



Extroverts tweet differently from introverts in Weibo

Zhenkun Zhou¹, Ke Xu¹ and Jichang Zhao^{2,3*} 

*Correspondence:

jichang@buaa.edu.cn

²School of Economics and Management, Beihang University, Beijing, China

³Beijing Advanced Innovation Center for Big Data and Brain Computing, Beijing, China

Full list of author information is available at the end of the article

Abstract

As dominant factors driving human actions, personalities can be excellent indicators to predict the offline and online behavior of individuals. However, because of the great expense and inevitable subjectivity in questionnaires and surveys, it is challenging for conventional studies to explore the connection between personality and behavior and to gain insight in the context of a large number of individuals. Considering the increasingly important role of online social media in daily communications, we argue that the footprints of massive numbers of individuals, such as tweets on Weibo, can be used as a proxy to infer personality and further understand its function in shaping online human behavior. In this study, a map from self-reports of personalities to online profiles of 293 active users on Weibo is established to train a competent machine learning model, which then successfully identifies more than 7000 users as extroverts or introverts. Systematic comparison from the perspectives of tempo-spatial patterns, online activities, emotional expressions and attitudes to virtual honors show that extroverts indeed behave differently from introverts on Weibo. Our findings provide solid evidence to justify the methodology of employing machine learning to objectively study the personalities of a massive number of individuals and shed light on applications of probing personalities and corresponding behaviors solely through online profiles.

Keywords: Personality; Extraversion; Social media; Machine learning

1 Introduction

The booming of online social media has made it an essential component of everyday life that reflects all aspects of human behavior. Millions of users have digitalized and virtualized themselves on popular platforms such as Twitter and Weibo, providing information ranging from basic demographics, status, and emotions to activities. These online profiles can be natural, detailed, long-term and objective footprints of massive numbers of individuals; thus, they are potential proxies to understand human personalities [1, 2]. As a sub-discipline of psychology, the study of human personalities has aimed at one general goal: to describe and explain the significant psychological differences between individuals. Revealing the connection between different personalities and the corresponding behavioral patterns, especially in the circumstance of online social media, is one of the most exciting issues [3] in recent decades. The growing body of evidence suggesting individual personality discrepancies in online social media further makes it imperative to probe online human behavior from the perspective of personality [4–6].

Personality is a stable set of characteristics and tendencies that specify similarities and differences in individuals' psychological behavior, and it is a dominant factor in shaping human thoughts, feelings and actions. However, personality traits, like many other psychological dimensions, are latent and difficult to measure directly. Self-report by asking subjects to fill out survey questionnaires based on personality is a classic way to assess respondents in conventional studies [7–9], but self-reporting has several limitations:

- **Cost.** Questionnaires in self-reports can be time consuming and costly, and the response rate might be unexpectedly low [10]. These concerns substantially reduce the number of valid participants, which is generally less than 1000 [11]. Persuasive and universal conclusions are hard to reach based on such a small number of samples.
- **Subjectivity.** Respondents fill out questionnaires mainly based on their cognition, memory or feelings, and they can hide their true responses or thoughts consciously or unconsciously. Particularly for self-reports on personality, an individual might not even recollect the circumstances exactly in a controlled laboratory environment. These sources of response bias may have a large impact on the validity of the conclusions [12, 13] and may even lead to a re-interview [14].
- **Low flexibility.** Questionnaires are generally designed based on the study assumptions before conducting experiments, and it is difficult to obtain insights outside of the scope of the previously established goals, i.e., existing self-reports might be much less inspiring due to the lack of extension.

To some extent, these limitations can be overcome by the emergence of crowdsourcing marketplaces, such as Amazon Mechanical Turk (MTurk), which offer many practical advantages that reduce costs and make massive recruitment feasible [15] and thus have become dominant sources of experimental data for social scientists. In the meantime, new concerns are presented in [16, 17]. For example, researchers worry that volunteers are less numerous and less diverse than desired, while Turkers complain that the reward is too low. In addition, MTurk has suffered from growing participant non-naivety [18]. To account for these shortcomings, recent progress in machine learning, especially the idea of computation-driven solutions in social sciences [19], has received increasing interest in the modelling and understanding of human psychological behavior, such as personality.

Indeed, the popularity of online social media provides a great opportunity to examine personality inference using significant amounts of data. Twitter, the most popular social media and micro-blogging service, enables registered users to read and post short messages called tweets. At the beginning of 2016, Twitter had reached 310 million monthly active users (MAU). As of the third quarter of 2017, Twitter averaged 330 million MAU. However, more people are now using Weibo, the Chinese variant of Twitter. According to the Chinese company's first quarter reports, it has 340 million MAU, up 30% from the previous year. These numbers imply that the availability of vast and rich datasets of active individuals' digital footprints on Weibo will unprecedentedly increase the scale and granularity in measuring and understanding human behavior, especially personality, because the cost of the experiment will be substantially reduced, the objectivity of the samples will be guaranteed and the flexibility of the data will be adequately amplified. At the same time, there are new opportunities to combine social media with traditional surveys in personality psychology. Kosinski et al. demonstrate that available digital traces in Facebook can be used to automatically and accurately predict personality [20]. With the help of developments in machine learning, computer models can make valid personality predictions and

even outperform self-reported scores [21]. In this study, we argue that from the perspective of computational social science, the profiles of active users on Weibo are excellent proxies in probing the interplay between personality and online behavior.

An online page with a 60-item version of the Big Five Personality Inventory is established first in our study to collect personality trait scores [22], and a total of 293 active users on Weibo are asked to complete the self-report on this page to provide a baseline for the study. Focusing on extraversion, the scores mainly follow a Gaussian distribution, and the subjects are accordingly divided into three groups of high, neutral and low extraversion scores. Then, by collecting online profiles of self-reporters from Weibo, a map between the self-reports of extraversion and the online profiles is built to train machine learning models to automatically evaluate the extraversion of more individuals without the help of self-reports. Three types of features, including 13 basic, 33 behavioral and 84 linguistic, are comprehensively considered in the support vector machine (SVM) model, and its performance is verified by cross-validations. With more than 7000 users labelled as extroverts or introverts by the model, we attempt to systematically study the differences in online behavior due to extraversion by investigating the following seven research questions:

- RQ1. Do extroverts and introverts tweet temporally differently on Weibo?
- RQ2. Do extroverts and introverts tweet spatially differently on Weibo?
- RQ3. What types of information do extroverts and introverts prefer to share?
- RQ4. Who is more socially active online?
- RQ5. Who pays more attention to online purchasing and shopping?
- RQ6. Do extroverts and introverts express emotions differently on Weibo?
- RQ7. Who cares more about online virtual honor, extroverts or introverts?

Unexpected differences in the online behavior of extroverts and introverts is observed based on these questions. Introverts post more frequently than do extroverts, especially during the day. However, extroverts visit different cities rather than staying in one familiar city as introverts do. The spatial discrepancy becomes more unintuitive as we zoom in to increased resolution. For example, introverts tend to check in while shopping, whereas extroverts enjoy posting from their workplace. In addition, a tiny fraction of introverts might attempt to camouflage their own loneliness by tweeting from a large number of different areas (> 20). Extroverts enjoy sharing music and selfies, whereas introverts prefer retweeting news. In terms of online interactions, extroverts mention friends more often than introverts do, implying higher social vibrancy. By presenting a purchasing index to depict online buying intention, we find that compared to extroverts, introverts devote more effort to posting shopping tweets to relieve loneliness due to a lack of social interaction with others. We also categorize the emotion delivered by tweets into anger, disgust, happiness, sadness, and fear [23] and find that introverts post more angry and fearful (high arousal) tweets while extroverts post more sad (low arousal) tweets. Finally, extroverts attach greater meaning to online virtual honor than introverts do, implying that they might be ideal candidates for online promotion campaigns with virtual honor. To the our best knowledge, this is the first study to thoroughly compare the online behavior of extroverts and introverts in a large-scale sample, and our findings will be helpful to understand the role of personality in shaping human behavior.

The contributions of the present study can be summarized into three aspects. Firstly, it evidently demonstrates that machine learning models can be employed to properly reach

larger populations in personality study. Great expense and low efficiency always constrains the sample size of self-reports in conventional methodology, however, a machine learning model of competent performance can automatically identify a massive number of samples from social media to essentially enhance the reliability of study. Extracting multiple features from social media to predict personality is not new [24, 25], but employing the prediction model to produce new samples and facilitate further explorations is rarely touched. We build the extraversion classifier based on small samples, which then identifies more than 7000 users as extroverts or introverts. The larger dataset benefits us to study detailed differences of behaviors without social-desirability bias, for instance, we analyze users' POIs to find spatial behavioral pattern of extroverts and introverts. However, most previous studies [20, 26, 27] concentrated on the predictive power of personality prediction models, instead of properly applying the models into the further research.

Secondly, through footprints in Weibo instead of Twitter, the personality differences and resulted behavioral patterns are investigated in the context of Chinese culture. Varying average scores to the same personality test imply different personality landscapes across cultures [28, 29]. Specifically, it is found that Chinese tend to be more formal than the westerners [30]. Due to the fact that Twitter and Facebook blocked in China, existing findings from them [31] cannot be directly extended to understand Chinese users or predict their behaviors, which makes our exploration from Weibo, the most popular Twitter-like service in China, necessary and novel. In Facebook and Twitter, a positive link exists between extraversion and engagement in activities and extroverts use the social networking sites to relieve their anxiety [32, 33]. Here we find in Weibo, one of popular social media in China, however, lowly extraverted users are characterized by higher levels of activity. Specifically, introverts post tweets more frequently than extroverts, implying their stronger dependency on Weibo. As for online emotion expression, introversion, when compared to extroversion, is associated with higher arousal emotions like anger and fear. The forementioned findings demonstrate that the behavioral pattern of extroverts and introverts in online social media of China is indeed different from the west. They also suggest that culture might play a profound role in understanding online behaviors. For example, Chinese often appear shy and self-conscious to westerners, and generally may not express their feelings well [34], especially when they are around strangers or in circumstances that they are not familiar to. According to this theory, towards to the online and open social platform like Weibo, Chinese users might tend to hide "real themselves" from being exposed to strangers. This may explain why extroverts in Weibo surprisingly appear inactive and quiet. And on the contrary, introverts seem to be more active in the cyberspace so as to relieve excessive stress and loneliness from the real world.

