Predicting User Personality with Social Interactions in Weibo

Yuting Jiang

Aalto University, Helsinki, Finland

Shengli Deng

School of Information Management, Wuhan University, Wuhan, China

Hongxiu Li

Tampere University, Tampere, Finland, and

Yong Liu

Aalto University, Helsinki, Finland

Abstract

Purpose - The purposes of this paper are to 1) explore how personality traits pertaining to the dominance influence steadiness compliance model manifest themselves in terms of user interaction behavior on social media; and 2) examine whether social interaction data on social media platforms can predict user personality.

Design/methodology/approach - Social interaction data was collected from 198 users of Sina Weibo, a popular social media platform in China. Their personality traits were also measured via questionnaire. Machine learning techniques were applied to predict the personality traits based on the social interaction data.

Findings - The results demonstrated that the proposed classifiers had high prediction accuracy, indicating that our approach is reliable and can be used with social interaction data on social media platforms to predict user personality. "Reposting," "being reposted," "commenting," and "being commented on" were found to be the key interaction features that reflected Weibo users' personalities, whereas "liking" was not found to be a key feature.

Originality/value -The findings of this study are expected to enrich personality prediction research based on social media data and to provide insights into the potential of employing social media data for the purpose of personality prediction in the context of the Weibo social media platform in China.

Keywords

Social interaction, personality, DISC, social media, Weibo

Introduction

"Personality" refers to a stable set of characteristics that can induce individual tendencies toward thoughts and behaviors (Maddi, 1989). Identifying the personality of an individual not only helps in understanding his or her potential needs in different contexts, it also provides hints about how the individual might respond to different situations. Personality prediction has been touted as extremely useful in behavioral research in diverse areas, such as information systems (IS), information management (IM), psychology, sociology, and marketing (Bansal *et al*., 2016; Deng, Liu *et al*., 2013).

Social media has dramatically changed the way people interact with each other in present-day society (Kim and Kim, 2018; Liu *et al*., 2018; Stavros *et al*., 2014). People utilize social media to share their opinions, feelings, and thoughts about different subjects and to advertise their activities on social media platforms, such as Facebook and Twitter. As a result, a large volume of data pertinent to human behavior, including social interaction, is available from social media platforms (Gloor *et al.*, 2013). Pan, Chen, and Guo (2016) argued that the social interaction data (,i.e. the cumulative amount of interaction between users like "liking number" and "commenting number") available through social media platforms provides the possibility of identifying personality traits in addition to the traditional approaches to personality research based on interviews and survey questionnaires. Moreover, according to Ortigosa *et al*. (2014), as personality is stable over time and in changing circumstances, an individual's personality traits will be evident in both their real and virtual lives, although they might manifest in different ways. Therefore, social media offers a reasonable approximation of the offline lives of social media users.

The availability of social media data has recently attracted the attention of researchers. For example, researchers have employed social media data to predict users' personalities for the purpose of understanding their behavior (Ong *et al*., 2017; Park *et al*., 2015). Although it has been argued that social interaction data from social media platforms is a useful resource for predicting personality traits (Ortigosa*et al*., 2014), limited studies have actually employed such data for personality prediction. Instead, prior research has primarily been based on users' textual messages exchanged over social media platforms to predict their personalities, overlooking the importance of social interaction data. In addition, prior studies have primarily focused on Facebook and Twitter, to the exclusion of other popular platforms. It has been argued that behavior on social media is influenced by users' cultural backgrounds (Men, 2015). Thus, it becomes necessary to validate whether social media data relating to users with different cultural backgrounds can also be used for personality prediction research.

Notably, while different personality models have been employed to predict personality traits mainly based on a limited amount of self-reported survey data, little research has delved into social interaction data to predict personality based on the dominance, influence, steadiness, compliance (DISC) personality model (Karasek, 1979). Likewise, previous personality prediction studies have tended to focus on the use of the Big Five Model (McCrae and Costa, 1987), a personality measurement model, to measure user personality based on social media data. Indeed, the Big Five Model and the DISC personality model are both the dominant theories in the area of personality traits. However, it remains unclear whether personalities measured by the DISC model can be modeled via social media data. As a result, there is a need for research that explores personality prediction with social interaction data based on different personality model with different cultural contexts to clarify the prediction accuracy of social interaction data.

To address this identified research gap, this study proposed a research model that integrated the DISC personality model and a social media personal interaction model

(SMPIM). The model was empirically tested with the social interaction data collected from 198 Weibo users and their personality trait data as collected via questionnaire. Five different machine learning techniques were applied to predict the users' personality traits. The objective of this research was to develop a research model that can be used to predict personality based on social interaction data with machine learning techniques. Based on this, we propose three research questions

1. Whether personalities measured by the DISC model can be modeled via social media data?
2. Whether personality on social media with Chinese cultural background can be effectively predicted through social interaction data?
3. How to build a dynamic social media personal interaction model?

In this way, this study offers several contributions. For theoretical implication, first, this paper verifies the feasibility of DISC personality modeling for predicting personality. To be specific, the main personality measurement models currently used are the Big Five Personality Measurement Model and the MBTI Model, while there is a lack of research on the effectiveness of DISC, another major personality measurement model. Making up this gap can provide a wide range of applicability for personality prediction based on social media data. Secondly, this study provides a theoretical basis for the similarity in the expression of social interaction information under different cultural backgrounds. To be specific, the current research on personality prediction based on social media data mainly focuses on social platforms with European and American cultural backgrounds such as Facebook and Twitter, while there is a lack of research on platforms with Asian cultural backgrounds such as Weibo in China. Since users from different cultural backgrounds may have different expressions of their personalities on social media, filling in the research gaps of Asian cultural backgrounds is helpful to provide effective evidence for the consistency of personality expressions from different cultural backgrounds. Finally, this study provides a further reference to the literature on personality modeling, personality prediction, cultural, and social interaction. For practical implication, firstly, this paper constructs a dynamic personality prediction model. Compared with the existing models, this paper considers the effect of time intervals on personality prediction and provides a new insight for understanding personality prediction based on social media interaction data from the perspective of interaction time. In addition, we propose a structured feature screening method that can help researchers obtain important features that are difficult to observe. In addition, we found that the combination of the WrapperSubsetEval feature selection algorithm and 1-Nearest Neighbor classification algorithm can achieve the highest prediction accuracy, which provides a reference for feature selection engineering of personality prediction. Finally, this study also provides practical guidance for social media service providers. Both the social interaction activities and the number of social interactions on social media platforms can reflect the personality characteristics of users, which provides an effective analysis method for a personalized recommendation.

The paper is organized as follows: the next section provides an overview of the commonly used personality models, online social interaction behavior, and personality prediction research using social media data. Next, a detailed description of the proposed research approach is presented. Subsequently, a description of the data collection, the personality classifiers, and the data analysis is introduced, followed by a discussion on research findings. Finally, the paper highlights the contribution of the research, points out the potential limitations of the study, and makes suggestions for future research.

Literature Review

Personality research

Personality can be defined as a set of attributes that characterize an individual's behavior, emotions, temperament, and mind (Mairesse *et al*., 2007). Personality research has been widely applied to explain different phenomena, such as probability of illness (Kotov *et al*., 2010), social hazard level (Edens *et al*., 2008), cognitive ability (Griffin *et al*., 2015), occupational performance (Zhao *et al*., 2015), and social network use (Chen *et al*., 2016). In the literature, the three most popular models for structuring personality are the Big Five Model, the Myers Briggs Type Indicator (MBTI; Briggs-Myers and Briggs, 1985), and the DISC personality model.

The Big Five Model classifies personality into five personality traits, including openness to experience, conscientiousness, extroversion, agreeableness, and neuroticism (McCrae and Costa, 1987). The Big Five Model emphasizes the stability of personality traits and has been considered as one of the most important general models for describing personality structure and measures (Hall *et al*., 2010).

The MBTI model is popular in the field of occupational assessment. According to the MBTI, people behave differently and their behaviors can be classified into a fixed number of categories (Boyle, 1995). The MBTI model suggests four categories of people behavior, each of which is composed of two opposite poles: extraversion-introversion (E-I), thinking-feeling (T-F), sensing-intuition (S-N), and judging-perception (J-P; Briggs-Myers and Briggs, 1985; Boyle, 1995). The MBTI model offers a forced-choice, self-reporting measure to describe the patterns of mental activity and personality of people in obtaining information, making decisions, and dealing with life (Stumpf and Parker, 2000).

Likewise, the DISC personality model offers an important model for predicting behavioral tendencies of people with different personality traits when facing different situations (Beamish, 2005; Bell *et al*., 2012). Behavioral tendency is an outcome of the interaction between context and personality, which reflects a person's thinking and behavior patterns. The DISC personality model includes four main measures of personality traits, which are dominance, influence, steadiness, and compliance. People with the dominance trait place an emphasis on shaping the environment by overcoming opposition to accomplish results, while those with the influence trait place an emphasis on shaping the environment by influencing or persuading others. Steadiness highlights cooperating with others within existing circumstances to carry out tasks. Compliance refers to working conscientiously within existing circumstances to ensure both quality and accuracy. The DISC model provides a basic framework for people to understand themselves and adapt their behaviors toward others in different contexts, such as within a work team (Reynierse *et al*., 2000), in a sales relationship (Deviney *et al*., 2010), in a leadership position (Bouchard, 2018), or in other relationships (Ogunyemi *et al*., 2011).

