ARTICLE IN PRESS YJCSS:2688

Journal of Computer and System Sciences ••• (••••) •••–•••

Contents lists available at SciVerse ScienceDirect

Journal of Computer and System Sciences

www.elsevier.com/locate/jcss

Predicting user personality by mining social interactions in Facebook

Alvaro Ortigosa *, Rosa M. Carro, José Ignacio Quiroga

Computer Science Department, Universidad Autónoma de Madrid, Madrid, Spain

A R T I C L E   I N F O

Article history:
Received 25 July 2012
Received in revised form 13 December 2012
Accepted 14 March 2013
Available online xxxx

Keywords:
Data mining in social networks
User modeling
Personality inference

A B S T R A C T

Adaptive applications may benefit from having models of users' personality to adapt their behavior accordingly. There is a wide variety of domains in which this can be useful, i.e., assistive technologies, e-learning, e-commerce, health care or recommender systems, among others. The most commonly used procedure to obtain the user personality consists of asking the user to fill in questionnaires. However, on one hand, it would be desirable to obtain the user personality as unobtrusively as possible, yet without compromising the reliability of the model built. On the other hand, our hypothesis is that users with similar personality are expected to show common behavioral patterns when interacting through virtual social networks, and that these patterns can be mined in order to predict the tendency of a user personality. With the goal of inferring personality from the analysis of user interactions within social networks, we have developed TP2010, a Facebook application. It has been used to collect information about the personality traits of more than 20,000 users, along with their interactions within Facebook. Based on all the collected data, automatic classifiers were trained by using different machine-learning techniques, with the purpose of looking for interaction patterns that provide information about the users' personality traits. These classifiers are able to predict user personality starting from parameters related to user interactions, such as the number of friends or the number of wall posts. The results show that the classifiers have a high level of accuracy, making the proposed approach a reliable method for predicting the user personality

1. Motivation

User modeling is essential in the context of adaptive systems. Depending on the goal of each adaptive environment, it can be useful to represent and use some information about the user features, preferences, needs, behaviors, etc. [1]. Regarding the characteristics that can be considered with adaptation purposes, personality is one interesting feature. The personality of an individual can be defined as a set of features that induces a tendency on the behavior of the individual; this tendency is stable through time and situations [2]. Knowing the personality of a given person provides hints about how he would probably react when facing different situations.

Identifying a user's personality can contribute to know, for example, his potential needs in different contexts. Therefore, adaptive applications may benefit from having models of users' personality to adapt their behavior accordingly. There is a wide variety of domains in which this can be useful, i.e., assistive technologies, e-learning, e-commerce, health care or recommender systems, among others. For example, in the context of e-commerce, the type of products to be offered to a user may vary depending on his personality with respect to *Impulsive Sensation Seeking*. Another application area would

* Corresponding author.
E-mail addresses: alvaro.ortigosa@uam.es (A. Ortigosa), rosa.carro@uam.es (R.M. Carro), jose.quiroga@estudiante.uam.es (J.I. Quiroga).

0022-0000/$ – see front matter
http://dx.doi.org/10.1016/j.jcss.2013.03.008

Please cite this article in press as: A. Ortigosa et al., Predicting user personality by mining social interactions in Facebook, J. Comput. System Sci. (2013), http://dx.doi.org/10.1016/j.jcss.2013.03.008

be education, since personality influences the way in which students learn and use their knowledge [3], and affects the way in which students accomplish their tasks [4]. In adaptive educational systems it would be useful to know the student personality to propose him the most suitable tasks in each context accordingly (for example, if he has only 10 minutes available and connects to the system, it could be inappropriate to propose him to develop a complex task, even if it is not so time-consuming, if he has a high level of *Neuroticism-Anxiety*).

Therefore, eliciting user personality can contribute significantly to user modeling for adaptive systems. The most commonly used procedure to obtain this information consists of asking the user to fill in questionnaires. However, users can find this task too time-consuming, since most of the personality questionnaires include many questions to answer in order to obtain an accurate user profile [5,6].