Thirdly, the dimensions of online behavior are greatly enriched, ranging from locating, shopping, emotions to virtual honors. A more comprehensive map between personality difference and resulted behavioral patterns can be thus established. And this map will inspire realistic applications in scenarios like social media marketing.

2 Literature review and theoretical background

Several well-studied models have been established for personality traits, among which the Big Five model is the most popular [35, 36]. In this model, human personality is depicted on five dimensions, namely, openness, neuroticism, extraversion, agreeableness and conscientiousness, and personality type is determined based on an individual's behavior over

time and under different circumstances. The Internet, currently one of the most pervasive scenarios, has profoundly changed human behavior and experience. With its explosive development, abundant research efforts have been devoted to investigating the relation between personality and Internet usage [37]. For example, research findings have demonstrated distinctive patterns of Internet use and usage motives of individuals with different personality types, where extroverts made more goal-oriented use of Internet services [38]. Focusing on online social media as a vital component of the Internet, extraversion and openness to experiences are found to be positively related with social media adoption [39] and introverted and neurotic people locate their “real me” through social interaction [40].

Moreover, it was also shown that users’ psychological traits could be inferred from their digital fingerprints on online social media [41, 42]. Golbeck et al. attempted to bridge the gap between personality research and social media and demonstrated that social media (Facebook and Twitter) profiles reflect personality traits [43, 44]. They suggested that the number of parentheses used is negatively correlated with extraversion; however, an explanation of the correlation was not provided and probing the correlations in a larger dataset remains necessary. Quercia et al. employed numbers of followees, followers and tweets to determine personality and suggested that both popular users and influentials are extroverts with stable emotions [45]. Additionally, patterns in language use on online social media, such as words, phrases and topics, also offer a way to determine personality [46]. For example, using dimensionality reduction for the Facebook Likes of participants, Kosinski et al. proposed a model to predict individual psycho-demographic profiles [20]. In terms of social media in China, Weibo and RenRen are the ideal platforms for conducting personality research [26, 47]. Considering the recent progress that has enabled computer algorithms to outperform humans in judging personality [21], online social media offers unprecedented opportunities for personality inference and understanding human behavior.

Each bipolar dimension (like extraversion) in the Big Five model summarizes several facets that subsume several more specific traits (extraversion vs. introversion). In this paper, we focus on extraversion, which is an indispensable dimension of personality traits. Many previous studies have revealed the connection between extraversion and online behavior and can be roughly reviewed from the following perspectives.

Locations. For decisions referring to human spatial activity, the most fundamental features are arguably the personality traits, given that these are relatively persistent dispositions. Researchers argue that this is supported by evidence of personality trait correlation with diverse human activities, ranging from consumer marketing to individual tastes [48]. As for Foursquare users, extroverted people would be more sociable and outgoing and so visit more venues [49]. However, the findings of [50] for extroverts were unexpected in that they did not provide evidence of individuals preferring the same locations.

Social interactions. Highly extroverted individuals tend to have broad social communications [26] and large network size [51]. For instance, extraversion is generally positively related to the number of Facebook friends [52, 53]. Gosling et al. also found particularly strong consensus about Facebook profile-based personality assessment for extroverts [54]. However, Ross et al. showed that extroverts are not necessarily associated with more Facebook friends [4], which is contrary to the later results of Bachrach et al. [52] and Hamburger et al. [53]. Through posting tweets, extroverts are more actively sharing their lives and feelings with other people, and personality traits might shape the language styles used

on social media. In English, extroverts are more likely to mention social words such as ‘party’ and ‘love you’, whereas introverts are more likely to mention words related to solitary activities such as ‘computer’ and ‘Internet’ [46]. In Chinese, extraversion is positively correlated with personal pronouns, indicating that extroverts tend to be more concerned about others [55].

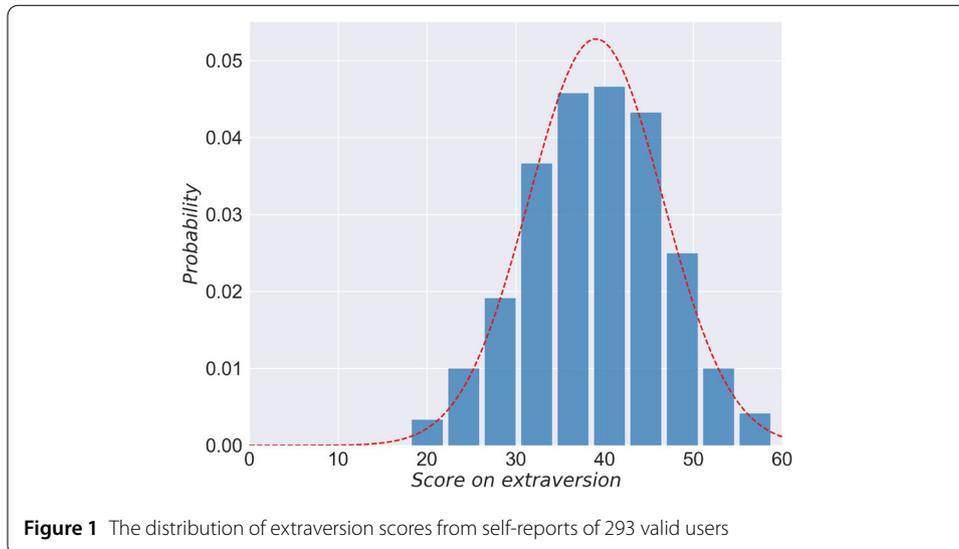
Buying intention. The personality trait of extraversion is one of the main factors driving online behavior, including buying; therefore, exploration of the relationship between extraversion and shopping is valuable. DeSarbo and Edwards found that socially isolated individuals tended to perform compulsive buying in an effort to relieve feelings of loneliness due to a lack of interactions with others [56]. However, the results of subsequent studies on the relationship between compulsive buying and extraversion are inconsistent [57, 58].

Emotion expression. In psychology, it is widely believed that extraversion is associated with higher positive affect, namely, extroverts experience increased positive emotions [59, 60]. Extroverts are also more likely to utilize the supplementary entertainment services provided by social media, which bring them more happiness [61]. Qiu et al. suggested that highly extroverted participants use these services to relieve their existential anxiety about social media [62]. Thus, it is necessary to investigate the relation between various emotions and extraversion rather than only positive affect. However, most existing studies based their conclusions on self-reports from very small samples, and the lack of data and objectivity leads to inconsistent or even conflicting results. Moreover, a comprehensive understanding of how extroverts and introverts behave differently in the context of online social media remains unclear. Hence, in this study, we employ machine learning models to identify and establish a large group of samples and then investigate the behavioral differences from diverse aspects to obtain solid evidence and comprehensive views.

3 Identification of extraversion

3.1 Dataset and participant population

The Big Five model is the most accepted and commonly used framework to depict human personality [36], and several measuring instruments have been developed to assess the Big Five personality traits. As for Big Five, personality inventories have both longer (NEO PI-R) and shorter versions (NEO-FFI). The NEO PI-R consists of 240 items and takes 45 to 60 minutes to complete, whereas the shorter NEO-FFI has only 60 items and takes 10 to 15 minutes to complete. In addition, the revision of the NEO-FFI involved the replacement of 12 ones of the 60 items. The revised edition is thought to be more suitable for younger individuals [22]. In the existing literature, the NEO-FFI is employed more often [26, 27]. Meanwhile, the resulting scales of NEO PI-R showed modest improvements in reliability and factor structure [63]. In this study, considering the cost and age of subjects, we build a web page with revised NEO-FFI to collect self-reported scores of different personality traits. We target Weibo users for voluntary participant recruitment, and both online and offline invitations were sent from December 1, 2014 to March 31, 2015. All the participants are manually checked, and only valid Weibo users (identified by Weibo ID, a unique identification for each user) are considered. The sample size generally ranges from 60 and 600 in existing studies, such as 62 in [64], 176 in [65], 250 in [66], 470 in [55] and 647 in [26]. Considering a better trade-off between cost and the reliability of the study, here we set the sample size close to 300. Finally, a total of 293 valid participants are selected for the following study (144 men and 149 women), and participants range in age from 19 to 25



years. It is worth noting that according to the official report of Weibo in 2015, users aged 17 to 33 years account for approximately 80% of total users, indicating that our refined samples of self-reports sufficiently represent the majority of Weibo users.

We focus on the Big Five personality trait of extraversion, which measures the tendency to seek stimulation in the external world and the company of others and to express positive emotions [35]. People who score high in extraversion (called extroverts) are generally outgoing, energetic and friendly. By contrast, introverts are more likely to be solitary and seek environments characterized by lower levels of external stimulation. The distribution of scores from the 293 valid samples (Weibo users) on extraversion is shown in Fig. 1. The scores follow a typical Gaussian distribution with a μ (mean value) of 39.03 and a σ (standard deviation) of 7.55. Figure 1 shows that the probability of scores near the mean value is higher than the occurrence of both high scores and low scores, indicating that a significant fraction of samples report neutral scores on extraversion and can be intuitively categorized as having no significantly distinct personality, i.e., neither extroverts nor introverts.