Mcelroy *et al*. (2007) argued that the current IS usage research is mainly based on the concept of a personalized user (e.g., perceived usefulness) rather than a personality-based concept. Personality traits should be taken into account in future IS usage research because personality is a more advanced predictor of IS use than cognitive variables (Mcelroy *et al*., 2007). In recent years, personality traits have attracted the attention of IS researchers, and some research has adopted the above three personality models to predict user personality based on survey data (Ortigosa *et al*., 2014). Although social media platforms have accumulated a huge amount of user interaction data, there is still a lack of personality prediction research on social media users based on social interaction log data in the virtual world. There is a call for research to use social media data in personality prediction research (Lima and de Castro, 2014).

Arguably, the application of different theoretical models in personality prediction using social media data is meaningful through an incorporation of machine learning methods. This endeavor would offer further evidence with regard to personality prediction accuracy using social interaction data, especially when prior research has mainly been based on use of the Big Five Model. We deem that a knowledge of whether personalities pertinent to the DISC model can be reliably modeled would offer new insights toward the applicability of this method in different contexts of personality prediction.

Online social interaction behavior

Online social interaction, also known as network interaction, refers to users' online interaction behavior on a network platform via information exchange with each other (Chen and Lin, 2018; Wiertz and Ruyter, 2016). Online social interaction has become an important part of individuals' lives, playing a key role in support communications between media, people, and society in the Internet era (Blazevic *et al*., 2014).

From a macro perspective, online social interaction in a social network environment includes different interaction activities, such as human-human interaction, human-computer interaction, and human information interaction. Human-computer interaction can be defined as "an exchange of information between participating agents through sets of information channels (interfaces) … where each has the purpose of using the exchange to change the state of itself or one or more others" (Storrs, 1994). Online social interaction has been found to benefit the development of human relationships, such as through building trust, establishing friendship, and promoting interpersonal interaction (Bock *et al*., 2005). Fidel (2012) suggested using human information interaction as an umbrella to bring many fields and subfields dealing with humans, information, and technology into a unified meta-discipline.

From a micro perspective, social interaction consists of different activities in the social media context, including following, reposting, commenting, liking, and sharing (Allen *et al*., 2014; Lu *et al*., 2014). Taking Facebook for instance, people make connections with friends, relatives, colleagues, and others whom they would like to keep in touch with. Facebook users can manipulate their personal "timeline" at will; "like" the status, photos, or comments of others; "subscribe" to public posts by others without adding him/her as a

Facebook contact; and share information with others. Twitter is less social friendship-oriented than Facebook. Twitter users can also subscribe to other users' tweets (similar to "following"), "like" tweets, update their profiles, and have items forwarded by other users to their own feed ("retweeting").

Personality prediction based on social media data

The availability of huge amounts of social media data provides opportunities for researchers to investigate social media user behavior and personality traits. Some scholars have attempted to use social media data to predict user personality and to explore users' behavior patterns and tendencies, and the relationships between user personality and such behavior patterns (Lee *et al.*, 2014; Tandera et al., 2017; Tsai *et al.*, 2017). Lee *et al.* (2014) have explored the relationship between user personality and self-presentation based on data from Facebook users, and found that extroversion is positively related both to self-presentation-related social interaction on Facebook walls, and also to commenting and sharing behaviors on Facebook. Highly competitive narcissists often update their statuses and positions on their walls, whereas neurotic and cautious people are less likely to write comments (Lee *et al.*, 2014). Golbeck *et al.* (2011) have applied the Big Five Model to make personality predictions based on analyzing the message texts and subjects posted on Facebook via machine learning, and the research outcome shows a high level of accuracy in predicting Facebook users' personalities with Facebook data. Tandera *et al.* (2017) also applied the Big Five Model to predict personality based on Facebook user information. Tsai *et al.* (2017) conducted a 2-phase experiment among 111 university students for 2 months to investigate the usage patterns of Facebook users based on the Big Five Model. In their research, they found that extraverts are associated with higher levels of using Facebook functions, such as wall posting and messaging, as well as with more interaction with real-life friends and less with strangers, whereas users with low emotional stability are more likely to participate in events with family and relatives, and to stop using apps than those with high emotional stability. Table 1 summarizes the recent studies on personality prediction based on social media data.

Author	Data	Data analysis techniques	Theory	Research findings
Golbeck, Robles, & Turner (2011)	The text and subjects posted on Facebook and personality data	Machine learning	The Big Five Model	Social media behavioral features on Facebook are useful in predicting personality and performance.
Lee, Ahm, & Kim (2014)	Self-presentational behavior data and personality data	Hierarchical multiple regression analyses; Correlation analysis	The Big Five Model	Extroversion is positively related to self-presentation-related social interaction on Facebook walls and commenting and sharing on Facebook. Highly competitive narcissists often update their statuses and positions on their walls, whereas neurotic and cautious people are less likely to write comments.

Ortigosa, Carro, & Quiroga (2014)	Social interaction data from Facebook and personality data	Machine learning	The alternative five model	The proposed research approach shows a high level of prediction accuracy. Social media data are a reliable data source for personality prediction.
Tandera, Hendro, Suhartono, Wongso, & Prasetio (2017)	The personality data and Facebook status data of 250 Facebook users and the status data of 150 Facebook users	Machine learning	The Big Five Model	The research results show high accuracy of personality prediction with an average accuracy of 74.17%.
Tsai, Chang, Chang, & Chang (2017)	Facebook usage data and personality data	Correlation analysis; Difference analysis	The Big Five Model	Extraverts are associated with higher levels of using Facebook functions, such as wall posting and messaging, as well as more interaction with real-life friends and less with strangers, whereas users with low emotional stability are more likely to participate in events with family and relatives, and to stop using apps than those with high emotional stability.

Table 1: Literature review of personality prediction based on social media data here

Past studies of personality prediction using social media data have mainly focused on the two most popular social media platforms, Facebook, and Twitter. In this study, Weibo, a popular social media platform in China, was selected as the research context. Evidently, applying the log data-based personality prediction research to user groups with a different cultural background will extend the scope of the research on personality prediction and offer new evidence concerning the generalizability of the research findings.

Furthermore, the research adopted the DISC personality model as the basic theoretical framework for personality prediction. We constructed a model based on the user features and user interaction features on Weibo, via different feature selection algorithms and classification algorithms, to perform personality prediction.

Method

This paper aims to solve three research problems. First, we sent questionnaires on personality measurement to users, and in order to solve research question 1, we chose the DISC personality measurement model. In addition, we asked whether we could get their social media links. In order to solve the research question 2, we chose Weibo with Chinese cultural background as the research platform. The questionnaire data included users' personality types, while the microblog data included users' social interaction data. Then, in order to solve the research problem 3, we built a dynamic social media personal interaction model. Finally, we used a combination of multiple cross-cutting feature selection methods and classification algorithms to predict the user's personality through social interaction data, which provided the final research results for research questions 1 and 2.

Data collection

In this study, we collected both users' interaction data via social media (Weibo) and users' personality data using a questionnaire survey. The questionnaire included some basic demographic information of the respondents and the measures developed based on the DISC model by Marston (Marston, 2013) (see Appendix). The DISC model can not only be used for personality trait prediction, but also for the analysis of the behavioral tendencies of people with different personalities in face of particular situations. At the end of the questionnaire, the respondents were asked to provide their Weibo nicknames or links and permission for the researchers in this project to collect their social interaction data from Weibo. After that, we collected the interactive data of users from their social media accounts.

We sent invitations to some university students at a university in Central China to join in this project via social media channels, and the online link of the questionnaire survey was attached to the invitation. A total of 293 students enrolled at this university participated in the research, and 240 valid questionnaires were obtained. From the 240 respondents, permission to access 198 Weibo links was authorized by users, and the survey data of the 198 respondents were used as a valid data set for the questionnaire on personality.

Based on the Weibo links and Weibo nicknames provided by the 198 respondents, we employed the crawler software to search the microblog data of the 198 Weibo users. Because of the complexity of the Weibo web page and the difficulty of data crawling, in this research, we decided to crawl the data via the mobile version of Weibo. We collected the social interaction log data of all the 198 respondents using the web crawler. We gathered all the social interaction data up until 12:50pm on April 19, 2017, including following lists, followers lists, and all kinds of behavior lists of the 198 Weibo users. Since the number of microblogs for each user was not the same, in this study, we adopted a unified approach and collected only 30 pages of microblogging data per user (the microblogging mobile version is based on the flow of waterfall information). Due to the limitations of the Weibo official Application Programming Interface (API), only the information about the top 100 people following (following) and people being followed (followers) of each user was obtained, and we did not obtain any interaction data regarding the commenting on others and being liked by others.

Sina Weibo user interaction

Sina Weibo is a Chinese website launched in 2009 by Sina Corporation. It is one of the most popular social media platforms in China. At the end of 2018, Weibo had over 446 million monthly active users. Figure 1 shows the increase in its user population over the past six years.

