Predicting Big Five Personality Traits of Microblog Users

Shuotian Bai, Bibo Hao, Ang Li, Sha Yuan, Rui Gao, Tingshao Zhu*
{*Institute of Psychology, University of Chinese Academy of Sciences*},
Chinese Academy of Sciences
Beijing, China
tszhu@psych.ac.cn

***Abstract*—Personality can be defined as a set of characteristics which makes a person unique. The study of personality is of central importance in psychology. Conventional personality assessment is performed by self-report inventory, which costs much manual efforts and cannot be done in real time. To solve these problems, this research aims to measure the Big-Five personality from the usages of Sina Microblog objectively. By conducting a user study with 444 users, this paper proposes multi-task regression and incremental regression algorithms to predict the Big-Five personality from online behaviors. The results indicate that personality can be predicted with a high accuracy through online Microblog usage.**

***Keywords*-personality; Sina microblogs; regression; prediction; Big-Five personality**

I. Introduction

Personality uniquely characterizes an individual, and profoundly influences user's mental status and social behaviors [1]. One of the mostly influential and generally accepted personality theories is the Big-Five personality theory [2][3], which includes five basic traits: Agreeableness (A), Conscientiousness (C), Extraversion (E), Neuroticism (N), and Openness (O), to form human personality [4][5]. From the point of view of behavior, agreeableness reflects the individual behavioral characteristics, such as conducting help, cooperation and sympathy for others. Conscientiousness includes elements of self-discipline, organization and thoroughness of planning, as well as the need for achievement. Extraversion is directly related to social skills, talkative ability and personal charm. Neuroticism reflects the degree of emotional stability, and has a close link to mental health (depression and anxiety). Openness reflects the richness of the individual imagination, aesthetic feelings, degree of dedication, and curiosity about new things [4].

Conventional personality assessments use self-reported inventory [3]. Although the self-report method has a profound theoretical basis, it still has many deficiencies. Firstly, self-report method is inefficient, as it costs a great deal of manpower and material resources in large-scale experiments. Secondly, it takes long time to complete inventory survey, and it cannot be done in real time. The returned questionnaires require manual processing and labeling calculation. Usually, it requires the second party review to ensure the precision of calculation which increases the experimental periods.

To address these issues, much research focuses on the online environment [6][7][8][9], such as microblog sites. The development of information technology provides a new method to conduct the research on personality and behavior. Microblog becomes a popular tool for social communication [10][11]. According to CNNIC, the number of Chinese microblog registered users is 370 million at the end of 2013 [12]. Since its initial launch in December 2005, Sina Microblog has been the leading microblogging service provider in China, and has become the destination for both uploading and downloading resources by status and photo/video-sharing. Besides, Sina Microblog also releases its Application Program Interfaces (APIs) for third party application development (http://open.weibo.com). After acquiring the authorization of APIs, users' behavior data can be downloaded automatically.

Most current research [13][14] pays attention to variation relationship between behavior and personality. The prediction of personality is still in the beginning. Therefore, this paper proposes to predict the big-five personality traits from user's microblogging behaviors. We use Sina Microblog APIs to download the user behaviors data on a large scale. There are 29 Microblogging based features designed to predict and compute user's big-five personality traits. In order to evaluate the quality of the dataset, we compare the predicting effect between incremental regression and linear regression, and prove that our processed dataset is reliable. On the dataset, we show that relevant factors exist between the five dimensions of personality [15], and establish a personality prediction model based on multi-task regression model. In the practical application of machine learning, it is very often to encounter multiple related tasks learning problem. The tasks share the same training set, and relatively independent but with some contact. The traditional idea trains model respectively for each task. This method considers only the information on each task, and ignores the correlation or shared information between tasks. Therefore, in order to improve learning outcomes, the multi-task learning is used in this paper. We predict the user's personality with the model and verify that the multi-task regression model work better than the other models.

The main contributions of this paper are as follows:

1. This paper manage to predict Sina Microblog users' Big Five personality through analyzing online behavioral characteristics. 2. The multi-task regression algorithm is used to predict the Big Five personality, and improve the prediction accuracy.

