



Article

A Method for Identifying the Mood States of Social Network Users Based on Cyber Psychometrics

Weijun Wang ^{1,*}, Ying Li ^{1,2}, Yinghui Huang ^{1,2}, Hui Liu ^{1,2} and Tingting Zhang ^{1,2}

¹ Key Laboratory of Adolescent Cyberpsychology and Behavior, Ministry of Education, Central China Normal University, Wuhan 430079, China; yingli@mails.ccnu.edu.cn (Y.L.); yinghui0121@gmail.com (Y.H.); hlccnu@mails.ccnu.edu.cn (H.L.); zhangtingting@mails.ccnu.edu.cn (T.Z.)

² School of Information Management, Central China Normal University, Wuhan 430079, China

* Correspondence: wangwj@mail.ccnu.edu.cn; Tel.: +86-027-6786-7649

Academic Editor: Nicola Lettieri

Received: 19 April 2017; Accepted: 12 June 2017; Published: 16 June 2017

Abstract: Analyzing people's opinions, attitudes, sentiments, and emotions based on user-generated content (UGC) is feasible for identifying the psychological characteristics of social network users. However, most studies focus on identifying the sentiments carried in the micro-blogging text and there is no ideal calculation method for users' real emotional states. In this study, the Profile of Mood State (POMS) is used to characterize users' real mood states and a regression model is built based on cyber psychometrics and a multitask method. Features of users' online behavior are selected through structured statistics and unstructured text. Results of the correlation analysis of different features demonstrate that users' real mood states are not only characterized by the messages expressed through texts, but also correlate with statistical features of online behavior. The sentiment-related features in different timespans indicate different correlations with the real mood state. The comparison among various regression algorithms suggests that the multitask learning method outperforms other algorithms in root-mean-square error and error ratio. Therefore, this cyber psychometrics method based on multitask learning that integrates structural features and temporal emotional information could effectively obtain users' real mood states and could be applied in further psychological measurements and predictions.

Keywords: mood states; cyber psychometrics; Profile of Mood State (POMS); sentiment analysis; microblog

1. Introduction

In the psychological field of emotion science, emotions play a crucial role in people's decision-making [1]. Emotions constitute potent, pervasive, predictable, harmful, and beneficial drivers of judgment and decisions through multiple mechanisms [2]. Because mood states comprise of a particularly important set of sentimental factors [3], they may have significant effects on cognition control [4], consumer behavior [3], and investment decisions [5]. Differing from immediate emotions, mood state is a long-term, continuous, emotional state and is reflected in the process of cognitive behavior, including individual perception, learning, and decision-making [3,6].

With the prevalence of Web 2.0, user-generated content (UGC) has exploded on social networks, especially on microblogging platforms such as Sina Weibo, Facebook, and Twitter, which provide data sources for social studies. As the technology of sentiment analysis and natural language processing (NLP) has gradually developed, many researchers have investigated sentiment analysis, applying it to opinion mining [7,8] and prediction of stock performance [5]. Research attempts to acquire people's emotion or mood information by analyzing text data on social networks and extracting emotional words [9]. However, due to the anonymity of the internet, the self-disclosure of social

network users is different from that in real life [10], which means the emotion expressed on the internet may not be in agreement with the users' real sentiments. Therefore, this study considered a stable and continuous emotional variable, mood state, trying to build a model to identify the mood state of social network users automatically. The mood state was defined as an indirect evaluation state that allows one to explain things and produce corresponding behaviors for the time being, according to their emotional content [11]. The Profile of Mood State (POMS) is a widely-used scale for the measurement of mood state and the abbreviated POMS used in this study consists of seven mood scales, measuring different aspects of psychological health of individuals [12]. The POMS has been used to study the influence of different mood states on the survival rate of long-term cancer patients [13]. It has been proven that there is a correlation between online behavior and psychological traits through empirical study [14]. In order to automatically identify the mood state of network users, this paper collected users' online behavior, extracted related features and adopted a multitask regression model to build the relationship between mood state and users' online behavior. The online information was used to calculate the psychological variables and this cyber psychometrics method focuses on making linkages from online behavior to psychological traits.

Based on cyber psychometrics, this study proposed a multitask model to automatically calculate the long-term, continuous mood state, and a natural language processing technology was used to build the mood state lexicon in order to extract emotional information. Identifying the mood state of social network users may provide some information for advertisers to improve the recommendation performance. For example, a combination of users' mood with a mood-based music recommendation could possibly make a better effect [15]. Merchants may develop different strategies since people in different moods behave differently with respect to risky decision-making [16,17].

Sentiment analysis techniques have been widely applied to the prediction and observation of political events [18], such as monitoring day-to-day electoral campaigns [19]. Thus, the automated information extraction method proposed by this study could be used to explore or track the preferences of social network users and provide reference information for political decisions as a growing number of citizens choose to express opinions and sentiments online.

On the other hand, this automatic calculation method may be widely applied in predicting other psychological variables or building pictures of users considering the rich information generated by social network users [20]. The remainder of this paper is structured as follows. The related work of sentiment analysis and cyber psychometrics method is discussed in Section 2. The experimental material and method is presented in Section 3. The feature extraction and regression result is described and discussed in Section 4. Finally, the conclusion is presented in Section 5.

2. Related Work

2.1. Sentiment Analysis and Emotion Recognition in Social Networks

With advances in machine learning and the emergence of big data sets, the computational detection of opinions, emotions, and subjectivities in unstructured UGC (from free-formatted texts, reviews, and blogs) has been applied to research over the past decade [21]. The linguistic and machine learning approaches are two of the main methods used to detect sentiments or emotions. The linguistic method uses dictionaries or lexicons that contain pre-determined affective words to calculate the frequency of a word and determine the emotional attributes of text [22]. The machine learning method uses computer algorithms to automatically learn text sentiment, given the trained data set [23]. This study adopted a lexicon-based method to construct the sentiment-related features and a lexicon of mood state was constructed.