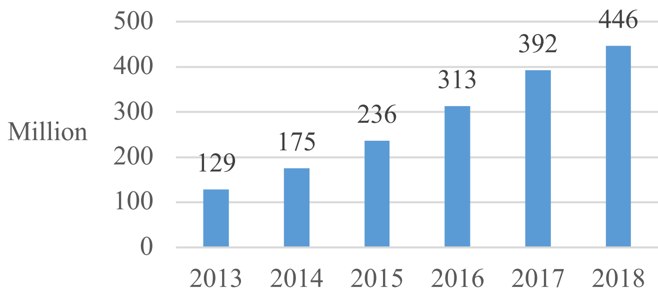

Figure 1: Monthly active users of Weibo over the past six years

Based on the core business of Weibo, we summarized the social interaction types and social interaction relationship types on Weibo (as shown in Figure 2). Regarding the social interaction types, the following exist: like, repost, comment, be liked, be reposted, and be commented on. With regard to the social interaction relationships between the users, the subsequent ones can be identified: following, non-following, self, follower, and non-follower. Table 2 provides an explanation of the terms used to indicate the different interaction parameters.

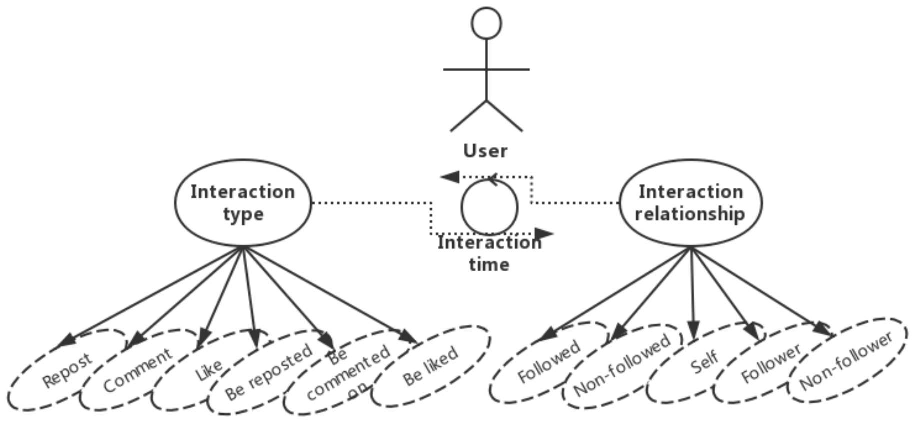

Figure 2: Weibo user interaction structure diagram

Name	Explanation with examples
Like	David liked Henry's photo.
Be liked	Henry's photo was liked by David.
Repost	David reposted Henry's blog.
Be reposted	Henry's blog was reposted by David.
Comment	David commented on Henry's status.
Be commented on	Henry's status was commented on by David.
Following	David followed Henry's Weibo, so Henry is David's following.
Non-following	David didn't follow Henry's Weibo, so Henry is David's non-following.
Self	David liked/reposted/commented on his own blog.
Follower	David followed Henry's Weibo, so David is Henry's follower.
Non-Follower	David didn't follow Henry's Weibo, so David is Henry's non-follower.

Table 2: Terms used to indicate different interaction parameters of Weibo

User interactive model on Weibo

According to the definition of interaction, we find that the interaction behavior consists of three elements: the information channel and the two users connected by the information channel. We extracted the parts of the platform architecture of Weibo (see Figure 3) that are consistent with the concept of interaction as the three principal elements of the SMPIM: actor user, interaction behavior and target user. In addition, the Relationship element in Figure 3 represents the relationship between two interacting users, and different interactions lead to different relationships. The characteristic counts refer to the quantification of interactions.

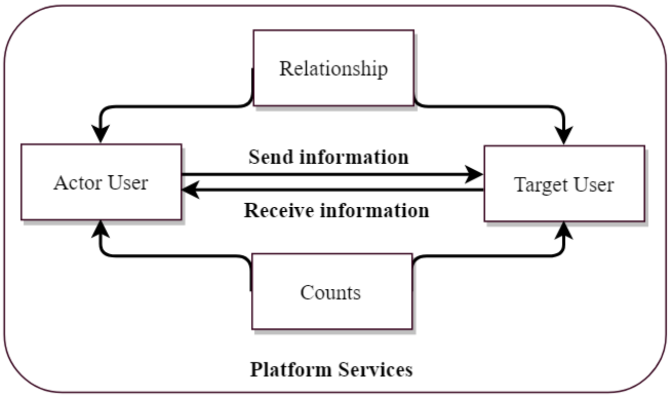

Figure 3: Weibo platform architecture diagram

Interaction is a two-way behavior which involves both the "sender of information" and the "receiver of information." Thus, we divide the interaction behavior into active interaction and passive interaction depending on the difference in the initiator. Active interaction behavior includes three main actions of the user him/herself: liking, reposting, and commenting. Passive interaction behavior consists of the three actions carried out by other users: being liked, being reposted, and being commented on. For example, "User A likes User B's microblog." For User A, the liking behavior is his/her active behavior, whereas User B's microblog is liked by another, and so it is passive behavior for him/her.

Accordingly, interactive users are divided into active interactive users and passive interactive users. An active interactive user refers to the interactive object of the user's active behavior, including followed users, non-followed users, and the user him/herself. A passive interactive user refers to the interactive object of the user's passive behavior, including followers and non-followers. In addition, in order to comprehensively quantify the interaction behavior, we take the "global" variable into account, which refers to the overall feature value of behavior, and is calculated only based on the "behavior," without categorizing it. For example, global variables of active behaviors represent the overall eigenvalues of all active behaviors such as liking, reposting, and commenting.

Based on the social interaction activities on Weibo, we propose a SMPIM for social media users, as shown in Figure 4. In the SMPIM, there are three layers:

a) The "actor-tier" refers to the performer of the interactive behavior;

b) The "behavior-tier" represents the interactions on Weibo from the Weibo user perspective, including passive behavior and active behavior; the interaction behavior is divided into six behavior types based on active and passive ones;

c) The "target user-tier" refers to the participants of the interactive behavior, including the passive target user, active target user, and global user.

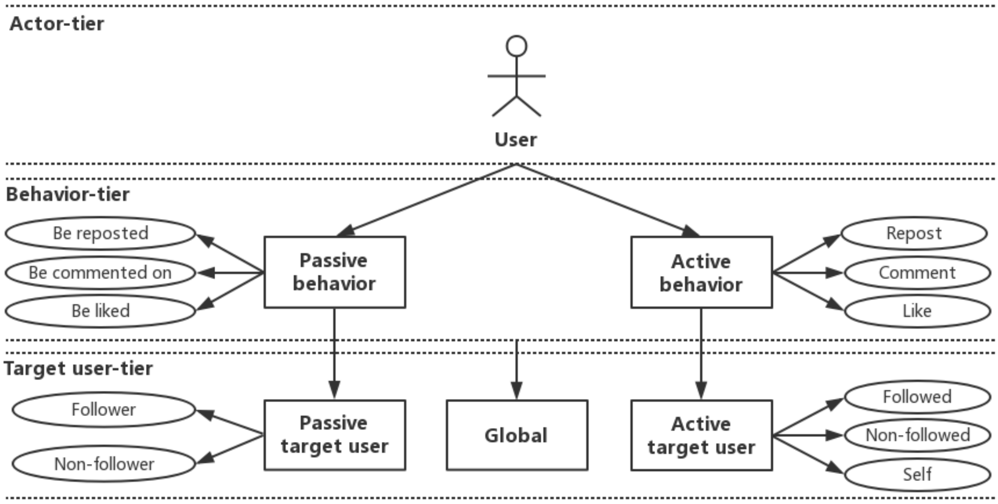

Figure 4: SMPIM framework for Weibo

In the SMPIM framework, only certain combinations of conditions can occur. Table 3 lists the possible combinations of interactions.

Behavior		Following (Af)	Non-following (ANf)	Self (AS)	Follower (Bf)	Non-follower (BNf)
Active Behavior	Like (L)	X	X	X		
	Repost (R)	X	X	X		
	Comment (C)	X	X	X		
Passive Behavior	Be liked (BL)				X	X
	Be reposted (BR)				X	X
	Be commented on (BC)				X	X
Notes: X means there is a combination of the two different user interactions.						

Table 3: Possible user interaction on Weibo

Default interactive feature generation

Feature selection is the first task to carry out when features are used as taxonomies. In terms of feature selection, some researchers have chosen useful ones based on previous research and their own understanding of the classification goal. This method might be inaccurate and subjective and also makes it difficult to find new useful features. In this study, we take the following steps to select the features based on the research context, which helps reduce the cost of manual decisions in feature selection: first, we generate all the default features (DFs) according to the logical structure based on the SMPIM, and then, we automatically select the key features (KFs) that are valid for the classification algorithm from the DFs via some filter approaches.

The logical structure tree that generates the default feature set is designed based on the

SMPIM framework. The root node represents the user in the actor-tier, and the next two layers are consistent with the SMPIM framework. Each path from the root node to the target user node represents an interaction mode. For example, the path of User → L → Fed indicates that a user sends out the active interactive behavior "like," and the recipient of the behavior is the user's follower. There are 21(3*4+3*3) interaction types. In order to quantify the degree of these interactions, we designed three quantitative characteristics, including count, time, and weight. Counting is a feature derived from the architecture diagram of the Weibo platform. Statistical variables measured in this study include basic number (N), average number (AvgN), and standard deviation (StdN). Each social interaction activity happens at a specific time. Previous studies mainly applied static interaction data in personality prediction and have ignored the temporal changes involved in social interaction. For example, users with different personalities might repost the same microblog with different reposting intervals. Thus, in this study, we take interaction times into consideration in our proposed research model, including the maximum time range (MaxT), minimum time range (MinT), average time range (AvgT), and standard deviation of time range (StdT). Finally, due to the different counts and lengths of account records in the sample, we consider converting the numbers (such as basic number [N], average number [AvgN], and standard deviation [StdN]) to weights, which means presenting them as proportions of the parent body (Nw).