On one hand, we think that user personality should be obtained as unobtrusively as possible, yet without compromising the reliability of the model built. Some research has already been done related to automatic or semi-automatic user modeling acquisition. Several works have tried to infer some personality characteristics, as it will be discussed in the state of the art section.

On the other hand, we think that personality can be inferred by analyzing how users interact in online social networks (OSNs). In this context, an initial consideration would be whether virtual interactions reflect user personality in "real" or offline life. For example, [7] and [8] verify that Facebook,[1] currently the most popular OSN, is, apparently, a good approximation to the user offline life. If a given user has occasional interactions with another user through the Web, it will not be the case that they have many more offline interactions. In addition, it has been shown that people use virtual social networks to support already existing relationships rather than looking for new ones (for Facebook users, around 77% of their social relationships in the real world are replicated in the virtual environment).

In the same line, the user interactions on social networks are a good measure of user behavior in real life [9]. They verify this hypothesis analyzing user comments on pictures and publications of Facebook wall. For example, they show that even if a user has a lot of relations ("friends" in Facebook terminology) he only interacts regularly with a small portion of them. That is because there are, like in real life, constraints that make interacting with all friends impossible (e.g., available time).

Furthermore, [10] shows that, contrary to the idea that people built idealized virtual-identity, Facebook profiles reflect actual personality. The same idea is supported by [11], where it was found out that personality impression based on Facebook profiles generally show strong patterns of convergence with real personality.

However, it is important to bear in mind that this work does not presume that users would act exactly in the same way both in real and in virtual life. What is known is that personality, by definition, is stable through time and situations. That means that if a user has tendency to be, for example, sociable, that tendency would show up both on real and virtual life, maybe in different ways. What this work assumes is that users with similar personality features will show common behavior patterns when interacting through virtual social networks, that those patterns can be mined, and that they can be used, afterwards, to identify the tendency of a user personality. In this direction, we think that it may be possible to infer personality from the analysis of user interactions within online social networks, such as the number of user's friends or the number of wall posts.

There exist a growing number of online social networks available through the Web. From these applications, Facebook is by large the more popular around the world. On June 2012, it reached 955 million of active users (that is, without considering the profiles created but not used) [12]. Fig. 1a shows the growing rate: in the last three months it increased 54 million users (almost 6%). More than a half of Facebook users are daily active users (552 millions), as shown in Fig. 1b.

Fig. 2 also shows that more than 543 million users access and update their profiles through mobile devices. Besides, according to [13], Facebook users have an average of 130 "friends" (the term used within Facebook to denote people related to the user) and are connected, on the average, to 60 items, counting groups, pages and events. The Facebook application was translated to more than 70 languages; it has more than one million of developers and more than 500,000 available applications.

Among all the social networks available through the Internet nowadays, we have chosen Facebook not only because of the huge amount of users connected to it, but also because it offers the possibility of building applications to be incorporated and used by all these users. In this context, we have developed TP2010 [14], a Facebook application intended to obtain information about the user personality through a personality test, as well as to collect all the data available from the user interactions within the social network. The application was delivered on December 2009, and it implements the ZKPQ-50-cc personality test [5].

TP2010 goal is to discover relationships between the users' results on the personality test and those attributes describing their interactions within Facebook. Our work attempts to find rules to predict user personality (that is, the tendency of the user on each personality dimension being considered), starting from data about his interactions in Facebook, without asking him to fulfill specific personality tests. With this aim, we have applied data mining techniques in order to build classifiers of user personality. These classifiers were trained based on the analysis of data from more than 20,000 users of the TP2010 application. Most of the related works found in the literature that try to find relationships between user personality and user interactions in social networks focus on analyzing how single features correlates, on the average, with personality traits. We intend to go beyond and analyze the collected data by looking for patterns of user interactions that correspond to

[1] www.facebook.com.

