Personality Recognition based on User Generated Content

Cuixin Yuan[1,2], Junjie Wu[1,3], Hong Li[1,*], Lihong Wang[4]
[1]School of Economics and Management, Beihang University, Beijing, China
[2]Beijing Key Laboratory of Emergency Support Simulation Technologies for City Operations, Beijing, China
[3]Beijing Advanced Innovation Center for Big Data and Brain Computing, Beihang University, Beijing, China
[4]National Computer Network Emergency Response Technical Team/Coordination Center of China, Beijing, China
* Corresponding author: hong_lee@buaa.edu.cn

Abstract—**Personality recognition has lots of applications, like business recommendation systems, human resource management, author identification and psychological treatment. Traditionally, identifying one's personality traits depends on questionnaire scales. With the development of the Internet, people can express their opinions, emotions, attitudes, etc., on social networks. From the viewpoint of psychology, people's behaviors and decisions are influenced by their personality characteristics, so it's convenient for us to recognize one's personality based on his social network content. In this paper, we build a new personality recognition model based on a standard dataset, i.e., the Facebook update status content. Firstly, we extract the language features via the Linguistic Inquiry and Word Count (LIWC) dictionary tool. Then all the words in the status content are trained by Google news pre-trained word2vec embedding with 300 dimensions. Next, we construct the personality recognition model by using the Convolutional Neural Network (CNN) method with the LIWC features. We conduct five binary classifications on the most widely used personality model, i.e., the big five personality that contains five traits: extroversion, neuroticism, agreeableness, conscientiousness, and openness. Finally, our experimental results demonstrate that the method we present in this paper achieves very good performances, especially on the openness trait.**

Keywords- Personality Recogniton; Feature Extraction; Big Five; Convolutional Neural Network

I. Introduction

Personality is a psychological construct reflecting an individual's emotion, thought-pattern, motivation, and behavior characteristics. Personality has a great importance on our life. For examples, we can make a decision to choose who we like to make friends through judge his personality. From the aspect of commercial company, it's vital to conduct the product recommendation based on personality recognition. Because if we have the same personality, we may have the same preference for the goods. Personality always influences our behavior. In turn, company can provide the targeted service according to customer's personality characteristic. In the other sides, personality recognition makes a contribution to human resource management for recruiting the competent person. In addition, sometimes personality trait is related to human beings psychological problems, and personality recognition is the first step in treatment. Therefore, it's necessary to study the personality recognition.

Personality measurement is derived from words study [1]. There has been a fundamental assumption, which is lexical hypothesis. In our life, important things are always given descriptive words. Thus finding personality traits from vocabulary has become an important way to study personality. Gordon W. Allport is the first one to do this work. He had counted how many vocabularies describing personality difference in the English dictionary. As a result, it reached to 17,953 vocabularies. After that he then selected 4500 relative important words. And last he choose 35 particularly important words from that. Finally he conduct the factor analysis of these. Based on the results, he found that there are always five factors in the result. It is what we know today the origin of the big five factor model. While recently the most popular model to describe the personality traits is still the "big five personality model". It contains five dimensions:

1) Extroversion (outgoing/energetic vs. solitary/reserved). People with high extroversion trait are full of energy, positive to emotions, surgency, assertiveness, sociability, tendency to seek stimulation in the company of others and talkativeness. 2) Neuroticism (sensitive/nervous vs. secure/confident). Neuroticism identifies certain people who are more prone to psychological stress. The tendency to experience unpleasant emotions easily, such as anger, anxiety, depression and vulnerability. 3) Agreeableness (friendly/ compassionate vs. challenging/detached). People with high agreeableness trait are tendency to be compassionate and cooperative rather than suspicious and antagonistic towards others. It is also a measure of one's trusting and helpful nature, and whether a person is generally well-tempered or not. 4) Conscientiousness (efficient/organized vs. easy-going/careless). People with high conscientiousness trait are tendency to be organized and dependable, show self-discipline, act dutifully, aim for achievement, and prefer planned rather than spontaneous behavior. 5) Openness (inventive/curious VS. consistent /cautious). People with high openness trait are appreciated for art, emotion, adventure, unusual ideas, curiosity, and variety of experience. Openness reflects the degree of intellectual curiosity, creativity and a preference for novelty and variety a person has.