In view of the proposed SMPIM, a default feature set (DFS) was designed that includes the largest number of DFs. The DFS was applied to the personality prediction. Figure 5 shows the tree-like logical structure that generates the DFs and visualizes the process of DF generation.

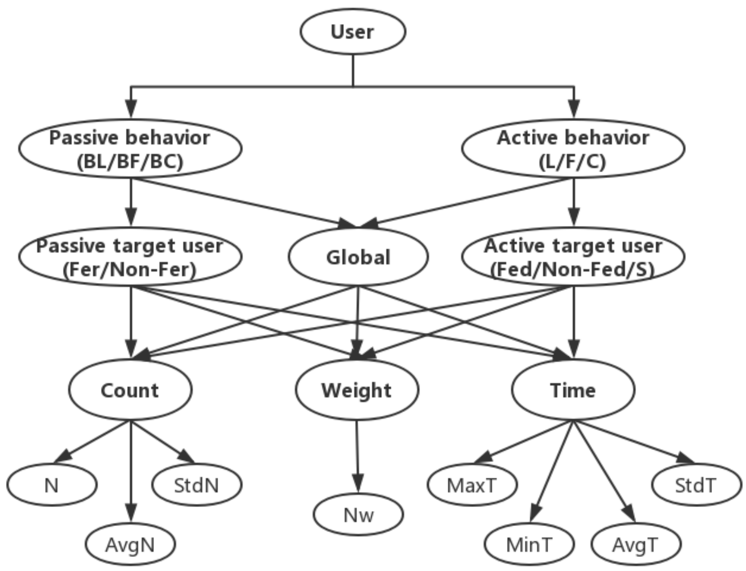

Figure 5: The logic diagram of default interactive features

A logical structure table of the generation process of the DFs is shown in Table 4. Theoretically, we can get 168 ((3*4+3*3)*8=168) default interaction features in total.

Behavior types	Relationship type	Explanation		Abbreviation
Active behavior (Like/Repost/Comment)	Global	Time	The time interval of interactions between the user and target user.	MaxT
				MinT
	Following			AvgT
				StdT
	Non-following	Count	The number of interactions between the user and target user.	N
				AvgN
	Self			StdN
		Weight	The weight of the number of interactions between the user and target user.	Nw
Passive behavior (Be liked/Be reposted/Be commented on)	Global	Time	The time interval of interactions between the user and target user.	MaxT
				MinT
				AvgT
				StdT
	Follower	Count	The number of interactions between the user and target user.	N
				AvgN
				StdN
	Non-follower	Weight	The weight of the number of interactions between the user and target user.	Nw

Table 4: A logical structure table of the default interactive features

Table 5 explains how each statistic is calculated in the proposed model, taking User → Repost → Global as an example.

Behavior type	Relationship type	Quantitative characteristics	Statistical variables	Description	Abbreviation
Repost	Global	Count	N	The total number of reposts by User A	N_R
				The number of users whose microblogs are reposted by User A	N_UR
				The number of all active interactions of User A	N_A
				The number of target users for all active interactions of User A	N_UA
			AvgN	The average number of reposts User A made to individual users (N_R/ N_UR)	AvgN_R
			StdN	The standard deviation of User A's reposts to different users	StdN_RU
		Weight	Nw	N_R/ N_A	Nw_R
				N_UR/ N_UA	Nw_UR
		Time	MaxT	The maximum time interval between adjacent reposts by User A	MaxT_R

			MinT	The minimum time interval between adjacent reposts by User A	MinT_R
			AvgT	The average time interval between adjacent reposts by User A	AvgT_R
			StdT	The standard deviation of the time intervals between adjacent reposts by User A	StdT_R

Table 5: The calculation method for each statistical variable

Key interactive feature selection

For a particular classification algorithm, not all the features are valid, so it is necessary to select the relevant features that are beneficial to the learning algorithm from all the basic features. In fact, a high number of feature dimensions often leads to a reduction in algorithm performance. Therefore, if some features are selected from all the features to build a model, the training time for learning the algorithm can be greatly reduced, whereas the interpretability of the algorithm can also be increased. Therefore, in this study, we adopted the following two most commonly used and standard feature selection methods to select some features from the 168 interaction features:

The filter approach: The main idea of the filter method is to weigh each feature. The heavier it is, the more important the feature. The main methods used for weighing each feature are information gain and the chi-square test. It is worth noting that the filter method first filters the features and then trains the feature subset. Therefore, in the filter method, the selection of features and the learning of the classification algorithm is separated.

The wrapper approach: Wrapper methods integrate feature selection and algorithm learning. Specifically, the basic feature set is generated into different combinations, and the final classification algorithm is directly taken as the evaluation function of the feature selection. The optimal feature subset is selected for a specific classifier. It turns the selection of subsets into the problem of finding the optimal solution. Particle swarm optimization (PSO) or artificial bee colony algorithms is commonly used methods in this approach.

In order to ensure the accuracy of the research results, this paper uses five feature selection methods in turn: CfsSubsetEval, GainRationAttributeEval, InfoGainAttributeEval, OneRAttributeEval (Holte, 1993), and WrapperSubsetEval (Kohavi & John, 1997). Then we choose the best feature selection method by evaluating the prediction accuracy of the five feature extraction methods.

Interaction feature classifier

In this research, we took the DISC model as the theoretical framework for personality prediction, and personality was classified based on the dominance, influence, steadiness, and compliance types following the model; users with different interactive features were classified using these four personality types. Therefore, a classification algorithm in the field of data mining was used to generate the classifier model and test it.

The process of interaction feature classifying mainly uses the classification algorithm to train the key feature matrix to obtain the optimal model. First, the key feature subset was obtained by different filtering algorithms. Second, the different key feature subsets were sequentially sorted by Bayesian classification, Bayesian network classification, neighbor classification, decision tree classification, and a support vector machine. Third, a total of five kinds of classification algorithm were used for classifier training. Finally, we selected the classification algorithm and feature selection method with the highest classification accuracy.

In machine learning, classification prediction is a supervised learning method. First, a training model is used based on a certain number of samples. The data input for the training model includes the attributes of each sample and the corresponding categories. The training model is then tested against another set of test data. The data for the test data set includes attributes for each test sample. The output is the predicted category of training samples. The principle of classification prediction is shown in Figure 6.

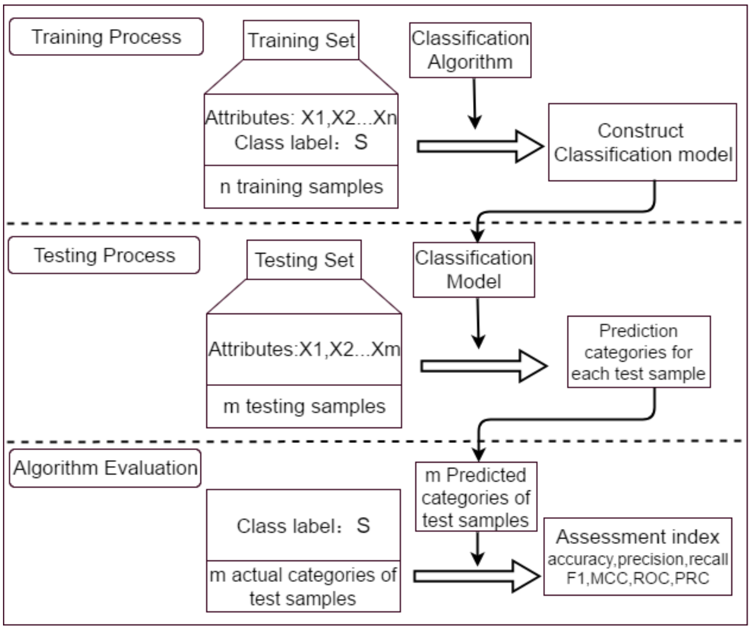

Figure 6: Schematic diagram of classification prediction

Evaluation of the prediction

The evaluation of the prediction results of the personality prediction model takes into account indexes, such as classification accuracy and prediction error. In this paper, the specific evaluation indicators are the accuracy rate, recall rate, F-measure, Matthews coefficient (MCC), Receiver Operating Characteristic (ROC) curve, and Precision recall curve (PRC). The evaluation indicators in this study are explained as follows:

(1) The accuracy rate indicates the proportion of users whose true personality is also A, among the users who are classified as having the same personality A at the time of forecasting;

(2) The recall rate indicates the proportion of users who are also categorized as personality A among all users with a true personality A;

(3) The F-measure is the weighted harmonic mean of Precision and Recall:

$$F_1 = \frac{2 * PR}{P + R}$$

(4) The MCC is mainly used to measure an imbalanced data set:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

(5) The area under the ROC curve measures the relationship between the specificity of the personality prediction model and the sensitivity of the model, the value of which is between 0 and 1, and the closer to 1 it is, the higher the accuracy of the model;

(6) The area under the PRC measures the accuracy of the relationship between the recall rate and the index; an index close to 1 shows a better classification model.