In contrast to emotion, which reflects short-term affects and is connected with occurred experience, the mood state reflects medium-term affects [24]. In NLP research, affects, feelings, emotions and sentiments are often considered similarly and used interchangeably [25]. In general, moods differ from emotions, feelings, or sentiments in that they are less specific, less intense, and less likely to

be triggered by a particular stimulus or event [24,26]. They are different in many ways, such as duration and time mode, and the comparison is shown in Table 1. Thus, in general, the mood state is a relative stable variable that may influence one's behavior over a long time and could be reflected in online social behavior.

Table 1. Comparison table of mood and sentiment [24,26].

Emotion Expression	Mood State	Sentiment
Duration	Relatively long	Relatively short
Time mode	Gradual, continuous	Rapid, episodic
Strength	Relatively weak	Relatively strong
Cause	Caused by specific events	Caused by specific events
Function	Shows current state information of users	Shows current state information of context
Orientation	No orientation	Points to specific event

2.2. Cyber Psychometrics Method

Relying on big data, the cyber psychometrics method can help predict the users' psychological characteristics through online behavior and is extensively used by researchers. It is acknowledged that microblogging information could be used to build a picture of users and Wald et al. made a prediction of psychopathy using twitter content [20]. Considering the effect of personality on real life [27], Bai et al. has conducted research to predict the Big-Five personality of Chinese Weibo users based on user behaviors at social network sites using the decision tree method [28]. Zhang et al. used linguistic features to predict suicide probability of Weibo users through linear regression [29]. Golbeck et al. collected information from Twitter and made a prediction of personality using the Gaussian process [30]. Researchers have made use of UGC data to predict psychological variables while the mood state is rarely studied. In addition, when dealing with multiple related tasks, studies always model every prediction task separately [31]. However, psychological variables are always interrelated and a joint prediction method such as multitask learning could improve generalization performance by training tasks together and capturing their intrinsic correlation [32]. Therefore, a multitask regression method was adopted in this study to predict the mood state of Weibo users.

This paper obtained the mood states of social network users through the POMS, then used the cyber psychometrics method to build a model to establish the relationships between online behavior and mood state, and finally calculated the psychological variables automatically based on both the structured information and unstructured text. A lexicon of mood state was constructed in this paper and used to detect the sentiment in unstructured text. The multi-task regression model was adopted to calculate users' mood state automatically.

3. Materials and Methods

3.1. Data

For the assessment of mood states, the POMS has been widely used. It was compiled by McNair et al. in 1971 and was initially used to measure the psychological health of individuals. The POMS consists of 65 items and six mood scales (Tension-Anxiety, Depression-Dejection, Anger-Hostility, Vigor-Activity, Fatigue-Inertia, and Confusion-Bewilderment). R. Grove et al. later developed an abbreviated POMS, which contains 40 items and seven mood scales, including an esteem-related affect, and verified its reliability and validity [12]. Participants rated each item on a five-point answer scale. The score of each mood dimension equals the sum of scores of several particular questions. The total score of abbreviated POMS, called total mood disturbance, is calculated by subtracting the scores of two positive dimensions from the five negative dimensions and adding 100. Therefore, the higher the total score is, the more negative mood one may have.

With the permission of participants, researchers gained access to their Weibo home address and then obtained detailed information and blog messages using a Web crawler. The experiment

was conducted between June 1 and 3, 2016 and comprised 224 users, including 110 males and 114 females. The experiment was conducted in a campus and the participants were college students. The subjects ranged in age from 17 to 28 and there were 27 individuals younger than 20, 101 individuals between 20 and 22, 92 individuals in the range of 23 to 25 years of age, and 4 individuals older than 25. The participants consisted of 72 individuals with a bachelor's degree, 126 individuals with a masters, five PhD degree, and 21 individuals with education below an undergraduate level. The inactive user was defined as someone who had no more than 50 friends and no blog updates within three months prior to the experiment. The inactive users were considered invalid and excluded and 132 active users were selected. The questionnaire result showed that these individuals were under normal psychological state considering that the extreme value of POMS score did not reach the bounds.

The design of this study is shown in Figure 1. The mood states of social network users were obtained through the Profile of Mood State. The structured data such as registration date and number of followers could be captured from the Weibo home page directly and comprised the structural features. The text information of microblogging was processed via a lexicon-based method to generate the sentiment related features. After the feature extraction, a multitask regression model was adopted and trained.

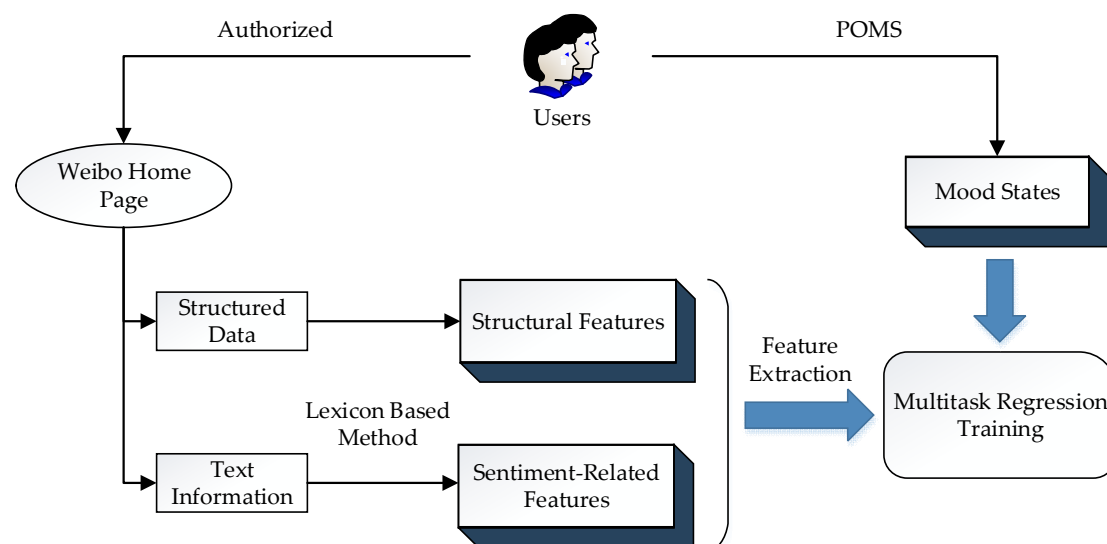


Figure 1. Constructing the model for identifying mood states. POMS: Profile of Mood State.