II. Related Work

Due to the popularity of the network, social networking has become a social reality corresponding to online virtual society. Based on the study of personality on Microblog, most of the foreign research are based on Facebook or Twitter.

Kosinski *et al.* [16] researched on psychological factors on Facebook. Their analysis was based on a dataset of over 58,000 volunteers who provided their Facebook Likes, detailed demographic profiles, and the results of several psychometric tests. They used logistic/linear regression to predict individual psycho-demographic profiles from Likes. The model correctly discriminated between homosexual and heterosexual men. For the personality trait "Openness", prediction accuracy was close to the test-retest accuracy of a standard personality test.

Gosling *et al.* [17], conducted experiments towards the manifestations of personality in Facebook usage. They delivered a mapping between personality and SNS online behaviors, and examined the personality with self-reported Facebook usage and observable profile information. They provided the correlation factor between personality and online behaviors. In their experiment, 11 features were used, including the number of friends, weekly usage and other features. Although their research verified the correlation of online characteristics and personality labels, they did not further establish prediction model of personality, or give the quantitative indicators of personality and online behaviors.

Correa *et al.* [18] researched the relationship between use of social networking media and the user's Big Five personality. By calculating the Pearson correlation factors, they found the use of social networking media was significant positive correlated with openness and extroversion, and negatively related with neuroticism. However, their study did not further consider the prediction of personality based on the social media usage.

Kelly *et al.* [19] worked on the influence of social networking usage based on users' Big Five personality. They invited 219 graduate students as a group, and collected user data through self-presentation recording. Results were meaningful, but the self-presentation of online behavior brought the Subjective orientation of the identification of online behavior. It was not strict enough for setting up the sample set in such a way.

All in all, the current researches on network users personality are more concentrated on correlation analysis of personality and network behavior. The prediction of user's personality based on recording user behavior is still in the beginning. By calling Sina Microblog APIs, this paper automatically obtains users' microblogging data. Then, the paper build a multi-task personality prediction model based on the Big Five personality theory and multi-task learning algorithm.

Multi-task learning method was firstly proposed by Caruana [20]. He used feedforward neural network model, and broke the limit of training one task each time. Training results showed that the connection weights of the input node and hidden layer node contained the sharing of information between the tasks. The specific information for each task was included between the hidden layer nodes and the output nodes. Although the method was not complicated, this way of modeling inspired scholars the idea of multi-task learning, and has been applied in machinery automation, medical diagnostics and other fields [21].

III. Method

An online experiment platform is implemented as shown in Figure 1. It is a Web Access Connection to Sina Microblog, in which participants could log in the platform with Sina Microblog account and authorize the platform to download their Microblog behavior data (profile and status). They are also instructed to complete the Big-Five Inventory to get their personality scores. There are totally 29 features extracted and labeled by the personality scores. Finally, the platform outputs the feature vectors of each participant.

A. Participants

The survey was conducted during April 18, 2012 to May 12, 2012. Participants were invited to log in our platform online and finish the personality inventory test. From the 563 Sina Microblog participants, 444 (171 males and 273 females) with an average age of 23.8 were recruited, while others were dropped because of the inactive usage for Sina Microblog or carelessness in finishing the inventory. This research defines qualified participants as those that have published more than 500 pieces of status, and have new status published within the last three months.

B. Measures

In this study, Big-Five personality is measured as ground truth labels. The score of each dimension is within -1 to +1. We use John O.'s 44-item Big Five Inventory, which is a 44-item questionnaire measuring Agreeableness (Agre.), Conscientiousness (Cons.), Extraversion (Extr.), Neuroticism (Neur.), and Openness (Open.). BFI is one of the most widely used brief measures of the Big-Five personality [22]. The questionnaire has been generally used for its high reliability and validity. The inventory has five subsets, and each one corresponds to one personality dimension. The Chinese version of BFI (http://www.ocf.berkeley.edu/ johnlab/bfi.htm) is used in the study. The items in the inventory consist of short and easy understanding phrases to assess the