Results

Weibo interaction analysis

The interactive behavior data from Weibo included both basic attributes (including the number of user microblogs, the number of followers, and the number of following) and interaction attributes (the number of comments, likes, reposts, and times being reposted). Due to the lack of data about being commented on and liked, these two interactions were not included in this research.

Figure 7 shows the number of tweets posted by 198 users and ranked from low to high.

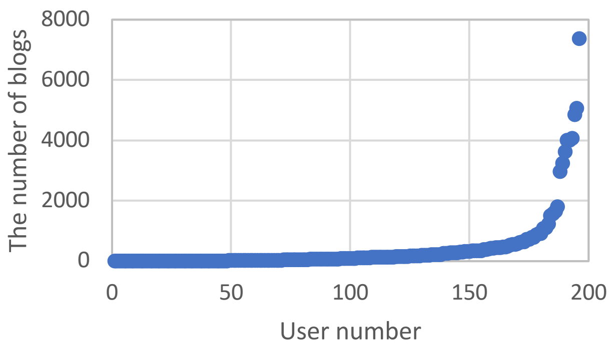

Figure 7: Blog number distribution of users

User Personality Analysis

The personality distribution of the 198 respondents is shown in Figure 8. There are 105 users with the steadiness personality, 44 with the influence one, 33 with the compliance one, and 16 with the dominance personality. The personality distribution is not balanced, which is consistent with prior studies in the literature on the distribution of personality (Imran, Faiyaz, and Akhtar, 2018; Tadesse et al., 2018). As in the study of Imran et al. (2018), the quantitative distributions of the big five personalities were 1268 Openness, 1488 Conscientiousness, 2085 Extroversion, 944 Agreeableness, and 603 Neuroticism, respectively. And, Tadesse et al. (2018) also found that the quantitative distributions of the five personalities were 176 Openness, 130 Conscientiousness, 96 Extroversion, 134 Agreeableness, and 99 Neuroticism in their study, respectively.

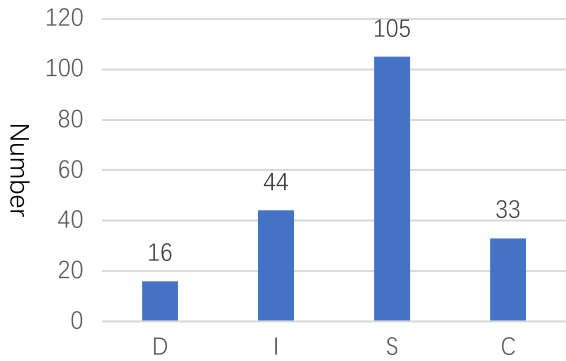

Figure 8: User personality distribution based on the DISC model

Prediction results and analysis

Based on the collected user interaction data and the generated set of DFs, we conducted calculations on the DFs. A DF value table was generated for each object, and a topic eigenvalue matrix was generated for all users and DFs (as shown in Figure 9). Each element in the matrix Value i, j is the eigenvalue of Feature j for user Subject i.

$$\begin{array}{c} \\ Subject_1 \\ Subject_2 \\ \vdots \\ Subject_m \end{array} \begin{array}{c} \begin{array}{cccc} Feature_1 & Feature_2 & \cdots & Feature_n \end{array} \\ \begin{bmatrix} Value_{1,1} & Value_{1,2} & \cdots & Value_{1,n} \\ Value_{2,1} & Value_{2,2} & \cdots & Value_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ Value_{m,1} & Value_{m,2} & \cdots & Value_{m,n} \end{bmatrix} \end{array}$$

Figure 9: Eigenvalue matrix of the user's default features

The matrix content was converted into comma-separated values (CSV) files for use as Weka input data. Figure 10 shows the key features of cross-feature filtering using Weka's different feature selection methods to obtain different recommended KFs. After obtaining KFs, we used different built-in classification algorithms in Weka to train the personality classification prediction model.

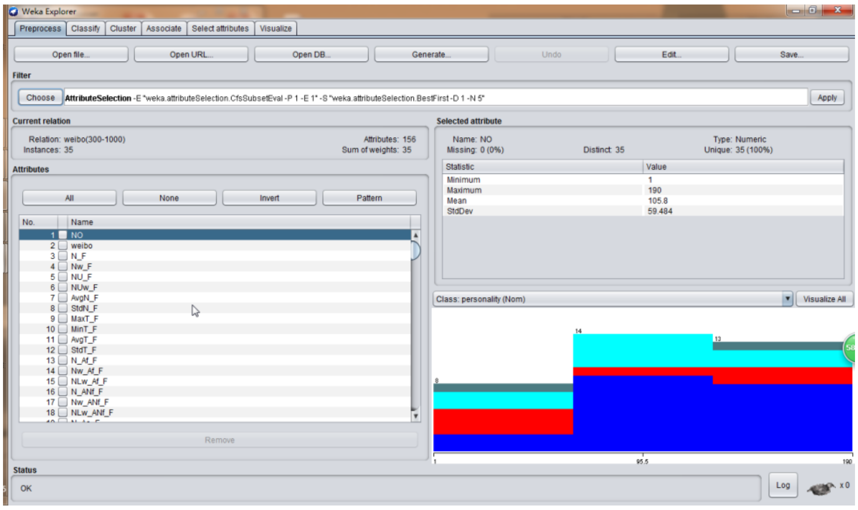

Figure 10: Weka data import interface

During data cleaning, we found that in the data set there were missing values for different social interaction activities for the users with less than 300 microblogs. As too many missing values in the data set would affect the prediction result deviation, we only considered the user interaction data from users with more than 300 microblogs. Similarly, when the number of microblogs is greater than 1000, the order of magnitude will increase significantly, and these outliers will affect the predicted results. In order to ensure the accuracy of this experiment, this paper mainly applied the data from users with 300 to 1000 microblogs to construct the personality prediction model. A data set from 35 users, whose microblogging numbers were between 300 and 1000, was generated to train the model. Finally, a default interaction characteristic value table was calculated according to the interaction data (see Table 6). We randomly selected the key feature data set of 28 users as the training set, and those of the remaining 7 users as the test set.

In order to verify the effectiveness of the sample size, we increased the sample size from 2 to 35, and used the 1-nearest neighbor classification algorithm to successively calculate the accuracy. As shown in Figure 11, when the sample number reaches 27, the prediction accuracy tends to stabilize, reaching 0.792. Therefore, the number of 35 samples is reasonable and effective for the classification algorithm in this study.

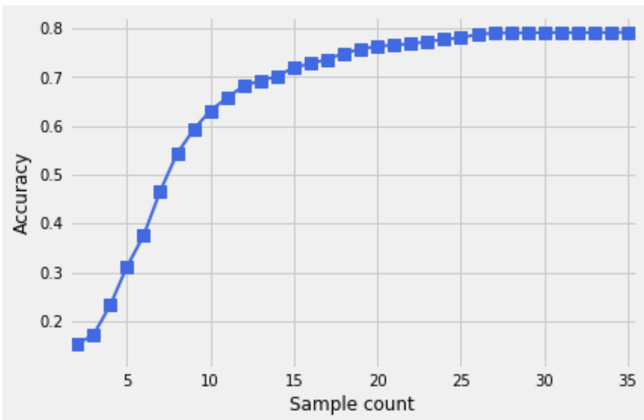

Figure 11: The prediction accuracy of different samples

The experiment employed a ten-fold cross-validation method. The data set was divided into 10 subsets at random. Each of 9 subsets was used for model training, and the remaining subset was reserved for testing, followed by 10 training sessions and tests. The final result is the average of these 10 validations.

Number	1	..	3	8	13	24	…
Weibo	337	..	332	349	445	305	…
N_R	28	..	160	257	181	90	…
Nw_R	0.444444444	..	0.3669724771	0.8923611111	0.6053511706	0.2743902439	…
NU_R	15	..	111	174	87	47	…
NUw_R	0.384615385	..	0.4703389831	0.8923076923	0.4754098361	0.5164835165	…
AvgN_R	1.866666667	..	1.441441441	1.477011494	2.08045977	1.914893617	…
StdN_R	1.707499797	..	1.015456749	3.637986214	5.013915727	1.470703801	…
MaxT_R	380.8604167	..	107.3486111	63.74791667	427.3104167	131.5659722	…
MinT_R	0	..	0.000694444	0	0	0.000694444	…
AvgT_R	19.36324588	..	7.372733229	4.707066515	7.855763889	8.224578652	…
StdT_R	74.9987061	..	15.87114847	7.74314639	45.97821502	18.60946868	…
N_Af_R	18	..	25	107	75	43	…
Nw_Af_R	0.45	..	0.1101321586	0.8425196850	0.5395683453	0.1660231660	…
NRw_Af_R	0.642857143	..	0.15625	0.416342412	0.414364641	0.477777778	…
N_ANf_R	10	..	135	150	106	47	…
Nw_ANf_R	0.158730159	..	0.3096330275	0.5208333333	0.3545150502	0.1432926829	…
NRw_ANf_R	0.357142857	..	0.84375	0.583657588	0.585635359	0.522222222	…
N_As_R	0	..	0	1	2	0	…
Nw_As_R	0	..	0	0.0048309179	0.0127388535	0	…
NRw_As_R	0	..	0	0.003891051	0.011049724	0	…
NU_Af_R	6	..	9	18	22	15	…
NUw_Af_R	0.4	..	0.081081081	0.103448276	0.252873563	0.319148936	…
NU_ANf_R	9	..	102	155	64	32	…

