reactions and chronic health impairment,'' *Scandin. J. Work Environ. Health*, vol. 32, no. 6, pp. 482–492, Dec. 2006, doi: 10.5271/sjweh.1053.

[4] S. Sonnentag, L. Venz, and A. Casper, ''Advances in recovery research: What have we learned? what should be done next?'' *J. Occupational Health Psychol.*, vol. 22, no. 3, pp. 1–16, 2017, doi: 10.1037/ocp0000079.

[5] K. Wentz, K. Gyllensten, J. K. Sluiter, and M. Hagberg, ''Need for recovery in relation to effort from work and health in four occupations,'' *Int. Arch. Occupational Environ. Health*, vol. 93, no. 2, pp. 243–259, Feb. 2020, doi: 10.1007/s00420-019-01476-7.

[6] C. Fritz, S. Sonnentag, P. E. Spector, and J. A. McInroe, ''The weekend matters: Relationships between stress recovery and affective experiences,'' *J. Organizational Behav.*, vol. 31, no. 8, pp. 1137–1162, Nov. 2010, doi: 10.1002/job.672.

[7] S. Sonnentag and S. A. E. Geurts, "Methodological issues in recovery research," in *Current Perspectives on Job-Stress Recovery: Research in Occupational Stress and Well-Being*, S. Sonnentag, P. L. Perrewé, and D. C. Ganster Eds. Oxford, U.K.: Emerald Publishing Group, 2009, pp. 1–36.

[8] S. Sonnentag and E. Natter, "Flight attendants' daily recovery from work: Is there no place like home?" *Int. J. Stress Manage.*, vol. 11, no. 4, pp. 366–391, 2004, doi: 10.1037/1072-5245.11.4.366.

[9] S. Sonnentag and C. Fritz, "The recovery experience questionnaire: Development and validation of a measure for assessing recuperation and unwinding from work," *J. Occupational Health Psychol.*, vol. 12, no. 3, pp. 204–221, 2007, doi: 10.1037/1076-8998.12.3.204.

[10] F. Gao, Q. Ding, F. Wang, and P. Wang, "Recovery from work: A necessary process helps regain the good work condition," *J. Psychol. Sci.*, vol. 39, no. 1, pp. 207–213, 2016, doi: 10.16719/j.cnki.1671-6981.20160131.

[11] M. Molino, C. G. Cortese, A. B. Bakker, and C. Ghislieri, "Do recovery experiences moderate the relationship between workload and work-family conflict?" *Career Develop. Int.*, vol. 20, no. 7, pp. 686–702, Nov. 2015, doi: 10.1108/CDI-01-2015-0011.

[12] A. I. Sanz-Vergel, J. Sebastián, A. Rodríguez-Muñoz, E. Garrosa, B. Moreno-Jiménez, and S. Sonnentag, "Adaptación del 'cuestionario de experiencias de recuperación' a una muestra española [adaptation of the 'recovery experience questionnaire' in a Spanish sample]," *Psicothema*, vol. 22, no. 4, pp. 990–996, 2010.

[13] A. Shimazu, S. Sonnentag, K. Kubota, and N. Kawakami, "Validation of the Japanese version of the recovery experience questionnaire," *J. Occupational Health*, vol. 54, no. 3, pp. 196–205, May 2012, doi: 10.1539/joh.11-0220-oa.

[14] S. Sonnentag, "Psychological detachment from work during leisure time," *Current Directions Psychol. Sci.*, vol. 21, no. 2, pp. 114–118, Apr. 2012, doi: 10.1177/0963721411434979.

[15] F. Wang, "The interaction between recovery experience and work engagement of primary school teachers: The role of A/B personality," M.S. thesis, School Psychol., Shandong Normal Univ., Ji'nan, China, 2017.

[16] E. Demerouti, A. B. Bakker, S. A. E. Geurts, and T. W. Taris, "Daily recovery from work-related effort during non-work time," *Res. Occupational Stress Well Being*, vol. 7, pp. 85–123, May 2009, doi: 10.1108/S1479-3555(2009)0000007006.

[17] A. J. Watson and E. Brunswik, "Perception and the representative design of psychological experiments," *Phil. Quart.*, vol. 8, no. 33, pp. 382–383, 1958, doi: 10.2307/2216617.

[18] T. Zhu, *Psychological Research and Applications in the Age of Big Data*. Beijing, China: Science Press, 2016.