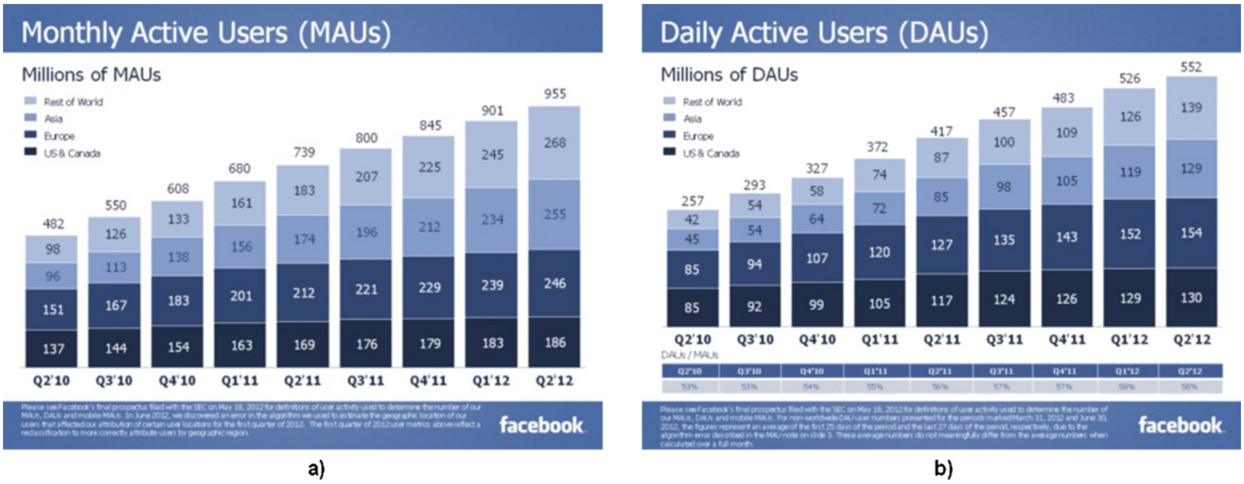

Fig. 1. Monthly and daily active users in Facebook.

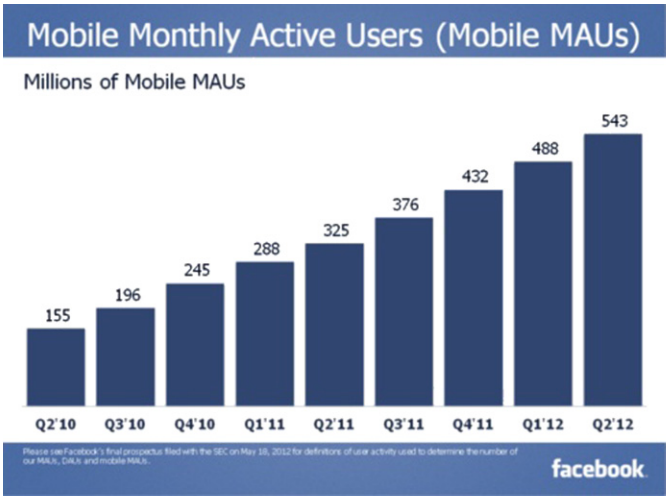

Fig. 2. Mobile monthly active users in Facebook.

specific types of user personality. In that way, having data of a given user's interactions would make it possible to predict his personality, at least regarding some personality traits.

The paper is structured as follows. Section 2 presents the state of the art. Section 3 shows how data about user personality and interactions is collected though the Facebook application TP2010. Section 4 describes the data collected from more than 20,000 users. Section 5 shows how personality classifiers are built. Section 6 presents the results of personality prediction. And, finally, Section 7 comprises conclusions and future work.

2. State of the art

The growing popularity of online social networks, especially Facebook, has given rise to an also growing number of psychological and sociological studies based on them. A recent survey [15] identifies more than 400 articles fulfilling three requirements: a) they specifically investigate Facebook; b) they have been published in a peer-reviewed academic journal or peer-reviewed conference proceedings; and c) they report empirical findings. On this survey, Wilson et al. classify the articles in five categories, according to the question they try to answer: i) Who is using Facebook and what are users doing while on Facebook? ii) Why do people use Facebook? iii) How are people presenting themselves on Facebook? iv) How is Facebook affecting relationships among groups and individuals? v) Why are people disclosing personal information on Facebook despite potential risks?