3.2. Detecting Emotional Features of Weibo Text

There are many methods for handling semantic orientation recognition in natural language processing, including manual marking and the lexicon-based approach. Manual tagging has the highest accuracy but costs a great amount of time and resources. The lexicon-based method is widely used in sentiment analysis and many researchers have adopted approaches like word embedding to bootstrap lexicons for one particular domain or language [33,34]. When constructing a lexicon, a seed word set is created using tools like WordNet, HowNet, and a synonymy thesaurus [35]. The list of words is expanded based on seed words according to synonymous or antonymous relations. In this paper, a mood state lexicon was constructed on the basis of abbreviated POMS. The 40 adjectives in abbreviated POMS are regarded as seed words and expanded via Tongyici Cilin (a Chinese synonym thesaurus) and Word2Vec.

The Tongyici Cilin contains both synonymous and related words. The Harbin Institute of Technology Information Retrieval Laboratory produced the extended version of Tongyici Cilin, composed of 77,343 words [36]. In this thesaurus, words are organized into a five-layer tree structure. In the fifth layer, every category contains several primitive words, which refer to words that have

only one meaning. Every group of primitive words has a code of five layers, which is represented by 11 letters and numbers. Tian et al. has provided a method to calculate the similarity between primitive words in Tongyici Cilin according to their codes [37].

The 40 adjectives in abbreviated POMS are expanded in Tongyici Cilin to obtain seed words. Considering that this study was executed based on a Chinese thesaurus, a similar study was repeated using an English thesaurus to demonstrate the process. The result is shown in Table 2.

Owing to the lack of updates for Tongyici Cilin and the diversity and variability of the network language, the vocabulary used by Weibo users is usually different from that used in written language. Thus, relying only on the general thesaurus cannot cover enough words under the specific network environment.

Table 2. Extracting seed words from an English thesaurus.

POMS Subscales	Items	Seed Words
Tension	Restless, Nervous, On-edge, Tense, Uneasy, Anxious	Agitated, Anxious, Disturbed, Nervous, Edgy, Antsy, Uneasy, Concerned, Tense, Apprehensive, Strained, Impatient, Insecure, Afraid, Distressed
Anger	Peeved, Bitter, Resentful, Grouchy, Angry, Furious, Annoyed	Annoyed, Enraged, Heated, Resentful, Indignant, Outraged, Sullen, Bitter, Furious, Irate, Irritable, Offended, Mad
Fatigue	Worn out, Weary, Bushed, Fatigued, Exhausted	Bored, Bushed, Disgusted, Exhausted, Fatigued, Overworked, Jaded, Drained, Weakened, Disabled, Sleepy, Overtired
Depression	Hopeless, Helpless, Sad, Worthless, Miserable, Discouraged	Desperate, Forlorn, Helpless, Sad, Tragic, Dejected, Discouraging, Gloomy, Pathetic, Wretched, Useless, Despairing, Depressed, Dismayed, Pessimistic
Vigor	Cheerful, Vigorous, Full of pep, Active, Energetic, Lively	Animated, Bright, Optimistic, Lively, Rosy, Chirpy, Cheery, Perky, Merry, Buoyant, Active, Zealous, Potent, Dynamic, Brisk, Spirited, Vital, Peppy, Frisky
Confusion	Bewildered, Forgetful, Confused, Concentration, Uncertain	Addled, Puzzled, Perturbed, Distracted, Dazed, Bewildered, Befuddled, Confused, Disorganized, Forgetful, Inattentive
Esteem	Embarrassed, Ashamed, Proud, Competent, Satisfied	Ashamed, Abashed, Apologetic, Embarrassed, Humbled, Honored, Noble, Satisfied, Appreciative, Content, Competent, Qualified, Decent

The word embedding method can map words into high-dimensional numeric vectors and has been applied in computing similarity, lexical analogy, and machine learning. Word2Vec is an open source natural language processing tool developed by Google in 2013, which translates texts into word vectors. Applying the deep neural network algorithm, Word2Vec simplifies the processing of textual content into numeric operations in the high dimensional vector spaces. The similarity of word vectors represents the semantic similarity of textual words. Based on the semantic relevance of words, this paper utilized Word2Vec to expand the word set from 5 million corpuses on Weibo [38]. Before the training of Word2Vec model, a Chinese segmenting tool, Jieba, was used to cut words and then the stop words were removed. For the train of Word2Vec, here are some parameter settings. The skip-gram model was adopted and hierarchical softmax method was used to train the skip-gram. The window size was set to be 6 and the dimensionality of the feature vectors was 100.