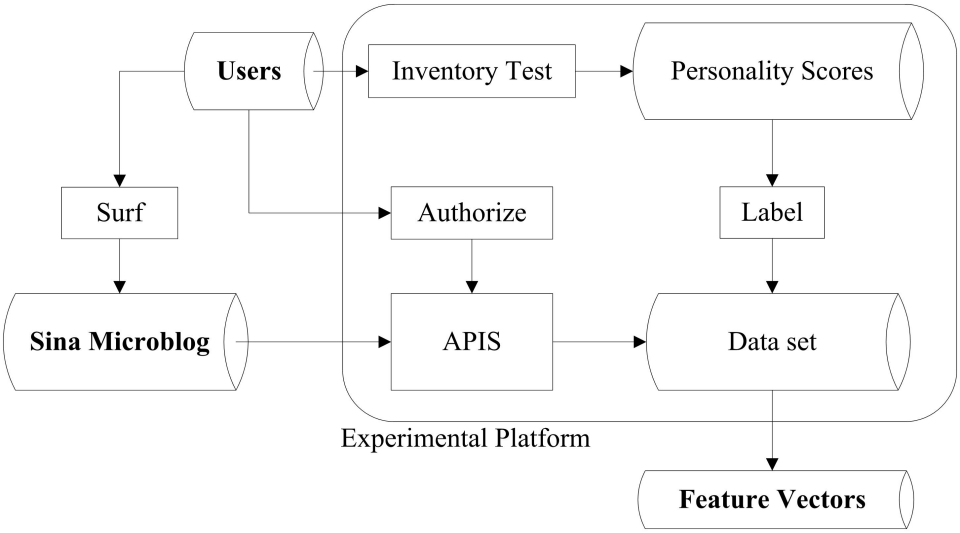

Figure 1. Flowchart of data collection.

prototypical traits of the Big-Five dimensions. Participants evaluate themselves on each question with a 5-point Likert scale, ranging from Disagree strongly(1 point) to Agree strongly(5 points). Then their scores are linearly mapped to interval -1 to +1.

C. Behavioral Features

Twenty nine features are extracted from the Sina Microblog dataset, which can be categorized into 4 groups. The first group includes 4 features about the *profile* information of participants. Seven features are used to figure up how the user presents to others and form the *self-presentation* group. Group *security settings* contains 7 features about the privacy settings. At last, the *social networking* group, having 11 features, is defined as the online interaction of the user. The detailed information of features are shown in table I.

IV. Regression Models

The prediction of personality can be regarded as a model fitting process. In this paper, we adopt two modeling approaches, incremental regression and multitask regression. Through sorting the sample points, incremental regression amends the existing model or creates a new model according to the error of the next sample in the existing model. In two dimension space, it can fit a complex curve with several simple lines. Nevertheless, Big Five personality labels are multivariate vectors. There exists weak correlation between tasks. In this study, multi-task regression learning [23] is used to predict the personality. Different from typical regression, multi-task regression puts all labels together, and tries to model all tasks in a whole. Therefore, the model consists both the specific information of each task and the shared information among tasks.

A. Incremental Regression Model

Incremental regression is a linear regression which can be used to fit complex non-linear problems. As shown in Algorithm 1, it needs to sort the training samples first. Facing to the different situations, the sorting strategy varies a lot. Normally, it sorts the samples according to the normalized magnitude of each sample vector from small to large. Then it starts with one end of the sorted sample set and builds a simplest local regression model with fewest(n) samples. n is selected as the minimum sample amount for modeling. For example, if the target is a regression problem in two-dimensional space, n will be set as 2. Next, it tests the error of the model with the next sample in the sample queue. If the error is less than the threshold, the algorithm will refit the model with this new sample. Otherwise, the local model together with its domain will be saved into the line array.

Algorithm 1 The Incremental Regression Algorithm

```
Require: Ranked Points, Parameter n, error threshold
  Initialize point array, line array, normalize dataset
  Normalize dataset
  Rank the sample points
  repeat
    Put n samples from ranked Points into point array
    Model = Line Regression(point array)
    while error>error threshold do
      Put the next one sample into point array
      error=(y(next sample)-Model(next sample))
    end while
    Put the last one sample back to the ranked Points
    Save the Model into line array
  until No samples exist in ranked Points
```

Table I
ONLINE FEATURES IN SINA MICROBLOG

Group	Number	Example
Profile	4	hometown ID(province ID, city ID), gender, register date
Self-presentation	7	length of screen name and self-description, whether user's domain is the same as his blog address
Security settings	7	comments available or not, type of verification, whether user's domain is default
Social networking	11	number of friends, followers, mutual followers and statuses, and the proportion of the original status

B. Multi-task Regression Model

The main objective of the multi-task learning is to use multiple task modeling strategies to improve performance beyond the single-task learning in the same scene.