The big five personality traits are often abbreviated as "OCEAN" model. And personality recognition has widespread applications. Matthews et al. thought that it is reliable to research personality traits, for personality traits are usually stable in a short time[2]. In fact, there is a lot of study on personality recognition. Traditionally if we want to judge one's personality, we need the testee to finish a personality measuring scale on paper. This costs lots of time and money. And we don't ensure

978-1-5386-5178-0/18/$31.00 © 2018 IEEE

all the people complete the questionnaire truly. With the development of internet, more and more people like to communicate with others on the social network, like Facebook, Twitter, Instagram, Sina Weibo, blogs and Wechat. They can express their thoughts, ideas, opinions, emotions, and comments online freely. From the viewpoint of psychology, people's behavior is related to their personality, and the behavior online is consistent to the offline in our daily time. What's more, the online digital record is easy to be obtained. Therefore, scholars begin to study the personality problems through social network datasets, containing the post content, favorite number, comment, or emoticons.

Many researches consider the personality recognition problem as a binary classification. The personality score in each dimension is divided into two parts. If the score is lower than the mean of this dimension trait score, its label is "n", and if the score is higher than the mean of this dimension trait score, its label is "y". Then based on the social network content, different authors extract features from the text, and build five binary classifier using different methods, such as support vector machine (SVM), Naïve Bayes (NB), Logistic Regression (LR), K-Nearest Neighbor (KNN) and Back-Propagation neural network (BP), and so on. Finally, the aim of experiment is to compare the accuracy on the five personality traits. Now the best effect of personality recognition is openness dimension. As for the feature extraction engineering, some researches are based on syntax, semantics, categories and the Linguistic Inquiry and Word Count (LIWC) dictionary tool to extract feature from the content. Those studies depend on the assumption that vocabulary use is correlative with personality traits. But current research is limited to the classifier and the performance is not well. Now, with the rise of deep learning algorithms, some researchers have adopted the deep learning methods to study the personality recognition, such as Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), and Long Short Term Memory (LSTM). This also improves the personality recognition study, and drives the development of computational psychology. Given that, in this paper we propose a new method through deep learning method to build the personality recognition model in public open dataset Facebook. We employ the convolutional neural network (CNN) to build the model and add new features extracted from the content through LIWC into the model. As a result, we find the classification accuracy outperformance than previous research, especially on openness personality trait.

The remainder of the paper is organized as follows: In Section II, we present the related work. In Section III, we describe the methodology. The experimental results and demonstration are conducted in Section IV, and we finally conclude this paper in Section V.

II. Related Work

Nowadays there are lots of achievements about the personality recognition. It provides the foundation for our research. In the field of personality recognition, there are different datasets and classification methods. We will from three aspects to review the related work.

A. Datasets

The earliest research about the personality is based on some written data and blogs. Oberlander J and Nowson S et al. carried on the personality study on blog dataset. They extracted the word frequency features through N-grams, then used the methods of Naïve Byes and Support Vector Machine to classify the personality traits [3][4]. Shlomo A et al. worked on the personality recognition through 2236 essays, which were written by psychology students in 20 minutes and they could write whatever they thought. They also extracted the categories and frequency of the words as the input to classifier SVM, then to distinguish their extroversion and neuroticism [5]. Mairesse also used the same dataset to do the personality research [6]. F Alam depended on the Facebook dataset to identify the personality, which was adopted by mypersoanlity app on Facebook platform. Users can complete the big five personality scale from the International Personality Item Pool (IPIP), and then decide whether they agree to get their Facebook data voluntarily [7]. Skowron made use of the combination of Twitter and Instagram datasets to go on the personality experiment [8]. D Estival et al. employed the corpus of emails, nearly 9800 messages, to analysis the big five personality traits [9]. In China, there are also some researches about personality recognition according to Chinese social network datasets. Bai exploited 209 Ren Ren dataset to identify personality score by C4.5 decision trees [10]. Li used 547 users' Weibo data to go on the personality study. It included statistic data, user portrait data, self-description, social behavior and update status [11]. Peng collected 222 Chinese users' Facebook data, and used SVM to classify personality traits [12]. To sum up, different datasets are all labeled with the personality score. And the studies compared the classification results. Maybe the results compared with each other are lack of consistent due to the datasets difference. So our study is based on the public open standard datasets, Facebook, in order to increase the research consistency.

