A PERSONALITY PREDICTION METHOD OF WEIBO USERS BASED ON PERSONALITY LEXICON

Yuanyuan Feng and Kejian Liu

Department of Computer and Software Engineering,
Xihua University, Chengdu, China

ABSTRACT

Personality is the dominant factor affecting human behavior. With the rise of social network platforms, increasing attention has been paid to predict personality traits by analyzing users' behavior information, and pay little attention to the text contents, making it insufficient to explain personality from the perspective of texts. Therefore, in this paper, we propose a personality prediction method based on personality lexicon. Firstly, we extract keywords from texts, and use word embedding techniques to construct a Chinese personality lexicon. Based on the lexicon, we analyze the correlation between personality traits and different semantic categories of words, and extract the semantic features of the texts posted by Weibo users to construct personality prediction models using classification algorithm. The final experiments shows that compared with SC-LIWC, the personality lexicon constructed in this paper can achieve a better performance.

KEYWORDS

Personality Lexicon, Machine Learning, Personality Prediction.

1. INTRODUCTION

With the rapid development of the Internet, the number of users using social network platforms is gradually increasing. As one of the largest Chinese social network platform, Weibo has a huge mount of users who will express and share their feelings, expectations and experiences on Weibo. These information not only reflects the status of users, but also may affect the spread of social public opinion.

Psychology study shows that [1] , the way an individual speaks and writes often reflects his or her personality. A person's mindset is based on behaviour, emotion, psychology and motivation, which are collectively called personality and have a great influence on individual behaviour. Weibo contains a large number of texts which can well reflect the psychological activities and personality characteristics of users. For example, people with high extroversion tend to use more words related to positive emotions, while people with high neuroticism tend to use more words with negative emotions [2]. Although there are extensive researches on personality prediction of microblog users at present, most of them are conducted on the basis of users' behaviours. There are not many researches on the construction of special microblog personality lexicon. Therefore, in this paper, the author proposes a personality prediction method based on microblog text. In order to explain the differences of different personalities, this paper adopts the personality analysis method based on lexicon. As a psychological lexicon, Linguistic Inquiry and Word Count (LIWC) [2] is a good choice, but it is not very effective in personality identification. On

David C. Wyld et al. (Eds): NLP, MLTEC, CLBD, SEAPP, NeTIoT, VLSIE, ITCS, ARIA, DMS - 2021
pp. 149-159, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.112312

the one hand, the lexicon is more suitable for some formal documents, perhaps because the Weibo version is short and highly colloquial. On the other hand, the lexicon lacks specificity and is not specifically used for personality calculations. Therefore, this paper proposes a method to construct personality lexicon based on “Big Five” model for microblog text. Firstly, word term frequency-inverse document frequency (TF-IDF) is used to extract keywords. Then the clustering algorithm and Word2vec were combined to divide different clusters according to semantics, and then the personality lexicon was formed by word expansion on this basis. Finally, Support Vector Machine (SVM), Random Forest (RF) and Naive Bayes (NB) algorithm are used to achieve personality prediction.

2. RELATED WORK

2.1. Evaluation Methods

There are four main methods to assess personality: questionnaire, dictionary-based and machine learning-based. Questionnaires are the most intuitive method and are designed to ask participants to rate their own behaviour using Likert scales. The most popular questionnaires include the revised NEO-Personality Scale (240 items), NEO Five-factor Scale (60 items) and BFI-44 Big Five Scale (44 items), which are designed for the Big Five personality traits[3]. Although questionnaire is a commonly used method to assess personality traits, participants may be reluctant to fill in the questionnaire or fill it out randomly because some questions involve private information. Therefore, the number of questionnaires collected by this method is usually small and the quality is difficult to guarantee. Therefore, some new methods of predicting human personality are needed. The dictionary-based approach is to detect semantic similarity in the text, that is, semantic similarity equals score. The similarity between different words can be calculated to predict personality. LIWC is a dictionary used to analyse English texts, but Chinese and English have very different grammatical rules. To solve this problem, Liu Qiu et al.[4] determined speech classes and factor structures related to personality traits by analysing the expression modes of Chinese users, and compared with LIWC, they found that in the expression of personality, There are great differences in functional words used in Chinese and English contexts. The method of personality recognition based on machine learning mainly trains the model according to the extracted personality characteristics, so as to achieve personality prediction. As most personality prediction ignores the relationship between different personality traits, individual personality prediction fails to achieve ideal results. In this regard, Gao et al.[5] took different personality traits as a quantitative whole and constructed a multi-objective regression model based on GBDT-Multitarget stacking and BP neural network. The correlation between personality traits was incorporated into the model calculation to predict the entire personality structure of users. Shu Xiaomin et al.[6] predicted Personality based on RAkEL-PA(RAkEL-Personality Analysis), an improved model of RAkEL(Multi-label integration method - random K-label set), and the results showed that there were more people with multiple Personality traits. It suggests that personality traits depend on each other. NavonilMajumder et al.[7] proposed a method of using convolutional neural network (CNN) to extract personality characteristics from articles on stream of consciousness and train different models for different personality traits, so as to achieve the effect of personality recognition.