The cosine similarity of word vectors is computed here to select words added into the corpus. Given that $word_1 = (v_1^{word_1}, v_2^{word_1}, \dots, v_m^{word_1})$, and $word_2 = (v_1^{word_2}, v_2^{word_2}, \dots, v_m^{word_2})$, where m is the dimension of word vector, the cosine similarity could be calculated as:

$$sim(word_1, word_2) = \cos \theta = \frac{\sum_{k=1}^m v_k^{word_1} \times v_k^{word_2}}{\sqrt{\sum_{k=1}^m (v_k^{word_1})^2 \times \sum_{k=1}^m (v_k^{word_2})^2}} \quad (1)$$

The numerator represents the dot product of two-word vectors and the denominator represents the modular product of two-word vectors. This research set 0.8 as the threshold of the similarity value, and traversed the corpus to obtain the top-10 most similar words with seed words. After filtering through the threshold, the extended word was added into candidate word sets. Based on the new candidate word sets, the same procedure was carried out repeatedly and the next iteration operated back and forth until there were no new words extracted. Finally, manual verification was adopted to remove words with large deviation or no meaning. As a result, approximately 3000 candidate words were selected and a mood state lexicon containing seven dimensions of mood state words (tension, anger, fatigue, depression, vigor, confusion, and esteem) was constructed.

The semantic distance between words in tweets and the seed words in the lexicon was computed and the score of the word in each mood state dimension could be obtained. For every user, the word set of each tweet was represented as $Logwords = \{word_1, word_2, \dots, word_n\}$, and the seed words set of mood states was $L_{SeedWordsets} = \{L_{s1}, L_{s2}, \dots, L_{s7}\}$, where L_{si} contains a set of seed words in one dimension. By computing the similarity between word in tweets and seed words and finding the maximum similarity in each dimension, a seven-dimension attribute vector for every word was obtained, denoted as $W_i = \{S_{i1}, S_{i2}, \dots, S_{i7}\}$, where $S_{ij} = \max\{sim(word_i, L_{sj})\}$. For example, for the word “restless” in a tweet, calculate the similarity between this word and the words in L_{s1} . If the word “anxious” in L_{s1} has the maximum similarity with “restless”, then this similarity is regarded as the value of the first dimension in the attribute vector, denoted as S_{i1} . Same procedures could be repeated in other dimensions.

For each tweet, this paper defined seven-dimensional variables, $V_{text} \in V^7$, to represent its sentiment-related feature, representing tension, anger, fatigue, depression, confusion, vigor, and esteem, respectively. This value function combined all the word vectors contained in the text and obtained the final score of mood states.

$$V_{text} \in V^7 = \sum_{i=1}^n W_i \quad (2)$$

3.3. The Construction and Evaluation of the Feature Set

This research collected 52 features of the Weibo users and divided them into four categories, which are listed in Table 3. The first category, denoted as D_i , covers basic information about the users, including the length of Weibo ID, gender, registration date, number of labels, and length of personal profile. The second category, denoted by S_j , contains the number of followers, number of follows, and total tweets, demonstrating the social characteristics of users. The third category reveals prosperities of users’ tweets, including the average length of tweets, average times for the use of the @ signal, monthly average, number of tweets, and other related features. These features are denoted as T_m . The last category indicates the sequential sentiment features of Weibo users’ blog text, including the sentiment-related features of tweets in five specific periods, denoted as M_n . The five periods include the last week, month, 3 months, 6 months, and 1 year. Therefore, the users’ feature set can be represented as: $MoodStates = \{Di, Sj, Tm, Mn\}$.

The Pearson correlation coefficient was used in the process of feature extraction and was calculated as follows:

$$\rho = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (3)$$

where ρ represents the degree of correlation between the Weibo users' features and their sentimental scores. If $\rho > 0$, it means that the higher the value of features, the higher would be the user's sentimental score. The larger the value of ρ , the larger the growth rate of their sentimental scores.

Table 3. Features Divided into Four Categories.

Symbol	Explanation	Features Contained
D_i	Basic information of users	Length of Weibo ID, gender, registration date, number of labels and length of personal profile
S_j	Social characteristics of users	Number of followers, number of follows and number of total tweets
T_m	Prosperities of users' tweets	Average length of tweets, average times for the use of the @ signal, monthly average number of tweets and other related features
M_n	Sequential sentiment-related features of blog text	Sentimental vector of tweets in five specific periods including 1 week, 1 month, 3 months, 6 months and 1 year

3.4. Prediction of Mood States of Weibo Users Using Multitask Regression

Conventional network psychometric research usually adopts regression, neural networks and other methods to model every prediction task separately to fit the specific information of each regression task. Recent studies have shown that psychological variables are interrelated and that they play an important role in the joint prediction of different psychological variables [39]. Thus, this study adopted the multitask learning method to improve performance. The multitask learning method not only saved the specific information between tasks, but also integrated the information of multiple tasks to establish a more effective prediction model for the calculation of the different dimensions of mood states.

In the process of predicting the Weibo users' mood states, the multitask regression method was introduced to build the calculation model. Aimed at predicting mood state variables from eight sentimental dimensions (seven sub-dimensions and one total score), this paper set eight learning tasks to build a regression model. In the regression model, the number of regression tasks was denoted as T . For each task t , there was an independent training set $\{(x_{tn}, y_{tn})\}$, where $t = 1, 2, \dots, T, n = 1, 2, \dots, N$, and N is the number of instances, assuming that all tasks have the same number of instances. The learning function is $f_t: R^d \rightarrow R$, and the training set is $\{(x_1^t, y_1^t), (x_2^t, y_2^t), \dots, (x_N^t, y_N^t)\}$. The linear function is $f_t(X) = w^t X$, where w^t represents the model coefficient in task t , and X is the input column vector. The loss function is $L^t(w^t, X, y^t)$, where W is a coefficient matrix with w^t as the row vector and w_k as the column vector so that w_k is the coefficient related to the feature k .

$$W = \begin{bmatrix} w_1^1 & \cdots & w_K^1 \\ \vdots & \ddots & \vdots \\ w_1^T & \cdots & w_K^T \end{bmatrix} \quad (4)$$

To filter the global features, w_k is constrained as follows:

$$\min_w \sum_{t=1}^T L^t(w^t, X, y^t) + \lambda \sum_{k=1}^K \|w_k\|_2^2 \quad (5)$$