Therefore, it is reasonable to divide the samples into three groups: extroverts (high scores, labelled as 1), neutrals (scores around the mean, labelled as 0) and introverts (low scores, labelled as -1). Specifically, extroverts are subjects with scores greater than 42.81 ($\mu + \sigma/2$), introverts are users with scores less than 35.25 ($\mu - \sigma/2$) and neutrals are users whose scores range from 35.25 to 42.81. The thresholds ($\mu \pm \sigma/2$) are set to balance the sizes of the three categories to avoid bias in the machine learning models. By classifying 293 valid samples into three categories, we can obtain a training set to establish and evaluate machine learning models that do not require self-reports.

After obtaining permission from the valid users, we continuously collect their online profiles until March 1, 2016, including demographics and tweets posted through Weibo's open APIs. To guarantee the quality of the data, only active users with more than 100 tweets, including 45 extroverts (1), 44 introverts (-1) and 56 neutrals (0), are included in the training set. The training data are generally balanced with respect to the three classification labels, especially for extroverts and introverts, which is helpful to avoid bias in the machine learning models.

3.2 Extraversion classifier

As stated in the previous section, many aspects of online profiles have been found to be related to users' personalities. For machine learning, a basic but practical rule is to collect all the possible features in training models, especially for nonlinear ones such as support vector machine and random forest [67]. In the process of training, though mostly in a manner of black box, models will keep effective features or their combinations that play dominating roles in predicting the target and ignore weak features that functions trivially. Hence, to establish a competent classifier to identify the three categories of extraversion without the help of self-reports, we attempt to extract as many features as possible from the digital and textual records. These features are then roughly grouped into basic, interactive and linguistic features. The details of the different types of features are as follows.

Basic features. Basic features are selected to reflect the user's demographics, preliminary statuses and elementary interactions on social media, including gender, tweeting patterns and privacy settings. Specifically, tweeting patterns contain $\log(\text{AUW} + 1)$ (where AUW is the age of a user on Weibo in units of days), $\log(\text{NT} + 1)$ (where NT is the total number of tweets the user posted), $\log(\text{NT}/(\text{AUW} + 1))$ (the frequency of posting), $\log(\text{NFER} + 1)$ (where NFER is defined as the number of the user's followers), $\log(\text{NFEE} + 1)$ (where NFEE denotes the number of the user's followees), $\text{NT}/(\text{NFER} + 1)$, and $\text{NT}/(\text{NFEE} + 1)$. With respect to the privacy settings, corresponding binary features indicate whether a user allows comments from others, whether the user allows private messages sent from others and whether the user allows Weibo to track their real-time location. In addition, we consider the length of self-description as a feature. Investigation of the relationship between users' nicknames [26] and extraversion scores in our samples demonstrate that nicknames are not significantly associated with the extraversion in Weibo ($\rho > 0.05$) and functions trivially in extroversion classification (with accuracy of 30.7%). As for the status of being verified or not, because only celebrities, institutions and governments in Weibo are verified and they only occupy a small proportion of the entire population. Meanwhile, most of our samples with self-reports are not verified users. Hence, we do not select users' nicknames and verified status as features in personality prediction.

Interactive features. Interactive features are designed to reflect the sophisticated patterns of social interactions on Weibo at different time granularities of days or weeks. Here, social interaction includes posting, mentioning, and retweeting, which have been verified to be key behaviors related to extraversion in previous research. Specifically, for a certain time granularity (daily or weekly) and a certain social interaction, a vector composed of the average occurrence of the interactions (over the entire life of a user in our collection) at different hours or days of a week is calculated. Then, the following features are extracted from this vector: (1) the average number of interactions, (2) the hour or day with the most interactions, (3) the maximum occurrence of hourly or daily interaction, (4) the hour or day with the lowest occurrence of the interaction, (5) the variance of the interaction occurrence on different hours or days. Additionally, the proportions of tweets containing mentions and retweets are also considered as features to reflect a user's interactive intensity.

Linguistic features. Previous efforts to explore extraversion have demonstrated that language styles on social media can be effective indicators of personality traits. For English (like in Twitter or Facebook), linguistic variables can be derived from LIWC (Linguistic Inquiry and Word Count), a text analysis software widely used in study of language us-

age in kinds of dimensions. And specific tools can also be built to tokenize and filter out linguistic features [27]. Therefore, we collect 261 keywords which are derived from terms or phrases that frequently used in personality tests, including both Chinese and English keywords, to linguistically model the tweets posted by users of different groups. After preprocessing the text, all tweets posted by a user are combined to form a document that represents the user's language style, and all users' documents compose the corpus. Then, the classic tf-idf scores are employed to evaluate the 261 keywords. Tf-idf stands for term frequency-inverse document frequency, which is a weight often used in text mining and features selection. This weight is a statistical measure used to evaluate how important a word is to a document in a collection or corpus [68]. Through adjusting the threshold of tf-idf, the top 84 keywords (tf-idf > 500) are selected to extract linguistic features (these keywords are found to produce the best prediction in latter experiments). Specifically, for any term within the 84 selected keywords, if it occurs in a document (corresponding to a user) its feature value is 1; otherwise, it is 0. This method, called bag-of-words, is typically utilized in natural language processing [69]. We also consider the average length of tweets posted by the user.

It is worth noting that in our dataset of online profiles, there are significant differences in the scales of the extracted features. To train unbiased machine learning models, feature standardization is required. We perform min-max normalization and transform each feature into a range between zero and one. The transformation is given by

$$X_i = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}}, \quad (1)$$

where X_i is the i th item in feature set X , X_{\max} is the maximum value of X , and X_{\min} is the minimum value of X .

In summary, we extract a total of 130 features for each Weibo user, including 13 basic features, 32 interactive features and 85 linguistic features, which are used as the input of the machine learning models.

3.3 Models and accuracy

Based on the training data and feature set obtained from the previous sections, three popular machine learning models, including random forest, naive Bayes and support vector machine (SVM), are employed to perform the 3-category classification of extraversion. As for SVM, we choose C-SVM (multi-class classification) as the solution and RBF as the kernel function.

We adapt k -fold ($k = 10$ in this paper) cross-validation to estimate our model. In k -fold cross-validation [70], the original sample is randomly partitioned into k equal sized subsamples. Of the k subsamples, a single subsample is retained as the validation data for testing the model, and the remaining $k - 1$ subsamples are used as training data. The cross-validation process is then repeated k times, with each of the k subsamples used exactly once as the validation data. The baseline of accuracy for 3-category classification is 33.33%. As shown in Table 1, our 10-fold cross-validation results indicate that the random forest model cannot adequately classify extraversion (with accuracy close to the baseline). The naive Bayes and SVM models outperform the baseline solutions significantly, especially the SVM model, whose accuracy for both extroverts and introverts is approximately 50%.

Table 1 The average accuracy and F1 score of the machine learning models

Model	Random forest	Naive Bayes	SVM
Accuracy of inferring extroverts	36.98%	47.11%	52.28%
Accuracy of inferring introverts	37.75%	46.44%	49.49%
Accuracy	34.73%	39.17%	42.31%
F1 score	0.393	0.406	0.451

We also measure the average F1 score by calculating the precision and recall for each label i and find that the unweighted mean, which is defined as

$$F1\ score = \frac{1}{3} \cdot \sum_{i=-1,0,1} \frac{2 \cdot (precision_i \cdot recall_i)}{precision_i + recall_i}, \quad (2)$$

is 0.451, indicating that both precision and recall confirm the good performance of SVM.

As previous studies markedly varied in the methods employed to study the relationship between digital footprints and personality traits, Azucar et al. used Pearson's r to compare the accuracy of prediction models, and suggested this evaluation method is independent of the types of models [31]. We follow this work [31] and measure the correlation ($r = 0.26$, $\rho < 0.01^{**}$) between the actual and predicted value (1 for extroverts, 0 for neutrals and -1 for introverts) for our model. Although the accuracy is lower than some excellent results ($r = 0.40$ in [20] and $r = 0.54$ in [26]), our model is still competitive of which the performance is close to and even better than many important studies ($r = 0.19$ in [27] and $r = 0.28$ in [71]).

Therefore, we train an SVM model as the extraversion classifier to identify extroverts and introverts on Weibo without the help of self-reports. Based on the good accuracy and F1 score, we argue that machine learning models such as SVM can overcome the limitations of conventional approaches, such as self-reports, greatly extend the scope of personality exploration and offer an opportunity to comprehensively assess the behavioral differences between extroverts and introverts on social media.

4 Differences between extroverts and introverts

Employing the obtained SVM classifier, we attempt to identify extroverts and introverts from a large population of Weibo users whose online publicly available profiles were collected through Weibo's open APIs between November 2014 and March 2016. Users with less than 100 tweets were omitted to avoid sparsity. After converting each user into a representative feature set, our SVM classifier can automatically categorize the user as an extrovert, neutral or introvert. From 16,856 users, we identify 4920 extroverts and 2329 introverts. To establish a comprehensive spectrum of the behavioral discrepancy for extroverts and introverts on social media, patterns in time, geography, online activity, emotional expression and attitude to virtual honor are investigated according to our seven research questions.