2.2. Personality Model

Myers Briggs Type Indicator (MBIT) [8] is a popular personality model that has a large audience and is widely used in enterprise human resources. This model describes personality from four dimensions: introversion and extroversion, sense and intuition, logic and emotion, judgment and cognition. According to the scores of different dimensions, MBTI is divided into 16 personality types. However, due to the overly idealistic assumption, MBTI has long been controversial in academic circles, so there are not many in-depth studies on MBTI at present [9] .

The Big Five Model is widely used in predicting personality in the field of psychology and has academic significance [10, 11]. This Model describes personality from Five dimensions: The common cause is openness, conscientiousness, extroversion, agreeableness, neuroticism. Openness describes a person's cognitive style. Conscientiousness reflects the way in which an individual controls, manages, and regulates his or her own behaviour, assessing the organization, persistence, and motivation of an individual in goal-directed behaviour. Extroversion describes the degree of interpersonal interaction and the need for and response to external stimuli. Agreeableness assesses an individual's attitude towards others. Neuroticism reflects the individual's emotional regulation and control. Table 1 shows the performance of each personality traits in general.

Table 1. Personality traits.

Personality Traits	Scores	
	High	Low
Openness	imaginative, aesthetic, creative	pragmatic, compliant, conventional
Extroversion	enthusiastic, lively, good at socializing	implicit, euphemistic, not good at socializing
Agreeableness	trusting, direct, helpful and generous	suspicion, apathy, isolation
Conscientiousness	self-discipline, persistence, achievement, prudence, restraint	laziness, carelessness, weak willpower
Neuroticism	calm, calm, sense of security	vulnerability, depression, insecurity

The Big Five model is the most widely used personality measurement model, no matter in the study of network social behaviour and personality or the study of text and personality. Zhenkun Zhou et al.[12] found that users with lower scores of extraversions tend to post more posts with anger and fear, while users with higher scores tend to post more posts related to sadness. Qiu et al. [13] proved that linguistic features of Twitter can be used to judge agreeableness and neuroticism. Quercia et al. [14] understood the personality characteristics of Twitter users based on the number of followers and followers, and found that popular users and influential users were highly extroverted with relatively stable emotions. In general, the Big Five model is more suitable for academic research, therefore, the author in this paper based on the Big Five personality model to achieve personality prediction.

3. ESTABLISHMENT OF PERSONALITY LEXICON

The overall structure of the personality lexicon construction method for Weibo users in this paper is shown in Figure 1.

