

Mining large-scale smartphone data for personality studies

Gokul Chittaranjan · Jan Blom · Daniel Gatica-Perez

Received: 16 September 2011 / Accepted: 14 October 2011 / Published online: 16 December 2011
© Springer-Verlag London Limited 2011

Abstract In this paper, we investigate the relationship between automatically extracted behavioral characteristics derived from rich smartphone data and self-reported Big-Five personality traits (extraversion, agreeableness, conscientiousness, emotional stability and openness to experience). Our data stem from smartphones of 117 Nokia N95 smartphone users, collected over a continuous period of 17 months in Switzerland. From the analysis, we show that several aggregated features obtained from smartphone usage data can be indicators of the Big-Five traits. Next, we describe a machine learning method to detect the personality trait of a user based on smartphone usage. Finally, we study the benefits of using gender-specific models for this task. Apart from a psychological viewpoint, this study facilitates further research on the automated classification

and usage of personality traits for personalizing services on smartphones.

Keywords Smartphones · Big-Five · Personality · Lausanne data collection campaign

1 Introduction

Mobile phones have increasingly become an indispensable part of our daily lives. In light of the rapid growth of mobile phones [15], studying the psychological, social, and economic implications of mobile telephony has gained an increased importance. Smartphones provide a new lens to investigate this phenomenon [26]. Since they are programmable, they enable the development of data collection tools to record various behavioral aspects of the user, ranging from how the device is used across different contexts to analyzing spatial and social dimensions of the everyday life of the user through sources such as GPS, call logs, and Bluetooth.

This data intensive framework provides a wealth of new opportunities as it allows us to understand the impact of context on user behavior as well as to study individual differences such as personality of the users. In turn, it can enable the design of communication features and multiple mobile applications that are tailored to the individual needs and preferences of a user.

On the other hand, personality has been found to influence the behavior of an individual in social interactions. In personality psychology, personality traits play a central role in describing a person [21]. This topic has also been found to be of vital importance in computing. Several recent studies have investigated personality traits and their relationship to the use of Internet and forms of social media

Based on “Who’s who with Big-Five: Analyzing and Classifying Personality Traits with Smartphones” by Gokul Chittaranjan, Jan Blom and Daniel Gatica-Perez which appeared in the Proceedings of the International Symposium on Wearable Computers, San Francisco, California, June 2011. © 2011 IEEE.

G. Chittaranjan · D. Gatica-Perez (✉)
Idiap Research Institute, Centre du Parc, PO Box 592,
Rue Marconi 19, 1920 Martigny, Switzerland
e-mail: gatica@idiap.ch

G. Chittaranjan
e-mail: ghatta@idiap.ch

G. Chittaranjan · D. Gatica-Perez
École Polytechnique Fédérale de Lausanne (EPFL),
1015 Lausanne, Switzerland

J. Blom
Nokia Research Center Lausanne, PSE-C, EPFL,
1015 Lausanne, Switzerland
e-mail: jan.blom@nokia.com

such as Youtube, blogs, Facebook, and other social networks [1, 3, 9, 28, 32].

Since mobile phones also mediate social interactions, phone usage could reflect an individual's personality [5]. However, in contrast to the significant amount of research in the web and social media context, surprisingly, few studies have been carried out in the past to investigate the connection between mobile phone usage and personality of individuals. In particular, the following points have not been adequately addressed: First, there is a clear need for scalability of studies to both a large and diverse feature set and a user base. This has not been possible in the past because of the burden on the user, who is often a customer, in answering lengthy questionnaires. Second, the rich contextual information that can be extracted with current smartphones has not been studied from the perspective of personality. Third, the automatic inference of usage or traits, based on features that can be reliably extracted from continuously collected data, has not been explored.

Determining the personality of mobile phone users, besides being important solely from the psychological point of view, can also provide an interesting framework for mobile computing. The ability to draw connections between personality and behavioral aspects derived through contextual data collected by mobile phones could lead to designing and applying machine learning methods to classify users into personality types. Such understanding could be used in various ways in the context of mobile applications. For instance, prior research has shown that personality is linked to user interface preferences, like the surface color of an application [4]. Certain personality traits, like extraversion/introversion, have also been found to be linked to preferences pertaining to visual esthetics of web sites [16]. The personality of a user might also determine the kind of functions the individual is disposed to use on the phone, for example, of place recommenders that could match the preferences of people with specific traits [14]. Individual differences in personality may also correlate with the impact of context on the user. For instance, when faced with idle time, is an extravert likely to use the device in a different way as compared to an introvert? The preferred interaction modalities may also differ across personality types. Conscientious persons, for example, may be more likely to switch their devices to a silent mode in a socially sensitive situation. Although the examples given above are hypothetical, they nevertheless indicate that expending efforts on establishing a link between personality and behavior can be justified by the wealth of design opportunities such a discovery would enable.

Our previous work on this problem [6], on a smaller dataset of 83 users and a period of 8 months, enabled us to establish that several smartphone usage cues were

predictive of the Big-Five personality traits. We were also able to show that they could be potentially used to predict the Big-Five personality traits.

In this paper, we build upon the previous work, by studying smartphone usage and its relationship to the Big-Five personality model [21]. We also enhance our experimental framework and method to classify users according to self-perceived personality, using features that are by nature privacy sensitive and extracted from anonymous usage logs and phone sensors on the Nokia N95 smartphone. Our experiments are based on subset of the Lausanne Data Collection Campaign [17] and contain data continuously collected from 117 participants for a duration of 17 months.

First, we show that significant relationships exist between personality traits and automatically aggregated smartphone usage cues. Next, we discuss the differences that arise across genders and establish the need to build gender-specific models for personality prediction. Finally, we describe an automated method to address the difficult task of classifying users according to their personality traits.

The paper is organized as follows. Section 2 describes previous work on personality measurement by direct or indirect means. The dataset used along with details about feature extraction is given in Sect. 3. The statistical analysis of the features and personality along with a discussion of differences observed across genders is described in Sect. 4. Subsequently, a machine learning method for the classification of users based on their Big-Five traits is described in Sect. 5. Finally, we conclude in Sect. 6.

2 Related work

The Big-Five personality framework [21] has received considerable support in psychology, although there has not been a universal acceptance of the concept. This framework is a hierarchical model of personality traits that represent personality at the broadest level of abstraction [13]. It consists of five bipolar factors, namely extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience [21]. These factors, described in Table 1, summarize several more specific traits and are believed to capture most of the individual differences in human personality [13].

Given the objectives of this work, it is useful to contrast personality assessment methods into questionnaire and behavior based. The questionnaires used in many Big-Five personality studies are typically lengthy. This can be a limitation when a large number of participants at geographically spread areas have to complete questionnaires online. Therefore, efforts have been made to develop brief

Table 1 The Big-Five traits and examples of adjectives describing them [21]

Trait	Examples of adjectives
Extraversion (E)	Active, assertive, energetic, enthusiastic, outgoing, talkative
Agreeableness (A)	Appreciative, forgiving, generous, kind, sympathetic
Conscientiousness (C)	Efficient, organized, planful, reliable, responsible, thorough
Neuroticism (N)	Anxious, self-pitying, tense, touchy, unstable, worrying
Openness to experience (O)	Artistic, curious, imaginative, insightful, original, wide interests

scales in psychology [13], so as to minimize the time required by the participants to fill in a survey as well as the cost associated with the process of filling in questionnaires. In this context, Gosling et al. introduced the Ten Item Personality Inventory (TIPI) [13] that includes, as the name suggests, ten questions to determine the Big-Five personality traits. It has been shown that the TIPI instrument reaches adequate convergence with the Big-Five measures in self-reported ratings [13]. Hence, in our study, we use TIPI to measure self-perceived personality.

On the other hand, in relation to assessing personality indirectly through behavioral characteristics, Pianesi et al. showed that personality traits in a meeting environment can be detected using audio–visual features and supervised learning [24]. In this case, personality of the participants was revealed by how participants spoke and interacted in the experimental situation. Similarly, Mairesse and Walker describe an automatic procedure using NLP and audio features to detect the Big-Five traits from conversation extracts [19, 20]. While the above examples highlight that behavioral characteristics can be indicative of the personality of an individual, the role of the mobile phone in revealing this behavior remains a relatively unexplored territory. This is surprising given that there is plenty of prior research pertaining to modeling users and their mobile phone usage patterns. To name a few examples, Eagle and Pentland described the concept of eigenbehavior and its usefulness in predicting behavioral patterns and ties in a network of people [11]. Farrahi and Gatica-Perez have illustrated ways of determining routines of users by modeling sensor data pertaining to location collected from mobile phones using topic models [12]. Further, Do and Gatica-Perez [10] recently presented an analysis of application usage in smartphones for the purpose of user retrieval. Similarly, Verkasalo et al. studied the reasons and motivation behind using applications across users and nonusers [31]. These studies tie well with the thriving “app-usage” culture established by smartphone manufacturers—through services like the Apple App Store¹, Nokia Ovi Store², and the Android Market³.

¹ <http://www.apple.com/iphone/apps-for-iphone/>.

² <http://store.ovi.com/>.

³ <http://market.android.com/>.

However, very few studies have directly addressed the relationship between smartphone usage and personality, although personality plays a vital role in social science and psychology.