4.1 Temporal differences

Users with different personality traits might post tweets unevenly at different hours of the day, i.e., hourly posting patterns, which can be reflected by the distribution of tweets by hour of the day. As shown in Fig. 2, introverts prefer to post tweets from 8:00 to 18:00, whereas extroverts are active and excited from 19:00 to 1:00 of the next day, indicating that

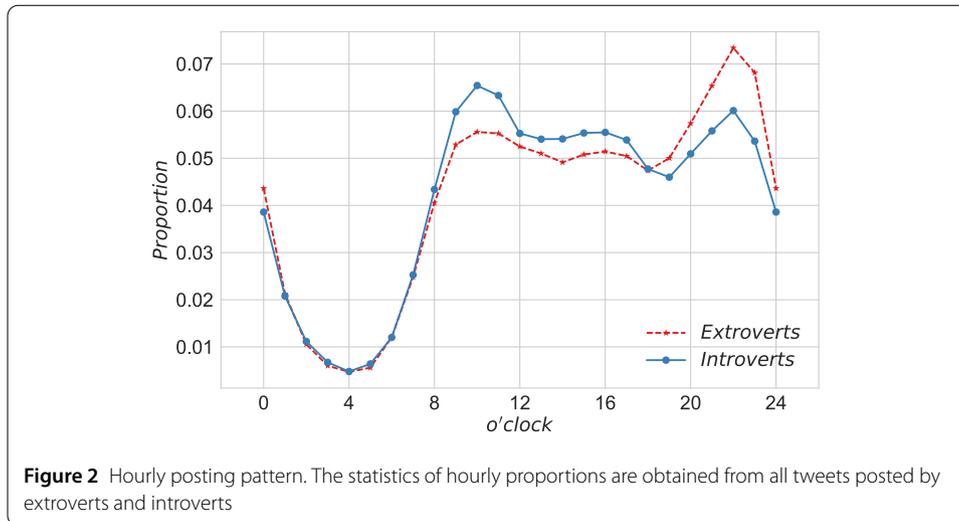


Table 2 Tweeting habits during different periods of a day at the individual level

Time	1:00–8:00	8:00–19:00	19:00–1:00
Extroverts	0.085	0.557	0.358
Introverts	0.087	0.608	0.305

Table 3 Average intervals and standard deviations of posting tweets for extroverts and introverts

Time interval (hour)	Mean	Standard deviation
Extroverts	28.10	74.36
Introverts	19.09	62.25

Time interval (hour) within 24 hours	Mean	Standard deviation
Extroverts	6.41	6.76
Introverts	5.61	6.19

extroverts are more vibrant than introverts at night. Further evidence at the individual level is presented in Table 2: the proportion of extroverts tweeting during the day (from 8:00 to 19:00) is 0.557 and that of introverts is 0.608; the proportion of extroverts tweeting at night (from 19:00 to 1:00 of the following day) is 0.358 while that for introverts is 0.305. The active posting of extroverts at night suggests that their nightlife is more diverse than that of introverts.

The interval between two temporally consecutive tweets of an individual Weibo user is an excellent indicator of the degree of preference and dependency on social media. We calculate the average interval (in units of hour) for extroverts and introverts from the timestamps of their tweets. As shown in Table 3, introverts post tweets more frequently than do extroverts, suggesting stronger dependency on social media. Specifically, the mean interval of introverts is 19.09 hours while that for extroverts is 28.10 hours. This finding is consistent with and can be explained by the previous finding that individuals who are socially isolated tend to depend on and indulge in social media to relieve loneliness due to a lack of interactions with others in the real world [56]. Meanwhile, differences in the standard deviations (respectively, 62.25 and 74.36 hours) reveal that the posting frequency of introverts on Weibo is more consistent than that of extroverts. Moreover, if considering

Table 4 ANOVA of tweeting habits and average intervals

	Sum of squares	Mean square	F-value	p-value
1:00–8:00	1	0.7087	0.812	0.367
8:00–19:00	137	137.17	160.7	***
19:00–1:00	200	199.84	236.5	***
Time interval	294	293.75	353.1	***
Time interval within 24 hours	94	93.89	109.2	***

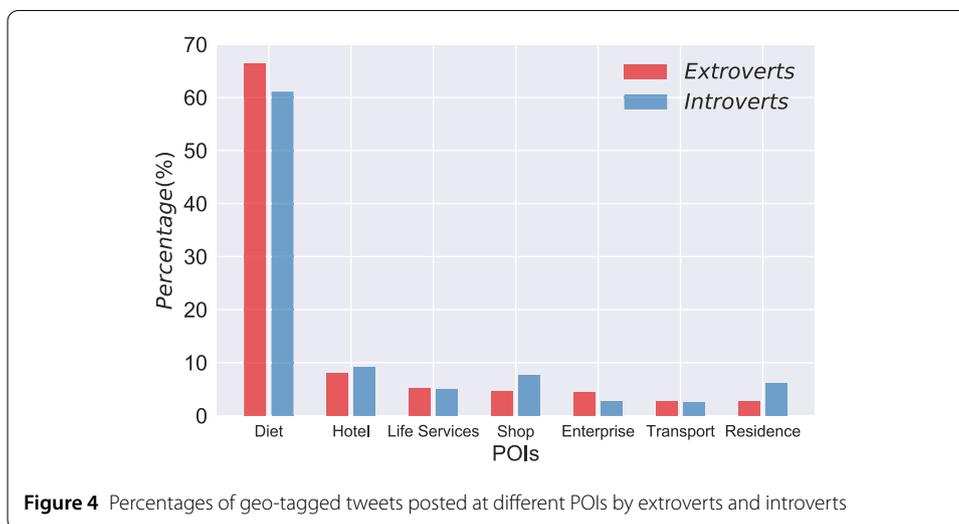
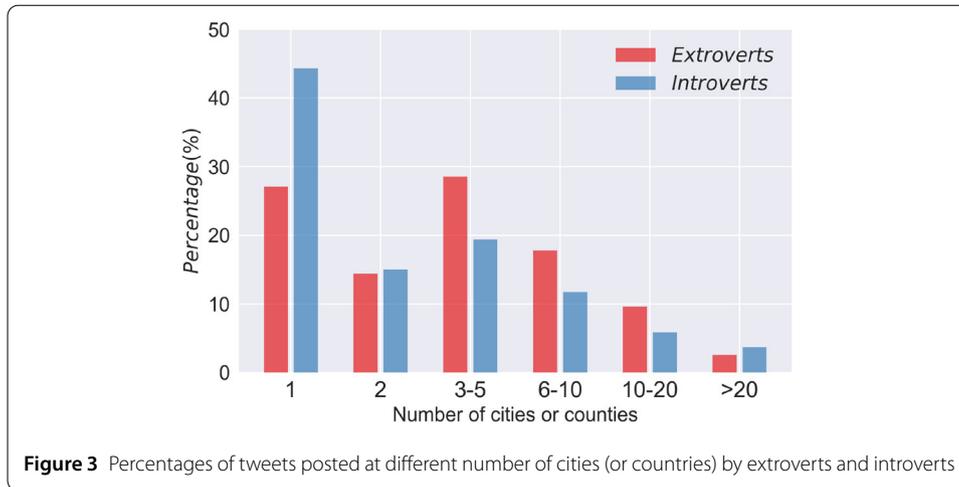
Note: ** $p < 0.01$, *** $p < 0.001$.

only the time interval within one day (i.e., ignoring intervals greater than 24 hours), the mean interval of extroverts decreases to 6.41 hours and that of introverts is 5.61 hours. In this case, the standard deviation of the time interval of extroverts is 6.76 hours, which is greater than that of introverts (6.19 hours). This results further justifies the finding that introverts post more frequently than extroverts do and illustrates the greater preference for social media usage. In addition, we perform analysis of variance (ANOVA), which analyses the differences among group means and their associated procedures, to further verify the results. Table 4 shows that the differences between the two groups are statistically significant in tweeting habits and average intervals. Note that most users are inactive on Weibo during the night-period between 1:00AM–8:00AM [23, 72], hence it is expected that the p -value of testing referring to tweeting during this period is above 0.01, i.e., the difference between extroverts and introverts cannot be sensed when they are all inactive.

4.2 Spatial differences

An individual user can post geo-tagged tweets (or checkins) containing the latitude and longitude of their current location. This information provides a proxy to explore geographical differences between extroverts and introverts. To perform the geo-analysis, we extract the geographical locations of 57,710 tweets, of which 38,729 tweets are posted by extroverts and 18,981 by introverts.

For each geo-tagged tweet, we transform the longitude and latitude to the corresponding city (or county) through GeoPy project [73]. Then, for each user, we can obtain the list of cities or counties where they posted tweets. The results of the comparison of extroverts and introverts are shown in Fig. 3, which surprisingly illustrates significant differences in the spatial characteristics of users with different personality traits. The results show that 44.32% of introverts post tweets from a single city or county, perhaps their residence, suggesting that nearly half of introverts prefer to stay in a familiar city or county. By contrast, only 27.12% of extroverts are located in a single city or county, which is far less than that of introverts. More specific, 14% of extroverts and 15% of introverts post tweets from two cities (or counties), 29% of extroverts and 19% of introverts post tweets from 3–5 cities or counties, and the trend of extroverts tweeting at more places holds persistently as the city (or county) number increases from 6 to 20. It is noteworthy that the number of extroverts posting tweets in 3–5 cities is even more than the number posting from a single city, which implies that extroverts prefer going to or visiting more cities than do introverts, for whom posting from a single city or county dominates. When the number of cities (or counties) is greater than 20, the ratio within introverts is unexpectedly significantly greater than that of extroverts, implying that a tiny fraction of introverts might attempt to camouflage their own loneliness by posting tweets from a large number of different places [56].



Beyond the city granularity, we can also perform a geographical comparison with better resolution, such as point-of-interest (POI), which is a detailed description of the featuring function of small regions or point locations within cities. Specifically, POIs exclude private facilities, such as personal residences, but include many public facilities that seek to attract the general public, such as retail businesses, amusement parks, and industrial buildings. Government buildings and significant natural features are also POIs. POIs also refer to hotels, restaurants, fuel stations and other categories in automotive navigation systems and recommendation systems [74]. In this study, nine types of POIs, namely, restaurants, hotels, life services, shops, enterprises, transportation, entertainment, neighborhoods and education, are considered and the percentages of the six most-visited POIs by extroverts and introverts are shown in Fig. 4. Most geo-tagged tweets are posted from restaurants, accounting for 66.38% of tweets within the nine POIs we select by extroverts and 61.10% by introverts. There are significant differences between extroverts and introverts in visiting shops and enterprises. The percentage of tweets from shops is 4.58% for extroverts and 7.68% for introverts, indicating that introverts prefer to checkin or post tweets while shopping. Furthermore, the percentage of tweets from companies and enterprises is 4.46% for

extroverts and 2.59% for introverts. Since companies and enterprises are the workplaces of individuals, this result suggests that extroverts tend to post while they are working.