B. Feature Extraction

In theory, understanding the relationship between words use and personality traits completely has a significant impact on human behavior [13]. Thus many scholars established the classification model through feature extraction from the words use frequency or words attribute. F. Iacobelli et al. chose the features from four aspects based on a blogs dataset, including stemming, proper noun, laughter word and punctuation. Then they computed the Term Frequency-Inverse Document Frequency (TF*IDF) to represent the importance of the word as the input to the personality classification model [14]. G Farnadi et al. utilized the Linguistic Inquiry and Word Count (LIWC) dictionary to extract features from Facebook content [15]. This dictionary tool have been widely used in personality study to find the features through vocabularies. The linguistic features contain words count, psychological process, relevance, personal pronoun, language dimension and so on. Y Bachrach verified the relationship between personality traits and Facebook features, such as the numbers of friends, groups, photos, status and tags [16]. This research extracted features were based on the social network features. In addition to that, SM Mohammad thought emotion words in the content also reflect different personality traits. And he found that using emotion features, like excitement, guilt, desire and admire, indeed can improve the classification performance [17]. In conclusion, considering the technical aspect of feature extraction, generally it includes lexical features and semantic features. There are also some studies on the feature strategies to analysis the relationship between personality and the use of punctuation, emotion, topic, and the potential semantic analysis. Besides, the basic tool for feature extraction is LIWC.

C. Classification Methods

As mentioned above, current research about the personality recognition focus on the binary classification in five personality traits. Alam worked on the personality recognition by a variety of methods, including support vector machine (SVM), Byes Logistic Regression (BLR), and Multinomial Naïve Byes (MNB). As a result, he found the method of MNB has the best classification performance [18]. Jon Oberlander acquired the user's blogs through URL, and verified which classification algorithm done well in identify personality. The experiment result proved that the method of Bayes was with higher accuracy than SVM [19]. F. Iacobelli et al considered that previous research about personality recognition is a binary question, so he put forward to analyze the dependency in the different personality traits by the method of Conditional Radom Fields (CRF) and found that agreeableness and emotion stable could improve the classification performance on agreeableness dimension trait [20]. Despite these methods, there are lots of researches adopting neural network methods. Kalghatgi et al. proposed a neural network model by means of Multi-Layer Perception (MLP) and took the semantic and social behavior features by manual annotation to be the input, then give the personality prediction label as the output [21]. Su et al used the Recurrent Neural Network (RNN) to explore the rotation characteristic of dialogue to predict personality. He put the LIWC and grammar features as input, and each user's personality prediction value as output [22]. Ling et al proposed the hierarchical word features. And his study based on the assumption that grammar and semantics of words are sequence features, so he used the Long Short-Term Memory (LSTM), variable RNN, to construct word vectors. When language model and lexical labeling are considered, the model significantly improved the personality prediction accuracy [23]. Besides the content features, M Kosinski used 58000 voluntaries Facebook favorite list data to predict user psychology traits through logistic regression to extract predict indicator. And this proved that it's credible to identify personality based on social network data [24].

In conclusion, the majority of current research on personality recognition problem is based on the SVM algorithm, and relies on the semantic and vocabulary features. Later some improved methods are integrated the features engineering or classifiers. But using deep learning algorithms to identify personality is rare, like convolutional neural network, which has good effect in image classification. Therefore, we will adopt the CNN method to establish the personality recognition model on the Facebook update status. This new attempt would contribute to personality study development.

III. METHODOLOGY

In this section, we present primary algorithm and steps used in our personality recognition model, including data set, feature extraction and convolutional neural network method.

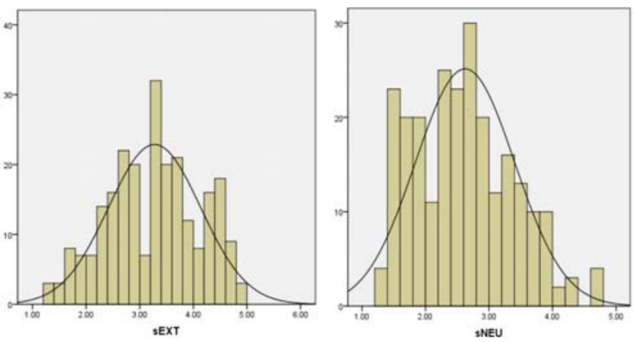
A. Data Set

In our paper, we use the gold standard labelled dataset, mypersoanlity. It has been collected by David Stillwell and Michal Kosinski through the means of a Facebook application that implements the big five test (Costa &McCrae's NEO-PI-R domains and facets) and other psychological tests. The volunteers can choose whether they are agree to link their Facebook homepage to get the content. At last, there are about 58000 samples of volunteers from the United States, obtained through the mypersoanlity Facebook application. This dataset we used is a subset, which contain 250 users and about 9900 status updates. Besides the dataset also contains the gold standard personality score and labels (self-assessments obtained using a 100-item long version of the IPIP personality questionnaire). The labels have been derived from scores with a median split.