NUw_ANf_R	0.6	..	0.918918919	0.890804598	0.735632184	0.680851064	…
NU_As_R	0	..	0	1	1	0	…
NUw_As_R	0	..	0	0.005747126	0.011494253	0	…
AvgN_Af_R	3	..	2.777777778	5.944444444	3.409090909	2.866666667	…
AvgN_ANf_R	1.111111111	..	1.323529412	0.967741935	1.65625	1.46875	…
StdN_Af_R	2.699794231	..	3.566166902	11.3109588	9.970692349	2.603325751	…
StdN_ANf_R	2.204372759	..	1.059309419	3.854515745	5.845836566	1.782375298	…
MaxT_Af_R	380.8604167	..	392.8236111	90.32430556	1030.594444	152.1194444	…
MaxT_ANf_R	4.203472222	..	107.3486111	122.6291667	459.8972222	131.5659722	…
MaxT_As_R	0	..	0	0	15.64375	0	…
MinT_Af_R	0	..	0.984027778	0	0	0.000694444	…
MinT_ANf_R	0	..	0.000694444	0	0.000694444	0.001388889	…
MinT_As_R	0	..	0	0	15.64375	0	…
AvgT_Af_R	30.75339052	..	42.83093171	11.34851939	19.05886824	17.09442791	…
AvgT_ANf_R	0.876697531	..	8.257053275	8.075745712	11.60547619	15.78734903	…
AvgT_As_R	0	..	0	0	15.64375	0	…
StdT_Af_R	92.64697526	..	77.98056309	16.40878688	118.7316487	35.47500531	…
StdT_ANf_R	1.413804302	..	16.52344624	14.00386288	59.24869791	26.00934777	…
…	…	..	…	…	…	…	…
Personality	Compliance	..	Compliance	Influence	Dominance	Steadiness	…

Table 6: Default interaction feature table

In this paper, we applied five attribute selection methods to select attributes, and then employed five classification algorithms for model testing and training, including Bayesian classification, Bayesian network classification, K nearest neighbor classification, decision tree classification, and a support vector machine classification algorithm. The final experimental results are shown in Table 7.

Classification algorithm	Not filtered	CfsSubset Eval	GainRation Attribute Eval	InfoGain Attribute Eval	OneR Attribute Eval	Wrapper Subset Eval
Naïve Bayes	34.29%	57.14%	34.29%	34.29%	34.29%	80%

Bayes Net	42.86%	51.43%	42.86%	42.86%	42.86%	54.29%
1-IBK（knn）	54.29%	57.14%	54.29%	54.29%	54.29%	82.86%
3-IBK（knn）	51.43%	57.14%	51.43%	51.43%	51.43%	74.29%
5-IBK（knn）	60%	68.58%	60%	60%	60%	80%
J48	57.14%	60%	60%	60%	57.14%	74.29%
LibSVM	54.29%	51.43%	54.29%	54.29%	54.29%	71.43%

Table 7: Classification accuracy

The use of the wrapper approach for the extraction of KFs can improve the classification model prediction accuracy as the values for personality prediction accuracy are higher than in the other five approaches for all the classifications (see Table 7). Obviously, 1-nearest neighbor classification is superior to the classification of decision trees as 1-nearest neighbor classification has higher prediction performance and better prediction results. The best classification accuracy is 82.86% with the WrapperSubsetEval and 1-nearest neighbor classification algorithm.

A subset of the KFs is shown in Table 8 (User A is set as the subject of the personality prediction).

Feature Abbreviation	Feature Meaning
AvgN_ANf_R	The average of the reposts A made to non-following.
MinT_As_R	The minimum time interval between the adjacent reposts A made to self.
StdT_Af_R	The standard deviation of the time interval between the adjacent reposts A made to following.
MinT_BNf_BR	The minimum time interval between the adjacent reposts A is reposted to non-follower.
AvgN_BC	The average of the reposts A is reposted. (N_BC / NU_BC)
NLW_Bf_BC	The weight of the number of comments A is commented on to follower. (Scope: All comment interactions)
StdN_Bf_BC	The standard deviation of the number of comments A is commented on to follower.
MinT_Bs_C	The minimum time interval between the adjacent comments A made to self.

Table 8: The best key feature subset

As shown in Table 8, reposting, being reposted, commenting, and being commented on are key interaction features reflecting Weibo users' personalities, whereas liking is not listed as a key feature. Table 8 presents the eight key features included in our personality prediction model (AvgN_ANf_R, MinT_As_R, StdT_Af_R, MinT_BNf_BR, AvgN_BC, NLW_Bf_BC, StdN_Bf_BC, and MinT_Bs_C). In other words, these eight key features are the determinants of personality in our proposed model. Among the eight key features, seven of them are related to the interaction relationship (AvgN_ANf_R, MinT_As_R, StdT_Af_R, MinT_BNf_BR, AvgN_BC, NLW_Bf_BC, and StdN_Bf_BC), four to the interaction time (MinT_As_R, StdT_Af_R, MinT_BNf_BR, and MinT_Bs_C), five to the reposting behavior (AvgN_ANf_R, MinT_As_R, StdT_Af_R, MinT_BNf_BR, and AvgN_BC), and three to the commenting behavior (NLW_Bf_BC, StdN_Bf_BC, and

MinT_Bs_C).

Ten-fold cross-validation was used to verify the effect of the model classification with percentages indicating the correct classification accuracy. As shown in Table 9, both five-neighbor classification and J48 (decision tree classification) have good general-purpose performance and good classification results for different filtering algorithms. Except for the Bayesian network classification algorithm, the accuracy values of the other classification algorithms are all above 70%. The classification accuracy indicators for the five classification algorithms are presented in detail in Table 9.

Classification accuracy indicators	Naïve Bayes	Bayes Net	1-IBK （knn）	3-IBK （knn）	5-IBK （knn ）	J48	LibSVM
TP	0.800	0.543	0.829	0.743	0.800	0.743	0.714
FP	0.159	0.543	0.099	0.229	0.159	0.305	0.286
Precision	0.753	0.295	0.792	0.733	0.751	0.768	0.684
Recall	0.800	0.543	0.829	0.743	0.800	0.743	0.714
F1-Measure	0.773	0.382	0.805	0.703	0.768	0.700	0.673
MCC	0.662	0.000	0.709	0.551	0.661	0.560	0.495
ROC	0.796	0.344	0.837	0.722	0.711	0.698	0.714
PRC	0.698	0.346	0.738	0.597	0.619	0.628	0.573
Correct classification accuracy	80%	54.29%	82.86%	74.29%	80%	74.29%	71.43%

Table 9: Classification accuracy indicators

Conclusions and Discussions

There were some interesting findings from our study, which also answered three of our research questions.

First of all, like other personality measurement models, DISC model can also effectively construct the personality of users in online social media. In this study, we found that personality prediction based on the DISC personality model via machine learning has a high level of accuracy. The finding further validated the previous research finding of Chen et al. (2016), which showed that the DISC personality model can be employed as a personality framework to predict Facebook users' personal traits. The research finding shows that the DISC personality model can be applied in personality prediction research based on social media data within different contexts, such as for social media users with different cultural backgrounds.

Second， although the cultural backgrounds of users vary, such as in the case of the users of Weibo, Facebook, and Twitter, our research finding shows that social media interaction data from different social media platforms and from different cultural backgrounds, can be used to predict user personality. However, there are still some different findings from different cultures. In this study, liking behavior on Weibo has not been listed as a key feature in the personality prediction model compared to other reposting and commenting behavior and interaction time and patterns. The finding is in contrast to the research finding in the work of Ghavami et al. (2015). In their research, Facebook users' liking behavior was found to reveal user personality. This might be due to the different cultural backgrounds of Facebook users and Weibo users. In this study, we targeted Chinese Weibo users. Chinese users have a different cultural background and personality compared to people with a Western cultural background.

Third, among the eight valid features extracted, five are related to reposting interaction and three are related to commenting interaction. The research findings show that the reposting and commenting behaviors of Weibo users reflect users' personality characteristics. In addition, four features are related to the interaction time, which shows that interaction time is another factor reflecting Weibo users' personality traits. In other words, users with different personalities have different interactivity intervals.

Our research also confirms the deep relationship between social interaction data and users' personality. That is to say, people's personality can be reflected in their interactions on social media. The way people behave online and offline can reflect a user's personality to some extent. We found that social interaction data from social media use, especially Weibo, can be mined to predict user personality as our results show high prediction accuracy based on both users' social media interaction data and their self-reported personality data based on the DISC personality model. The research finding is consistent with prior research findings based on Facebook and Twitter based on the Big Five model, which showed that social interaction data can be applied to predict user personality (Moore and Mcelroy, 2012; Ortigosa, Carro, and Quiroga, 2014).