4.3 Online activities

Diverse online activities, such as sharing, interacting and buying, on Weibo can be identified only through tweets posted by users. Hence, by mining the text of tweets posted by extroverts and introverts, we attempt to illustrate the behavioral difference landscape of online activities.

Sharing. Each tweet in Weibo is labelled with a tag to indicate its posting source. For example, if a user logs into Weibo and posts one tweet, the source could be a mobile device (e.g., iPhone) or web browser (e.g., Chrome). Weibo users share news, videos, and music with their friends or the public on social media, and the diverse sources of this shared information are contained in the posted tweets, for example, news websites, mobile applications or other social platforms that offer a sharing interface to Weibo. Additionally, tweets shared from selfie mobile software are tagged as selfies, which contain a self-portrait typically taken by a camera phone. Because of these features, we utilize the source label of each tweet to analyse the sharing behavior of extroverts and introverts. The contributions of the above four types of sharing in terms of all tweets are shown in Fig. 5. The fraction of news sharing of introverts (0.612%) is three times greater than that of extroverts (0.194%). By contrast, extroverts enjoy sharing more videos, music and selfies on social media than do introverts, especially selfies, e.g., the fraction of selfie tweets for extroverts is 0.354%, which is much higher than that of introverts (0.128%). It is widely believed that selfies are related to individual narcissism [75–77], and our findings further suggest that extraversion is positively coupled with selfies on social media.

Interacting. Interacting patterns, especially the mentions and retweets, are comprehensively considered in the feature set we extract in section *Extraversion classifier* and are used as the input of the extraversion classifier. Intuitively, performing the analysis of behavioral differences of interacting patterns on extroverts and introverts identified by the classifier would be meaningless because the differences have already been latently considered in the classifier. To avoid the biased comparison and provide solid evidence, we perform the difference analysis directly on the training set, i.e., user self-reports. We use

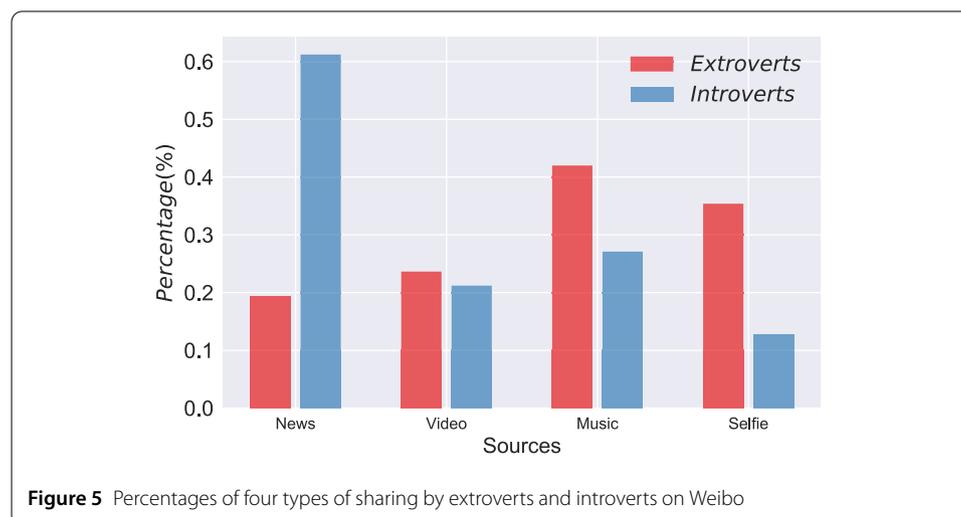


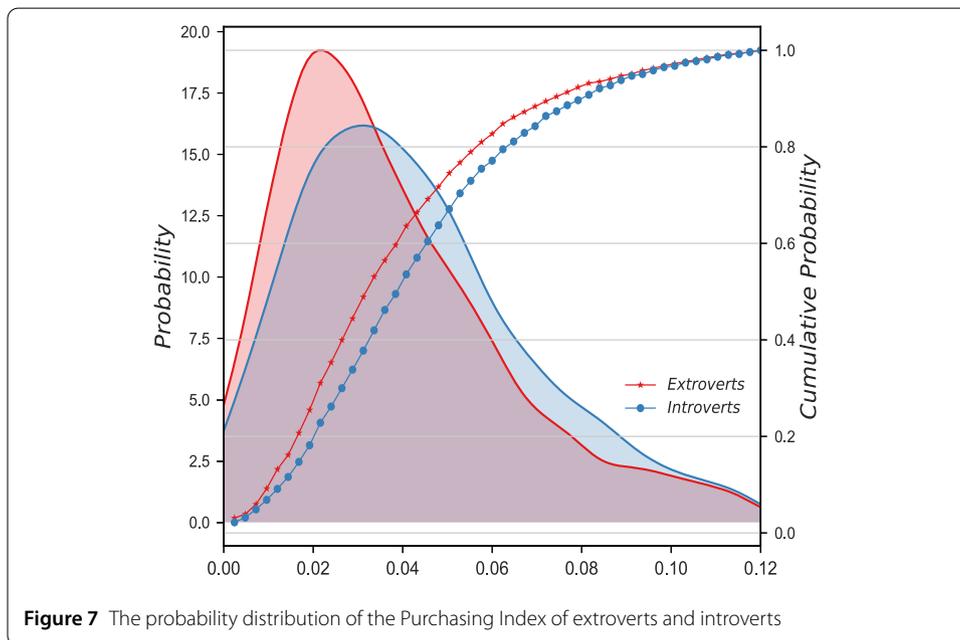
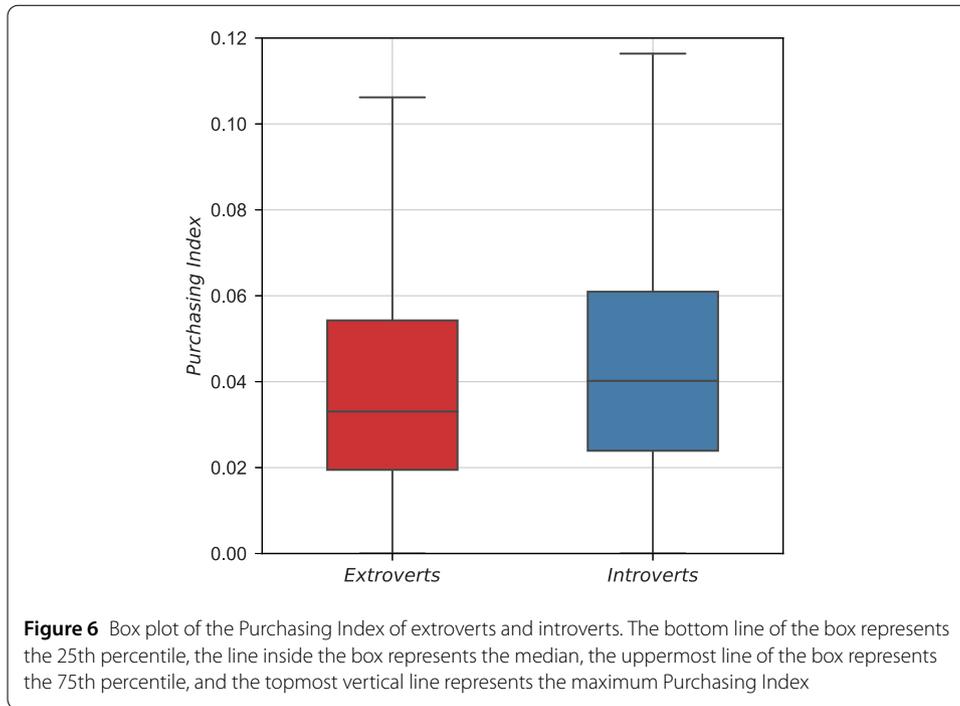
Table 5 Interacting features with relatively high Pearson correlations with extraversion scores

Feature	Coef.
Rate of @ in tweets	0.172**
Average number of posting most tweets with @ within one hour	0.150**
Variance of posting tweets with @ by hours of the day	0.135**
Variance of posting tweets with @ by day of the week	0.131**

Note: ** $p < 0.01$, *** $p < 0.001$.

the Pearson correlation to measure the linear dependence between the interaction features and the extraversion scores of participants. Features with relatively high Pearson correlations (Coef. > 0.13) with respect to extraversion scores are listed in Table 5. It is interesting to find that the features related to mentioning behavior (@ behavior) are positively correlated with extraversion scores. Mentioning behavior is regarded as one of the most important forms of online interactions. Specifically, both the rate of @ in all tweets and the average rate of tweets with @ within one hour on one day reflect the frequency of interaction with other users. Meanwhile, the variance of tweets posted with @ by hour of the day or by day of the week illustrate the irregularity and randomness of the mentioning behavior of users. Therefore, on the basis of Table 5, we conclude that extroverts are more socially active and interactive than introverts on social media; however, their interactions are more casual and temporally less regular than those of introverts.