In the context of assessing the relationships between behavioral characteristics of a mobile phone user and personality, recently, Poschl and Doring presented an analysis relating usage patterns in phones to users clustered on the basis of Big-Five personality traits into two discrete groups. All information in this study was gathered using questionnaires [25]. Similarly, Butt and Phillips presented a study of personality and its relationship to mobile phone usage [5]. The detailed NEO-FFI personality test [8] in conjunction with the Coopersmith self-esteem inventory [7] was administered to participants of the study. Factors describing levels of phone usage were obtained from another questionnaire. The features used in this study were related to phone calls and SMS usage. Many of the comparisons made in the study were motivated by previous work investigating the link between personality traits and Internet usage [5]. In this study, disagreeable individuals tended to be more likely to report receiving more calls and also a higher proportion of calls as “unwanted.” Outgoing calls were not significantly explained by the traits. Extraverted, neurotic, and non-conscientious individuals were reported to have spent more time sending/receiving SMS, and extraverted and disagreeable individuals were found to spend more time changing the ringtone or wallpapers. In a similar work, Phillips et al. also found that disagreeable individuals were more likely to play games on their phone [23]. Further, Lane and Manner have recently studied the effects of smartphone ownership and usage on the Big-Five traits [25]. This study was also questionnaire based. Several participants that were a part of this study did not own a smartphone, and this study had the limitation of being subjected to participants’ reliance on memory and biases. In the context of predicting personality traits using machine learning methods, Oliveira et al. have investigated the possibility of extracting features from phone call logs to predicting the Big-Five personality traits using regression methods [22]. This dataset used in this study comprises 6 months of call records from 39 users in Mexico.

Our study differs from past work in several ways. Firstly, we utilize information available in today's smartphones, such as the usage of apps and proximity information derived from Bluetooth in addition to the traditional call and SMS usage information. All cues are automatically extracted from usage logs, without intervention or input from users. Therefore, we do not rely on personal recall of these usage cues that can be prone to human errors and biases. Secondly, we use a short personality questionnaire that makes the project scalable to a large population. We also devise an automatic classification method, using supervised learning to classify users according to the Big-Five traits.

3 Description of the dataset

In this work, we use smartphone data of 117 participants of the Lausanne data collection campaign [17], a people sensing project organized in the French-speaking region of Switzerland. We use data collected for a continuous period of 17 months (between October 2009 and February 2011) using a continuous, non-intrusive data collection software running on Nokia N95 phones. This software collected anonymized logs of calls (*Call Logs*), SMS (*SMS Logs*), Bluetooth scans (*BT Logs*), calling profiles (*Profile Logs*), and application usage (*App Logs*).

As a part of the exit survey in the campaign, participants were administered an online questionnaire in English and French, based on their language of preference, requesting information about their demographics, gender, age, and personality. In our dataset, 61 and 56 participants chose to answer in English and French, respectively. From these questionnaires, we found that of the 117 participants, 73 were men and 39 were women, 5 participants chose not to disclose their gender. The mean age was 30.2 years with a standard deviation of 7.3 years. The minimum and maximum ages were 19 and 63 years, respectively; 84 of the 117 participants had at least a university degree. The dataset contained 45 Asians, 4 North Americans, 65 Europeans, one South American, and a user marked "other" indicating that he/she did not belong to any of the above places. All users were previous mobile phone users, but most of them had not owned a smartphone before the study. Therefore, they discovered most of the features of the N95 during the data collection process.

Self-perceived personality was measured using the TIPI questionnaire [13] (given in Table 2). The questionnaire comprises two questions per dimension (one of which is negatively scored) of the Big-Five personality. For example, questions 1 and 6 correspond to extraversion. Question 6 needs to be reverse scored, since it refers to introversion. Therefore, the value for extraversion, for a given user, is

Table 2 The TIPI questionnaire

I see myself as ...	
1. Extraverted, enthusiastic	(1)–(7)
2. Critical, quarrelsome	(1)–(7)
3. Dependable, self-disciplined	(1)–(7)
4. Anxious, easily upset	(1)–(7)
5. Open to new experiences, complex	(1)–(7)
6. Reserved, quiet	(1)–(7)
7. Sympathetic, warm	(1)–(7)
8. Disorganized, careless	(1)–(7)
9. Calm, emotionally stable	(1)–(7)
10. Conventional, uncreative	(1)–(7)

Each user assigned a value between 1 (agree strongly) and 7 (disagree strongly) for the questions given below, based on how he/she perceived himself/herself

computed as the average of question 1 and question 6 (reversed). Please note that for brevity, in the discussions to follow, we refer "Openness to experience" as "Openness."

3.1 Extraction of features

Continuously collected data from a software running on the phones of participants were uploaded every night to a server. The data were made available after anonymization, thereby making the features used in this study, by nature, privacy sensitive. Details of the data collection process are detailed in previous work [17].

The features were extracted from five modalities, which are enlisted in Table 3. Those features extracted from communication and application logs (*Call Logs*, *SMS Logs*, *App Logs*) were based on all events recorded when the data collection software was running on the phones. Therefore, these features were relatively "clean" and captured various aspects of communication and applications usage on the phone.

Further, features pertaining to calling profiles were obtained by first extracting *events* that represented segments of time for which a calling profile was active by parsing the *Profile logs*. These logs recorded the phone profile state approximately every minute. There were 5 different calling profiles that were identified (Normal, Silent, Beep, Ascending, and Ring Once).

On the other hand, features from *Bluetooth* logs were based on scans done approximately every 3 minutes. Defining a time slot as one Bluetooth scan, *events* that captured the duration for which an ID was observed were computed. These *events* were meant to capture a rough description of the social context (such as crowded or solitary environments). Since the number of Bluetooth devices in the vicinity of a user is only a noisy proxy for the crowdedness of a place, features extracted from this

Table 3 Table of features aggregated from the Nokia N95 smartphone

Modality	Feature name	
<i>SMS Logs</i>	Avg. SMS length (Inbox)	
	Avg. word length (Inbox)	
	No. words of length >6 (Inbox)	
	Messages with unique ID (Inbox)	
	Avg. SMS length (Sent)	
	Avg. word length (Sent)	
	No. words of length >6 (Sent)	
	Messages with unique ID (Sent)	
	<i>Call Logs</i>	Outgoing (O) Calls
		Avg. duration (O Calls)
Total duration (O Calls)		
Incoming (I) Calls		
Avg. duration (I Calls)		
Total duration (I Calls)		
Unique contacts (O Calls)		
Unique contacts (I Calls)		
I/O Calls		
Avg. duration (I + O Calls)		
Total duration (I + O Calls)		
Unique contacts in call logs		
Missed (M) Calls		
Unique contacts (M)		
O to I ratio [†]		
M to (I + O) ratio [†]		
SMS received		
Unique contacts (SMS received)		
SMS sent		
Unique contacts (SMS sent to)		
SMS Incoming/Outgoing Ratio [†]		
<i>App Logs</i>	Office	
	Internet	
	Video/Audio/Music	
	Maps	
	Mail	
	Youtube	
	Calendar	
	Camera	
	Chat	
	SMS	
	Games	
<i>BT Logs</i>	Unique BT IDs	
	Common BT ID seen count	
	BT IDs in 50% of IDs seen	
	BT IDs seen more than 4 slots	
	BT IDs seen more than 9 slots	
	BT IDs seen more than 19 slots	
	Max time a BT ID seen	
BT IDs seen for ≥ 5 slots		

Table 3 continued

Modality	Feature name
<i>Profile Logs</i>	Probability {Normal (N), Silent (S) Beep (B), Ascending (A), Ring Once (RO)}
	Longest segments (N, S, B, A, RO) [†]
	Shortest segment (N, S, B, A, RO) [†]
	Avg. segment length (N, S, B, A, RO) [†]
	No. Segments (N, S, B, A, RO) [†]
	Dominant profile count (N, S, B, A, RO)
	Avg. changes in profile (daily)
	SD of no. changes in profile (daily)

[†] The subset of features not used in the statistical analysis

modality inherently contain a certain degree of uncertainty as compared to the features derived from communication or application logs.

In summary, the selection of these features was based on previous work enlisted in Sect. 2 and on the choice of features that could reasonably characterize levels of individual and relational activity.

3.2 Aggregation of features

Since our dataset contains longitudinal smartphone data, we had to aggregate the features at a timescale that would be long enough to capture the usage of a smartphone feature, while giving enough data points to conduct statistical analysis and train our machine learning model. Therefore, the features used in our studies are aggregated from the logs on a monthly level. In other words, all users were split across months, which gave us 1121 user-months. From each of the user-months, features describing different aspects of smartphone usage were computed automatically by parsing the logs, as summarized in Table 3. All features except those from *BT Logs* and *Profile Logs* were obtained by aggregating events (such as the opening of an Office or Internet application) as and when they happened. Features pertaining to Bluetooth and calling profiles were based on the duration of the Bluetooth and calling profile events, respectively. The aggregated features from *BT Logs* captured the number of times and the duration for which BT IDs were seen. In the case of *Profile Logs*, from the events that represent time segments, the probability of observing a segment in a month, its duration, and the most dominant profile type in a day was computed. The number of changes in the calling profile for each day and its standard deviation were also calculated as features.