Assume there are T tasks and n instances. Each instance is represented as a column vector $x \in \mathcal{R}^m$ (m features) and paired with a multi-dimensional output vector $y \in R^T$ (T tasks). Therefore,

$$X = [\overrightarrow{x_1}, \overrightarrow{x_2}, \ldots, \overrightarrow{x_n}] \tag{1}$$

$$Y = [\overrightarrow{y_1}, \overrightarrow{y_2}, \ldots, \overrightarrow{y_n}] \tag{2}$$

The objective is to find an $T \times m$ coefficient matrix W such that $\mathbf{W} = \arg\min_{W}\{\hat{Y} - WX\}$.

We start with an MTL formulation that jointly considers T regressors, in a way similar to the primal form of the loss function as

$$\min_{W}(L(X, Y, W; 1:T) + \lambda\Omega(W)) \tag{3}$$

where $L(X, Y, W; 1:T)$ denotes the empirical loss function, $\Omega(W)$ is the regularization term, and λ is a trade-off constant.

In this study, $L(X, Y, W; 1:T)$ is set as the least square loss and the regularizer is set as Frobenius norm. That is,

$$L(x, y, \mathbf{W}; 1:T) = \sum_{t=1}^{T}\sum_{n=1}^{N}(y_{tn} - \sum_{h} w_{th}x_{hn})^2 \tag{4}$$

$$\Omega(W) = tr(W^T W) \tag{5}$$

Substitute into equation (3), it has a unique optimizer,

$$W^* = (X^T X + \lambda I)^{-1}(X^T Y) \tag{6}$$

The trade-off constant λ can be selected in many ways. Here we choose the bias-variance decomposition [24] which minimizes the expected loss $(bias)^2 + variance$. This model with the optimal predictive capability is the one that leads to the best balance between bias and variance.

V. RESULTS

This paper experiments on how personality reflects Sina Microblog behaviors. On one hand, we try to find the associated modes of users' personality and network characteristics. On the other hand, through different machine learning algorithms, computational models of personality based on network characteristics are founded. In order to test the performance of the different models, we use 5-fold cross validation for training and Mean Absolute Error (MAE) as the assessment criteria.

A. Correlation Analysis

We try to take all personality dimensions as a whole and use multitask regression to build a novel model. However, a basic premise is that there exists correlation (even weak correlation) between the Big Five personality dimensions. Therefore, we firstly analyze the Pearson Correlation Coefficient between personality dimensions in Table II. In an intuitive understanding, Pearson correlation coefficient describes the degree of tightness between two fixed variables and defined as

$$\rho = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \tag{7}$$

where n is the total sample size, X_i and Y_i are observations, $\bar{X}$ and $\bar{Y}$ are mean values. ρ describes the degree of linear correlation between the two variables. The value of ρ ranges between -1 and +1. If ρ is positive, the two variables are positively correlated (the greater the value of one variable, the greater the value of another variable). If ρ is negative, the two variables are inversely related (the smaller the value of a variable, the greater the value of another variable). The absolute value of ρ stands for the strength of the correlation.

Table II
PEARSON CORRELATIONS OF PERSONALITY SCALES

	Agree.	Cons.	Extr.	Neur.	Open.
Agree.	-	.247**	.108*	-.406**	.174**
Cons.	.247**	-	.260**	-.403**	.281**
Extr.	.108*	.260**	-	-.345**	.318**
Neur.	-.406**	-.403**	-.345**	-	-.161**
Open.	.174**	.281**	.318**	-.161**	-

**. Correlation is significant at the 0.01 level (2-tailed).
*.Correlation is significant at the 0.05 level (2-tailed).

For the above table, agreeableness, conscientiousness, extraversion, and openness pairwise show a significant positive correlation, especially for extraversion and openness($\rho = .318$). However, neuroticism is significantly negative correlated with the other four dimensions, especially for the

correlation factors with agreeableness and conscientiousness(above 0.4).