The linear least squares fit (LLSF) is used here to make the fitting. The loss function is denoted as:

$$L^t(w^t, X, y^t) = \sum_{n=1}^N (y_n^t - w^t \cdot x_n)^2 \quad (6)$$

Therefore, the objective function is:

$$W = \operatorname{argmin} \sum_{t=1}^T \sum_{n=1}^N (y_n^t - w^t \cdot x_n)^2 + \lambda \sum_{k=1}^K \|w_k\|_2^2 \quad (7)$$

Finally, it can be concluded that:

$$W = \left(\lambda I + \sum_n x_n x_n^T \right)^{-1} \left(\sum_n x_n y_n^T \right) \quad (8)$$

The parameter $\lambda > 0$ is the regularization parameter, controlling the trade-off between the regression loss and the size of the weight vector, as measured by the ℓ_2 norm in Equations (5) and (7) [40,41]. The objective function is smooth and convex and can be minimized by standard method such as coordinate descent [41].

4. Results and Discussion

This paper used correlation analysis to select features that are closely related to mood states in order to establish the feature set. Through the feature extraction, 25 significant features were obtained, including 5 features from the first 3 categories and 20 sentiment-related features from the fourth category. Further, the non-sentimental features contained registration time, the number of users that one follows, the average number of comments, the number of tweets that are reposted by Weibo users, and the average comments of blogs that are reposted by Weibo users. The mood states feature comprises different textual sentiment-related features within 1 week, 1 month, 3 months, 6 months, and 1 year. This indicates that user's mood state is related to both the text expression and stable statistical characteristics. This endorses the psychological conclusion that the user's mood state is a kind of psychological characteristic that has a correlation between the personality and the affect [42,43], considering that studies have shown that the personality could be automatically acquired through users' social network behavior [28]. The result also indicates that up to 80% of the features are emotional features, demonstrating that text expression plays an important role in calculating users' mood state.

Table 4 displays the correlation coefficients between the extracted features and the different dimension of mood states, including the total score of POMS. In addition, 1 week, 1 month, 3 months, 6 months, and 1 year represent the average correlation coefficient of the sentimental score of the text during these periods.

Table 4. Correlation coefficients of selected features and mood states.

Features	Tension	Anger	Fatigue	Depression	Vigor	Confusion	Esteem	Total Score
Registration date	−0.145	−0.084	−0.22 **	−0.089	0.07	−0.074	−0.068	−0.144
Number of follows	0.124	0.135	0.118	0.041	−0.01	0.119	0.04	0.169 *
Average comments	−0.016	−0.166 *	0.131	−0.018	−0.197 **	0.025	−0.14	−0.081
Average reposts of reposted tweets	0.349 ***	0.191 *	0.085	0.175 *	−0.055	0.223 **	−0.006	0.234 **
Average comments of reposted tweets	−0.079	−0.087	−0.187 *	−0.046	0.026	−0.044	−0.027	−0.104
1 week	−0.041	0.044	−0.154 *	−0.15 *	0.130 *	−0.066	0.089 *	−0.044
1 month	−0.043	0.048	−0.158 *	−0.153 *	0.133	−0.061	0.078	−0.044
3 months	−0.057	0.004	−0.145 *	−0.163 *	0.109	−0.035	0.027	−0.067
6 months	−0.033	−0.017	−0.094	−0.167 *	0.068	−0.006	−0.049	−0.067
1 year	0.136	0.162 *	−0.048	−0.007	−0.016	0.09	−0.04	0.077

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

These results show that the average repost number of blogs reposted by Weibo users has significant positive correlation with several mood state dimensions, including tension, anger, depression, confusion and total score, indicating an important non-sentimental feature. The dimension of fatigue and depression are both negatively correlated with several sentimental score of tweets in different time span while other dimensions do not have the same level of correlations.

Figure 2 shows the correlation trend diagram regarding the mood states of tweets within each time span. The correlation coefficient trend of sentimental features within 1 year is obviously different from other time spans. Perhaps this is just because 1 year is too long for users' mood state to be stable. The sentimental features of tweets within 1 year have a much higher correlation with the dimensions of tension and anger. While considering the dimensions of vigor and esteem, the sentimental features of tweets have a higher correlation within a much shorter time span.