Since this leads to a very large number of features, in the discussions in Sect. 4, we consider a subset of these features. However, for our classification task described in

Table 4 Descriptive statistics for the independent variables (Big-Five traits)

Predictors	μ	σ	Median	Min	Max	Skew
Entire population						
Extraversion	4.24	1.29	4.0	1	7	-0.25
Agreeableness	4.85	1.34	5.0	2	7	-0.46
Conscientiousness	5.01	1.60	5.5	1	7	-0.80
Emotional stability	4.53	1.36	4.5	1	7	-0.46
Openness to experience	4.71	1.48	5.0	1	7	-0.46
Female population						
Extraversion	4.10	1.41	4.0	1	6.5	-0.45
Agreeableness	5.45	1.30	6.0	2	7	-1.23
Conscientiousness	5.62	1.51	6.0	1	7	-1.44
Emotional stability	4.88	1.40	5.0	2	7	-0.26
Openness to experience	4.72	1.55	4.5	2	7	-0.36
Male population						
Extraversion	4.29	1.22	4.5	1.5	7	-0.09
Agreeableness	4.55	1.29	4.5	2	7	-0.27
Conscientiousness	4.77	1.57	5.0	1	7	-0.75
Emotional stability	4.37	1.34	4.5	1	6.5	-0.71
Openness to experience	4.70	1.46	5.0	1	7	-0.57

Sect. 5, we consider the entire feature set in the feature selection step.

4 Statistical analysis

In this section, we use commonly used statistical analysis techniques to understand the relationship between smartphone usage and the Big-Five traits. We begin by analyzing the Big-Five trait dimensions by examining its descriptive statistics and intra-trait correlations. Next, we give an overview of the statistical techniques (correlation and multiple regression analysis) that will be used in this paper. Subsequently, we describe our observations through these techniques. Finally, we discuss the observed results in light of existing literature in psychology.

4.1 Analysis of independent variables

The descriptive statistics for the TIPI questionnaire data for entire population and different subsets of it is given in Table 4. The table shows that a higher skew is observed for the agreeableness and conscientiousness traits in women, in addition to a higher mean. A higher mean has been observed in the established norm as well [30].

In order to address the high negative skewness for the agreeableness and conscientiousness traits in the female population, they were inverted and log transformed for the statistical analyses. However, when presenting the results and discussion, numbers are negated to show the effects for a same measure, across all populations (i.e., agreeableness

Table 5 Correlations between independent variables (the Big-Five traits) for the entire population ($N_{users} = 117$, $N_{months} = 1,121$)

	A	C	ES	O
Extraversion (E)	0.04	0.08	-0.19*	0.20*
Agreeableness (A)		0.65**	0.64**	0.45**
Conscientiousness (C)			0.60**	0.43**
Emot. stability (ES)				0.41**
Openness to exp. (O)				

* $p < 0.05$, ** $p < 0.01$

and conscientiousness instead of disagreeableness and non-conscientiousness).

As a next step, we present the inter-trait correlations in Table 5. Several significant correlations exist among the traits. Agreeableness, emotional stability, and conscientiousness are strongly positively correlated. Similar correlations were also seen in our previous study with a smaller dataset [6]. However, all the correlations seen in Table 5 are below the selection criteria used in the test for multicollinearity in previous work [29].

4.2 Overview of the analysis of dependent variables

All smartphone features barring two (which were not skewed) were strongly positively skewed. Therefore, a log transformation was applied to the feature space prior to conducting statistical analysis. Further, features derived from the *App Logs* were sparse due to the low frequency of usage of some of the applications. Therefore, for analysis involving this source, we chose only those user-months for

which there had been some use of the application. Finally, for all features, only those user-months were chosen for which there were at least 7 days of usage. This was done to avoid user-months that might contain little or no data due to various reasons such as vacations and problems with phone usage.

In psychology literature, Pearson’s correlation coefficient is commonly used as a bounded measure of correlation, or linear dependence between two variables. For two random variables X and Y , it is given by:

$$r = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

where $\text{cov}(X, Y)$ is the covariance between the random variables X and Y and σ_X denotes the variance of a random variable X . $r = 1$ denotes a positive sloped linear relationship, and $r = -1$ denotes a negative sloped linear relationship. Values in-between indicate sublinear relationships between the variables.

In our work, we compute the Pearson’s correlation coefficient between the Big-Five traits and the smartphone features. The correlation analysis results are given in Tables 6, 7 and 8 for those features that showed a correlation significant to a level of $p < 0.01$ for the entire population, the male or female populations.

Additionally, regression analysis is also a commonly used tool to study the relationships between dependent and independent variables.

In linear regression, the dependent variable (Y) is expressed as a linear combination of the independent variables (\mathbf{X}) in the form given below:

$$Y = b_0 + \sum_{i=1}^{N=5} b_i x_i$$

where N is the number of independent variables and $\mathbf{X} = \{x_1 \dots x_N\}$. Here $\mathbf{B} = \{b_0 \dots b_5\}$ denote the regression coefficients. If the variables used for performing regression analysis are normalized, then the regression coefficients thus obtained are called standardized regression coefficients (β). The use of these “standardized” coefficients ignores the independent variables’ scale of units and therefore makes results comparable.

In order to determine the goodness of fit of the regression model, to the given data, the coefficient of determination R^2 is normally used. It indicates the proportion of variability in the feature that has been accounted for by the regression model. An F test is then used to determine the statistical significance of the overall fit, followed by a t test of the individual β -coefficients.

In our case, we are interested in the relationships between features (dependent variables), as a function of the Big-Five traits (independent variables). Therefore, we also

conducted multiple regression analysis with the features as the dependent variables and the Big-Five traits as the independent variables. Subsequently, we considered the R^2 , F values and β values (in cases where the t test indicates a significant coefficient).

The goodness of fit (R^2) and its significance are summarized in Tables 9 and 10.

In the sections to follow, we first make observations across the entire population. Next, we bring out the differences observed across genders. Finally, we interpret the observations based on previous work.

4.3 Observations for the entire population

In the sections to follow, we first describe the observations from the correlation analysis structuring the discussion around each of the Big-Five traits, followed by the multiple regression analysis, where the discussion is structured around data types.

4.3.1 From correlation analysis

4.3.1.1 Extraversion (Table 6) Extraversion was found to be positively correlated to the use of the Office and Calendar apps. However, significant negative correlations were seen for the use of the Internet, Games, and Camera. Extraverts were more likely to spend more time on incoming calls and also receive more calls. The total duration of calls and the number of unique contacts associated with voice calls were likely to be higher as well. A slight positive correlation was also seen for the number of SMS messages received and extraversion. Interestingly, other SMS features did not significantly correlate with this trait. Lastly, it was found that extraverts had a higher probability of setting the phone on the *Ring Once* mode and were less likely to use *Silent* as the most dominant profile.

4.3.1.2 Agreeableness (Table 6) Agreeableness was found to be negatively correlated to the use of several applications, including Office, Internet, Video/Audio/Music, Mail, Calendar, and SMS apps. The SMS length in the sent folders was more likely to be longer for agreeable users. From the *Call Logs*, no significant correlations were seen between the duration and number of voice calls and agreeableness. Further, it was seen that the number of BT IDs seen for long duration of time was likely to be higher for disagreeable users. Finally, this trait was also found to be correlated positively to the use of the *Normal* profile and negatively to the use of all other calling profiles.

4.3.1.3 Conscientiousness (Table 7) This trait was found to be negatively correlated to the use of Video/Audio/Music and Youtube applications. They also were more likely to

Table 6 Features exhibiting correlation with $p < 0.01$ (in bold) across different populations (A All, M Male, F Female) for the Big-Five traits

Features	Correlation		
	A	F	M
<i>Extraversion</i>			
Office	0.12	0.09	0.18
Internet	-0.13	-0.40	0.01
Video/Audio/Music	-0.03	-0.26	0.03
Maps	-0.00	-0.31	0.03
Mail	0.09	-0.06	0.21
Calendar	0.09	0.01	0.14
Camera	-0.11	-0.29	-0.05
Games	-0.43	0.13	-0.49
Avg. SMS length (Inbox)	0.03	-0.00	0.11
Avg. SMS length (Sent)	-0.05	0.06	-0.10
Incoming (I) Calls	0.14	0.16	0.11
Avg. duration (I Calls)	0.20	0.29	0.11
Total duration (I Calls)	0.21	0.29	0.13
Unique contacts (O Calls)	0.14	0.10	0.11
Unique contacts (I Calls)	0.13	0.09	0.11
Avg. duration (I + O Calls)	0.09	0.03	0.13
Total duration (I + O Calls)	0.09	0.07	0.12
Unique contacts in call logs	0.15	0.09	0.13
Unique contacts (M)	0.07	0.06	0.02
SMS received	0.09	0.07	0.04
BT IDs seen more than 4 slots	-0.02	-0.19	0.08
BT IDs seen more than 9 slots	0.00	-0.22	0.12
BT IDs seen more than 19 slots	0.01	-0.21	0.13
BT IDs seen for ≥ 5 slots	0.01	-0.21	0.13
Probability silent (S)	-0.01	-0.19	0.06
Probability Ascending (A)	-0.06	0.12	-0.13
Probability Ring Once (RO)	0.19	-0.03	0.26
Dominant profile count (N)	0.08	0.17	0.08
Dominant profile count (S)	-0.11	-0.32	-0.03
Dominant profile count (B)	0.07	0.11	0.05
Dominant profile count (A)	-0.04	0.10	-0.15
Dominant profile count (RO)	0.06	-0.01	0.16
SD of no. changes in profile (daily)	-0.01	-0.14	-0.00
<i>Agreeableness</i>			
Office	-0.14	-0.03	-0.20
Internet	-0.11	-0.01	-0.17
Video/Audio/Music	-0.08	0.08	-0.07
Mail	-0.18	-0.45	-0.23
Youtube	0.24	-0.00	0.37
Calendar	-0.15	-0.17	-0.15
Chat	-0.14	-1.00	-0.07
SMS	-0.11	0.18	-0.32
Avg. SMS length (Sent)	0.06	0.13	-0.09
Avg. word length (Sent)	0.08	0.24	-0.10