B. Regression Accuracy

This paper extracts 29 features, as a result at least 30 samples are needed to train a local model. Therefore, we set $n = 30$. We sort the dataset according to the normalized magnitude of each sample, and choose the error threshold as 0.10. We build a local model with the first 30 samples, and use the local model to test the error of the following sample. If the error is smaller than the threshold, we put these 31 samples together and rebuild a new local model. We repeat this process until the error of a sample is larger than the threshold. After traversing all the samples, a group of local model will be returned and compose a global prediction model. Table III shows the mean absolute error of predicting personality on Sina Microblog.

The above analysis proved that personality dimensions have moderate correlation relationship with each other. When predicting psychological properties, this relationship needs to be taken into account to improve the performance. Therefore, multitask regression is used here. Two comparable prediction systems are set up. One is the single task linear regression model which uses the online behaviors of participants to predict personality, which can be used as the baseline. The other one is the multi-task regression model which can figure up the common information between tasks as well as the special information of each task. After training the multi-task regression algorithm, λ is selected as 1.6989 ($\ln(\lambda) = 0.53$) for each dimension. The Mean Absolute Errors are shown in Table III and Figure 2.

Table III
MEAN ABSOLUTE ERROR OF DIFFERENT ALGORITHMS FOR EACH DIMENSION.

Dimension	Single task	Incremental reg.	Multitask reg.
Agre.	0.1881	0.1818	0.1314
Cons.	0.1899	0.1765	0.1442
Extr.	0.2086	0.1920	0.1389
Neur.	0.2196	0.2037	0.1302
Open.	0.2021	0.1891	0.1473
Aver.	**0.2017**	**0.1886**	**0.1384**

VI. DISCUSSION

The previous experiment tried to predict personality on Sina Microblog. Three models were established with one as the prediction baseline.

A. Data Reliability

As shown in the last row in table III, the simplest single task linear regression works worst (percentage MAE is 20.17). Although incremental regression is still a broad linear regression, it can train and build several local optimal models, which can be more powerful than the single task regression model (percentage MAE is 18.86). However, incremental regression has its own limitations. First, it relies too much on the sorting strategy. Since sample sorting is the first step of the algorithm, and the following local model establishment depends on the sorted sample, without a better sorting method, the algorithm will not get a responsible result. Second, the algorithm is quite sensitive with the noisy samples. When a noisy sample is coming to the test of the local model, the test error will get an extremely large value. Even though the next samples are all regular, the domain of this local model will still get an end.

We try to illustrate noise sensibility between linear regression and incremental regression with an example in Figure 3. In this example, samples are in two dimensional coordinate with $\langle x_i, y_i \rangle$ pairs. The two sub-pictures in the upper row demonstrate the difference between the two methods in an ideal dataset. Although there is no noisy sample, linear regression can still not build the pattern, since independent variable X is not linearly related with dependent variable Y. Yet, incremental regression focuses on local optimum, and uses several linear model to fit the nonlinear pattern, which works relatively better than simple linear regression. Even so, incremental regression is too much sensitive with noisy samples as shown in the lower two sub-pictures in Figure 3. Therefore, in the personality prediction problem, it is extremely important to identity the noisy samples and remove them out of the dataset.

In the actual process of conducting online survey, we records the time nodes of answering each question of the personality inventory. Time nodes can be used for noisy samples checking which can not be achieved by the previous offline psychological investigations. We manually check the time used for each question and the distribution of the answers. We do the following operation on the raw data: 1. If the time spent for one question is less than one second, remove the sample. Usually, it is impossible to answer the question in such a short interval. 2. If the answer distribution or answer array shows a clear regular pattern, remove it. We denote the answer option with English letters A to E, and remove the samples with "AAAA...","AABBCC...", "ABCD...","DCBA" or "ABCDCBA" patterns. In order to keep the quality of the samples, we inform the participants about our demands. Once they decide to take part in our experiment, their answers will be carefully checked and a fee will be only given to the participants who passed the noise check. The MAE of the incremental regression gets a nearly 2 percent reduction which proves that our dataset is reliable.