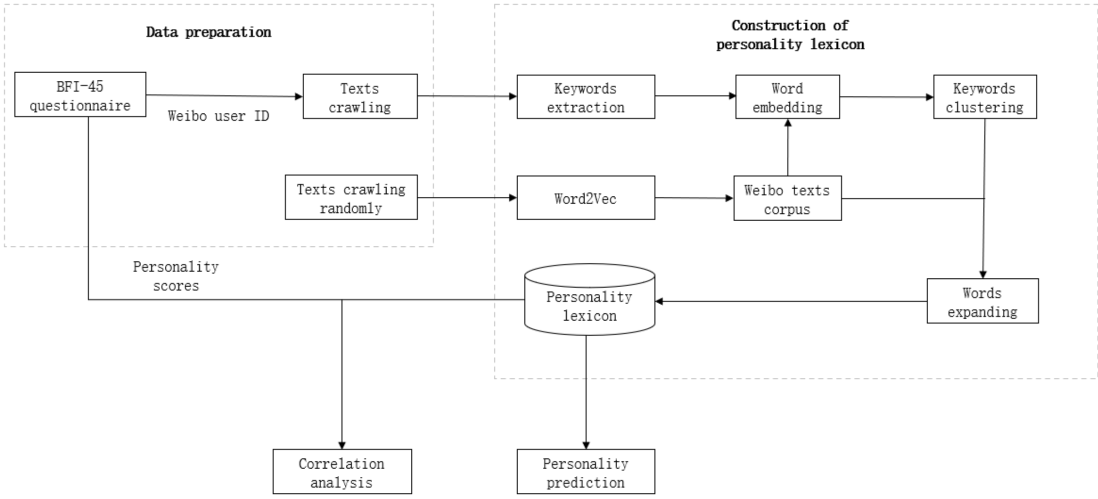

Figure 1. Personality prediction framework

3.1. Data Preparation

3.1.1. Questionnaire Processing

In order to collect users' personality data, the author takes BFI-44 [15] as a questionnaire, which contained a total of 44 questions, including 8 or 9 questions for different personality dimensions. The short questions is convenient for filling in and collecting the questionnaire. In the questionnaire, the user is required to fill in his/her Weibo ID number, so as to facilitate the subsequent retrieval of relevant Weibo text. It took about two months to publish the questionnaire on the Internet. A total of 461 questionnaires were collected, and 379 were valid. Each question was then converted into a score of 1 to 5 on a Likert scale. The personality scoresare shown in Table 2.

Table 2. Personality score statistics.

	Mean	S.D.	L(%)	M(%)	H(%)
A	32.67	4.58	20.69	58.13	21.18
C	27.79	5.08	19.89	57.34	22.77
E	22.85	5.41	22.92	57.19	19.89
N	25.90	4.62	27.83	43.02	29.15
O	30.69	4.90	19.89	57.45	22.66

The personality score obtained from the questionnaire is a continuous variable, which needs to be converted into discrete variables before it can be used in the personality classification model. In this paper, personality scores are converted according to. Table 1 shows the proportions of different personality dimensions.

3.1.2. Data Acquisition and Processing

First of all, crawl the 379 participants' Weibo texts according to their Weibo IDs filled in the questionnaires to form personality dataset. In addition, a Weibo corpus of about 1.3G is constructed by randomly crawling Weibo user texts for subsequent keyword clustering and expansion. Then, the two data sets are cleaned to remove punctuation marks, special symbols and other invalid contents. Finally, use word segmentation tools to cut words.

3.2. Construction of Personality Lexicon

In this paper, the construction of personality lexicon is mainly divided into two steps: extraction of personality keywords and construction of personality lexicon. The first step is to extract keywords related to personality traits from users' texts. The second step is to divide words into multiple clusters using k-means algorithm and Word2vec. By analysing each cluster, the semantic name is given to it. Then, the keywords extracted from each cluster are extended to obtain a personality lexicon containing words of different semantic categories.

3.2.1. Keywords Extraction

In information retrieval, TF-IDF, as the most widely used method, reflects the importance of a word to documents in a textual corpus. It assigns weights to each term in the document based on its term frequency and inverse document frequency. Items with higher weight scores were considered more important. Therefore, the author uses TD-IDF method to calculate the weight of each word in the user's texts, and extracts a certain number of high-weight words. Although TF-IDF algorithm can extract keywords, not all keywords are related to personality. Therefore, this paper uses Chi-square to select the first N keywords related to personality based on users' personality scores.

3.2.2. Construction of Personality Lexicon

In order to construct a personality lexicon, k-means clustering algorithm and Word2vec are used to put keywords with similar semantics into the same cluster. Before clustering, the appropriate number of clustering K should be determined first. For this reason, the author did a series of experiments with K values between 5 and 30, and finally selected K=18, which performed better, as shown in Figure 2. And we list some keywords of categories in appendix at the end of this paper. These keywords are words close to the cluster centre, which can well describe the overall characteristics of each category.