We explore a deeper question about the relationship between online and offline interaction driven by personality traits. Online interaction behaviors, such as like, forwarding and comments, reveal user behavior patterns and tendencies. Extroversion, for example, is positively related both to self-presentation-related social interaction on Facebook walls, and also to commenting and sharing behaviors on Facebook. Highly competitive narcissists often update their statuses and positions on their walls, whereas neurotic and cautious people are less likely to write comments (Lee et al., 2014), and this is the principal mechanism by which online interaction behavior is used to predict personality. Personality traits drive online interactions to follow the same behavior patterns as offline interactions. However, there are significant differences between the two patterns. Specifically, the main forms of offline interaction are conversation and observation, which can directly reflect people's personality characteristics. For example, in the process of getting along with a person, it is easy to judge whether he is stable or impulsive, extroverted or introverted. While online interaction behavior is reflected by some interactive data, such as likes, comments, and shares, and in the online environment, it is difficult to judge the user's personality characteristics through direct perception. However, we can use machine learning techniques to find clues to personality traits in online interaction data. Specifically, machine learning can learn behavioral patterns of online interactions and establish relationships with individual tags. This study can provide a reference for studying the relationship between online and offline interactive behavior patterns.

Implications

This study takes Weibo as the research context to examine whether social interaction behavior on social media platforms can help predict user personality. The theoretical contribution of this study can be summarized as follows.

First, this paper verifies the feasibility of DISC personality modeling for predicting personality. To be specific, the main personality measurement models currently used are the Big Five Personality Measurement Model and the MBTI Model, while there is a lack of research on the effectiveness of DISC, another major personality measurement model. Making up this gap can provide a wide range of applicability for personality prediction based on social media data. This study showed that social media interaction data can be used to predict user personality based on the DISC personality model. This research enriches previous research in this field, from the Big Five personality model to the DISC personality model. The research findings further validated that social media data can be applied to predict user personality based on the different theoretical lens of personality traits and in different cultural contexts.

Secondly, this study provides a theoretical basis for the similarity in the expression of social interaction information under different cultural backgrounds. To be specific, the current research on personality prediction based on social media data mainly focuses on social platforms with European and American cultural backgrounds such as Facebook and Twitter, while there is a lack of research on platforms with Asian cultural backgrounds such as Weibo in China. Since users from different cultural backgrounds may have different expressions of their personalities on social media, filling in the research gaps of Asian cultural backgrounds is helpful to provide effective evidence for the consistency of personality expressions from different cultural backgrounds.

Finally, this study provides a further reference to the literature on personality modeling, personality prediction, cultural, and social interaction.

For practical implication, firstly, this paper constructs a dynamic personality prediction model. Compared with the existing models, this paper considers the effect of time intervals on personality prediction and provides a new insight for understanding personality prediction based on social media interaction data from the perspective of interaction time. The research results confirmed that the time interval cannot be ignored in personality prediction models as it can, to some degree, reflect one's personality. Previous studies mainly apply static social media data in personality prediction, ignoring the role of time intervals. Specifically, user personality will be reflected not only by what a user does through social interaction on a social media platform, but also by when a user performs the social interaction.

Second, this study proposed a structured method for social media interaction feature selection that can generate many DFs in a structured way and help obtain KFs by applying some filter approaches. The structured method enriches the feature selection methods and helps researchers to find some features that are hard to find by observation or experience or that have not been researched in the prior literature.

Third, we found that the combination of the WrapperSubsetEval feature selection algorithm and 1-Nearest Neighbor classification algorithm can achieve the highest prediction accuracy, which provides a reference for feature selection engineering of personality prediction.

Finally, this study also provides practical guidance for social media service providers. Both the social interaction activities and the number of social interactions on social media platforms can reflect the personality characteristics of users, which provides an effective analysis method for a personalized recommendation. Both social interaction activities and social interaction times on social media platforms can reflect user personality traits, which indicates that social media platforms should take both user interaction activities and social interaction times into consideration in order to get a comprehensive understanding of user personality. In addition, social media platforms should understand and cluster users based on their social interaction activities, such as commenting and reposting, in order to provide customized advertisements to and recommendations for individual users that can meet their needs.

Limitations and Future Study

This study has several limitations that should be acknowledged. First, in this study, the social interaction data collection was only carried out on the Weibo platform, and the Weibo interaction data are for a limited amount of active Weibo users due to the limitation of official microblogging API. Thus, future research should investigate personality prediction with as much data as possible collected across different social media platforms. Second, this study mainly considered the interaction pattern and interaction time and did not take the exchanged content and emotion in content on social media platforms into account in the proposed personality prediction model. Further studies should consider both the exchanged content and emotion and the interaction time and patterns in personality prediction models via combining text mining, machine learning, and sentiment analysis tools. Third, this study was conducted in the context of a Chinese social media platform. Thus, cautions should be exercised when applying the research findings in this study to other cultural backgrounds; this study should be replicated for other international social media platforms, such as Facebook and Twitter, to generalize the proposed research approach for different social media contexts as well as to investigate whether there is a cultural difference in personality prediction across international social media users. Finally, this study only focuses on the mode and degree of social interaction, without considering the text content and emotional expression of the interaction which is also important cues to reflect users' personality characteristics. Therefore, in future studies, we should consider including textual content, such as language style and affective score, into our research model.

References

Amichaihamburger, Y., & Vinitzky, G. (2010). "Social network use and personality".

Computers in Human Behavior, Vol. 26 No. 6, pp. 1289–1295.

Bansal, G., Zahedi, F. M., & Gefen, D. (2016). "Do context and personality matter? Trust and privacy concerns in disclosing private information online". *Information and Management*, Vol. 53 No. 1, pp. 1–21.

Beamish, G. (2005). "How chief executives learn and what behaviour factors distinguish them from other people". *Industrial & Commercial Training*, Vol. 37 No. 3, pp. 138–144.

Bell, R. M., Fann, S. A., Morrison, J. E., & Lisk, J. R. (2012). "Determining personal talents and behavioral styles of applicants to surgical training: a new look at an old problem, part II". *Journal of Surgical Education*, Vol. 69 No. 1, pp. 23–29.

Blazevic, V., Wiertz, C., Cotte, J., Ruyter, K. De, & Keeling, D. I. (2014). "GOSIP in Cyberspace: Conceptualization and Scale Development for General Online Social Interaction Propensity". *Journal of Interactive Marketing*, Vol. 28 No. 2, pp. 87–100.

Bock, G. W., Zmud, R. W., Kim, Y. G., & Lee, J. N. (2005). "Behavioral intention formation in knowledge sharing: examining the roles of extrinsic motivators, social-psychological factors, and organizational climate". *Mis Quarterly*, Vol. 29 No. 1, pp. 87–111.

Bouchard, L. A. (2018). "AAHPM Leadership Forum: Ignite—Utilizing DISC Behavioral Styles to Increase Leadership and Team Effectiveness in Palliative Care and Hospice Settings (P14)". *Journal of Pain & Symptom Management*, Vol. 55 No. 2, pp. 554.

Boyle, G. J. (1995). "Myers-Briggs Type Indicator (MBTI): Some Psychometric Limitations". *Australian Psychologist*, Vol. 30 No. 1, pp. 71–74.

Briggs-Myers, & Briggs, K. C. (1985). "Myers–Briggs Type Indicator (MBTI)". *Palo Alto, CA: Consulting Psychologists Press*, pp. 2413--2413.

Chen, C. C., & Lin, Y. C. (2018). "What drives live-stream usage intention? The perspectives of flow, entertainment, social interaction, and endorsement". *Telematics & Informatics*, Vol. 35 No. 1, pp. 293–303.

Chen, T. Y., Tsai, M. C., & Chen, Y. M. (2016). "A user's personality prediction approach by mining network interaction behaviors on Facebook". *Online Information Review*, Vol. 40 No. 7, pp. 913–937.

Correa, T., & Hinsley, A. W. (2010). "Who interacts on the Web?: The intersection of users' personality and social media use". *Computers in Human Behavior*, Vol. 26 No. 2, pp. 247–253.

Deng, S., Liu, Y., Li, H., & Hu, F. (2013). "How does personality matter? An investigation of the impact of extraversion on individuals' SNS use". *Cyberpsychology, Behavior, and Social Networking*, Vol. 16 No. 8, pp. 575–581.

Deviney, D., Mills, L. H., & Gerlich, R. N. (2010). "Environmental Impacts on GPA for Accelerated Schools: A Values and Behavioral Approach". *Journal of Instructional*

Pedagogies, Vol. 3 No. 1, pp. 1–15.

Edens, J. F., Poythress, N. G., Lilienfeld, S. O., Patrick, C. J., & Test, A. (2008). "Further evidence of the divergent correlates of the Psychopathic Personality Inventory factors: prediction of institutional misconduct among male prisoners". *Psychological Assessment*, Vol. 20 No. 1, pp. 86–91.

Ghavami, S. M., Asadpour, M., Hatami, J., & Mahdavi, M. (2015). "Facebook user's like behavior can reveal personality". In *2015 7th Conference on Information and Knowledge Technology (IKT)* (pp. 1–3). Urmia, IRAN: IEEE.

Gloor, P., Mustafaraj, E., Schoen, H., Takis Metaxas, P., Strohmaier, M., & Gayo-Avello, D. (2013). "The power of prediction with social media". *Internet Research*, Vol. 23 No. 5, pp. 528–543.