B. Personality and Online Features

After verifying the reliability of our dataset, we try to predict personality based on it. Since personality is multi-dimension, and there exist moderate correlations among the dimensions, we choose multitask regression to build the personality prediction model. Results show that the

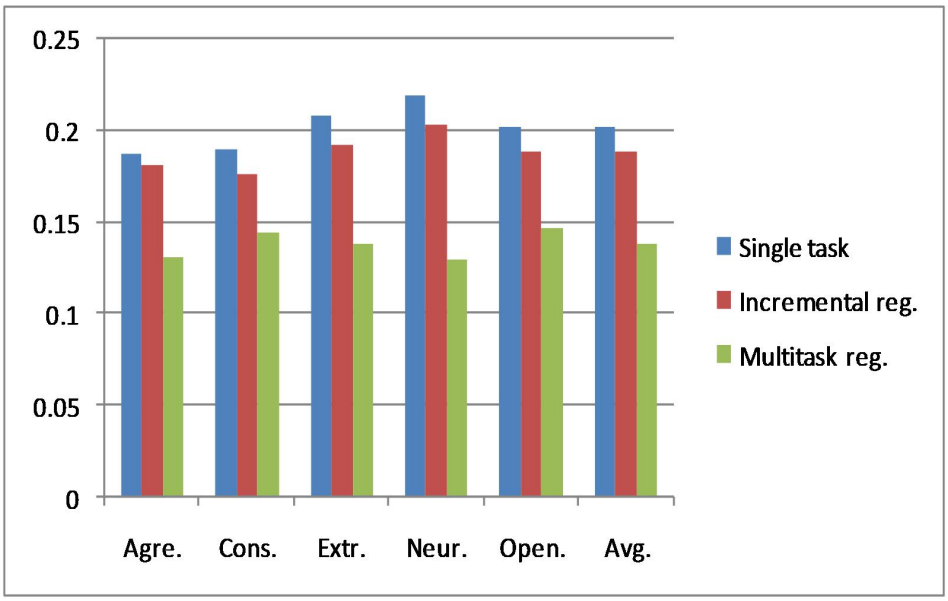

Figure 2. MAEs of different different algorithms on Big Five personality dimensions.

Table IV
PART OF REGRESSION COEFFICIENTS OF MULTITASK REGRESSION FOR EACH DIMENSION.

Features	Agre.	Cons.	Extr.	Neur.	Open.
Number of mutual followers	**0.1921**	**0.4508**	**0.4965**	-0.0720	**0.6041**
Number of favourites	-0.0473	**-0.1337**	-0.0121	-0.0470	**0.5999**
Number of followers	-0.0351	**-0.5066**	**-0.1208**	0.0726	**-0.3900**
Number of friends	-0.0116	**0.1157**	-0.0409	-0.0345	0.0122
Length of screen name	0.0699	**-0.1949**	**0.1082**	0.0254	0.0471
Number of statuses	0.0786	**-0.1310**	**-0.1979**	**0.1823**	0.0930
Percentage of the original status	**0.1013**	0.0331	0.0582	0.0244	-0.0172

average mean absolute error of multitask regression model is 13.84%, which gets about 5 percentage points reduction compared to incremental regression. Before training the model, we normalize each dimension of the features in dataset. We list a part of the regression coefficient matrix of the model in Table IV(not all the coefficients are meaningful), and analyze the relationship between online features and personality.

From the table, agreeableness is positively related with number of mutual followers (if A is the follower of B and B is also the follower of A, they are mutual follows), the mutual followers proportion in all followers and original status percentage [25]. That means users with high degree of agreeableness tend to be the mutual followers of other users. If one user follows them, they will follow back to this user more likely. At the same time, they also like to update status more often and most of their statuses are written by themselves (original statuses percentage, 0.1013) not the retweeting others' statuses.