Table 6 continued

Features	Correlation		
	A	F	M
No. words of length >6 (Sent)	0.10	0.27	-0.11
Messages with unique ID (Sent)	0.07	0.26	-0.07
Outgoing (O) Calls	0.03	0.10	0.15
Avg. duration (O Calls)	0.07	0.29	0.07
Total duration (O Calls)	0.06	0.21	0.14
Incoming (I) Calls	0.03	0.14	0.14
Unique contacts (O Calls)	0.03	0.08	0.19
Unique contacts (I Calls)	-0.01	0.04	0.13
I/O Calls	0.03	0.12	0.16
Avg. duration (I + O Calls)	0.03	0.30	0.01
Total duration (I + O Calls)	0.04	0.22	0.11
Unique contacts in call logs	0.01	0.07	0.17
Missed (M) Calls	0.04	0.10	0.16
Unique contacts (M)	0.07	0.05	0.19
SMS received	-0.07	0.09	-0.15
SMS sent	-0.01	0.21	-0.17
Unique contacts (SMS sent to)	-0.13	-0.09	-0.15
Common BT ID seen count	-0.06	0.01	-0.15
BT IDs in 50% of IDs seen	0.06	-0.07	0.14
BT IDs seen more than 4 slots	-0.12	-0.08	-0.11
BT IDs seen more than 9 slots	-0.13	-0.07	-0.13
BT IDs seen more than 19 slots	-0.12	-0.05	-0.11
Max time a BT ID seen	-0.08	-0.06	-0.14
BT IDs seen for ≥ 5 slots	-0.11	-0.08	-0.08
Probability Normal (N)	0.17	0.11	0.11
Probability silent (S)	-0.07	-0.07	-0.03
Probability Ascending (A)	-0.10	-0.06	-0.06
Probability Ring Once (RO)	-0.14	0.04	-0.15
Dominant profile count (N)	-0.01	-0.15	0.11
Dominant profile count (S)	-0.04	0.13	-0.18

Contd. in Table 7 ...

spend lesser time on incoming calls and have lesser number of missed calls and lesser number of unique contacts associated with their missed calls. When seen across the entire population, features pertaining to SMS did not significantly correlate with conscientiousness. It was also seen that the number of BT IDs seen for long durations of time was also lesser for conscientious users. Correlations seen for the profile usage were similar to those seen for agreeableness with conscientious users more likely to use the *Normal* calling profile and less likely to use the other ones.

4.3.1.4 Emotional stability (Table 7) This trait was found to be negatively correlated to the use of Office and Calendar apps. The word length of SMS in both inbox and sent items was more likely to be higher for Emotionally stable users. In the *Call Logs*, it was seen that the duration

Table 7 (Contd. from Table 6.) Features exhibiting correlation with $p < 0.01$ (in bold) across different populations (A: All, M: Male, F: Female) for the Big-Five traits

Features	Correlation		
	A	F	M
<i>Conscientiousness</i>			
Internet	-0.07	0.06	-0.14
Video/Audio/Music	-0.12	0.13	-0.17
Youtube	-0.44	-0.00	-0.33
SMS	-0.01	0.27	-0.18
Avg. SMS length (Inbox)	-0.00	0.10	-0.10
Avg. word length (Inbox)	-0.06	0.15	-0.14
No. words of length >6 (Inbox)	-0.04	0.19	-0.17
Messages with unique ID (Inbox)	-0.06	0.14	-0.12
Avg. SMS length (Sent)	0.08	0.18	-0.03
Avg. word length (Sent)	0.07	0.23	-0.05
No. words of length >6 (Sent)	0.07	0.28	-0.10
Messages with unique ID (Sent)	0.06	0.23	-0.03
Outgoing (O) Calls	0.04	0.18	0.10
Avg. duration (O Calls)	0.05	0.31	0.02
Total duration (O Calls)	0.05	0.27	0.08
Incoming (I) Calls	-0.04	0.13	-0.05
Avg. duration (I Calls)	-0.09	-0.03	-0.13
Total duration (I Calls)	-0.08	0.08	-0.11
Unique contacts (I Calls)	-0.09	-0.04	-0.03
I/O Calls	0.01	0.18	0.06
Avg. duration (I + O Calls)	-0.01	0.28	-0.06
Total duration (I + O Calls)	0.00	0.26	0.01
Missed (M) Calls	0.01	0.18	0.05
Unique contacts (M)	-0.08	-0.02	-0.05
SMS received	-0.02	0.17	-0.12
SMS sent	0.05	0.21	-0.08
Unique contacts (SMS sent to)	-0.13	-0.14	-0.14
BT IDs in 50% of IDs seen	-0.01	-0.17	0.08
BT IDs seen more than 4 slots	-0.08	-0.06	-0.04
BT IDs seen more than 9 slots	-0.08	-0.06	-0.03
Probability normal (N)	0.22	0.13	0.20
Probability silent (S)	-0.08	-0.09	-0.02
Probability Ascending (A)	-0.12	-0.15	-0.09
Probability Ring Once (RO)	-0.25	0.03	-0.29
Dominant profile count (N)	0.06	0.00	0.16
Dominant profile count (S)	-0.07	-0.01	-0.17
Dominant profile count (B)	0.09	0.12	0.06
Dominant profile count (A)	-0.12	-0.17	-0.13
Dominant profile count (RO)	-0.03	0.03	-0.12
SD of no. changes in profile (daily)	-0.06	0.03	-0.11
<i>Emotional stability</i>			
Office	-0.18	-0.07	-0.23
Video/Audio/Music	-0.05	0.23	-0.15
Calendar	-0.13	-0.20	-0.10

Table 7 continued

Features	Correlation		
	A	F	M
SMS	-0.05	0.16	-0.18
Games	0.37	-0.08	0.48
Avg. SMS length (Inbox)	-0.04	0.03	-0.14
Avg. word length (Inbox)	0.09	0.18	0.04
No. words of length >6 (Inbox)	0.08	0.22	-0.02
Messages with unique ID (Inbox)	0.10	0.18	0.09
Avg. word length (Sent)	0.13	0.23	0.02
No. words of length >6 (Sent)	0.12	0.25	-0.02
Messages with unique ID (Sent)	0.12	0.23	0.03
Outgoing (O) Calls	0.02	-0.11	0.14
Avg. duration (O Calls)	0.07	0.20	0.03
Total duration (O Calls)	0.05	0.02	0.11
Avg. duration (I Calls)	-0.09	-0.01	-0.11
Unique contacts (O Calls)	-0.06	-0.16	0.07
Unique contacts (I Calls)	-0.07	-0.09	0.00
I/O Calls	0.01	-0.06	0.12
Avg. duration (I + O Calls)	0.02	0.25	-0.05
Unique contacts in call logs	-0.07	-0.16	0.04
Missed (M) Calls	0.02	-0.07	0.13
Unique contacts (M)	-0.05	-0.15	0.03
SMS sent	0.05	0.20	-0.06
Unique contacts (SMS sent to)	-0.13	-0.11	-0.16
Common BT ID seen count	-0.09	-0.07	-0.11
BT IDs in 50% of IDs seen	0.08	-0.06	0.16
BT IDs seen more than 4 slots	-0.09	-0.12	-0.05
BT IDs seen more than 9 slots	-0.08	-0.08	-0.06
Max time a BT ID seen	-0.13	-0.13	-0.15
Probability silent (S)	0.08	0.03	0.15
Probability Ascending (A)	-0.11	0.00	-0.11
Probability Ring Once (RO)	-0.15	0.06	-0.17
Dominant profile count (N)	0.00	-0.18	0.11
Dominant profile count (S)	-0.04	0.20	-0.20
Avg. changes in profile (daily)	-0.01	0.18	-0.08
SD of no. changes in profile (daily)	-0.02	0.18	-0.10

Contd. in Table 8 ...

of incoming calls showed a small, but significant negative correlation to this trait. Further, it was seen that the number of BT IDs that were seen for long durations of time was likely to be higher for higher neuroticism. Finally, it was found that those scoring higher on this trait also were more likely to use the *Silent* profile, and less likely to use the *Ascending* and *Ring Once* profiles.

4.3.1.5 *Openness to experience* (Table 8) In the *App Logs*, the use of Office, Calendar, and SMS applications was found to be negatively correlated to openness. The

Table 8 (Contd. from Table 7.) Features exhibiting correlation with $p < 0.01$ (in bold) across different populations (A: All, M: Male, F: Female) for the Big-Five traits

Feature	Correlation		
	A	F	M
<i>Openness to Experience</i>			
Office	-0.19	0.14	-0.34
Internet	-0.06	-0.04	-0.11
Video/Audio/Music	0.05	0.25	-0.10
Mail	-0.03	0.32	-0.15
Calendar	-0.11	0.08	-0.18
SMS	-0.19	-0.07	-0.32
Avg. SMS length (Inbox)	-0.13	-0.23	-0.02
Avg. word length (Inbox)	-0.10	-0.05	-0.18
No. words of length >6 (Inbox)	-0.09	-0.07	-0.14
Messages with unique ID (Inbox)	-0.08	-0.01	-0.18
Avg. word length (Sent)	-0.09	-0.03	-0.14
No. words of length >6 (Sent)	-0.08	-0.01	-0.14
Messages with unique ID (Sent)	-0.09	0.01	-0.18
Outgoing (O) Calls	0.01	0.13	-0.05
Unique contacts (I Calls)	0.08	0.08	0.06
I/O Calls	0.02	0.15	-0.04
Unique contacts in call logs	0.07	0.03	0.08
Missed (M) Calls	0.01	0.15	-0.06
SMS received	-0.09	-0.06	-0.18
SMS sent	-0.11	-0.05	-0.20
Unique contacts (SMS sent to)	-0.09	-0.08	-0.12
Common BT ID seen count	0.03	0.19	-0.04
BT IDs in 50% of IDs seen	-0.01	-0.20	0.09
Max time a BT ID seen	-0.02	0.14	-0.09
Probability normal (N)	-0.03	-0.22	0.04
Probability silent (S)	-0.05	0.12	-0.14
Probability beep (B)	0.14	0.14	0.15
Probability Ascending (A)	0.10	0.14	0.10
Probability Ring Once (RO)	-0.09	0.11	-0.12
Dominant profile count (N)	0.03	-0.10	0.15
Dominant profile count (S)	-0.13	0.04	-0.28
Dominant profile count (B)	0.15	0.08	0.20

length of messages in both the inbox and sent folders was also found to be negatively correlated to openness. Few features exhibited significant correlations in the *Call Logs*. The number of unique contacts found in *Call Logs* was more likely to be higher for users scoring higher in this trait. Further, it was seen that the number of SMS sent or received was also negatively correlated to openness. Lastly, users scoring high on openness were more likely to use the *Beep* and *Ascending* calling profiles and less likely to use the *Ring Once* profile.