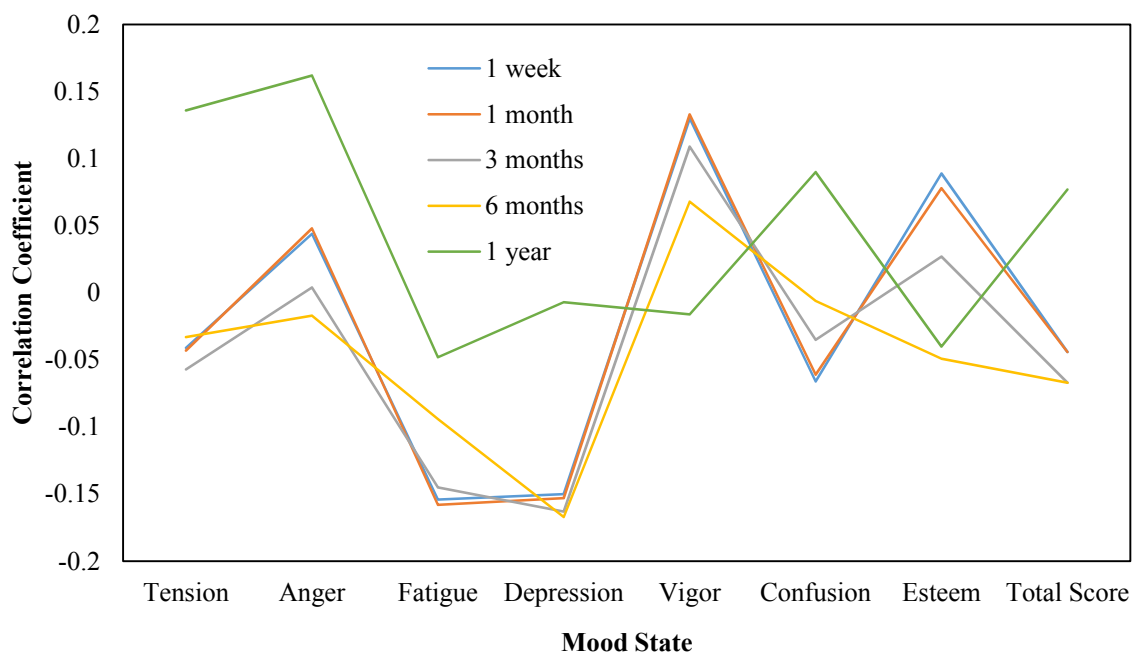


Figure 2. Correlation coefficients of sentimental scores within text during different time spans and real mood states.

This study employed the 10-fold cross-validation to train and validate the model, and the root-mean-square error (RMSE) was used as the evaluation index to evaluate the model. The RMSE is the most common evaluation index in the machine learning regression model, which is also called root-mean-square deviation (RMSD). It is defined as follows:

$$RMSE = \sqrt{\sum_{i=0}^n (y_i - \hat{y}_i)^2} \quad (9)$$

where y_i represents the true value of sample i , $\hat{y}_i$ is the predicted value of sample i and n is the number of samples.

Figure 3 shows the RMSE of the different regression models on the experimental data set. It could be seen that the regression of total score always reach the largest RMSE in each regression model, since total score is a comprehensive index and may cause more error. The Gaussian process has the largest prediction error on each task, followed by the linear regression and Lasso regression. The prediction error of multitask regression is slightly less than that of the Lasso regression method on each task. For the average RMSE of multiple tasks, the values of the Gaussian process, linear regression,

Lasso regression, and multitask regression is 23.919, 12.206, 8.336 and 6.273, respectively, indicating the smallest for the multitask regression model.

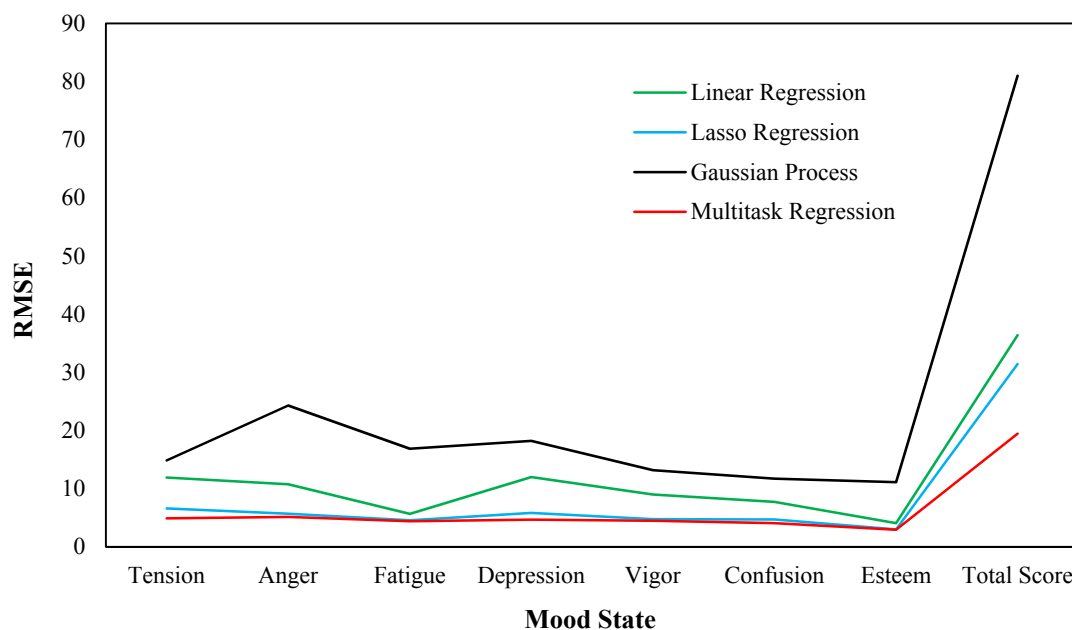


Figure 3. The root-mean-square error (RMSE) of different models.

Table 5 shows the prediction error rate of different algorithms on different tasks. In each task, specifically, for each dimension of mood state, the error rate of multi-task regression is the smallest. Therefore, the accuracy of the multi-task regression method is superior to other algorithms, i.e., the cyber psychometrics produces a good validity and accuracy rate. This demonstrates that multi-task regression could learn the correlation between different dimension of mood state and outperform the single-task regression method [44].

This study has proved the usefulness of this automated identification method and thus, could be applied to identifying other psychological variables or preferences of social network users using the rich UGC data in the social network. Businessmen and policy-makers could also possibly make use of the predictions to support decision-making.

Table 5. The error ratio of different models.

Mood State	Linear Regression	Lasso Regression	Gaussian Regression	Multitask Regression
Tension	0.784	0.434	0.978	0.324
Anger	0.736	0.391	1.661	0.353
Fatigue	0.411	0.331	1.219	0.319
Depression	0.909	0.442	1.380	0.355
Vigor	0.487	0.258	0.713	0.243
Confusion	0.600	0.365	0.908	0.315
Esteem	0.289	0.214	0.790	0.210
Total score	0.583	0.503	1.297	0.312

5. Conclusions

This study proposed cyber psychometrics method to calculate mood states of social network users automatically, adopting natural language processing and the machine learning method. Some conclusions can be drawn through analysis and discussion.