Conscientiousness can be regarded as self-discipline. When reflecting to Sina Microblog, users with high degree of conscientiousness have more mutual followers(0.4508) and more friends online(0.1157), although their follower number is not in large scale(-0.5066) [26]. They are well self-controlled and can carefully publish statuses without "watering" online (usually have short screen name, -0.1949), therefore, their status number is relatively small(-0.1310).

Extravert has more mutual followers(0.4965) online and a relatively long screen name(0.1082) [27]. However, since extraverts are more eager to show their charms, they tend to have a more strong sociability. Therefore, extraverts have more friends offline (online friends number is relatively small, -0.1208) and don't need to publish more statues to attract others(-0.1979).

Neuroticism is the stability of emotion. People with high degree in neuroticism are more likely to suffer mental health and do a lot of repetitive operations, therefore, they will update their status quite often(0.1823) [28].

Openness shows the richness of individual imagination and curiosity about new things. People with high score in openness tend to follow others very much and listen to what other users are talking about(0.6041) in spite of the shortness in followers(-0.3900). There are more resources in their favourites(0.5999) for their curiosity about online

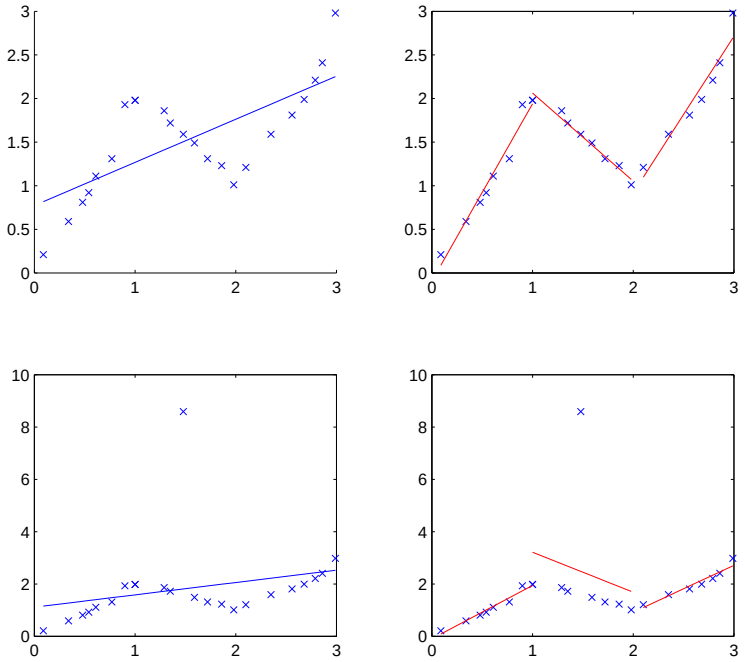

Figure 3. Incremental regression and linear regression.

information around them [29].

VII. Conclusion and Future Work

This paper analyzes the personality on Sina Microblog based on the big-five theory. We proposes an incremental regression model to prove the reliability of the dataset and the multi-task regression model to improve the predicting accuracy. We verify the reliability of our dataset, and concludes how personality influence and reflect the online behaviors.

In future, we will continue to collect users' data in Sina Microblog, and invite more participants to get a larger dataset. We also plan to design and extract other user network characteristics on Microblog, and further research on the behavior patterns of other psychological attributes, such as mental health and social attitudes. We will also try other typical multi-task learning algorithms to tune the performance of the predicting model.

Acknowledgment

The authors gratefully acknowledges the generous support from National High-tech R&D Program of China (2013AA01A606), NSFC (61070115), Institute of Psychology (113000C037), Key Research Program of CAS (KJZD-EW-L04), Strategic Priority Research Program (X-DA06030800) and 100-Talent Project (Y2CX093006) from Chinese Academy of Sciences.

References

[1] Burger J. (2008). Personality. Belmont: Thomson Wadsworth.

[2] Costa P., McCrae R. (1992). Neo personality inventorycrevised (neo-pi-r) and neo five-factor inven-tory (neo-ffi) professional manual. Odessa, FL: Psychological Assessment Resources.

[3] Goldberg L. (1993). The structure of phenotypic personality traits. American psychologist 48(1):26.

[4] Funder D. (2001). Personality. Annu Rev Psychol 52: 197-221.