In most of these works, mainly the older ones, data are collected by requiring the users to fulfill surveys, many times offline. For example, [16] presents a study about online privacy. The users were asked, through e-mail in this case, to answer questions about how long they have been using Facebook, the number of friends they have, how many minutes they were online every day, whether their profile was restricted only to friends, their gender and their ethnic group. The dataset

Please cite this article in press as: A. Ortigosa et al., Predicting user personality by mining social interactions in Facebook, J. Comput. System Sci. (2013), http://dx.doi.org/10.1016/j.jcss.2013.03.008

contained information about 5000 users. Some of the conclusions reached are that users tend to restrict the profiles only to friends when they reach a given number of friends and that users with restricted profiles have, on the average, 25% more friends than those keeping their profile public.

On a different line, [17] and [18] show that, while university students fulfill most of the profile slots (59%), a sample composed of university and non-university users only complete 25% of the information. More interestingly, they highlight that knowing whether the user has filled in their age and sentimental status, allows to inferring, with great accuracy, whether he keeps public or private most of his profile. In the same way, they found out that the age and the amount of public data correlates negatively.

More recent studies collect data (semi) automatically. There are two possible approaches: data crawling and Facebook applications. Data crawling involves mechanisms similar to those used by search engines for indexing Web pages: data from public Facebook user pages are collected, and pages (users) are related through the links found in the pages. The advantage of this approach is that it does not require any user action; the main drawback is that, as the time goes on, less information can be collected in such a way, since Facebook implements more restrictive privacy policies. Following this approach, [17] finds a correlation between the number of friends and the number of profile slots completed by the user. Specifically, the correlation is greater with reference data (place of birth, studying institution, etc.), followed by contact data (sentimental state, address, etc.) and lastly interest data (music, cinema, books, etc.). Another work [19] analyzes relationships between users from a different perspective: they measure how similar two given users are, by considering those networks, groups and friends in common. Besides, they analyze two types of interactions between pairs of users: whether they have written one in the wall of the other and whether they have cross-tagged themselves in photos.

Nevertheless, the authors of these works have strong limitations for collecting some data like, for example, the number of friends. The problem is the growing tendency of Facebook users to restrict the access to their data, with most of the information unavailable for public access. Moreover, currently Facebook itself does not support collecting data by automatic means without explicit authorizations.

The second automatic approach consists of using Facebook applications. The idea is to build Facebook applications and ask the users to give them permission to access to their data. In order to encourage them to give it, most of these applications offer them something back. For example, an application looking to collect data about user tastes can ask a new user to fulfill a questionnaire about his tastes and, afterwards, provide him recommendations based on the information available from those users with similar tastes. In general, these applications use viral market techniques to reach a large number of users. Through this Facebook application approach, a lot of user data can be obtained.

The most noticeable application implementing this approach is myPersonality [20], up to our knowledge. This application offers the user a detailed explanation of his personality once he has installed the application and fulfilled one or more of the proposed questionnaires. According to the data reported by the application designers, they have collected data from more than 5 million users. Part of these data is publicly available and can be used for different studies about Facebook users and their personality. There are two works related to the current proposal that make use of data from myPersonality.

The first one uses personality profiles (Big Five model [6]) and Facebook profile data of 180,000 users [21]. They examine correlations between the users' personality and the properties of their Facebook profiles, such as the size and density of their friendship network, the number of photos uploaded, the number of events attended, the number of group memberships, or the number of times user has been tagged in photos. The results show significant relationships between personality traits and various features of Facebook profiles, like, for example, the relationship between *Openness* and the number of associations with Facebook groups. In the same line of our proposal, they propose to use machine-learning techniques for combining signals from different Facebook features in order to predict personality of individuals in a reliable way. This is an important difference with most of the previous works, which usually focused on correlating Facebook profiles with personality traits averaged over large groups, but were inaccurate when dealing with individuals. Using multivariate linear regression they are able to predict, with reasonable accuracy, two of the five traits of the Big Five personality model; for the other three traits the prediction has low or very low accuracy. They tried more sophisticated machine-learning methods for predicting traits, including tree based rule-sets and support vector machines. However, for all of the personality traits, accuracy changed very little when using these machine-learning methods. Finally, another remark about this work that should be taken into account is the dataset size. The original dataset was very large: 180,000 individuals. However, because of privacy and implementation issues, not all the data were available for all the individuals. For the correlation analysis, the authors state that they had "at least 15,000 data points per feature and over 50,000 data points for most of the features". Furthermore, when applying machine-learning techniques they only used data from 5000 individuals. Nevertheless, this is a very large dataset, mainly if compared with similar works like [22], which applies machine-learning techniques to data from 167 users. As we describe in the next sections, we also had to discard a great portion of the original dataset as well, but at the end we applied data mining techniques to a sample of more than 20,000 users.