The scale of big five tested scores, mean value and standard deviation of the Facebook dataset are present in Table 1. The five personality traits are Extroversion (sEXT), Neuroticism (sNEU), Agreeableness (sAGR), Conscientiousness (sCON), and Openness (sOPN). We can see that the average score of openness is higher than the other personality traits. And the Neuroticism score is lower compared with others. The stand deviation is generally lower. This reflect most of the volunteers on Facebook are open, appreciated art and ideas, imaginative, and stable emotions. The distribution of personality scores are showed in Fig. 1. From that, we can conclude that the personality scores of Facebook users conform to normal distribution.

Table 1. Statistics of the Big Five Tested Scores

Personality traits	sEXT	sNEU	sAGR	sCON	sOPN
Mean	3.28	2.62	3.60	3.53	4.08
Stand deviation	0.054	0.049	0.043	0.047	0.036



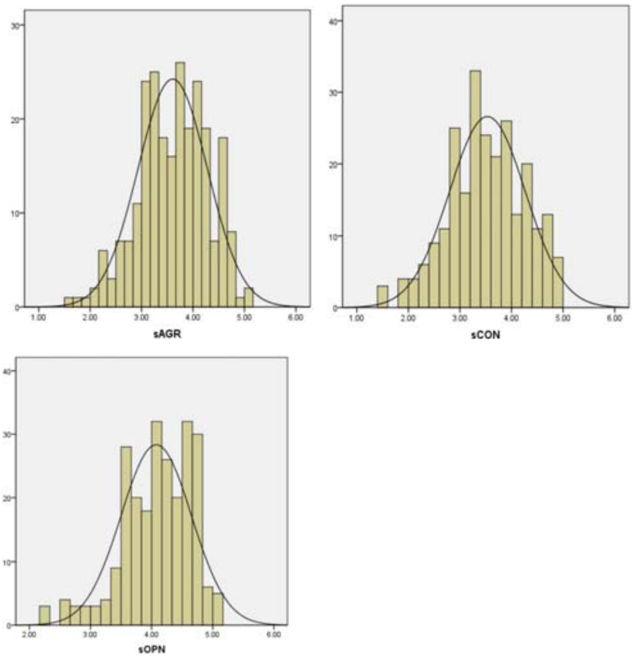


Fig. 1. Distribution of personality scores on the Facebook dataset

B. Feature Extraction

Classification effect sometimes depends on the feature extraction. As we all know, there are lots of manners to extract features, for example, linguistic feature, words frequency, semantics, grammar, emotion and punctuation. The most commonly used is LIWC dictionary. In our research, we also employ the LIWC tool to extract features.

LIWC 2015 is the gold standard in computerized text analysis. Learn how the words we use in everyday language revealing our thoughts, feelings, personality, and motivations. Based on years of scientific research, LIWC2015 is more accurate, easier to use, and provides a broader range of social and psychological insights. The LIWC2015 analysis contain many different output dimensions. Here we use two type features, the traditional LIWC dimension and summary variables. The calculated variable is the word count (WC), which is just the raw number of words within a file. Traditional LIWC dimension contains five specific aspects, such as I-words (I, ME, MY), social words, positive emotions, negative emotions, cognitive processes. Traditional LIWC dimensions reflect percentage of total words within the text we provided. The summary variables are research-based composites that have been converted to 100-point scales where 0=very low along the dimension and 100=very high. Analytic refers to analytical or formal thinking. Clout taps writing that is authoritative, confident, and exhibits leadership. Authenticity refers to writing that is personality and honest. Emotional tone is scored such that higher numbers are more positive and upbeat, and lower numbers are more negative.