[5] Landers R., Lounsbury J. (2006). An investigation of big-five and narrow personality traits in relation to internet usage. Computers in Human Behavior 22(2): 283-293.

[6] Amiel T., Sargent S. (2004). Individual difference in internet usage motives. Computers in Human Behavior 20: 711-726.

[7] Goby V. (2006). Personality and online offline choices: Mbti profiles and favored communication modes in a singapore study. Cyber Psychology and Behavior (9): 5-13.

[8] Amichai-Hamburger Y., Wainapel G., Fox S.(2002). "on the internet no one knows i'm an introvert": extroversion, neuroticism, and internet interaction. CyberPsychology and Behavior 5: 125-128.

[9] Correa T., Hinsley A. W., De Zuniga H. G. (2010). Who interacts on the web? the intersection of users' personality and social media use. Computers in Human Behavior 26(2): 247-253.

[10] Ahn Y. Y., Han S., Kwak H. (2007). Analysis of topological characteristics of huge online social networking services. International conference on World Wide Web : 835-844.

[11] Mucha P. J., Richardson T., Macon K. (2010). Community structure in time-dependent, multi-scale, and multiplex network. Science 328(5980): 876-878.

[12] The 31st Chinese Internet Development Statistics Report. http://wenku.baidu.com/view/bef8d307f12d2af90242e615.html.

[13] Kathryn Wilson S. F., White K. M. (2010). Psychological predictors of young adults use of social networking sites. Cyber-psychology, Behavior, and Social Networking 13(2): 173-177.

[14] Ma C. J., Zhang M. (2011). Application of social network analysis in psychology. Advances in Psychological Science 19(5): 755-764.

[15] Hamburger Y. A., and Ben-Artzi E. (2000). The relationship between extraversion and neuroticism and the different uses of the internet. Computers in Human Behavior, (16):441-449.

[16] Kosinski M., Stillwell D., Graepel T. (2013). Private traits and attributes are predictable from digital records of human behavior. Proceedings of the National Academy of Sciences, 2013.

[17] Gosling S. D., Augustine A. A., Vazire S. (2011). Manifestations of personality in online social networks: Self-reported facebook-related behaviors and observable profile information. Cyberpsy-chology, Behavior, and Social Networking 14(9): 483-488.

[18] Correa T., Hinsley A., and Ziga H. (2010). Who interacts on the web? The intersection of users' personality and social media use. Computers in Human Behavior, 26(2):247-253.

[19] Kelly M., and James C. M. (2012). The influence of personality on Facebook usage, wall postings and regret. Computers in Human Behavior, (28), 267-274.

[20] Caruana R. (1997). Multitask learning. Machine Learning, (28) 41-75.

[21] Caruana R. (1997). Multitask learning. Ph. D. thesis, School of Computer Science, Carnegie Mellon University.

[22] Gosling S. D., Rentfrow P. J., Swann W. B. (2003). A very brief measure of the Big-Five personality domains. Journal of Research in Personality, (37), 504-528.

[23] Argyriou A. , Evgeniou T., Pontil M. (2008). Convex multitask feature learning.Machine Learning, 73(3):243-272.

[24] Bishop M. (2006). Pattern Recognition and Machine Learning. Springer.

[25] Yang Y., Lei L. (2007). The relationship between adolescents' extraversion/agreeableness, Internet service preference, and Internet addiction. Psychological Development and Education, 42-48.

[26] Peter A. R., Donald H. K. (2008). The impact of the big five personality traits on the acceptance of social networking website. Americas Conference on Information Systems.

[27] Kathryn W., Stephanie F., Katherine M. W. (2010). Psychological predictors of young adults' use of social networking sites. Cyberpsychology, Behavior, and Social Networking, 13(2):173-177.

[28] Luo T., Ding D. (2006). Relationships among personality traits, motive of internet use, and tendency of internet addiction. Chinese Journal of Clinical Psychology, 14 (4):365-367.

[29] Nie F., Jiang P., Wu Y., Chen S. (2007). Relationship between network communication and personality traits of teenagers. Nervous Diseases and Mental Health, 7(6):468-471.