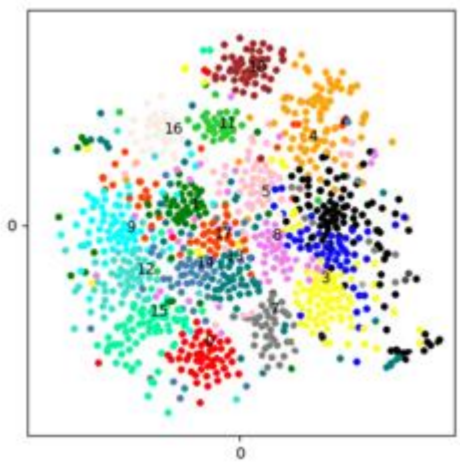

Figure 2. 18 word categories

In appendix, Categories 0 and 11 are related to evaluation, expressing attitudes towards people and things, including positive and negative evaluation; Categories 1 is related to time; Categories

2 is words related to daily life. Categories 3 is related to relationship, including family, friends and strangers; Category 4 is about to places or attractions; Categories 5 is the description related to cognitive process; Categories 6 is blessing words; Categories 7 is related to platform activities, such as Weibo lucky draw, Weibo red packets and activities sharing on other platforms; Categories 8 and 15 describe people's emotional states, including positive and negative emotions. Category 9 relates to physical health, mainly describing body parts and health conditions. Category 10 is related to social events; Category 12 is job-related; Category 13 is related to values; Category 14 relates to school life; Category 16 has to do with sports activities; Category 17 relates to food. The personality lexicon constructed in this paper not only has targeted classification, but also incorporates colloquial words and buzzwords on the Internet, which is of great help to predict the personality of Weibo users.

4. CORRELATION ANALYSIS

In order to understand the correlation between text features and users' personality scores, the author extracts keywords to represents the microblogs with semantic categories or topics. With help of personality lexicon, the author calculates the number of each semantic category of keywords from users' microblogs and take them as the personality lexicon features of the user. Then Pearson correlation is performed based on the personality scores and the personality lexicon features. According to the analysis results, personality traits can be explained from the perspective of text. As can be seen from Table 3, for example, agreeableness is negatively correlated with work, and users with high agreeableness express more blessings to others; Extroversion is positively correlated with relationship, indicating that users with higher extroversion pay more attention to communication with others. Neuroticism is positively correlated with both negative and positive emotions, indicating that users with high neuroticism are emotionally unstable. Openness are positively correlated with locations, cognition and values, indicating the creativity, imagination and exploration ability of them.

Table 3. Correlation coefficient between personalities and categories

Label	Categories	A	C	E	N	O
0	Positive evaluation	0.017*	-0.019	-0.02*	0.014**	0.012
1	Time	0.02*	0.049**	0.053	-0.07	-0.041
2	Daily life	0.081	0.04	0.031	0.049*	0.096
3	relationship	-0.055	0.039	0.053**	0.012	-0.108
4	Locations	-0.07	0.021	-0.073	0.016	0.153***
5	Cognition	-0.085	0.029	-0.001	0.004*	0.052**
6	Blessing	0.081*	-0.048	-0.08	-0.078	-0.013
7	Platform activities	0.015	-0.101	0.012**	-0.012	0.084
8	Positive emotion	0.064	-0.027*	0.065	0.013*	-0.039
9	Health/Body	0.057*	0.072	-0.06*	0.06*	0.101**
10	Social event	-0.105	0.031	-0.024**	0.041*	0.078
11	Negative evaluation	-0.054**	-0.05*	-0.069	0.055**	-0.025
12	Job	-0.026***	0.077**	-0.014	-0.005	-0.08
13	Value	-0.015**	0.021	-0.031*	-0.027	0.044**
14	School life	0.035*	0.021*	0.057**	0.009	0.043
15	Negative emotion	-0.089*	0.032	-0.044	0.018***	-0.073
16	Sports activities	0.129*	-0.11	0.035	0.015	0.057
17	Food	0.041	0.055	0.034	0.029	0.069***

* Denote significance at 10% level; ** Denote significance at 5% level; *** Denote significance at 1% level.

5. PERSONALITY PREDICTION MODEL

From personality questionnaires, we get Big-five personality scores that are represented as high, medium and low grade, and make use of machine learning algorithm for training the five classifiers. Due to deep learning is based on the neural network algorithm, and require a lot of data for training so as to have a better performance. However, through the questionnaire only 379 effective questionnaires were collected, which is not enough for deep leaning training, so the traditional machine learning algorithm is a good choice. Therefore, SVM, RF and NB models are used as classifiers in this paper to make predictions based on the constructed personality lexicon and SC-LIWC [16] , simplified Chinese version of LIWC. The experimental results are shown in Table 4.