[19] M. Kosinski, S. C. Matz, S. D. Gosling, V. Popov, and D. Stillwell, "Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines," *Amer. Psychologist*, vol. 70, no. 6, pp. 543–556, Sep. 2015, doi: 10.1037/a0039210.

[20] M. Kosinski, D. Stillwell, and T. Graepel, "Private traits and attributes are predictable from digital records of human behavior," *Proc. Nat. Acad. Sci. USA*, vol. 110, no. 15, pp. 5802–5805, Apr. 2013, doi: 10.1073/pnas.1218772110.

[21] G. Park, H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, M. Kosinski, D. J. Stillwell, L. H. Ungar, and M. E. P. Seligman, "Automatic personality assessment through social media language," *J. Personality Social Psychol.*, vol. 108, no. 6, pp. 934–952, Jun. 2015, doi: 10.1037/pspp0000020.

[22] W. Youyou, M. Kosinski, and D. Stillwell, "Computer-based personality judgments are more accurate than those made by humans," *Proc. Nat. Acad. Sci. USA*, vol. 112, no. 4, pp. 1036–1040, Jan. 2015, doi: 10.1073/pnas.1418680112.

[23] D. Quercia, M. Kosinski, D. Stillwell, and J. Crowcroft, "Our Twitter profiles, our selves: Predicting personality with Twitter," in *Proc. IEEE 3rd Int. Conf. Privacy, Secur., Risk Trust IEEE 3rd Int. Conf. Social Comput.*, Boston, MA, USA, Oct. 2011, pp. 9–11, doi: 10.1109/PASSAT/SocialCom.2011.26.

[24] C. Segalin, A. Perina, M. Cristani, and A. Vinciarelli, "The pictures we like are our image: Continuous mapping of favorite pictures into self-assessed and attributed personality traits," *IEEE Trans. Affect. Comput.*, vol. 8, no. 2, pp. 268–285, Apr. 2017, doi: 10.1109/TAFFC.2016.2516994.

[25] G. Nave, J. Minxha, D. M. Greenberg, M. Kosinski, D. Stillwell, and J. Rentfrow, "Musical preferences predict personality: Evidence from active listening and Facebook likes," *Psychol. Sci.*, vol. 29, no. 7, pp. 1145–1158, Jul. 2018, doi: 10.1177/0956797618761659.

[26] J. J. Gladstone, S. C. Matz, and A. Lemaire, "Can psychological traits be inferred from spending? Evidence from transaction data," *Psychol. Sci.*, vol. 30, no. 7, pp. 1087–1096, Jul. 2019, doi: 10.1177/0956797619849435.

[27] L. Li, A. Li, B. Hao, Z. Guan, and T. Zhu, "Predicting active users' personality based on micro-blogging behaviors," *PLoS ONE*, vol. 9, no. 1, Jan. 2014, Art. no. e84997, doi: 10.1371/journal.pone.0084997.

[28] T. Zhu, J. Wang, N. Zhao, and X. Liu, "Psychological research changes in the age of big data," *J. Xinjiang Normal Univ. (Philosophy Social Sci.)*, vol. 36, no. 4, pp. 100–107, 2015, doi: 10.14100/j.cnki.65-1039/g4.2015.04.011.

[29] S. Bai, B. Hao, A. Li, D. Nie, and T. Zhu, "Depression and anxiety prediction on microblogs," *J. Univ. Chin. Acad. Sci.*, vol. 31, no. 6, pp. 814–820, 2014, doi: 10.1111/j.1365-2958.1991.tb00806.x.

[30] B. Hao, L. Lin, G. Rui, A. Li, and T. Zhu, "Sensing subjective well-being from social media," presented at the 10th Int. Conf. Active Media Technol., Warsaw, Poland, 2014.

[31] Q. Hu, "Identification of Internet users mental health based on sina microblog," M.S. thesis, School Comput. Inf. Eng., Henan Univ., Zhengzhou, China, 2015.

[32] Y. Zhou, X. Wang, S. Bai, N. Zhao, and T. Zhu, "Identifying social attitude based on network behavior," *Bull. Chin. Acad. Sci.*, vol. 32, no. 2, pp. 188–195, 2017, doi: 10.16418/j.issn.1000-3045.2017.02.009.

[33] B. Zhang, Y. Zhou, N. Zhao, and T. Zhu, "Social mentality analysis based on weibo big data—The case of 'Qingdao sky-high price shrimp,'" *Cities Disaster Reduction*, no. 2, pp. 26–29, 2018. [Online]. Available: https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CJFD&filename=CSJZ201802006, doi: doi: CNKI:SUN:CSJZ.0.2018-02-006.