Firstly, we obtain the features data through LIWC tool online. Then we verify the relationship between the LIWC variables and personality traits based on Pearson correlation coefficient. The result is shown on table 2. As we can see, there are four personality traits that have significant correlation with LIWC features, except agreeableness. Extroversion has positive relationship with social words, positive emotions and cognitive process significantly. Neuroticism has negative relationship with emotional tone. This is consistent with personality traits behavior. It also proves that words use in language can reflect different personality characteristic. Conscientiousness is negative related to cognitive process. Openness is positive related to social words and clout. Besides the LIWC tool, we also adopt the convolutional neutral network, which extract the features from the content automatically through deep learning. Combined these two category features, we can further improve the classification model performance.

Table 2. Pearson Correlation Coefficient

Variable	sEXT	sNEU	sAGR	sCON	sOPN
I-Words (I, ME, MY)	.026	.046	-.034	-.101	-.045
Social Words	**.126***	-.056	.083	-.022	**.139***
Positive Emotions	**.127***	-.093	.080	.049	-.017
Negative Emotions	.021	.052	.028	-.123	.018
Cognitive Process	**.132***	.053	.060	**-.127***	.060
Analytic	-.010	-.054	.020	.103	-.067
Clout	.072	-.119	.083	.067	**.133***
Authenticity	-.001	.089	-.025	-.061	-.075
Emotional Tone	.101	**-.176****	.089	.107	.029

** stands for significant at the $p < 0.01$ level

*stands for significant at the $p < 0.05$ level.

C. Convolutional Neutral Network

With the development of deep learning, Convolutional Neutral Network (CNN) has been used in many research fields, especially in image classification. Nowadays, it also has been used in text classification, and obtain great performance. Navonil Majumder has adopt the CNN method to complete personality recognition based on essays dataset [25]. On this basic, we add the LIWC features into CNN construction to establish a new personality recognition model in our paper based on the standard dataset, Facebook. And the classification effect is improved, especially on openness dimension trait. While currently there is little research employed the CNN method to identify personality traits on social network datasets. So this maybe make a contribution to the personality research.

Convolutional Neural Network is one feed forward neural network. Generally, the basic structure of CNN consists two layers, one is feature extraction layer. Each neuron input is connected with previous local receptive field, then extract the local feature. Once the local feature is extracted, its location relationship with others is also determined. The other one is feature map. Each computing layer in the network is composed multiple feature maps, and each feature map is a plane with the same weight in all neurons. In the feature map structure, sigmoid function is always used as the activation function, so that the feature map has invariant displacement. As for our design CNN model, we describe it from five process layers, including input layer, convolution layer, pooling layer, full connection layer and output layer.

(a) Input layer. We represent the dataset as a set of documents in which each user' status update on Facebook has been put in one document. Each document has a sequence of sentences, and each sentence has a sequence of words. In our experiment, we used the Google's pre-trained word2vec embedding to train our dataset with 300 dimensions. Each word has the same fixed length word vector to represent. In order to guarantee all sentences contain the same number of words, we pad the shorter sentence with dummy words. If a word is not found in Google news word2vec list, we assign it a 300 dimensional vectors with a uniform with distribution in [-0.25, 0.25]. Before we put all the data into the model, we have the data preprocessed, such as replace the abbreviation, lower characters.

(b) Convolution layer. After the words vector input, we use three convolutional filters to extract features from each sentence. Here we use 200 n-gram feature maps, n=1, 2, 3. We also add a bias to the output of the filter. To introduce nonlinearity, we apply the rectified linear unit (RLU) function to the feature maps. Last, we get a 600 dimension vectors to represent the sentences through concatenate the three types of n-grams. Next we apply the convolution and max polling to each sentence in the document. While all the parameters in the sentences are shared. When each sentence are processed, the document vector can be obtained based on the sentence vector.

(c) Pooling layer. In order to reduce the dimensions to a feature map, we adopt the max polling for that. Max pooling means that one filter extracts some values, we just retain the largest in the pooling layer and drop out all the others. The largest value always keep the strongest characteristics, and other weak characteristics are abandoned. It also reduces over fitting problems.

(d) Connection layer. From the feature extraction, we have introduce the LIWC features. At last, we all obtain 9 features from the language words use. So we concatenate those 9 features with the document vector. Finally, this leads to the 609 dimensions document vector. In addition the LIWC features are related to the personality traits.

(e) Output layer. After we get all the dataset feature vectors, we use a two layer perceptron consisting of a full connection layer. And last we apply two size softmax layers to give the probability for yes or no classes on five personality traits.