4.3.2 From regression analysis

4.3.2.1 App logs (Table 9) Several applications were found to significantly explain variance in traits. Upon examining the regression coefficients, it was found that the Office app was more likely to be used by conscientious participants ($\beta = 0.20$, $t = 3.23$) who score low on openness ($\beta = -0.18$, $t = -3.93$) and explained up to 7% of the variance in the traits. The Internet was found to be more likely to be used by Introverts ($\beta = -0.12$, $t = -3.63$) and disagreeable ($\beta = -0.17$, $t = -3.90$) users, while it explained only 3.6% of the variance. This is also shown by the significant negative pairwise correlation of -0.13 and -0.11 for extraversion and agreeableness, respectively. Further, the Mail app was also found to be more likely to be used by disagreeable ($\beta = -0.35$, $t = -4.69$) and conscientious ($\beta = 0.28$, $t = 3.64$) users and accounted for 8% of the variation in the traits. The Video/Audio/Music apps were more likely to be used by users who score higher on openness ($\beta = 0.14$, $t = 4.50$) and low on conscientiousness ($\beta = -0.16$, $t = -4.12$). Youtube on the other hand was found to be more likely to be used by Extraverts ($\beta = 0.49$, $t = 3.79$) and non-conscientious ($\beta = -0.64$, $t = -5.32$) participants. For the use of the Calendar app, regression showed that disagreeable participants were more likely to use it. This is reinforced by the significant negative correlation of -0.15 seen in Table 6. The SMS app was found to be more likely to be used by disagreeable individuals ($\beta = -0.15$, $t = -3.65$) who are conscientious ($\beta = 0.14$, $t = 3.75$) and less open ($\beta = -0.22$, $t = -6.9$). Finally, the Big-Five traits did not significantly explain the use of Camera and Chat apps. Similar results were observed in our previous study with a smaller dataset [6].

4.3.2.2 SMS logs (Table 9) It was found that the length of the inbox messages, measured in words, was more likely to be higher for unconscientious ($\beta = -0.16$, $t = -3.86$), emotionally stable ($\beta = 0.29$, $t = 6.57$) participants scoring low on openness ($\beta = -0.15$, $t = -4.54$). Emotionally stable participants with low openness were also more likely to send longer messages and have more messages in their sent items folder. This is also shown in our correlation analysis as significant positive correlations in Table 7, thus supporting this result. This indicates that emotionally stable users with low openness are likely to send longer SMS and receive more (and longer) responses.

4.3.2.3 Call logs (Table 9) The features corresponding to outgoing calls did not significantly explain the variation in traits. This is in concordance with our previous study [6]. It was found that the number of incoming calls and their average and total duration were more likely to be higher for

Table 9 Regression analysis across different populations (A All, M Male, F Female) for the Big-Five traits

Features	A		F		M	
	F	R ²	F	R ²	F	R ²
Office	9.1	0.07	2.9	0.07	20.0	0.21
Internet	8.3	0.04	18.3	0.22	6.9	0.05
Video/Audio/Music	8.6	0.03	10.7	0.11	6.5	0.04
Maps	2.4	0.02	3.9	0.17	0.5	0.01
Mail	6.1	0.08	8.2	0.40	10.3	0.17
Youtube	9.4	0.48	0.0	0.00	6.5	0.41
Calendar	6.7	0.04	5.2	0.09	8.8	0.07
Camera	2.4	0.02	5.2	0.12	0.5	0.01
Chat	0.8	0.13	0.0	0.00	1.1	0.20
SMS	15.1	0.05	8.0	0.09	24.7	0.13
Games	7.3	0.52	0.0	1.00	6.7	0.56
Avg. SMS length (Inbox)	5.8	0.02	7.9	0.09	5.0	0.03
Avg. word length (Inbox)	12.6	0.05	4.7	0.05	14.2	0.09
No. words of length >6 (Inbox)	9.1	0.04	7.0	0.08	9.9	0.06
Messages with unique ID (Inbox)	13.4	0.05	3.5	0.04	17.4	0.11
Avg. SMS length (Sent)	3.8	0.02	4.4	0.05	4.5	0.03
Avg. word length (Sent)	10.2	0.04	7.8	0.09	6.1	0.04
No. words of length >6 (Sent)	8.6	0.03	10.1	0.11	4.7	0.03
Messages with unique ID (Sent)	9.8	0.04	6.9	0.08	8.3	0.05
Outgoing (O) calls	0.7	0.00	8.1	0.09	13.1	0.07
Avg. duration (O Calls)	2.5	0.01	9.6	0.10	1.9	0.01
Total duration (O Calls)	1.9	0.01	7.0	0.08	8.8	0.05
Incoming (I) calls	8.3	0.03	6.2	0.07	14.3	0.08
Avg. duration (I Calls)	14.5	0.05	11.9	0.12	8.1	0.05
Total duration (I Calls)	16.2	0.06	12.4	0.13	14.1	0.08
Unique contacts (O Calls)	11.3	0.04	9.8	0.10	11.5	0.06
Unique contacts (I Calls)	12.8	0.04	5.5	0.06	9.0	0.05
I/O Calls	1.4	0.01	6.7	0.07	14.3	0.08
Avg. duration (I + O Calls)	4.7	0.02	11.8	0.12	5.3	0.03
Total duration (I + O Calls)	3.8	0.01	6.4	0.07	12.9	0.07
Unique contacts in call logs	12.2	0.04	9.2	0.10	10.5	0.06
Missed (M) calls	1.3	0.00	6.4	0.07	16.4	0.09
Unique contacts (M)	12.3	0.04	7.8	0.08	18.4	0.10
SMS received	10.7	0.04	5.9	0.06	10.8	0.06
Unique contacts (SMS received)	1.9	0.01	2.0	0.02	1.6	0.01
SMS sent	11.7	0.04	8.7	0.09	11.1	0.06
Unique contacts (SMS sent to)	6.3	0.02	2.3	0.03	5.2	0.03

Values indicated in bold are significant to a degree of $p < 0.01$. Refer Sect. 4 for details *Contd. in Table 10 ...*

extraverts and non-conscientious users. For example, for total duration of calls, β was 0.22 and -0.16 and t was 7.857 and -4.15 , respectively. Similar values were seen for the other features. Additionally, it was found that the number of unique contacts associated with outgoing calls was also more likely to be higher for extraverted ($\beta = 0.13, t = 4.48$), agreeable ($\beta = 0.14, t = 3.52$), and non-conscientious users ($\beta = -0.16, t = -4.31$). The number of incoming calls associated with unique contacts in a user’s address book was more likely to be higher for

users scoring high on extraversion ($\beta = 0.12, t = 4.137$) and less on conscientiousness ($\beta = -0.18, t = -4.73$) and openness ($\beta = 0.13, t = 4.04$). The total duration of calls was also found to more likely to be higher for extraverts. The number of missed calls did not significantly describe the variation in the traits. Further, we found that users scoring high on extraversion ($\beta = 0.15, t = 5.10$) and emotional stability ($\beta = 0.19, t = 4.65$) and low on agreeableness ($\beta = -0.14, t = -3.53$) and openness ($\beta = -0.13, t = -4.24$) were more likely to receive SMS.