The second work based on myPersonality uses the data to analyze the relationship between personality and different types of Twitter[2] users, including popular users and influential ones [23]. In this work, only the data of those 335 users who specified their Twitter accounts in their Facebook profiles could be used, out of more than 5 million of myPersonality users. However, the works presents some interesting findings: both popular and influential users are extroverts and emotionally

[2] www.twitter.com.

stable (low score in the *Neuroticism* trait); popular users are 'imaginative' (high score in *Openness*), while influential users tend to be 'organized' (high score in *Conscientiousness*). They also explore the possibility of predicting individual personality based the analysis of interaction data. However, for this purpose, not only is the dataset rather small, but also they can only access to very basic data: their following, followers, and listed counts. The authors claim to obtain predictors, from these data available, with the same accuracy than other existing recommender systems predicting user ratings for movies.

3. Personality and interaction Data Collection

As it was mentioned above, the goal of this work is to obtain a method for eliciting user personality without asking the user to fulfill a specific questionnaire. The approach consists of analyzing the user behavior while interacting with a social network and detecting the parameters or combination of parameters that best describe the user personality.

A reference framework must be defined in order to describe user personality. In other words, there is a need for a model of personality that describes its structure. Once a personality model has been chosen, interaction data from those users whose personality is known is needed. In this way, patterns of behavior related to certain personality features can be mined.

The next subsections explain some personality models, making emphasis on the one we have chosen for this study, and describe the method used for collecting data about user interactions and personality.

3.1. Personality models

Traditionally, in Psychology, personality structure is modeled in traits or factors. The personality of an individual would be described as the value of each trait for this individual. Three of the most accepted models for structuring personality are: the Eysenck three-factor model (also known as the P.E.N. model, which stands for Psychoticism, Extroversion Neuroticism) [24], the Big Five model [6] and the Alternative Five [25]. As their names suggest, the first model structures personality in three traits, while the second and third ones structure it in five traits. There is no consensus about which model describes personality better. However, it is usually accepted that their traits are equivalent most of the times; the three of them provide information about individual reactions to situations, and they provide information to decide, for example, which pedagogic approach is better according to different personalities.

In this work, the Alternative Five model has been chosen because of the free availability of the corresponding questionnaire, the ZKPQ-50-cc [5]. By using this questionnaire, it is possible to classify the personality of an individual according to five traits: *Sociability* (Sy), *Activity* (Act), *Aggression-Hostility* (Agg-Host), *Impulsive Sensation Seeking* (ImpSS) and *Neuroticism-Anxiety* (N-Anx).

In addition to the fifty questions of ZKPQ-50-cc corresponding to the five traits of the Alternative Five Model, ten additional questions contained in the original ZKPQ questionnaire [26] related with *infrequency* have been used in this work. *Infrequency* refers to the user's need of social acceptance. Therefore, a high *infrequency* reveals the possibility that the user has tried to obtain results corresponding to a "nice personality" and, therefore, it is quite probable that he has not been sincere. The answers corresponding to these ten questions are used to eliminate subjects with possibly invalid records: scores higher than 3 (out of 10) are considered as indicators of record questionable validity [26]. The set of questions from ZKPQ-50-cc questionnaire plus the ten *infrequency* questions have been named ZKPQ-60.