IV. Experimental Results

We complete the experiment based on the Facebook dataset with 9900 update status. Each user Facebook content is put in one document, and we get the LIWC features of each user online. For the personality recognition, we also established five binary classification model by the method of CNN with same configuration. We use stochastic gradient descent algorithm to tune the network parameters. In our experiments, the network is converged after 50 epochs. To evaluate our method, we compare all the results based on the accuracy with different methods and previous research results. The results are shown in table 3.

We compare our experiment result with the basic classifier SVM, BLR, MNB on the same dataset Facebook [7]. Besides that, we also compare the results by the same method, CNN, on the essays dataset, which is a written dataset used for personality recognition. And we also refer the baseline result summarized by Navonil Majumder [25]. It is the state of the art accuracy result published on the essays dataset. As we can see that whatever classification method we use, the personality recognition accuracy is about 0.56 on big five personality traits except the openness traits, which is always the highest value in accuracy, more than 0.62. But it is distinctly to find that our experiment result in this paper outperformance than others on openness dimension treat, which tends to 0.76. This is a significant improvement. And this is also illustrated the deep learning method CNN with the LIWC features we proposed in the paper indeed has a good effect on personality recognition. Specifically, in the other four personality traits, our method is similar to other methods on performance. For extroversion treat, the classification algorithm SVM and MNB all acquire the same high accuracy value, it is reach to 0.58, and our method result is tend to 0.57. For neuroticism treat, the best classification effect is MNB algorithm, which tends to 0.62. Our method ranks second among all methods. It is reach to 0.60. For agreeableness and consciousness traits, the recognition accuracy is same about 0.59, and our method reaches to 0.57 and 0.58. In all, different classifiers on the four personality traits recognition accuracy have little difference and not improve the performance obviously.

Table 3. Accuracies of Different Methods on Five Personality Recognition

Method	sEXT	sNEU	sAGR	sCON	sOPN
SVM	**0.58**	0.59	0.58	0.58	0.65
BLR	0.56	0.57	0.58	0.57	0.62
MNB	**0.58**	**0.62**	**0.59**	**0.59**	0.69
CNN Facebook	0.57	0.60	0.57	0.58	**0.76**
CNN essays	0.58	0.59	0.56	0.57	0.62
Others essays	0.56	0.58	0.55	0.55	0.60

V. Conclusion

Personality recognition have lots of important applications, and with the development of computing psychology, many personality research achievements have been applied in business recommendation system and healthy diagnosis. In this paper, we overview the researches on the personality recognition from three aspects to introduce the current study. Most research think the personality recognition as a binary classification. Based on the study used the CNN algorithm to identify personality traits, we propose a new model, which put the LIWC extracted features into the CNN method. Before this, we carry through correlation analysis between the LIWC features extracted from Facebook status update and personality traits. Then we conduct the experiment on the standard social network data, Facebook, with about 250 users, 9900 status update. As a result, our method outperformance the baseline, and especially on the openness trait, our method achieves the best effect. The identification accuracy is reach to 0.76. This is proved our design method combined the CNN algorithm and LIWC features has great performance. What's more, we can conclude that social network content could reflect the personality characteristic to a certain degree. Different personality traits are always used to prefer different emotion words, positive words, social process words or negative words. When we consider the language features and words use features, it's helpful to recognize and understand personality traits.

Given that the openness recognition accuracy is higher than others, we can make a contribution to the business recommend system based on this. Because users with high openness personality treat are full of curiosity to new things and product. They have a preference for novelty and variety. This is the target customers that business company is finding to recommend new product. And this type users are easy to accept that, so it helps to open new markets. Therefore, based on the personality recognition, we will further to apply it to the recommend system in the future work.

In a conclusion, there still other information we can get from the social network digital track, and maybe it also helpful to identify the personality. Our paper completes the experiment on the social network status update. In the future, we also can carry on some other experiments based on other social network dataset.

ACKNOWLEDGMENT

This work is partially supported by the grants from the National Natural Science Foundation of China (NSFC) (71471009,71725002,71531001).

REFERENCES

[1] Costa P T, Mccrea R B. Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI)[M]. Springer New York, 1992.

[2] Gerald Matthews, Ian J. Deary, and Martha C. Whiteman.2003. Personality Traits. Cambridge University Press, second edition. Cambridge Books Online.