[34] J. Sun, B. Zhang, and T. Zhu, "Government satisfaction assessment based on social media big data," *China Party Government Cadre Forum*, no. 2, pp. 75–77, Feb. 2017, doi: 10.14117/j.cnki.cn11-3331/d.2017.02.020.

[35] S. Li, Y. Wang, J. Xue, N. Zhao, and T. Zhu, "The impact of COVID-19 epidemic declaration on psychological consequences: A study on active Weibo users," *Int. J. Environ. Res. Public Health*, vol. 17, no. 6, p. 2032, Mar. 2020, doi: 10.3390/ijerph17062032.

[36] X. Han, J. Wang, M. Zhang, and X. Wang, "Using social media to mine and analyze public opinion related to COVID-19 in China," *Int. J. Environ. Res. Public Health*, vol. 17, no. 8, p. 2788, Apr. 2020, doi: 10.3390/ijerph17082788.

[37] B. Parkinson and P. Totterdell, "Classifying affect-regulation strategies," *Cognition Emotion*, vol. 13, no. 3, pp. 277–303, May 1999, doi: 10.1080/026999399379285.

[38] G. Rui, B. Hao, L. He, Y. Gao, and T. Zhu, "Developing simplified Chinese psychological linguistic analysis dictionary for microblog," in *Proc. Int. Conf. Brain Health Informat.*, 2013, pp. 359–368, doi: 10.1007/978-3-319-02753-1_36.

[39] N. Zhao, D. Jiao, S. Bai, and T. Zhu, "Evaluating the validity of simplified Chinese version of LIWC in detecting psychological expressions in short texts on social network services," *PLoS ONE*, vol. 11, no. 6, pp. 1–15, 2016, doi: 10.1371/journal.pone.0157947.

[40] P. Chu, "The relationship between activities and academic pressure recovery of high school students: The effect of recovery experience and pleasurable experience," M.S. thesis, School Psychol., Shandong Normal Univ., Ji'nan, China, 2018.

[41] Q. Bo, "Validity of the Chinese version of the perceived social self-efficacy scale for in-service staff," M.S. thesis, School Psychol. Cogn. Sci., East China Normal Univ., Shanghai, China, 2011.

[42] K.-T. Hau, Z. Wen, and Z. Cheng, *Structural Equation Model and Its Applications*. Beijing, China: Educational Science Publishing House, 2004.

[43] P. Wang, X. Li, X. Zhan, Y. Zhang, Y. Yan, and W. Meng, "Building consumer confidence index based on social media big data," *Hum. Behav. Emerg. Technol.*, vol. 1, no. 3, pp. 261–268, Jul. 2019, doi: 10.1002/hbe2.156.

[44] Z. Han, Y. Zhang, H. Zhang, Y. Wan, and J. Huang, "On effective short text tendency classification algorithm for Chinese microblogging," *Comput. Appl. Softw.*, vol. 29, no. 10, pp. 89–93, 2012, doi: 10.3969/j.issn.1000-386x.2012.10.025.

[45] H. Zou and T. Hastie, "Addendum: Regularization and variable selection via the elastic net," *J. Roy. Stat. Soc.*, vol. 67, no. 5, p. 768, Nov. 2005, doi: 10.2307/3647619.

[46] J. E. Yoo, "TIMSS 2011 student and teacher predictors for mathematics achievement explored and identified via elastic net," *Frontiers Psychol.*, vol. 9, pp. 317–327, Mar. 2018, doi: 10.3389/fpsyg.2018.00317.

[47] H. Li, *Statistical Learning Methods*. Beijing, China: Tsinghua Univ. Press, 2012.

[48] Y. Wang, "Research and application of visualization algorithms based on medical big data," M.S. thesis, School Comput. Sci. Technol., Tiangong Univ., Tianjin, China, 2018.

[49] Y. Zhou, L. Zhang, X. Liu, Z. Zhang, S. Bai, and T. Zhu, "Predicting the trends of social events on Chinese social media," *Cyberpsychol. Behav. Social Netw.*, vol. 20, no. 9, pp. 533–539, 2017, doi: 10.1089/cyber.2017.0142.

[50] G. J. Meyer, S. E. Finn, L. D. Eyde, G. G. Kay, K. L. Moreland, R. R. Dies, E. J. Eisman, T. W. Kubiszyn, and G. M. Reed, "Psychological testing and psychological assessment: A review of evidence and issues," *Amer. Psychologist*, vol. 56, no. 2, pp. 128–165, 2001, doi: 10.1037/0003-066X.56.2.128.