3.2. TP2010 Data Collection

In order to analyze the relationship between user behavior and personality, interaction data, along with personality traits of each subject, must be collected. With this goal, a Facebook application was built. These types of applications are built using the Facebook infrastructure and are freely available for Facebook users, which usually know about a new application through their contacts (viral marketing).

The application developed in this work was named TP2010 (TP is the acronym of the Spanish translation of "Personality Test") and is targeted to Spanish-speaker users. TP2010 consists mainly of the questionnaire ZKPQ-60. Before presenting the questionnaire, TP2010 asks the user for permission to:

- Post on the user's wall, in order to publish the test results.
- Access to the user's information, even when the user is not connected, in case it would be necessary to obtain new data.
- Send e-mails to the user, just in case TP2010 may need to send any information or, for example, to invite him to another application. Once the user finishes the test, and before sending the results back, the application suggests the user to inviting his friends to join TP2010. Finally, the user must configure his privacy preferences.

Once the user finishes the test, and before sending the results back, the application suggests the user to inviting his friends to join TP2010. Finally, the user must configure his privacy preferences:

- Whether the results can be published on his wall.
- Whether he wants to activate the "friend recommendation" option (more details about this will be given below).

Fig. 3. Snapshot of TP2010 showing the inferred personality data.

- Whether he allows his friends to compare the results they obtain in TP2010 with his.

Afterwards, TP2010 presents and explains to the user the results obtained, and then it offers the possibility of:

- Taking the test again: for example, if the system detects a low level of sincerity and recommends the user to take it again.
- Publishing the results on his wall: for example, in case that he had not selected this option in the previous menu or in case that he wants to publish them again.
- Exiting the application and going back to his Facebook profile.

Fig. 3 shows a snapshot of TP2010 showing the user personality according to his answers to the corresponding questionnaire.

The menu bar situated at the top of the interface allows the user to:

- Seeing the results again: in the case that the user is looking at another screen and wants to go back to the result window.
- Comparing the results obtained with those of other Facebook users (friends, people with similar age, people from the same country [27], all the people that has filled in the questionnaire, etc.). It also presents a "compatibility factor", aimed at adding entertainment to TP2010, in order to promote a wider use of the application. This factor is calculated as one hundred minus twice the sum of the distances in each dimension. Table 1 shows the calculation of the compatibility between the two users, which is graphically represented by TP2010 as shown in Fig. 4. The rationale behind this equation is that:
  - Users with maximum differences (e.g., one of them scoring 0 and the other one scoring 10 on each trait) would have a compatibility factor of 0; that is $100 - (2{*}10 + 2{*}10 + 2{*}10 + 2{*}10 + 2{*}10)$.
  - Users with no difference at all would have a compatibility factor of 100; that is $100 - (2{*}0 + 2{*}0 + 2{*}0 + 2{*}0 + 2{*}0)$.
- Using the "friend recommendation" option: the system shows, from all the people that have filled in the test, those whose results are more similar to the ones obtained by the user, that is, those whose compatibility factor is higher. The user can filter the results by sex. A snapshot of the friend recommendation made to a user is shown in Fig. 5.
- Watching the description of the test ZKPQ-50-cc.

Table 1
Example of compatibility calculation.

•	User 1	User 2	Difference
Act	3	5	2
Agghost	6	4	2
Sy	6	1	5
ImpSS	1	7	6
N-Anx	2	3	1
TOTAL	–	–	16

Compatibility: $100 - 2^{*}16 = 68\%$.