Buying. The most intrinsic nature of social media is status updates; thus, an experience such as buying or shopping can be identified by counting related keywords, namely, the word-count method, which has been employed extensively in the field of psychology [78] in recent decades. In this study, 14 buying keywords are selected to identify buying behavior (e.g., BUY and SHOPPING), response in sales promotion (e.g., DISCOUNT and 11.11, a famous day for promotion sales in China advocated by Taobao Inc.) and mentioning or sharing of online shopping malls (like AMAZON and TAOBAO). For tweets posted by extroverts or introverts, those containing one or more of the selected keywords are labelled as buying related, and the fraction of buying tweets of each user is defined as the Purchasing Index, which reflects the intensity of the buying behavior or purchasing intention. After calculating the Purchasing Index of each user, the comparison of the buying behavior of extroverts and introverts is depicted in Fig. 6. The mean Purchasing Index of extroverts is 0.0440 and that of introverts is 0.048, which is 10% larger than that of extroverts. The 25th percentile, median, 75th percentile and maximum of the Purchasing Index of extroverts are, respectively, 0.020, 0.033, 0.054 and 0.774, and those of introverts are 0.0239, 0.0402, 0.0609 and 0.8480. Figure 7 shows the cumulative distribution function (CDF) and probability distribution function (PDF) of the Purchasing Index of extroverts and introverts. The Purchasing Indexes of more than 95% of users are less than 0.1, and the significant difference between extroverts and introverts is mainly located in this region. Specifically, at the same Purchasing Index level (e.g., > 0, > 0.04, > 0.06, > 0.08), the probability of introverts is always greater than that of extroverts, suggesting that introverts prefer to publish tweets that refer to purchasing compared to extroverts. This conclusion could apply to the advertising and sales of commodities and other realistic scenarios, i.e., introverts might be ideal marketing targets for online promotions. We adapt ANOVA to investigate the presence of significant differences in the Purchasing Index. As shown seen Table 6, the p -value of less than 0.001 indicates that the differences between the extroverts and introverts are statistically significant.



4.4 Online emotional expression

Tweets on social media not only deliver factual information but also the feelings of users, and these feelings can be automatically classified into different emotions by mining the text of tweets [23]. Because extraversion is widely believed to be associated with higher positive affect, namely, extroverts experience more positive emotions [59, 60], we investigate the differences between extroverts and introverts from the perspective of emotional expression. By employing a previously built system named MoodLens [23], we categorize each tweet into one of five emotions: anger, disgust, happiness, sadness or fear. Note that

Table 6 ANOVA of purchasing indexes

	Sum of squares	Mean square	F-value	p-value
Between groups	0.029	0.029	14.497	***
Within groups	14.563	0.002		

Note: ** $p < 0.01$, *** $p < 0.001$.

tweets without significant emotional propensity are ignored. Then, for each individual, either extrovert or introvert, we calculate the emotion index for all five sentiments, which is defined as the fraction of corresponding emotional tweets in the tweeting history and quantitatively represents the user's emotional disposition on social media.

Figure 8 shows the CDFs and PDFs of the five emotion indexes of extroverts and introverts. At the same Anger Index level ((0.1, 0.4)), Fear Index level ((0, 0.25)) and Disgust Index level ((0.05, 0.15)), the probabilities of introverts are always greater than those of extroverts, indicating that introverts post more tweets associated with negative feelings than do extroverts. However, for the Sadness Index and Happiness Index, the probabilities of introverts are always less than those of extroverts, suggesting that extroverts post tweets associated with joy or sadness with greater likelihood. Note that as can be seen in Figs. 8(d) and 8(e), the differences in the Happiness Index and Disgust Index are subtle, but ANOVA (shown in Table 7) confirms the significance with p -value < 0.001 . Our findings are consistent with the previous statement that extraversion is associated with higher positive affect; however, we also provide evidence that introversion is associated with high arousal and negative emotions like anger, fear and disgust and that extraversion is positively correlated with sadness. Indeed, on the basis of the data-driven approach on a large sample, our study simultaneously confirms the existing conclusion and provides new insights.

4.5 Attitudes to virtual honor

Weibo grants many optional badges to users that can be obtained by completing necessary operations following the demand of social media. For instance, users can connect their Weibo account and Taobao account to obtain the "Binding-Taobao" badge. This behavior, exposing the Taobao account to social media, is a risk to property security and privacy. However, the badges that users obtain are displayed publicly to others and are treated as an honor in the virtual world. Therefore, a user's response to badges is an indicator of their attitude to virtual honor in social media. We investigate the difference in attitude to virtual honor of extroverts and introverts. The "Binding-Taobao" badge is regarded as a relevant badge to perform the difference analysis, and the distributions of extroverts and introverts in terms of "Binding-Taobao" badges is shown in Fig. 9. The percentage of extroverts with the "Binding-Taobao" badge is 60.7% and that without the badge is 39.3%. The percentage of introverts with the "Binding-Taobao" badge is 53.9% and that without the badge is 46.1%. Clearly, the proportion of extroverts who obtain "Binding-Taobao" badges is larger than that of introverts. Additionally, we examine various other badges in Weibo, including "Red envelope 2015", "Public welfare", "Travel 2013", and "Red envelope 2014". The badge statistics indicate that extroverts tend to prefer badges more than introverts do; in other words, extroverts attach more importance to online virtual honor. Furthermore, from a marketing perspective, users with the "Binding-Taobao" badge are viewed as ideal targets, i.e., extroverts should be recommended with greater odds. However, as mentioned above,

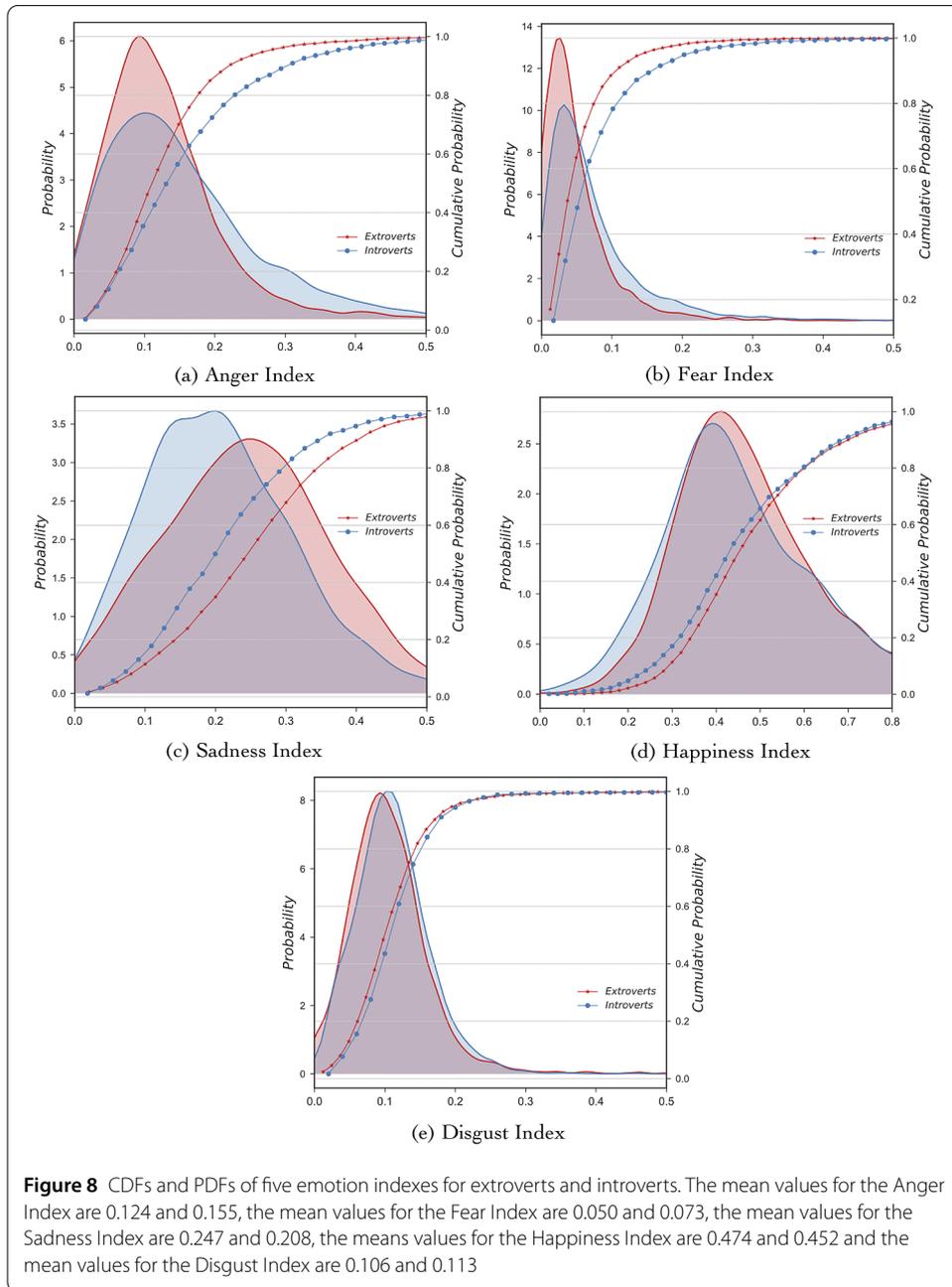
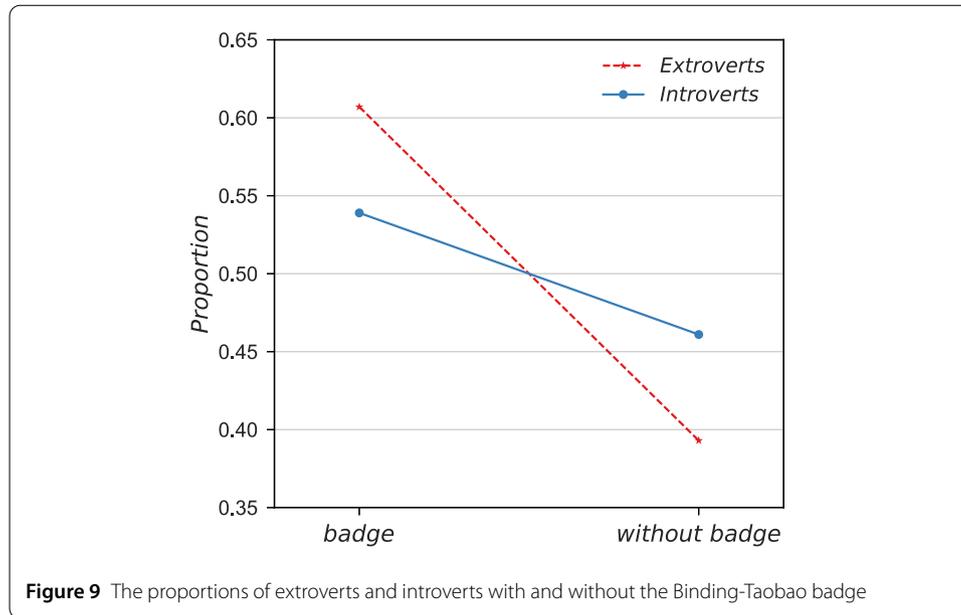


Table 7 ANOVA of mood indexes

	Sum of squares	Mean square	F-value	p-value
Anger	169	169.33	199.2	***
Fear	210	210.48	249.2	***
Sadness	156	156.48	183.8	***
Happiness	31	30.616	35.29	***
Disgust	22	21.949	25.27	***

Note: ** $p < 0.01$, *** $p < 0.001$.



introverts with higher Purchasing Index demonstrate stronger potential shopping intentions. Therefore, the badge could be a misleading signal, and our findings suggest that additional features should be comprehensively considered in marketing decisions.