[51] H. Dai and F. Zhang, *Psychological and Educational Measurement*. Guangzhou, China: Jinan Univ. Press, 2018.

[52] H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, L. Dziurzynski, S. M. Ramones, M. Agrawal, A. Shah, M. Kosinski, D. Stillwell, M. E. P. Seligman, and L. H. Ungar, "Personality, gender, and age in the language of social media: The open-vocabulary approach," *PLoS ONE*, vol. 8, no. 9, Sep. 2013, Art. no. e73791, doi: 10.1371/journal.pone.0073791.

PENG WANG was born in Shandong, China, in 1978. He received the Ph.D. degree in psychological statistics and measurement from Jiangxi Normal University, Nanchang, Jiangxi, China, in 2011.

He was engaged in data mining research at Tsinghua University as an Advanced Visiting Scholar, in 2012. Since 2018, he has been a Professor with the School of Psychology, Shandong Normal University. His research interests include big data psychology, modern measurement theory, and career planning. He is currently working on integrating career fields with big data. He has published various articles and chapters on these subjects, including the book *Looking at the World With Big Data: Middle School Students and Big Data Culture* (Wang 2018).

Prof. Wang was awarded the Young Talents of Dongyue Scholars at Shandong Normal University, for the period of 2018–2023. He is the Director of the Shandong Career Planning and Guiding Committee (SCPGC) and the Standing Director of the New Basic Career Education Research Center, Shandong Normal University.

YU SUN was born in Linyi, Shandong, China, in 1996. He received the B.S. degree from the Shandong University of Technology, Zibo, Shandong, in 2018. He is currently pursuing the master's degree with Shandong Normal University under the supervision of Prof. Peng Wang.

His current research interests include cyberpsychology and big data psychology.

WEIXUAN MENG was born in Linyi, Shandong, China, in 1995. She received the B.S. degree in psychology and the M.S. degree in applied psychology from Shandong Normal University, Jinan, Shandong, in 2017 and 2020, respectively. She is currently pursuing the Ph.D. degree with East China Normal University.

Since 2017, she has been with the Career Big Data Research Group under the guidance of Prof. Peng Wang and explored how to predict recovery through Weibo data. She has published a chapter on the subject, the book *Looking at the World With Big Data: Middle School Students and Big Data Culture* (Wang 2018). Her research interests include psychology of big data on Weibo, text mining, and recovery experience.

RUNSHENG PAN is currently pursuing the master's degree in applied psychology with Shandong Normal University under the guidance of Prof. Peng Wang. With a passion for psychology, currently, his research interests include pathological Internet use and adaptation of college students.

JUN WANG received the B.S. degree from Qingdao University, China, in 2016. He is currently pursuing the master's degree with Shandong Normal University. The research direction is big data psychology, which mainly uses the method of big data to collect, process, and analyze data to increase the understanding of social phenomena and public psychology. His research interest includes the psychology of big data. He followed his tutor Prof. Peng Wang to participate in several national and provincial projects.

XIANGPING ZHAN received the B.S. degree from Southwest University, China, in 2018. She is currently pursuing the master's degree with Shandong Normal University. Her research interests include big data psychology and text mining. She is also interested in big data method as a research tool, such as Weibo and WeChat in China, in analyzing people's online psychology and behavior through social media. She followed her tutor Prof. Peng Wang to carry on big data research.

XIAOYUE LI received the B.S. and M.S. degrees in applied psychology from Shandong Normal University, Jinan, Shandong, China, in 2016.

Her research direction is talent assessment and psychological measurement. She is currently engaged in talent evaluation and talent management, responsible for the preparation of a psychological scale. During the master's degree study, she researched computing the consumer confidence index (CCI) through Weibo, and some progress has been made in the study of her master's graduate article. This study took microblog text as the original data, built a consumer confidence index user dictionary, and calculated the consumer confidence index through different experimental conditions of the target words' frequency analysis.

DENGHAO ZHANG received the M.S. degree in management from Northwestern University, Xi'an, Shanxi, China, in 2000, and the Ph.D. degree in psychology from Peking University, Beijing, China, in 2008.

He is currently working as an Associate Professor at the Renmin University of China. His research interests include social psychology, personality psychology, and national psychology. His research focused on social exclusion, positive personality traits, and personality and mental health.