Through the correlation analysis of users' POMS scores and their sentiment expressed using Weibo text, this study found that the user's mood state is not only related to the text expression,

but also to stable statistical characteristics. Users' emotional expressions on the micro-blog text played an important part in the process of obtaining their true mood states. The correlation analysis results also showed that sentiment-related feature sets of different time span have different correlation characteristics with users' mood states, and the emotional features for 1 year evidently differ from those of other timespans.

Compared with other classic single-task machine learning methods, the cyber psychometrics method adopting multi-task regression performed better. The experiment compared the multi-task learning method and the classic single-task regression algorithms, and it was found that the multi-task machine learning method can enhance the performance in RMSE and error ratio by learning the correlation among different dimensions of mood states.

This method could be applied in psychological measurement and the prediction of the behavior of social network users. Such an automatic calculation method may be widely applied in predicting psychological variables and other characteristics of users considering the rapid development of internet, especially the social network. However, there were still some limitations in that the experimental subjects were college students of similar ages, and the characteristics of mood states concerning those with different identities and ages were not considered. In addition, this paper constructed a mood state lexicon (about 3000 words) to extract the users' sequential sentiment-related features. The scale is relatively small and the feature extraction process for the mood states of users under large-scale data requires further research.

Acknowledgments: This study is supported by Program of National Natural Science Fund "Service selection optimization of SaaS based on user preference-aware in cloud computing" (No. 71271099) and the National Natural Science Fund "User preference extraction and interactive personalized recommendation: research based on the screen hot spot" (No. 71571084).

Author Contributions: The study was carried out in collaboration between all authors. Weijun Wang and Yinghui Huang designed the research topic. Ying Li and Yinghui Huang conducted the experiment and wrote the paper. Weijun Wang, Hui Liu and Tingting Zhang examined the experimental data and checked the experimental results. All authors agreed to submission of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Loewenstein, G.; Lerner, J.S. The role of affect in decision making. *Handb. Affect. Sci.* **2003**, *619*, 3.
- Lerner, J.S.; Li, Y.; Valdesolo, P.; Kassam, K.S. Emotion and decision making. *Psychology* **2015**, *66*, 799–823. [[CrossRef](#)] [[PubMed](#)]
- Gardner, M.P. Mood states and consumer behavior: A critical review. *J. Consum. Res.* **1985**, *12*, 281–300. [[CrossRef](#)]
- Rowe, G.; Hirsh, J.B.; Anderson, A.K. Positive affect increases the breadth of attentional selection. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 383–388. [[CrossRef](#)] [[PubMed](#)]
- Ranco, G.; Aleksovski, D.; Caldarelli, G.; Grčar, M.; Mozetič, I. The effects of twitter sentiment on stock price returns. *PLoS ONE* **2015**, *10*, e0138441. [[CrossRef](#)] [[PubMed](#)]
- Martin, B.A.S. The influence of gender on mood effects in advertising. *Psychol. Mark.* **2003**, *20*, 249–273. [[CrossRef](#)]
- Maynard, D.; Gossen, G.; Funk, A.; Fisichella, M. Should i care about your opinion? Detection of opinion interestingness and dynamics in social media. *Future Internet* **2014**, *6*, 457–481. [[CrossRef](#)]
- Piryan, R.; Madhavi, D.; Singh, V. Analytical mapping of opinion mining and sentiment analysis research during 2000–2015. *Inf. Process. Manag.* **2017**, *53*, 122–150. [[CrossRef](#)]
- Mäntylä, M.V.; Graziotin, D.; Kuuttila, M. The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. *arXiv* **2016**, arXiv:1612.01556.
- Nguyen, M.; Bin, Y.S.; Campbell, A. Comparing online and offline self-disclosure: A systematic review. *Cyberpsychol. Behav. Soc. Netw.* **2012**, *15*, 103–111. [[CrossRef](#)] [[PubMed](#)]
- Parkinson, B.; Briner, R.; Reynolds, S.K.; Totterdell, P. *Changing Moods: The Psychology of Mood and Mood Regulation*; Addison Wesley Longman: Harlow, UK, 1996.