Golbeck, J., Robles, C., & Turner, K. (2011). "Predicting personality with social media". In *CHI'11 extended abstracts on human factors in computing systems* (pp. 253–262).

Griffin, Andrea, S., Guillette, Lauren, M., Healy, & Susan, D. (2015). "Cognition and personality: an analysis of an emerging field". *Trends in Ecology & Evolution*, Vol. 30 No. 4, pp. 207–214.

Hall, J. A., Park, N., Song, H., & Cody, M. J. (2010). "Strategic misrepresentation in online dating: The effects of gender, self-monitoring, and personality traits". *Journal of Social & Personal Relationships*, Vol. 27 No. 1, pp. 117–135.

Hoffman, D. L., & Novak, T. P. (1996). "Marketing in Hypermedia Computer-Mediated Environments: Conceptual Foundations". *Journal of Marketing*, Vol. 60 No. 3, pp. 50–68.

Holte, R. C. (1993). "Very simple classification rules perform well on most commonly used datasets". *Machine Learning*, Vol. 11 No. 1, pp. 63–91.

Hughes, D. J., Rowe, M., Batey, M., & Lee, A. (2012). "A tale of two sites: Twitter vs. Facebook and the personality predictors of social media usage". *Computers in Human Behavior*, Vol. 28 No. 2, pp. 561–569.

Imran, A., Faiyaz, M., & Akhtar, F. (2018). "An Enhanced Approach for Quantitative Prediction of Personality in Facebook Posts". *International Journal of Education and Management Engineering (IJEME)*, Vol. 8 No. 2, pp. 8–19.

Karasek, R. A. (1979). "Job Demands, Job Decision Latitude, and Mental Strain: Implications for Job Redesign". *Administrative Science Quarterly*, Vol. 24 No. 2, pp. 285–308.

Kim, N., & Kim, W. (2018). "Do your social media lead you to make social deal purchases? Consumer-generated social referrals for sales via social commerce". *INTERNATIONAL JOURNAL OF INFORMATION MANAGEMENT*, Vol. 39 No. 1, pp. 38–48.

Kohavi, R., & John, G. (1997). "Wrappers for feature subset selection". *Artificial Intelligence*, Vol. 97 No. 1, pp. 273–324.

Kotov, R., Gamez, W., Schmidt, F., & Watson, D. (2010). "Linking “big” personality traits to anxiety, depressive, and substance use disorders: A meta-analysis". *Psychological Bulletin*, Vol. 136 No. 5, pp. 768.

Lee, E., Ahn, J., & Kim, Y. J. (2014). "Personality traits and self-presentation at Facebook". *Personality & Individual Differences*, Vol. 69 No. 6, pp. 162–167.

Lima, A. C., & de Castro, L. N. (2014). "A multi-label, semi-supervised classification approach applied to personality prediction in social media". *Neural Netw*, Vol. 58 No. 5, pp. 122.

Liu, Q., Shao, Z., & Fan, W. (2018). "The impact of users’ sense of belonging on social media habit formation: Empirical evidence from social networking and microblogging websites in China". *INTERNATIONAL JOURNAL OF INFORMATION MANAGEMENT*, Vol. 43, pp. 209–223.

Liu, Y., & Shrum, L. J. (2002). "What Is Interactivity and Is It Always Such a Good Thing? Implications of Definition, Person, and Situation for the Influence of Interactivity on Advertising Effectiveness". *Journal of Advertising*, Vol. 31 No. 4, pp. 53–64.

Maddi, S. R. (1989). "*Personality theories: A comparative analysis (5th ed.)*". Homewood, IL, US: Dorsey Press.

Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). "Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text". *Journal of Artificial Intelligence Research*, Vol. 30 No. 1, pp. 457–500.

Marston, W. M. (2013). "*Emotions of normal people*". Routledge.

McCrae, R. R., & Costa, J. P. (1987). "Validation of the five-factor model of personality across instruments and observers". *Journal of Personality and Social Psychology*, Vol. 52 No. 1, pp. 81–90.

Mcelroy, J. C., Hendrickson, A. R., Townsend, A. M., & Demarie, S. M. (2007). "Dispositional Factors in Internet Use: Personality versus Cognitive Style". *Mis Quarterly*, Vol. 31 No. 4, pp. 809–820.

Men, W. H. S. T. L. R. (2015). "Consumer engagement with brands on social network sites: A cross-cultural comparison of China and the USA". *Journal of Marketing Communications*, 1–20.

Moore, K., & Mcelroy, J. C. (2012). "The influence of personality on Facebook usage, wall postings, and regret". *Computers in Human Behavior*, Vol. 28 No. 1, pp. 267–274.

Nie, D., Guan, Z., Hao, B., Bai, S., & Zhu, T. (2014). "Predicting Personality on Social Media with Semi-supervised Learning". In *Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)* (pp. 158–165). Warsaw, Poland: IEEE.

Ogunyemi, D. A., Mahller, Y. Y., Wohlmuth, C., Eppey, R., Tangchitnob, E., & Alexander, C. J. (2011). "Associations between DISC assessment and performance in obstetrics and gynecology residents". In *65th Annual Meeting of the Obstetrical and Gynecological Assembly of Southern California* (Vol. 56, pp. 398–404). Santa Monica, CA.

Ong, V., Rahmanto, A. D. S., Williem, & Suhartono, D. (2017). "Exploring Personality Prediction from Text on Social Media: A Literature Review". *Internetworking Indonesia*, Vol. 9 No. 1, pp. 65–70.

Ortigosa, A., Carro, R. M., & Quiroga, J. I. (2014). "Predicting user personality by mining social interactions in Facebook". *Journal of Computer & System Sciences*, Vol. 80 No. 1, pp. 57–71.

Pan, Y., Chen, X., & Guo, B. (2016). "The influence of personality traits and social networks on the self-disclosure behavior of social network site users". *Internet Research*, Vol. 26 No. 3, pp. 566–586.

Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., … Seligman, M. E. (2015). "Automatic personality assessment through social media language". *Journal of Personality & Social Psychology*, Vol. 108 No. 6, pp. 934.

Poropat, A. E. (2009). "A meta-analysis of the five-factor model of personality and academic performance". *Psychological Bulletin*, Vol. 135 No. 2, pp. 322–338.

Quercia, D., Kosinski, M., Stillwell, D., & Crowcroft, J. (2012). "Our Twitter Profiles, Our Selves: Predicting Personality with Twitter". In *IEEE Third International Conference on Privacy, Security, Risk and Trust* (pp. 180–185).

Reynierse, J. H., Ackerman, D., Fink, A. A., & Harker, J. B. (2000). "The Effects of Personality and Management Role on Perceived Values in Business Settings". *International Journal of Value-Based Management*, Vol. 13 No. 1, pp. 1–13.

Ross, C., Orr, E. S., Sisic, M., Arseneault, J. M., Simmering, M. G., & Orr, R. R. (2009). "Personality and motivations associated with Facebook use". *Computers in Human Behavior*, Vol. 25 No. 2, pp. 578–586.

Stavros, C., Meng, M. D., Westberg, K., & Farrelly, F. (2014). "Understanding fan motivation for interacting on social media". *Sport Management Review*, Vol. 17 No. 4, pp. 455–469.

Storrs, G. (1994). "A conceptualisation of multiparty interaction". *Interacting with Computers*, Vol. 6 No. 2, pp. 173–189.

Stumpf, H., & Parker, W. D. (2000). "A hierarchical structural analysis of perfectionism and its relation to other personality characteristics". *Personality & Individual Differences*, Vol. 28 No. 5, pp. 837–852.

Sumner, C., Byers, A., Boochever, R., & Park, G. J. (2013). "Predicting Dark Triad Personality Traits from Twitter Usage and a Linguistic Analysis of Tweets". In *2012 11th International Conference on Machine Learning and Applications* (pp. 386–393). Boca Raton, FL, USA: IEEE.

Tadesse, M. M., Lin, H., Xu, B., & Yang, L. (2018). "Personality Predictions Based on

User Behavior on the Facebook Social Media Platform". *IEEE Access*, *6*, 61959–61969.

Tandera, T., Hendro, Suhartono, D., Wongso, R., & Prasetio, Y. L. (2017). "Personality Prediction System from Facebook Users". In *2nd International Conference on Computer Science and Computational Intelligence (ICCSCI)* (Vol. 116, pp. 604–611). Bali, Indonesia.

Tsai, T. H., Chang, H. T., Chang, Y. C., & Chang, Y. S. (2017). "Personality disclosure on social network sites: An empirical examination of differences in Facebook usage behavior, profile contents and privacy settings". *Computers in Human Behavior*, Vol. 76, pp. 469–482.

Wiertz, C., & Ruyter, K. De. (2016). "Beyond the Call of Duty: Why Customers Contribute to Firm-hosted Commercial Online Communities". *Organization Studies*, Vol. 28 No. 3, pp. 347–376.

Wood, B. (2012). "Human Information Interaction: An Ecological Approach to Information Behavior". *International Journal of Information Management*, Vol. 64 No. 1, pp. 213–214.

Zhao, H., Seibert, S. E., & Lumpkin, G. T. (2015). "The Relationship of Personality to Entrepreneurial Intentions and Performance: A Meta-Analytic Review". *Entrepreneurship Research Journal*, Vol. 36 No. 2, pp. 381–404.