Table 10 (Contd. from Table 9. Regression analysis across different populations (A: All, M: Male, F: Female) for the Big-Five traits

Features	A		F		M	
	F	R ²	F	R ²	F	R ²
Unique BT IDs	3.0	0.01	7.9	0.08	1.9	0.01
Common BT ID seen count	14.1	0.05	11.3	0.12	8.2	0.05
BT IDs in 50% of IDs seen	7.1	0.03	6.7	0.07	5.5	0.03
BT IDs seen more than 4 slots	4.5	0.02	8.7	0.09	4.4	0.03
BT IDs seen more than 9 slots	5.8	0.02	8.5	0.09	6.8	0.04
BT IDs seen more than 19 slots	5.5	0.02	8.3	0.09	6.9	0.04
Max time a BT ID seen	10.8	0.04	7.7	0.08	8.6	0.05
BT IDs seen for ≥ 5 slots	6.0	0.02	13.9	0.14	7.2	0.04
Probability normal (N)	29.0	0.10	8.9	0.09	19.0	0.10
Probability silent (S)	15.4	0.05	6.5	0.07	24.0	0.12
Probability beep (B)	8.8	0.03	2.3	0.03	14.2	0.08
Probability Ascending (A)	18.3	0.06	9.5	0.10	18.0	0.10
Probability Ring Once (RO)	32.9	0.11	1.5	0.02	37.1	0.18
Dominant profile count (N)	3.7	0.01	6.3	0.07	5.8	0.03
Dominant profile count (S)	7.8	0.03	11.7	0.12	16.1	0.09
Dominant profile count (B)	7.8	0.03	2.6	0.03	12.1	0.07
Dominant profile count (A)	14.4	0.05	14.7	0.14	12.0	0.07
Dominant profile count (RO)	3.7	0.01	1.1	0.01	10.6	0.06
Avg. changes in profile (daily)	0.7	0.00	3.8	0.04	1.9	0.01
SD of no. changes in profile (daily)	1.2	0.00	3.6	0.04	2.4	0.01

Values indicated in bold are significant to a degree of $p < 0.01$. Refer Sect. 4 for details

This concurs with the results observed with the *SMS Logs* and with the pairwise correlations. Also, the number of SMS messages sent had a chance of being higher for extraverted ($\beta = 0.11$, $t = 3.69$), emotionally stable ($\beta = 0.17$, $t = 4.09$) users scoring low on openness to experience ($\beta = -0.19$, $t = -6.03$), which tallies with our previous results seen in the *SMS logs*.

4.3.2.4 BT logs (Table 10) None of the features from the BT Logs explained a large variation in the traits. It was generally seen that agreeable individuals had lesser number of BT IDs seen for long durations of time. Interestingly, the β values for emotional stability for his feature were not found to significantly contribute to the regression function, in contrast to our previous study on a smaller dataset [6]. Further, it was seen that extraverts ($\beta = 0.11$, $t = 3.81$) who are non-conscientious ($\beta = -0.14$, $t = -3.60$), emotionally stable ($\beta = 0.18$, $t = 4.21$) were more likely to have more BT IDs to account for 50% of the total BT IDs seen. On the other hand, introverted ($\beta = -0.15$, $t = -5.19$), conscientious ($\beta = 0.14$, $t = 3.54$), neurotic ($\beta = -0.25$, $t = -6.06$) users were likely to spend longer time around the most commonly seen BT ID.

4.3.2.5 Profile logs (Table 10) For features from this source, we organize our discussion into different calling profiles. The probability of the phone being in the *normal*

profile was more likely among agreeable ($\beta = 0.17$, $t = 4.39$), conscientious ($\beta = 0.31$, $t = 8.47$) users who score low on emotional stability ($\beta = -0.22$, $t = -5.65$) and openness ($\beta = -0.40$, $t = -4.61$). Disagreeable ($\beta = -0.16$, $t = -4.17$), non-conscientious ($\beta = -0.144$, $t = -3.84$), and emotionally stable ($\beta = 0.324$, $t = 8.13$) users were more likely to have their phone in the *silent* profile. Interestingly, extraverts were more likely to have *normal* ($\beta = 0.076$, $t = 2.61$), and those who additionally score low on openness ($\beta = -0.11$, $t = -3.37$) were less likely to have *silent* ($\beta = -0.11$, $t = -3.68$) as the most dominant profile. The *Beep* profile was more likely to be used by non-conscientious ($\beta = -0.11$, $t = -2.94$) users who score higher on openness ($\beta = 0.19$, $t = 6.16$). Also, users who were open were more likely to have *Beep* was the most dominant profile. The use of *Ascending* explained up to 6.2% of the variance in the traits, with introverted ($\beta = -0.12$, $t = -4.26$), neurotic ($\beta = -0.11$, $t = -2.58$), and open ($\beta = 0.24$, $t = 7.68$) users being more likely to use it. However, agreeable ($\beta = 0.26$, $t = 6.47$) and non-conscientious ($\beta = -0.27$, $t = -7.12$) users were more likely to have this profile as the most dominant one. The *Ring Once* profile explained upto 10.7% of the variance, with extraverted ($\beta = 0.231$, $t = 8.37$) and non-conscientious ($\beta = -0.32$, $t = -8.67$) users more likely to use it. The use of the *Ring Once* profile as the most dominant one explained a very small proportion of the variance

Table 11 Summary of observed gender differences in regression analysis for a subset of features that could be broadly representative of smartphone usage

Feature	All					Female					Male				
	E	A	C	ES	O	E	A	C	ES	O	E	A	C	ES	O
Office	-	-	↑	-	↓	-	-	-	-	-	↑	-	↑	↓	↓
Internet	↓	↓	-	-	-	↓	-	↓	-	-	-	↓	↓	↑	-
Video/Audio/Music	-	-	↓	-	↑	↓	-	-	-	↑	-	-	↓	-	-
Mail	-	↓	↑	-	-	↓	↑	-	-	-	↑	↓	-	-	-
Youtube	↑	-	↓	-	-	↑	↓	↓	↑	↑	↑	-	↓	-	-
Calendar	-	↓	-	-	-	-	-	-	↓	-	↑	-	-	-	↓
SMS	-	↓	↑	-	↓	-	-	↓	-	↓	-	↓	-	-	↓
Games	-	↓	-	↑	-	-	-	-	-	-	-	-	-	↑	-
Avg. word length (Inbox)	↑	-	↓	↑	↓	-	-	-	↑	↓	-	-	↓	↑	↓
Avg. word length (Sent)	-	-	-	↑	↓	-	-	-	↑	-	-	-	-	↑	↓
Outgoing (O) Calls	-	-	-	-	-	-	-	↓	↓	-	-	↑	-	-	↓
Avg. duration (O Calls)	-	-	-	-	-	-	-	↓	-	-	-	-	-	-	-
Incoming (I) Calls	↑	-	↓	-	-	↑	-	-	-	↑	↑	↑	↓	-	↓
Avg. duration (I Calls)	↑	-	↓	-	-	↑	-	-	↑	-	↑	↑	↓	-	-
Unique contacts (O Calls)	↑	↑	↓	-	-	-	↓	↑	↓	↑	↑	↑	-	-	-
Unique contacts (I Calls)	↑	-	↓	-	↑	-	↓	↑	↓	↑	↑	↑	↓	-	-
Unique contacts in call logs	↑	↑	↓	-	↑	-	↓	↑	↓	↑	↑	↑	-	-	-
Unique contacts (M)	-	↑	↓	-	-	-	↓	-	↓	-	-	↑	↓	-	↓
SMS received	↑	↓	-	↑	↓	-	-	↓	↑	↓	↑	↓	-	↑	↓
SMS sent	↑	-	-	↑	↓	-	-	-	↑	↓	↑	↓	-	↑	↓
BT IDs seen more than 4 slots	-	↓	-	-	-	↓	-	-	↓	-	-	↓	-	-	-
BT IDs seen more than 9 slots	-	↓	-	-	-	↓	-	-	↓	-	↑	↓	-	-	-
BT IDs seen more than 19 slots	-	↓	-	-	↑	↓	-	-	↓	-	↑	↓	-	-	-
Max time a BT ID seen	↓	-	↑	↓	-	-	-	-	↓	↑	-	-	↑	↓	-
Probability normal (N)	-	↑	↑	↓	↓	-	-	↓	-	↓	↓	↑	↑	↓	-
Probability silent (S)	-	↓	↓	↑	-	↓	-	-	-	-	↑	-	↓	↑	↓
Probability beep (B)	-	-	↓	-	↑	-	-	-	-	-	↓	-	↓	-	↑
Probability Ascending (A)	↓	-	-	↓	↑	↑	-	↑	-	↑	↓	-	-	↓	↑
Probability Ring Once (RO)	↑	-	↓	-	-	-	-	-	-	-	↑	-	↓	-	-
Dominant profile count (N)	↑	-	-	-	-	-	-	↓	-	-	-	-	-	-	-
Dominant profile count (S)	↓	-	-	-	↓	↓	-	-	-	-	-	-	-	-	↓
Dominant profile count (B)	-	-	-	-	↑	-	-	-	-	-	-	↓	-	-	↑
Dominant profile count (A)	-	↑	↓	-	-	↑	↓	↑	-	↑	↓	↑	↓	-	-
Dominant profile count (RO)	↑	-	↓	-	-	-	-	-	-	-	↑	-	↓	↑	-

(1.3%, $F = 3.66$) although significant, with extraverted and non-conscientious being more likely to score higher for this feature. Finally, the average number of daily changes in the profile or its standard deviation did not significantly explain the variance in the Big-Five traits.

4.4 Gender differences

Upon splitting up the data on the basis of gender, several interesting differences were observed. The differences in correlations are tabulated in Tables 6, 7 and 8. The differences in R^2 values and F -statistics are tabulated in

Tables 9 and 10. The differences in the observed trends, through regression analysis for a subset of the features, that were chosen such that the subset captures the overall usage of a smartphone by a user are summarized in Table 11. Please note that for clarity and conciseness of the discussion, the β and t values have been excluded.

4.4.1 App logs

Several differences were observed in the usage of applications across gender. Men were more likely to use applications like Games, Youtube, and Office. Hence, they

largely contributed to the regression coefficients in the analysis with the entire populations. The likelihood of the use of Internet apps by introverts was limited to the female population.

4.4.2 SMS logs

It was seen in the regression analysis that those scoring high on emotional stability and low on openness across both populations were more likely to send and receive longer SMS. However, lower conscientiousness was an additional coefficient that contributed to the regression of these features for the male population.