To summarize, from the perspective of tempo-spatial patterns, online activities, emotional expression and attitudes to virtual honor, we establish a comprehensive picture of how extroverts tweet differently from introverts in social media.

5 Conclusion

Personality traits, such as extraversion, are believed to play fundamental roles in driving human action; however, a detailed and comprehensive understanding of how people with different personality traits behave is missing, especially in the context of social media, which has become an indispensable part of daily life. Meanwhile, the lack of large samples and the unavoidable subjectivity cause conventional methods, such as self-reports, to produce bias. Hence, in this study, we argue that starting from a small-scale but refined voluntary sample and establishing a map between self-reports and online profiles can help to train a machine learning model to automatically infer the personalities of a massive number of individuals objectively without the costly expense of survey questionnaires. Specifically, a medium size of samples (active users with at least 100 posts in Weibo) are selected to finish self-reports and three types of features extracted from their profiles help train and optimize a model with competent performance (52.28% for extroverts and 49.49% for introverts) in extroversion prediction. Indeed, the SVM classifier helps us to filter out more than 7000 extroverts and introverts from Weibo and, to the best of our knowledge, build the first complete picture of how extroverts and introverts tweet differently on social media from the perspective of dimensions.

In addition to obtaining conclusions consistent with existing findings from conventional methods, new and insightful conclusions on Weibo are obtained. We demonstrate introverts on Weibo post more frequently than do extroverts, especially during the day, which is inconsistent with the findings on Twitter or Facebook [79, 80]. A tiny fraction of introverts locate themselves on a large number of different areas (> 20). It is necessary for scholars

to examine whether the unexpected phenomenon could appear on Twitter. As for online buying intention, introverts devote more efforts to posting shopping tweets. We also find that introverts post more high-arousal emotions. As a result, introverts tend to frequently use Weibo to assuage their isolation in the real world, which is consistent with early studies [39, 40]. Finally, we suggest that extroverted individuals may be optimal candidates for online promotion campaigns with virtual honor. Our findings offer solid evidence of the feasibility of machine learning approaches to personality research and shed light on realistic applications, such as online marketing and behavior analytics.

This study has inevitable limitations. The personality prediction model could be promisingly improved by adding new features including footprints of Likes [20], user-generated pictures [27], headshots and emoticons [81]. And exploring more details of how culture shapes online behavior will better interpret our findings. Besides, here we only investigate the personality difference and resulted behavioral patterns from the dimension of extroversion, ignoring other personality traits such as openness or conscientiousness. In fact, further explorations in these traits would essentially enrich the picture of how personality impact online behaviors. All these limitations will be promising directions in the future work.

Acknowledgements

This work was supported by the National Key Research and Development Program of China (Grant No. 2016QY01W0205), NSFC (Grant Nos. 71501005 and 61421003) and the fund of the State Key Lab of Software Development Environment (Grant No. SKLSDE-2017ZX-05).

List of abbreviations

MTurk, Amazon Mechanical Turk; MAU, monthly active users; SVM, support vector machine; NEO PI-R, Revised NEO Personality Inventory; NEO-FFI, NEO Five-Factor Inventory; AUW, age of a user on Weibo in units of days; NT, the total number of tweets the user posted; NFER, the number of the user's followers; NFEE, number of the user's followees; LIWC, Linguistic Inquiry and Word Count; POI, point-of-interest.

Availability of data and materials

The self-reports and online profiles of users mentioned in this study are publicly available to the research community after careful anonymization and can be downloaded freely through <https://doi.org/10.6084/m9.figshare.4765150.v1>.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

ZJ, ZZ and XK conceived of and designed the research. ZZ and ZJ conducted the experiments and analysed the results. ZJ, ZZ and XK wrote the paper. All authors read and approved the final manuscript.

Author details

¹State Key Lab of Software Development Environment, Beihang University, Beijing, China. ²School of Economics and Management, Beihang University, Beijing, China. ³Beijing Advanced Innovation Center for Big Data and Brain Computing, Beijing, China.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 23 November 2017 Accepted: 22 June 2018 Published online: 03 July 2018

References

1. Boyd DM, Ellison NB (2007) Social network sites: definition, history, and scholarship. *J Comput-Mediat Commun* 13(1):210–230
2. Barash V, Ducheneaut N, Isaacs E, Bellotti V (2010) Faceplant: impression (mis)management in Facebook status updates. In: ICWSM
3. Larsen RJ, Buss DM (2008) *Personality psychology: domains of knowledge about human nature*. McGraw-Hill Education, New York
4. Ross C, Orr ES, Sisc M, Arseneault JM, Simmering MG, Orr RR (2009) Personality and motivations associated with Facebook use. *Comput Hum Behav* 25(2):578–586
5. Ryan T, Xenos S (2011) Who uses Facebook? An investigation into the relationship between the Big Five, shyness, narcissism, loneliness, and Facebook usage. *Comput Hum Behav* 27(5):1658–1664

6. Simoncic TE, Kuhlman KR, Vargas I, Houchins S, Lopez-Duran NL (2014) Facebook use and depressive symptomatology: investigating the role of neuroticism and extraversion in youth. *Comput Hum Behav* 40:1–5
7. Costa PT, McCrae RR (1992) Revised NEO personality inventory (NEO PI-R) and NEO five-factor inventory (NEO-FFI). Psychological Assessment Resources, Odessa
8. Roberts BW, Chernyshenko OS, Stark S, Goldberg LR (2005) The structure of conscientiousness: an empirical investigation based on seven major personality questionnaires. *Pers Psychol* 58(1):103–139
9. Fast LA, Funder DC (2008) Personality as manifest in word use: correlations with self-report, acquaintance report, and behavior. *J Pers Soc Psychol* 94(2):334
10. Hoonakker P, Carayon P (2009) Questionnaire survey nonresponse: a comparison of postal mail and Internet surveys. *Int J Hum-Comput Interact* 25(5):348–373
11. Watt JH (2004) Internet systems for evaluation research. *New Dir Eval* 1999(84):23–43
12. Nederhof AJ (1985) Methods of coping with social desirability bias: a review. *Eur J Soc Psychol* 15(3):263–280
13. Furnham A (1986) Response bias, social desirability and dissimulation. *Pers Individ Differ* 7(3):385–400
14. Klausch T, Schouten B, Buelens B, Van Den Brakel J (2017) Adjusting measurement bias in sequential mixed-mode surveys using re-interview data. *J Surv Stat Methodol* 5(4):409–432
15. Paolacci G, Chandler J, Ipeirotis PG (2010) Running experiments on Amazon mechanical turk. *Judgm Decis Mak* 5(5):411–419
16. Wright KB (2005) Researching Internet-based populations: advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *J Comput-Mediat Commun* 10(3):JCMC1034
17. Bohannon J (2016) Mechanical turk upends social sciences. *Science* 352(6291):1263–1264
18. Peer E, Brandimarte L, Samat S, Acquisti A (2017) Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *J Exp Soc Psychol* 70:153–163
19. Lazer D, Pentland A, Adamic L, Aral S, Barabási A-L, Brewer D, Christakis N, Contractor N, Fowler J, Gutmann M, Jebara T, King G, Macy M, Roy D, Van Alstyne M (2009) Computational social science. *Science* 323(5915):721–723
20. Kosinski M, Stillwell D, Graepel T (2013) Private traits and attributes are predictable from digital records of human behavior. *Proc Natl Acad Sci USA* 110(15):5802–5805
21. Youyou W, Kosinski M, Stillwell D (2015) Computer-based personality judgments are more accurate than those made by humans. *Proc Natl Acad Sci USA* 112(4):1036–1040
22. McCrae RR, Costa PT (2004) A contemplated revision of the NEO five-factor inventory. *Pers Individ Differ* 36(3):587–596
23. Zhao J, Dong L, Wu J, Xu K (2012) MoodLens: an emoticon-based sentiment analysis system for Chinese tweets. In: *Proceedings of the 18th ACM SIGKDD*. ACM, New York, pp 1528–1531
24. Kedar SV, Bormane DS (2016) Automatic personality assessment: a systematic review. In: *International conference on information processing*, pp 326–331
25. Park G, Schwartz HA, Eichstaedt JC, Kern ML, Kosinski M, Stillwell DJ, Ungar LH, Seligman ME (2015) Automatic personality assessment through social media language. *J Pers Soc Psychol* 108(6):934
26. Li L, Li A, Hao B, Guan Z, Zhu T (2014) Predicting active users' personality based on micro-blogging behaviors. *PLoS ONE* 9(1):e84997
27. Liu L, Preotiuc-Pietro D, Samani ZR, Moghaddam ME, Ungar LH (2016) Analyzing personality through social media profile picture choice. In: *ICWSM*, pp 211–220
28. McCrae RR, Terracciano A (2005) Personality profiles of cultures: aggregate personality traits. *J Pers Soc Psychol* 89(3):407
29. Barceló J (2017) National personality traits and regime type: a cross-national study of 47 countries. *J Cross-Cult Psychol* 48(2):195–216
30. Lew WJ (1998) *Understanding the Chinese personality: parenting, schooling, values, morality, relations, and personality*. Edwin Mellen Press, Lewiston
31. Azucar D, Marengo D, Settanni M (2018) Predicting the big 5 personality traits from digital footprints on social media: a meta-analysis. *Pers Individ Differ* 124:150–159
32. Kuss DJ, Griffiths MD (2011) Online social networking and addiction—a review of the psychological literature. *Int J Environ Res Public Health* 8(9):3528–3552
33. Blackwell D, Leaman C, Tramosch R, Osborne C, Liss M (2017) Extraversion, neuroticism, attachment style and fear of missing out as predictors of social media use and addiction. *Pers Individ Differ* 116:69–72
34. Hays J (2015) Chinese personality traits: indirectness, pragmatism, competition and losing face. <http://factsanddetails.com/china/cat4/sub18/item116.html>. Accessed June 2015
35. Goldberg LR (1992) The development of markers for the Big-Five factor structure. *Psychol Assess* 4(1):26–42
36. Gosling SD, Rentfrow PJ, Swann WB (2003) A very brief measure of the Big-Five personality domains. *J Res Pers* 37(6):504–528
37. Orchard LJ, Fullwood C (2010) Current perspectives on personality and Internet use. *Soc Sci Comput Rev* 28(2):155–169
38. Amiel T, Sargent SL (2004) Individual differences in Internet usage motives. *Comput Hum Behav* 20(6):711–726
39. Correa T, Hinsley AW, de Zúñiga HG (2010) Who interacts on the Web?: the intersection of users' personality and social media use. *Comput Hum Behav* 26(2):247–253
40. Amichai-Hamburger Y, Wainapel G, Fox S (2002) "On the Internet no one knows I'm an introvert": extroversion, neuroticism, and Internet interaction. *Cyberpsychol Behav* 5(2):125–128
41. Ellison NB, Steinfield C, Lampe C (2007) The benefits of Facebook "friends": social capital and college students' use of online social network sites. *J Comput-Mediat Commun* 12(4):1143–1168
42. Ong EYL, Ang RP, Ho JCM, Lim JCY, Goh DH, Lee CS, Chua AYK (2011) Narcissism, extraversion and adolescents' self-presentation on Facebook. *Pers Individ Differ* 50(2):180–185
43. Golbeck J, Robles C, Turner K (2011) Predicting personality with social media. In: *International conference on human factors in computing systems, CHI 2011, extended abstracts volume*, pp 253–262
44. Golbeck J, Robles C, Edmondson M, Turner K (2011) Predicting personality from Twitter. In: *SocialCom/PASSAT*, pp 149–156