Table 4. Comparison of model accuracy.

Lexicon	Model	A	C	E	N	O	Average
Personality Lexicon	RF	**0.736**	**0.710**	**0.551**	**0.600**	**0.603**	**0.641**
	SVM	0.476	0.579	0.457	0.516	0.550	0.516
	NB	0.428	0.474	0.451	0.357	0.517	0.445
SC-LIWC	RF	0.523	0.518	0.491	0.493	0.479	0.501
	SVM	0.453	0.485	0.405	0.398	0.409	0.430
	NB	0.421	0.433	0.394	0.399	0.376	0.405

As can be seen from Table 4, the highest average accuracy of SC-LIWC is 0.501 of RF, and the lowest average accuracy of NB. The average accuracy of the personality lexicon constructed in this paper is 0.641 and 0.445, indicating that the personality lexicon is more effective for personality prediction. Compared with the different models of the two lexicons, the accuracy of RF is higher than the other two models, indicating that RF combined with the lexicons can improve the accuracy of personality prediction. In terms of personality dimension, RF model based on personality lexicon has the best performance, with relatively high accuracy of agreeableness and conscientiousness, reaching 0.736 and 0.710 respectively, which maybe because users with these two personality traits published a large number of microblog texts and learned more text features during training.

All in all, the personality lexicon proposed in this paper has a good applicability to Weibo. Although SC-LIWC is the simplified Chinese version of the lexicon, it still retains the grammatical and semantic features of English, leading to not good results in the analysis of Chinese microblog. In addition, the Weibo as Chinese social networking platform, users are usually published some texts of colloquial and network new words emerge in endlessly, SC-LIWC hasn't these special words. And the lexicon in this paper is build based on Weibo texts, more specific and detail, achieving a better performance when predicting personality.

6. CONCLUSIONS

Based on the "Big Five" personality theory, this paper first constructed a personality lexicon for Weibo texts to explain different personality characteristics from the perspective of text. Then, machine learning algorithms RF, SVM and NB are combined to achieve personality prediction of microblog users. The final experimental results show that the combination of RF model and personality lexicon proposed in this paper can achieve good results.

On social platforms, emojis in the text posted by users will also reflect a person's personality characteristics[17] . But due to the small amount of text containing emojis, we didn’t considered emojis in this paper. In the future, we will collect more texts with emojis, which can be added to establish a personality expression lexicon to improvethe accuracy of personality prediction.

APPENDIX

Table 5. Some keywords in categories.

Label	Name	Keywords
0	Positive evaluation	优秀/excellent，节奏/rhythm，性格/character，不错/good，外向/outgoing，完美/perfect，印象/impression，养成/form，表现/performance，理智/rational，不愧/worthy，成熟/mature，忍耐/patient，用功/hardworking，冷静/calm，良好/not bad，不含糊/unambiguous，大开眼界/eye-opening，肯吃苦/willing to bear hardships
1	Time	周一/monday，周末/weekend，早上/morning，第二天/next day，晚上/night，半夜/midnight，明年/next year，前些天/the day before，假期/holiday，以前/before，三天/three days，以后/after，目前/now
2	Daily life	叙旧/talk about the old days，聊天/talk，碰巧/happen to，路过/pass by，偷跑/sneak away，沏茶/make tea，过寿/celebrate the birthday，握手/shake hands，逛商场/go to the mall，购物/shopping，打折/discounts，酒友/drinking buddies，吸烟/smoking，天气/weather
3	Relation	爸爸/dad，爸比/daddy，姐妹/sisters，爸妈/dad and mom，朋友/friend，伙伴/partner，同事/colleague，对方/the opposite side，好朋友/good friend,，妈咪/mommy，姐姐/sister，爷爷奶奶/grandparents，邻居/neighbor，宝贝/baby，老公/husband，哥哥/brother
4	Locations	上海/Shanghai，南京/Nanjing，呼和浩特/Hohhot，济南/Jinan，泰国/Thailand，烟台/Yantai，列车/Train，武汉/Wuhan，城市/City，九华/Jiuhua，内蒙古/Inner Mongolia，镇江/Zhenjiang，中心/Center，成都/Chengdu，杭州/Hangzhou，北京/Beijing，美术馆/Art Museum
5	Cognitive	理解/comprehend，选择/choose，质疑/question，不解/wonder，迷惑/puzzle，搞清楚/make clear，明白/figure out，意识到/realize，渐渐/gradually，懂得/understand，熟悉/familiar，知道/know，英雄所见略同/great minds think alike
6	Blessing	祝愿/wish，生日快乐/happy birthday，迎接/to meet，鸿运/good luck，愿望/wishes，顺利/go smoothly，花好月圆/blooming flowers and full moon，家庭幸福/family happiness，牛年大吉/Happy Chinese New Year，红火，欢庆/celebrating，祝福/blessing，美好/beautiful
7	Platform activities	助力/help，投票/vote，超级/super，爱豆/idol，人气/popularity，投出，大赏/big reward，战队/team，直播/Live，盛典/grand ceremony，第一波/the first wave，喜获/congratulations to obtain，首档节目/the first TV show，红包/red packet，现金/cash，收到/received，抽奖/draw，关注/focus on，大奖/reward，惊喜/surprise，机会/opportunity，福袋/bless