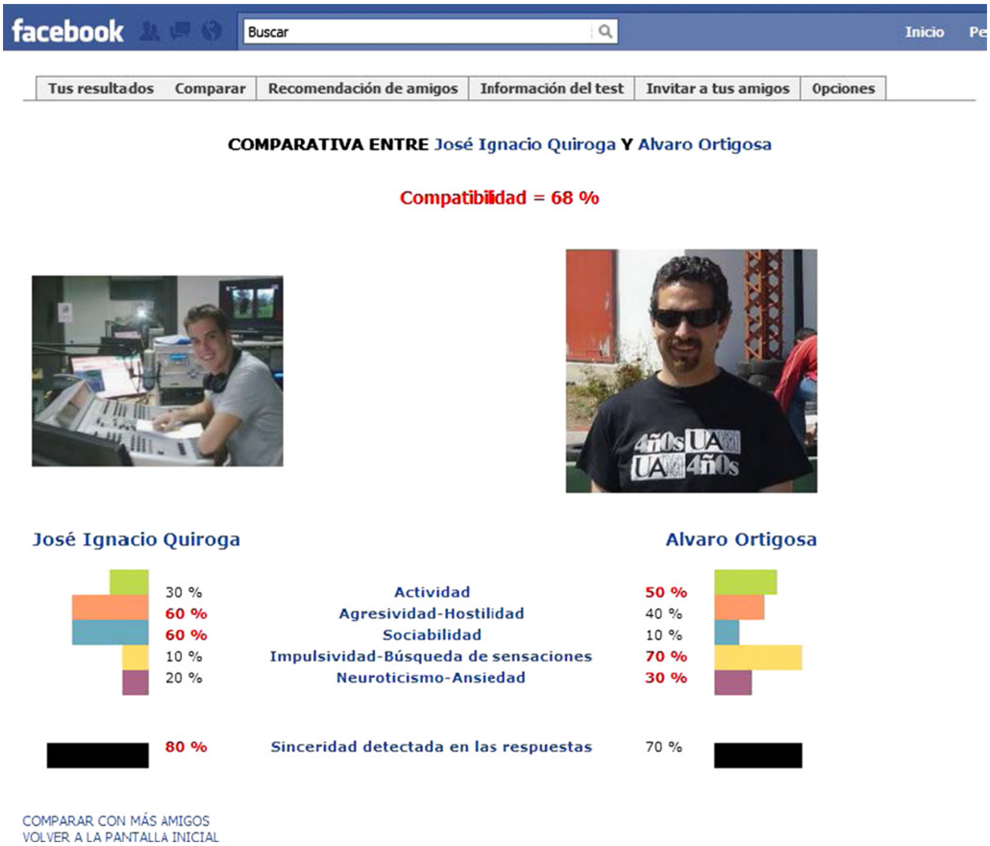

Fig. 4. Comparing two friends' personality in TP2010.

- Inviting the user's friends to join TP2010. If the user did not do this before, or in the case that he wants to invite more people to fill in the test, he can do it at any time. If his friends fill in the questionnaire and give the corresponding permission, they will be able to compare their results, which is usually encouraging.
- Option tab: the user can change privacy settings as well as the friend recommender option selection.

Many of these functionalities are offered only for the sake of encouraging the use of the application and facilitating its viral diffusion. As it has been stated before, the application goal is to collect data about Facebook usage and personality profiles. Every time a new user accesses TP2010, the application collects data about the user interaction within Facebook up to that moment and stores it in a database. Table 2 enumerates the data collected for each user that contributed to build the classifiers.

The first six values are computed directly from the user answers to the ZKPQ-60 test. The rest of them correspond to those parameters that can be obtained through the Facebook API by querying information about the user. The selection of these parameters was conditioned by two considerations. The first one deals with data accessibility through the Facebook API. Those items that could be retrieved within a reasonable amount of time were collected. For example, collecting information about the photos posted by a user took too much time at the time of the application delivery, which was impractical for the study goals. Secondly, the relevance of all the collected parameters for this study was analyzed, using attribute selec-

Fig. 5. Compatibility-based friend recommender in TP2010.

Table 2
Interaction parameters collected by TP2010 for each user.