45. Quercia D, Kosinski M, Stillwell D, Crowcroft J (2011) Our Twitter profiles, our selves: predicting personality with Twitter. In: *SocialCom/PASSAT*, pp 180–185
46. Schwartz HA, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones SM, Agrawal M, Shah A, Kosinski M, Stillwell D, Seligman ME et al (2013) Personality, gender, and age in the language of social media: the open-vocabulary approach. *PLoS ONE* 8(9):e73791
47. Bai S, Gao R, Zhu T (2012) Determining personality traits from RenRen status usage behavior. In: *Computational visual media: first international conference, CVM 2012. proceedings*, pp 226–233.
48. Chorley MJ, Whitaker RM, Allen SM (2015) Personality and location-based social networks. *Comput Hum Behav* 46(C):45–56
49. Chorley MJ, Colombo GB, Allen SM, Whitaker RM (2013) Visiting patterns and personality of Foursquare users. In: *2013 third international conference on cloud and green computing (CGC)*, pp 271–276
50. Noë N, Whitaker RM, Chorley MJ, Pollet TV (2016) Birds of a feather locate together? Foursquare checkins and personality homophily. *Comput Hum Behav* 58(C):343–353
51. Noë N, Whitaker RM, Allen SM (2016) Personality homophily and the local network characteristics of Facebook. In: *2016 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM)*, pp 386–393
52. Bachrach Y, Kosinski M, Graepel T, Kohli P, Stillwell D (2012) Personality and patterns of Facebook usage. In: *Proceedings of the 4th annual ACM web science conference. ACM, New York*, pp 24–32
53. Amichai-Hamburger Y, Vinitzky G (2010) Social network use and personality. *Comput Hum Behav* 26(6):1289–1295
54. Gosling SD, Gaddis S, Vazire S (2007) Personality impressions based on Facebook profiles. In: *ICWSM, vol 7*, pp 1–4
55. Qiu L, Lu J, Ramsay J, Yang S, Qu W, Zhu T (2016) Personality expression in Chinese language use. *Int J Psychol*. <https://doi.org/10.1002/ijop.12259>
56. Desarro W, Edwards E (1996) Typologies of compulsive buying behavior: a constrained clusterwise regression approach. *J Consum Psychol* 5(3):231–262
57. Mowen JC, Spears N (1999) Understanding compulsive buying among college students: a hierarchical approach. *J Consum Psychol* 8(4):407–430
58. Gohary A, Hanzaae KH (2014) Personality traits as predictors of shopping motivations and behaviors: a canonical correlation analysis. *Arab Econ Bus J* 9(2):166–174
59. McCrae RR, Costa PT (2003) *Personality in adulthood: a five-factor theory perspective*. Guilford Press, New York
60. Smillie LD, Wilt J, Kabbani R, Garratt C, Revelle W (2015) Quality of social experience explains the relation between extraversion and positive affect. *Emotion* 15(3):339–349
61. Deng S, Liu Y, Li H, Hu F (2013) How does personality matter? An investigation of the impact of extraversion on individuals' SNS use. *Cyberpsychol Behav Soc Netw* 16(8):575–581
62. Qiu L, Leung AK, Ho JH, Yeung QM, Francis KJ, Chua PF (2010) Understanding the psychological motives behind microblogging. *Stud Health Technol Inform* 154:140–144
63. Costa PT, McCrae RR (2008) The revised NEO personality inventory (NEO-PI-R). In: *The SAGE handbook of personality theory and assessment, vol 2*, pp 179–198
64. Skowron M, Tkaličič M, Ferwerda B, Schedl M (2016) Fusing social media cues: personality prediction from Twitter and Instagram. In: *Proceedings of the 25th international conference companion on world wide web*, pp 107–108
65. Gao R, Hao B, Bai S, Li L, Li A, Zhu T (2013) Improving user profile with personality traits predicted from social media content. In: *Proceedings of the 7th ACM conference on recommender systems. ACM, New York*, pp 355–358
66. Markovikj D, Gievska S, Kosinski M, Stillwell D (2013) Mining Facebook data for predictive personality modeling. In: *Proceedings of the 7th international AAAI conference on weblogs and social media (ICWSM 2013)*, pp 23–26
67. Kotsiantis SB, Zaharakis I, Pintelas P (2007) Supervised machine learning: a review of classification techniques. In: *Emerging artificial intelligence applications in computer engineering. Frontiers in artificial intelligence and applications, vol 160*, pp 3–24
68. Aizawa A (2003) An information-theoretic perspective of tf-idf measures. *Inf Process Manag* 39(1):45–65
69. Jiang J, Zhai C (2007) A systematic exploration of the feature space for relation extraction. In: *HLT-NAACL*, pp 113–120
70. Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *IJCAI*, pp 1137–1145
71. Qiu L, Lin H, Ramsay J, Yang F (2012) You are what you tweet: personality expression and perception on Twitter. *J Res Pers* 46(6):710–718
72. Hu Y, Zhao J, Wu J (2016) Emoticon-based ambivalent expression: a hidden indicator for unusual behaviors in Weibo. *PLoS ONE* 11(1):e0147079
73. ijl (nickname in GitHub): Geopy. GitHub (2016)
74. Ye M, Yin P, Lee W-C, Lee D-L (2011) Exploiting geographical influence for collaborative point-of-interest recommendation. In: *Proceedings of the 34th international ACM SIGIR conference on research and development in information retrieval. ACM, New York*, pp 325–334
75. Sorokowski P, Sorokowska A, Oleszkiewicz A, Frackowiak T, Huk A, Pisanski K (2015) Selfie posting behaviors are associated with narcissism among men. *Pers Individ Differ* 85:123–127
76. Weiser EB (2015) #Me: narcissism and its facets as predictors of selfie-posting frequency. *Pers Individ Differ* 86:477–481
77. Wang D (2017) A study of the relationship between narcissism, extraversion, drive for entertainment, and narcissistic behavior on social networking sites. *Comput Hum Behav* 66:138–148
78. Kramer AD (2010) An unobtrusive behavioral model of gross national happiness. In: *Proceedings of the SIGCHI conference on human factors in computing systems. ACM, New York*, pp 287–290
79. Michikyan M, Subrahmanyam K, Dennis J (2014) Can you tell who I am? Neuroticism, extraversion, and online self-presentation among young adults. *Comput Hum Behav* 33:179–183
80. Hughes DJ, Rowe M, Batey M, Lee A (2012) A tale of two sites: Twitter vs. Facebook and the personality predictors of social media usage. *Comput Hum Behav* 28(2):561–569
81. Wei H, Zhang F, Yuan NJ, Cao C, Fu H, Xie X, Rui Y, Ma W-Y (2017) Beyond the words: predicting user personality from heterogeneous information. In: *Proceedings of the tenth ACM international conference on web search and data mining. ACM, New York*, pp 305–314