		ing bag，集齐/collect all，门票/tickets，礼包/gift bag，会员/membership，免费/free，等级/level
8	Positive emotion	加油/fighting，人生/life，努力/effort，未来/future，生命/life，幸福/happiness，夏天/summer，感谢/thanks，力量/power，美好/beauty，能量/energy，青春/youth，拥有/own，感受/feelings，再见/goodbye，心中/In the heart，路上/on the road，长大/grow up，少年/teenager，感动/moved，值得/worth，期待/look forward to
9	Health/Body	闹肚子/stomach trouble，保健/health care，牙龈/gums，桑拿/sauna，按摩/massage，维生素/vitamins，布洛芬/ibuprofen，买药/buy medicine，止疼药/painkillers，皮肤科/dermatology，头孢/cephalosporin，甲沟炎/parchitis，疼死/really painful，腰疼/backache，酸痛/ache，头晕眼花/dizziness，手指/fingers，小腿/leg，腹肌/abs，脑子/brain
10	Social events	劫持人质/hostage-taking，罹难/death，染病/illness，避让/avoidance，跳窗/jumping out of a window，重创/trauma，重判/severe sentence，识破/see through，下药/drugging，恶性事件/vicious incident，急救车/ambulance，受害者/victim，性骚扰/sexual harassment，侮辱性/abuse，保姆/babysitter，罚跪/be punished to kneel，打骂/beat and scold，离婚冷静期/cooling-off period before divorce，捐献/donation，家暴/domestic violence
11	Negativeevaluation	讨厌/dislike，阴阳怪气/speak in a voice dripping with sarcasm，傻逼/sucker，恶心/disgusting，生气/angry，可恶/damn，玩笑/joke，不配/don’t deserve，卸载/uninstall，得寸进尺/The more one gets, the more one wants，自导自演/self-speech，无礼/rude，缺心眼儿/stingy
12	Job	会议/meetings，交接/handover，人力资源/human resources，小组/groups，迟到/tardiness，上班/to work，工资/salary，五险一金/five social insurance and one fund，退休/retirement，加班/work overtime，老板/boss，助手/assistant，业绩/KPI
13	Value	诚信/integrity，社会/society，为荣/honor，为耻/shame，国家/country，精神/spirit，文化/culture，权利/rights，清朗/Qinglang，文化观/cultural outlook，集体主义/collectivism，理想信念/ideals and beliefs，道德水准/moral standards，引导/guidance，纪律/discipline，守法/law-abiding
14	School life	大学生/college student，老师/teacher，学校/school，学习/study，作业/homework，论文/paper，同学/classmate，考研/prepare for postgraduate exams，开学/，上课/go to a class，毕业/graduation，考试/exam，宿舍/dormitory，公共课/optional course，考题/examination questions，没考上/failed in the exam，六级/College English Test-6
15	Negative emotion	事情/things，情绪/emotion，难过/sad，经历/experience，越来越/more and more，心情/mood，害怕/fear，想到/think of，失望/disappointment，身边/side，内心/inner，痛苦/pain，