12. Grove, J.R.; Prapavessis, H. Preliminary evidence for the reliability and validity of an abbreviated profile of mood states. *Int. J. Sport Psychol.* **1992**, *23*, 93–109.
13. Annunziata, M.A.; Muzzatti, B.; Flaiban, C.; Giovannini, L.; Carlucci, M. Mood states in long-term cancer survivors: An italian descriptive survey. *Support. Care Cancer* **2016**, *24*, 3157–3164. [[CrossRef](#)] [[PubMed](#)]
14. Hamburger, Y.A.; Ben-Artzi, E. The relationship between extraversion and neuroticism and the different uses of the internet. *Comput. Hum. Behav.* **2000**, *16*, 441–449. [[CrossRef](#)]
15. Kim, J.; Lee, S.; Yoo, W. Implementation and analysis of mood-based music recommendation system. In Proceedings of the 15th International Conference on Advanced Communication Technology (ICACT), Pyeongchang, Korea, 27–30 January 2013; pp. 740–743.
16. Buelow, M.T.; Suhr, J.A. Personality characteristics and state mood influence individual deck selections on the iowa gambling task. *Personal. Individ. Differ.* **2013**, *54*, 593–597. [[CrossRef](#)]
17. Morrongiello, B.A.; Stewart, J.; Pope, K.; Pogrebtsova, E.; Boulay, K.-J. Exploring relations between positive mood state and school-age children's risk taking. *J. Pediatr. Psychol.* **2015**, *40*, 406–418. [[CrossRef](#)] [[PubMed](#)]
18. Ceron, A.; Curini, L.; Iacus, S.M.; Porro, G. Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France. *New Media Soc.* **2014**, *16*, 340–358. [[CrossRef](#)]
19. Ceron, A.; Curini, L.; Iacus, S.M. Using sentiment analysis to monitor electoral campaigns: Method matters—Evidence from the United States and Italy. *Soc. Sci. Comput. Rev.* **2015**, *33*, 3–20. [[CrossRef](#)]
20. Wald, R.; Khoshgoftaar, T.M.; Napolitano, A.; Sumner, C. Using twitter content to predict psychopathy. In Proceedings of the 11th International Conference on Machine Learning and Applications (ICMLA), Boca Raton, FL, USA, 12–15 December 2012; pp. 394–401.
21. Yu, Y.; Duan, W.; Cao, Q. The impact of social and conventional media on firm equity value: A sentiment analysis approach. *Decis. Support Syst.* **2013**, *55*, 919–926. [[CrossRef](#)]
22. Taboada, M.; Brooke, J.; Tofiloski, M.; Voll, K.; Stede, M. Lexicon-based methods for sentiment analysis. *Comput. Linguist.* **2011**, *37*, 267–307. [[CrossRef](#)]
23. Tan, C.; Lee, L.; Tang, J.; Jiang, L.; Zhou, M.; Li, P. User-level sentiment analysis incorporating social networks. In Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, 21–24 August 2011; pp. 1397–1405.
24. Siegert, I.; Böck, R.; Wendemuth, A. Modeling users' mood state to improve human-machine-interaction. In *Cognitive Behavioural Systems*; Springer: Berlin, Germany, 2012; pp. 273–279.
25. Munezero, M.D.; Montero, C.S.; Sutinen, E.; Pajunen, J. Are they different? Affect, feeling, emotion, sentiment, and opinion detection in text. *IEEE Trans. Affect. Comput.* **2014**, *5*, 101–111. [[CrossRef](#)]
26. Katsimerou, C.; Redi, J.A.; Heynderickx, I. A computational model for mood recognition. In Proceedings of the International Conference on User Modeling, Adaptation, and Personalization, Aalborg, Denmark, 7–11 July 2014; pp. 122–133.
27. Barrick, M.R.; Mount, M.K. The big five personality dimensions and job performance: A meta-analysis. *Pers. Psychol.* **1991**, *44*, 1–26. [[CrossRef](#)]
28. Bai, S.; Zhu, T.; Cheng, L. Big-five personality prediction based on user behaviors at social network sites. *Comput. Sci.* **2012**, *8*, e2682.
29. Zhang, L.; Huang, X.; Liu, T.; Li, A.; Chen, Z.; Zhu, T. Using linguistic features to estimate suicide probability of Chinese microblog users. *Lect. Notes Comput. Sci.* **2014**, *8944*, 549–559.
30. Golbeck, J.; Robles, C.; Edmondson, M.; Turner, K. Predicting personality from twitter. In Proceedings of the 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT) and Social Computing (SocialCom), Boston, MA, USA, 9–11 October 2011; pp. 149–156.
31. Schwartz, H.A.; Johannes, C.E.; Margaret, L.K.; Lukasz, D.; Stephanie, M.R.; Megha, A.; Achal, S.; Michal, K.; David, S.; Martin, E.P.S.; et al. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS ONE* **2013**, *8*, e73791. [[CrossRef](#)] [[PubMed](#)]
32. Zhang, C.; Zhang, Z. Improving multiview face detection with multi-task deep convolutional neural networks. In Proceedings of the 2014 IEEE Winter Conference on Applications of Computer Vision (WACV), Steamboat Springs, CO, USA, 24–26 March 2014; pp. 1036–1041.
33. Castellucci, G.; Croce, D.; Basili, R. Bootstrapping large scale polarity lexicons through advanced distributional methods. In Proceedings of the Congress of the Italian Association for Artificial Intelligence, Ferrara, Italy, 23–25 September 2015; pp. 329–342.

34. Passaro, L.C.; Pollacci, L.; Lenci, A. Item: A vector space model to bootstrap an italian emotive lexicon. In Proceedings of the Second Italian Conference on Computational Linguistics CLiC-it, Trento, Italy, 3–4 December 2015; pp. 215–220.
35. Mäntylä, M.V.; Novielli, N.; Lanubile, F.; Claes, M.; Kuuttila, M. Bootstrapping a lexicon for emotional arousal in software engineering. In Proceedings of the 14th International Conference on Mining Software Repositories, Buenos Aires, Argentina, 20–28 May 2017.
36. Research Center for Social Computing and Information Retrieval. Available online: <http://ir.hit.edu.cn> (accessed on 1 June 2016).
37. Tian, J.L.; Wei, Z. Words similarity algorithm based on tongyici cilin in semantic web adaptive learning system. *J. Jilin Univ.* **2010**, *28*, 602–608.
38. Natural Language Processing & Information Retrieval Sharing Platform. Available online: <http://www.nlpir.org/?action-viewnews-itemid-299> (accessed on 20 May 2016).
39. Collobert, R.; Weston, J. A unified architecture for natural language processing: Deep neural networks with multitask learning. In Proceedings of the 25th International Conference on Machine learning, Helsinki, Finland, 5–9 July 2008; pp. 160–167.
40. Drakopoulos, G.; Megalooikonomou, V. Regularizing large biosignals with finite differences. In Proceedings of the 2016 7th International Conference on Information, Intelligence, Systems & Applications (IISA), Chalkidiki, Greece, 13–15 July 2016; pp. 1–6.
41. Koh, K.; Kim, S.-J.; Boyd, S. An interior-point method for large-scale l1-regularized logistic regression. *J. Mach. Learn. Res.* **2007**, *8*, 1519–1555.
42. Diener, E.; Lucas, R.E. *Personality and Subjective Well-Being*; Springer: Berlin, German, 2009; pp. 75–102.
43. Lucas, R.E.; Fujita, F. Factors influencing the relation between extraversion and pleasant affect. *J. Personal. Soc. Psychol.* **2000**, *79*, 1039–1056. [[CrossRef](#)]
44. Chen, L.; Zhang, Q.; Li, B. Predicting multiple attributes via relative multi-task learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1027–1034.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).