4.4.3 Call logs

While the outgoing calls did not explain a significant variation in the traits at the level of the entire population, interestingly, it explained 9 and 7% of the variance, significant to a degree of $p < 0.01$ in the female and male population, respectively. Among women, non-conscientious and neurotic users were more likely to make calls, while for men, agreeableness and low openness were the significant coefficients. Correlation analysis, however, revealed that conscientiousness itself is positively correlated to the number of outgoing calls in both the male and female population. The number and duration of incoming calls on the other hand were found to be likely to be higher for extraverted users across both male and female populations. In the male population, agreeableness always gave a significant positive β coefficient for the duration and number of incoming calls and the number of unique contacts associated with voice calls. Interestingly, among women, those scoring high on extraversion and openness were more likely to receive incoming calls.

4.4.4 BT logs

From the BT Logs, it was seen that the trend of extraverted and disagreeable users being more likely to have fewer BT IDs seen for longer durations was held true only for the male population. On the other hand, high scores on neuroticism and introversion for women were likely to decrease the value of this feature.

4.4.5 Profile logs

Features derived from this source exhibited a difference in the contribution of traits in explaining the traits, across genders. This is again seen in Table 11. Interestingly, the number of changes in profile per day did not significantly explain the traits among men. Among women, it was found that these features significantly explained the traits, albeit a

small amount. It was found that these features were higher valued for those who scored high on emotional stability.

4.5 Comparison of observations with previous studies

Previous studies [5, 23] have explored self-reported usage of phone calls, SMS, user profiles, and games based on surveys and personal recollection. No distinction between behaviors across genders was considered in these studies. To our knowledge, this is the first study that utilizes automatically extracted cues from multiple sources in a smartphone. In this section, we contrast our findings with previous work for each of the Big-Five traits.

4.5.1 Extraversion

Costa and McCrae associate extraverts with talkativeness, gregariousness, and outgoing nature [21, 8]. In a previous work on self-reported mobile phone use, it was found that extraversion was related to more time spent on incoming calls, although extraverts might not have liked to receive many calls [5]. Our results also show that extraverts are more likely to have more number of incoming calls and of longer duration. This could again be simply because extraverts have a more vibrant social life and because of their outgoing and talkative nature. In concordance with Butt and Philips work [5], outgoing calls were not good predictors of incoming and outgoing calls. This reinforces the hypothesis that extraverts might not receive incoming calls just because they have more friends, but because possibly people feel more comfortable to call extraverts. Finally, the higher usage of Internet among introverts, found for the entire population, has also been found in previous studies on Internet usage [18].

4.5.2 Agreeableness

Individuals who score less on this trait have been described in the past to be principally selfish, uncooperative, and not afraid to look for number one [8]. Conversely, Butt and Phillips found that disagreeable people were more likely to receive incoming calls. In our study, we found that disagreeableness was not a significant predictor in our regression analysis. However, pairwise correlations revealed that agreeableness in the male and female population was significantly (positively) correlated to the number of incoming calls. Also, agreeable men were found to communicate with more number of unique contacts through voice calls. This concurs with the descriptions of this trait in psychology [8] that agreeable people are more appreciative and generous in their behavior, which could make others perceive them as friendly. Interestingly, it was found in our study that higher agreeableness was associated

with longer and more number of SMS sent among women and an opposite trend among men.

4.5.3 Conscientiousness

Those who score high on conscientiousness can be characterized as being efficient, organized, planful, reliable, and responsible [21]. In our work, it was found that disagreeable and conscientious users were more likely to use the Mail application, while extraverts and non-conscientious participants were less likely to use the Youtube app. Since Mail could constitute essential communication, often used for professional purposes, conscientious people might be more likely to use this application. Correlation analysis also showed a negative correlation between conscientiousness and Internet usage for men. Further, it was found that the number of unique contacts associated with voice calls increased for non-conscientious users. This could be attributed to the responsible nature of conscientious users, who tend to contact fewer people in a prudent manner. This is further bolstered by the negative regression coefficients seen for both men and women for conscientiousness and the number of unique contacts associated with voice calls.

4.5.4 Emotional stability

It has been suggested in the past that those scoring low on emotional stability do not find mobile phone appealing [2]. It has also been suggested that neuroticism could explain time spent using SMS [5]. Interestingly, in our study, the frequency of opening the SMS app was not significantly correlated to emotional stability when observed for the entire population. Upon splitting it across genders, it was found that the frequency of usage of SMS was positively correlated to emotional stability among women and negatively among men. Further, it was found that for both populations, emotionally stable individuals with low openness to experience were more likely to send and receive longer SMS messages. However, the time spent on using the SMS app or writing messages is not directly captured by our features. Also, the ratio of SMS usage to voice calls has not been investigated in our study. Therefore, we were not able to validate whether neuroticism could explain the time spent using SMS as opposed to voice calls. Finally, it was found that overall, emotionally stable, extraverted individuals with low openness to experience were more likely to receive SMS. This could be due to the reason that users prefer to communicate with emotional stable individuals.

4.5.5 Openness to experience

Individuals with high openness tend to be imaginative, artistic, original, and with a wide range of interests [21].

Among women, this was associated with higher usage of Video/Audio/Music and Mail applications. Contrastingly, low openness was associated with higher usage of Office, Internet, Video, and Calendar applications among men. Also, the use of SMS was usually associated with low openness in both regression and correlation analyses. This could be due to users who are more open to experience relying on other forms of communication. Further, the number of unique contacts in call logs was also associated with higher openness in the entire population. Finally, it was found that high openness was associated with the use of *Beep* and *Ascending* user profiles. This may be attributed to the curious nature of such users, who try out the user of profiles other than the commonly used *Normal* or *Silent* profiles.

5 Classification of users based on personality

In the previous section, our analysis clearly showed that several smartphone usage cues significantly explained the variance and also displayed significant pairwise correlations with the Big-Five traits. Hence, as a next step, in order to put our analysis into a machine learning framework, we defined a binary classification task for each of the Big-Five traits, using the median value of the traits in a given population as a threshold to split the data into two classes. This was done with an intention to discriminate, for example, the more extraverted and less extraverted users in the given data.

As a first step, we chose those features, for each of the traits, that gave a significant pairwise correlation ($p < 0.01$). Next, we used a sequential backward feature selection algorithm and an SVM classifier with a radial basis function (RBF) kernel implemented in the Shogun Toolbox [27] in order to select features. We used leave-5-user-out cross-validation, as opposed to the leave-one-user-month-out cross-validation used in our previous work. This was done in order to classify personality types on 5 users who are completely unknown and not available in the training dataset. In each training/testing phase, the features were z-score normalized, using the mean and standard deviation computed with the training data.

We present the results averaged across a leave-1-user-out cross-validation (in order to have a different cross-validation scheme from the feature selection stage) in Table 12. Since the classes were often unbalanced (due to the discrete nature of the measured personality, with the TIPI questionnaire), we present both micro- and macro-averaged F-measures for the results. The micro-averaged F-measure gives equal weight to all classifications, so that F1 scores of the larger class influence the metric more. The macro-averaged F-measure on the other hand gives equal

Table 12 Average performance values (measured with F-measure) across all folds of leave-1-user-out cross-validation

Trait	Classifier		Majority		Random	
	macro	micro	macro	micro	macro	micro
(a) Entire Population						
Extraversion	0.58	0.77	0.38	0.77	0.49	0.67
Agreeableness	0.59	0.75	0.37	0.74	0.50	0.67
Conscientiousness	0.55	0.75	0.38	0.77	0.49	0.67
Emotional stability	0.54	0.71	0.35	0.70	0.50	0.67
Openness to experience	0.59	0.74	0.34	0.68	0.50	0.67
(b) Female population						
Extraversion	0.67	0.80	0.35	0.71	0.50	0.67
Agreeableness	0.49	0.81	0.42	0.84	0.47	0.67
Conscientiousness	0.62	0.77	0.34	0.67	0.50	0.67
Emotional stability	0.63	0.78	0.36	0.73	0.50	0.67
Openness to experience	0.54	0.72	0.36	0.72	0.50	0.67
(c) Male Population						
Extraversion	0.49	0.72	0.39	0.77	0.49	0.67
Agreeableness	0.69	0.83	0.37	0.75	0.50	0.67
Conscientiousness	0.58	0.75	0.37	0.74	0.50	0.67
Emotional stability	0.56	0.73	0.36	0.71	0.50	0.67
Openness to experience	0.60	0.76	0.38	0.75	0.49	0.67
(d) Hybrid Model						
Extraversion	0.59	0.77	0.38	0.77	0.49	0.67
Agreeableness	0.59	0.77	0.37	0.74	0.50	0.67
Conscientiousness	0.61	0.78	0.38	0.77	0.49	0.67
Emotional stability	0.60	0.75	0.35	0.70	0.50	0.67
Openness to experience	0.59	0.74	0.34	0.68	0.50	0.67

Traits in bold show cases where the trait is classified equal to or better than the majority baseline for the micro-averaged F-measure. Refer Sect. 5, for details

weight to the F1 scores of all labels, thus attributing equal weights to all classes.

In addition to the results from the SVM, the performance of two baselines, corresponding to random chance and majority class selection, is also given in Table 12. Finally, a hybrid model (Table 12d) that utilizes the best models to enhance the classification performance is also presented. This model was constructed by choosing the best models (men, women, or overall) to perform classification of all data points. A description of this model is described later in this section.

The results show that all traits can be classified better than chance albeit being a hard task [24]. Upon comparison of the micro-averaged F-measure results between the SVM and the majority baseline, we see that some traits are harder to classify than others. For the female-only population (Table 12b), with this performance metric, traits that were harder to classify were agreeableness (0.81 against 0.84) and openness to experience (0.72 against 0.72), while extraversion, emotional stability, and conscientiousness traits could be classified even better than the majority classifier.