		也许/maybe，放弃/give up，焦虑/anxious，悲伤/sorrow
16	Sports activities	国足/national football team，女篮/women's basketball team，乒乓球/table tennis，奥运会/The Olympic Games，锦标赛/championship，打球/play a ball game，跳水/diving，跑步/running，裁判/referee，场地/site，夺冠/take the crown，冠军/champion，金牌/gold medal，银牌/silver medal，口哨/whistle
17	Food	青稞/highland barley，牛奶.milk，味道/taste，鲈鱼/perch，火锅/hotpot，米饭/rice，橘子/orange，汉堡/hamburger，鸡蛋/egg，真香/really fragrant，蛋糕/cake，咖啡/coffee，好吃/delicious，海鲜/seafood，奶茶/milk tea

REFERENCES

[1] G. Stemmler and J. Wacker, "Personality, emotion, and individual differences in physiological responses," Biological Psychology, vol. 84, no. 3, pp. 541-551, 2010.

[2] J. W. Pennebaker and M. E. Francis, "Linguistic inquiry and word count: liwc {software program for text analysis}, 1999.

[3] P. John, L. P. Naumann, and C. J. Soto, "Paradigm shift to the integrative big five trait taxonomy: History, measurement, and conceptual issues", Handbook of personality: Theory and research (3rd edition), 2008.

[4] Q. Lin, L. Han, J. Ramsay, and Y. J. Fang, "You are what you tweet: Personality expression and perception on Twitter," vol. 46, no. 6, pp. 710-718, 2012.

[5] G. J. B, "Users’ Personality Prediction Model Based on Multi-Target Regression %J International Journal of Computational and Engineering," vol. 4, no. 4, pp. 25-28, 2019.

[6] S.Xiaoming and M.Xiaoning, " Research on user personality Analysis Model based on microblog text " , Software Guide, vol. 19, no. 11, pp. 25-28, 2020.

[7] N. Majumder, S. Poria, A. Gelbukh, and E. Cambria, "Deep Learning-Based Document Modeling for Personality Detection from Text." IEEE Intelligent Systems, vol. 32, no. 2, pp.74-79, 2017.

[8] Myers and M. Mccaulley, Mbti Manual: A Guide to the Development and Use of the Myers - Briggs Type Indicator. Consulting Psychologists Press, 1985.

[9] D. J. Pittenger, "Cautionary comments regarding the Myers-Briggs Type Indicator," Consulting Psychology Journal, vol. 57, no. 3, pp. 210-221, 2005.

[10] L. Goldberg and R. Lewis, "The development of markers for the Big-Five factor structure," Psychological Assessment, vol. 4, no. 1, pp. 26-42, 1992.

[11] S. D. Gosling, P. J. Rentfrow, and W. B. Swann, "A very brief measure of the Big-Five personality domains," Journal of Research in Personality, vol. 37, no. 6, pp. 504-528, 2003.

[12] Z. Zhou, K. Xu, and J. Zhao, "Extroverts Tweet Differently from Introverts in Weibo," EPJ Data Science, vol. 7, no.1, pp.18, 2018.

[13] L. Qiu, J. Lu, J. Ramsay, S. Yang, W. Qu, and T. Zhu, "Personality expression in Chinese language use," Journal of Research in Personality, vol. 52, no. 6, pp. 463-472, 2017.

[14] D. Quercia, M. Kosinski, D. Stillwell, and J. Crowcroft, "Our Twitter Profiles, Our Selves: Predicting Personality with Twitter," in IEEE Third International Conference on Privacy, 2012.

[15] O.P. John, S. Strivastava," The Big-Five trait taxonomy: History, measurement, and theoretical perspective," in L. A. Pervin and O. P. John(Eds.), Handbook of persinality: Theory and research, vol.2, pp. 102-138, 1999.

[16] C. L. Huang, C. K. Chung, N. Hui, Y. C. Lin, and J. W. Pennebaker, "Development of the Chinese linguistic inquiry and word count dictionary," Chinese Journal of Psychology, vol.54, no. 2, pp. 185-201, 2012.

[17] D. Marengo, F. Giannotta, M. J. P. Settanni, and I. Differences, "Assessing personality using emoji: An exploratory study," Personality & Individual Differences, vol. 112, no.c, pp. 74-78, 2017.

AUTHORS

Yuanyuan Feng is in department of Computer and Software Engineering, Xihua University, Chengdu, China. She is currently working on natural language processing, particularly ontexts classification.

Kejian Liu is a professor in department of Computer and Software Engineering at Xihua University. He is working on Computer network and wireless network technology, intelligent information processing technology, high-performance computing technology.