[3] Oberlander J, Nowson S. Whose thumb is it anyway?: classifying author personality from weblog text[C]// Coling/acl on Main Conference Poster Sessions. Association for Computational Linguistics, 2006:627-634.

[4] Nowson S, Oberlander J. Identifying more bloggers Towards large scale personality classification of personal weblogs[C]// 2007.

[5] Shlomo A, Dhawle S, Moshe K, et al. Lexical Predictors of Personality Type[C]// Joint Meeting of the Interface and the Classification Society of North America. 2005.

[6] Mairesse F, Walker M A, Mehl M R, et al. Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text[J]. Journal of Artificial Intelligence Research, 2007, 30(1):457--500.

[7] Alam F, Stepanov E A, Riccardi G. Personality Traits Recognition on Social Network -Facebook[C]// The Workshop on Computational Personality Recognition. 2013.

[8] Skowron M, Ferwerda B, Schedl M. Fusing Social Media Cues: Personality Prediction from Twitter and Instagram[C]// International Conference Companion on World Wide Web. International World Wide Web Conferences Steering Committee, 2016:107-108.

[9] Estival D, Gaustad T, Pham S B, et al. Author profiling for English emails[C]// Conference of the Pacific Association for Computational Linguistics. 2007:263--272.

[10] Bai S, Zhu T, Cheng L. Big-Five Personality Prediction Based on User Behaviors at Social Network Sites[J]. Computer Science, 2012, 8(2):e2682-e2682.

[11] Li L, Li A, Hao B, et al. Predicting Active Users' Personality Based on Micro-Blogging Behaviors[J]. Plos One, 2014, 9(1):e84997.

[12] Peng K H, Liou L H, Chang C S, et al. Predicting personality traits of Chinese users based on Facebook wall posts[C]// Wireless and Optical Communication Conference. IEEE, 2015:9-14.

[13] Pennebaker J W, Mehl M R, Niederhoffer K G. Psychological aspects of natural language. use: our words, our selves[J]. Annual Review of Psychology, 2003, 54(1):547-577.

[14] Iacobelli F, Gill A J, Nowson S, et al. Large Scale Personality Classification of Bloggers[M]// Affective Computing and Intelligent Interaction. Springer Berlin Heidelberg, 2011:568-577.

[15] Farnadi G, Zoghbi S, Moens M F, et al. Recognising personality traits using Facebook status updates[C]// 2013.

[16] Bachrach Y, Kosinski M, Graepel T, et al. Personality and patterns of Facebook usage[C]// Acm Web Science Conference. 2012:24-32.

[17] Mohammad S M, Kiritchenko S. Using Nuances of Emotion to Identify Personality[C]// ICWSM workshop on Computational Personality Recognition. 2013:319-323.

[18] Alam F, Stepanov E A, Riccardi G. Personality Traits Recognition on Social Network -Facebook[C]// The Workshop on Computational Personality Recognition. 2013.

[19] Oberlander J, Nowson S. Whose thumb is it anyway?: classifying author personality from weblog text[C]// Coling/acl on Main Conference Poster Sessions. Association for Computational Linguistics, 2006:627-634.

[20] Iacobelli F, Culotta A. Too Neurotic, Not too Friendly: Structured Personality Classification on Textual Data[C]// The Workshop on Icwsm. 2013.

[21] Mayuri Pundlik Kalghatgi, Manjula Ramannavar, and Nandini S. Sidnal. 2015. A neural network approach to personality prediction based on the bigfive model. International Journal of Innovative Research in Advanced Engineering (IJIRAE), 2(8):56–63.

[22] Su M H, Wu C H, Zheng Y T. Exploiting Turn-Taking Temporal Evolution for Personality Trait Perception in Dyadic Conversations[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2016, 24(4):733-744.

[23] Ling W, Luís T, Marujo L, et al. Finding Function in Form: Compositional Character Models for Open Vocabulary Word Representation[J]. Computer Science, 2015:1899-1907.

[24] Kosinski M, Stillwell D, Graepel T. Private traits and attributes are predictable from digital records of human behavior[J]. Proceedings of the National Academy of Sciences of the United States of America, 2013, 110(15):5802-5.

[25] Majumder N, Poria S, Gelbukh A, et al. Deep Learning-Based Document Modeling for Personality Detection from Text[J]. IEEE Intelligent Systems, 2017, 32(2):74-79.