On the other hand, for the male population (Table 12c), agreeableness, conscientiousness, and emotional stability

could be classified with a better micro-F-measure than the majority baseline. In order to exploit the high performance in classifying certain traits in gender-specific models, we finally present the hybrid method, in which the classification is done for the entire population (with the entire population being split across the median), but a gender-dependent model is used, if it has given a performance better than a gender-independent model (Table 12a) for the macro-averaged F-measure. For example, classification of extraversion was found to be good among women. Therefore, this model was used when a female was encountered during classification. But since the male-specific model did not perform well, it was not used in place of the gender-independent model. Therefore, finally, for men, the model built using the entire population was used instead. The hybrid method performed better than the single model based on the entire population.

6 Conclusions and future work

This paper lays the basis for research in the prediction and usage of personality traits for socially aware services on smartphones. Our study presents a detailed analysis of

the relationship between automatically extracted and aggregated smartphone usage features and the Big-Five personality traits. The methodology presented in this paper offers two main benefits. Firstly, the methods are easily scalable to large datasets and to a large number of features. Further, the features used are by nature privacy sensitive, which is of paramount importance in this area of research.

The results clearly show that several aggregated smartphone usage features could be predictive of the Big-Five personality traits. The analysis of smartphone usage features revealed several interesting trends. Many of these trends conform with past work in psychology literature. It was found that extraverts, who are characterized by talkativeness and outgoing nature, were more likely to receive calls and also spend more time on them. Features pertaining to outgoing calls were found not to be predictive of the Big-Five traits. Agreeableness among women was associated with an increase in the number of incoming calls. Agreeable men were found to communicate with more number of unique contacts through voice calls. On the other hand, conscientiousness was associated with higher usage of the Mail app, which could be used in a professional context, and with lower usage of the Youtube application, which is likely to be used for entertainment purposes. Conscientious users were also likely to contact lesser number of unique people through voice calls. This conforms with their characterization in the literature as responsible and organized individuals. Interestingly, emotional stability was linked to higher incoming SMS. And high openness was associated with increased usage of Video/Audio/Music apps in women and also with the usage of nonstandard calling profiles such as *Beep* and *Ascending* in the entire population. Lastly, we found that several differences between personality and smartphone usage existed across genders, which has not been explored in previous literature.

Subsequently, in Sect. 5, it was shown that a machine learning framework based on a supervised learning method can effectively classify an unknown user's Big-Five trait measures as belonging to either the higher half or lower half of the population.

Regarding future work, in our opinion, this work shows the potential for further research into how personality traits can be predicted from smartphone usage.

Today's smartphones, apart from the modalities described in this paper, can also capture information from other modalities. Utilizing this information, in the form of location traces from GPS, physical activity levels through the accelerometer/gyro-meter require further investigation. Since mobile phones mediate social interactions, studying the social networks and their relationship to personality traits of users is another topic of study. Finally, more work

is needed in the direction of predicting supplementary user characteristics such as gender and age and incorporating it into the existing prediction framework.

While this paper addresses the interplay between gender, personality, and smartphone usage, the paper also opens up several interesting questions. Extending the analysis to a data collection that goes beyond smartphone could shed further light on some of the findings in the present paper. For instance, could individuals with high introversion be keen to use web-based communication channels could be verified with such an overarching data set. Secondly, the interpretation of the findings presented in the present paper poses a methodological challenge. While quantitative data analysis methods used in this study are suitable for highlighting statistical regularities, qualitative techniques are likely to be needed in order to obtain more insights into the reasons for individuals with a certain personality profile behaving in a given way.

Acknowledgments This work was funded by the SNSF project "Sensing and Analyzing Organizational Nonverbal Behavior" (SONVB) and by Nokia Research Center (NRC) Lausanne. We thank Juha K. Laurila (NRC) and Trinh-Minh-Tri Do (Idiap) for valuable discussions.

References

1. Back MD, Stopfer JM, Vazire S, Gaddis S, Schmukle SC, Egloff B, Gosling SD (2010) Facebook profiles reflect actual personality, not self-idealization. *Psychol Sci* 21:372–374
2. Bianchi A, Phillips J (2005) Psychological predictors of problem mobile phone use. *CyberPsychol and Behav* 8:39–51
3. Biel J, Aran O, Gatica-Perez D (2011) You are known by how you vlog: Personality impressions and nonverbal behavior in youtube. In: Proceedings of international AAAI conference on weblogs and social media (ICWSM)
4. Brinkman WP, Fine N (2005) Towards customized emotional design: an explorative study of user personality and user interface skin preferences. In: Proceedings of annual conference on european association of cognitive ergonomics (EACE)
5. Butt S, Phillips JG (2008) Personality and self reported mobile phone use. *Comput Hum Behav* 24:346–360
6. Chittaranjan G, Blom J, Gatica-Perez D (2011) Who's who with big-five: Analyzing and classifying personality traits with smartphones. In: Proceedings of the 15th international symposium on wearable computers (ISWC), San Francisco, USA
7. Coppersmith S (1959) A method for determining types of self-esteem. *J Abnorm Soc Psychol* 59:87–94
8. Costa PT, J, McCrae RR (1992) Revised NEO personality inventory (NEO-PI-R) and NEO five-factor inventory (NEO-FFI) professional manual. Psychological assessment inventories
9. Counts S, Stecher K (2009) Self-presentation of personality during online profile creation. In: Proceedings of international AAAI conference on weblogs and social media (ICWSM)
10. Do T, Gatica-Perez D (2010) By their apps you shall understand them: mining large-scale patterns of mobile phone usage. In: Proceedings of international conference on mobile and ubiquitous multimedia (MUM)
11. Eagle N, Pentland A (2009) Eigenbehaviors: identifying structure in routine. *Behav Ecol Sociobiol* 63:1057–1066

12. Farrahi K, Gatica-Perez D (2010) Probabilistic mining of socio-geographic routines from mobile phone data. *IEEE J Sel Top Signal Process* 4:746–755
13. Gosling SD, Rentfrow PJ, Swann W (2003) A very brief measure of the big-five personality domains. *J Res Pers* 37:504–528
14. Graham L, Gosling S (2011) Can the ambiance of a place be determined by the user profiles of the people who visit it? In: *Proceedings of international AAAI conference on weblogs and social media (ICWSM)*
15. I.c.t. statistics. <http://www.itu.int/ITU-D/ict/statistics/ict/index.html> (Accessed 9 Feb 2011)
16. Karsvall A (2002) Personality preferences in graphical interface design. In: *Proceedings of the second nordic conference on human-computer interaction (NordICHI)*
17. Kiukkonen N, Blom J, Doussé O, Gatica-Perez D, Laurila J (2010) Towards rich mobile phone datasets: Lausanne data collection campaign. In: *Proceedings of ACM international conference on pervasive services (ICPS)*
18. Kraut R, Patterson M, Lundmark V, Kiesler S, Mukapadhyay T, Scherlis W (1998) Internet-paradox: a social technology that reduces social involvement and psychological well being?. *Am Psychol* 53:1017–1031
19. Mairesse F, Walker M (2006) Automatic recognition of personality in conversation. In: *Proceedings of human language technology conference of the NAACL, companion volume: short papers (NAACL-Short)*
20. Mairesse F, Walker M, Mehl MR, Moore RK (2007) Using linguistic cues for the automatic recognition of personality in conversation and text. *J Artif Intell Res* 30:457–500
21. McCrae RR, John OP (1992) An introduction to the five-factor model and its applications. *J Pers* 60:175–215
22. Oliveira R, Karatzoglou A, Concejero P, Armenta A, Olivier N (2011) Towards a psychographic user model from mobile phone usage. In: *Proceedings ACM CHI conference on human factors in computing systems (CHI) work in progress (WIP)*
23. Phillips JG, Butt S, Blaszczyński A (2006) Personality and self-reported use of mobile phones for games. *Cyberpsychol and Behav* 9:753–758
24. Pianesi F, Mana N, Cappelletti A, Lepri B, Zancanaro M (2008) Multimodal recognition of personality traits in social interactions. In: *Proceedings of International conf. on multimodal interfaces (ICMI)*
25. Poschl S, Doring N (2007) Integration and ubiquity. Towards a philosophy of telecommunications convergence, chap. Personality and the mobile phone: character-based differences of usage and attitudes towards mobile communication, pp. 161–168. Vienna, Passagen Verlag
26. Romero JJ (2011) Top 11 technologies of the decade. *IEEE Spectrum*, 24–27
27. Sonnenburg S, Raetsch G, Henschel S, Widmer C, Behr J, Zien A, Bona FD, Binder A, Gehl C, Franc V (2010) The shogun machine learning toolbox. *J Mach Learn Res* 11:1799–1802
28. Stecher K, Counts S (2008) Spontaneous inference of personality traits and effects on memory for online profiles. In: *Proceedings of international AAAI conference on weblogs and social media (ICWSM)*
29. Tabachnick B, Fidell L (2000) *Using multivariate statistics*, 4th edn. Allyn & Bacon, Boston
30. Appendix ii: Normative data for the ten-item personality inventory (tipi): Self-reported data. <http://homepage.psy.utexas.edu/homepage/faculty/gosling/tipisite/norms.pdf> (Accessed 9 Feb 2010)
31. Verkasalo H, Lopez-Nicols C, Molina-Castillo FJ, Bouwman H (2010) Analysis of users and non-users of smartphone applications. *Telemat Informat* 27:242–255
32. Yeo T (2010) Modeling personality influences on youtube usage. In: *Proceedings of international AAAI conference on weblogs and social media (ICWSM)*