Parameter name	Meaning
Act	Value in the Activity trait [0–10], calculated through the ZKPQ-60 test
Agg-Host	Value in the Aggression-Hostility trait [0–10], calculated through the ZKPQ-60 test
Sy	Value in the Sociability trait [0–10], calculated through the ZKPQ-60 test
ImpSS	Value in the Impulsive Sensation Seeking trait [0–10], calculated through the ZKPQ-60 test
N-Anx	Value in the Neuroticism-Anxiety trait [0–10], calculated through the ZKPQ-60 test
Inf	Value in the infrequency trait [0–10], calculated through the ZKPQ-60 test
Friends	Number of friends
Wall	Number of posts the user has in his wall
AfYear	Number of different friends that have written in the user's wall in the last year (Af is the acronym of Active friends)
AfMonth	Number of different friends that have written in the user's wall in the last month
PostsYear	Number of posts written in the user's wall in the last year
PostsMonth	Number of posts written in the user's wall in the last month
Months	Number of months since the user started using Facebook
AfPerMonth	Mean of different friends that have written in the user's wall per month
PostsPerMonth	Mean of posts written in the user's wall per month

tion techniques provided by the Weka tool box. Attributes like age and gender were discarded, since they did not contribute to improve the classifier accuracy. After the attribute selection phase, the attributes listed in Table 2 were the ones used to build the classifiers.

4. Collected data description

4.1. Dataset description

TP2010 was delivered on December 2009 and, amazingly, after 4 weeks of use, it reached 65,000 users. The three authors of this paper invited exclusively their friends to access TP2010 and, thanks to the network dynamics and viral marketing,

Table 3
Statistics of ZKPQ-50-cc and infrequency results.

Trait	Min	Max	Mean	Standard deviation
Act	0	10	4.54	2.71
Agg-Host	0	10	5.62	2.44
Sy	0	10	5.62	2.53
ImpSS	0	10	7.40	2.17
N-Anx	0	10	4.32	2.67
Infrequency	0	10	2.46	1.73

the invitation was spread to all those users. In July 2010, answers to ZKPQ-60 from 74,840 users were available. However, due to implementation issues, information about some interaction parameters was available only for a fraction of them (according to the documentation of Facebook available at that time, it seemed that we could access to all the information about a user offline, but it was only possible to do it when asking the user a special permission, which we incorporated afterwards). In a first stage of our study, we analyzed all the data collected in order to know which parameters we should consider to infer user personality with an acceptable level of accuracy. Once we decided which attributes were relevant, discarded the others (such as age or gender), and removed incomplete data, we finally worked with 20,988 instances, corresponding to the data of 20,988 new users, which constitute the kernel of the analysis presented in this work. Each instance includes the corresponding values for the attributes shown in Table 2 along with the results of the test ZKPQ-60 for one user.

4.2. Statistics

Table 3 shows a statistical description of the results for the five dimension personality questionnaire, along with the infrequency value, as collected by TP2010 from the 20,988 instances analyzed. Infrequency does not play the same role as the other features: as it was described above, a high value of infrequency reveals the probability that the user has not been sincere (scores higher than 3 lead to the elimination of the corresponding record).

Fig. 6 depicts the frequency distribution of the test answers for each dimension. It can be seen that all the personality traits, except *Impulsive Sensation Seeking*, are well balanced. As it will be explained later, this fact made no difference on the performance of the classifiers built, in terms of accuracy. However, it can reduce the confidence on the results for this trait, as it is easier to predict classes when a class *(High ImpSS* in this case) has many more instances than the others do.

Table 4 shows the statistical properties of the parameters that describe each user interactions. For example, it can be seen that the total number of friends (Friends) is quite different from the number of active friends (Af). It seems that Facebook users tend to have many friends but to interact only with a little portion of them, as described in [9]. As it can be seen in the table, the average number of friends per user are 186.03, while the mean number of different users with which one interacts per year is about the 20% of them (AfYear = 38.55). According to the average of the data obtained, we could characterize the typical user that joined TP2010 as a person with 186 friends, 39 of which have written on his wall during the last year (18 during the last month), with 88 messages on his wall, receiving 161 messages per year (60 in the last month), and having joined Facebook about 1 year ago.

5. Building the personality classifier

As it was stated above, most of the related works found in the literature that try to find relationships between user personality and user interactions in social networks focus on analyzing how single features correlates, on the average, with personality traits. From the data collected by TP2010 it can be observed that both *Sociability* (Fig. 7a) and *Impulsive Sensation Seeking* (Fig. 7b) correlate positively with the number of friends. Both figures also show a similar tendency, albeit less evident, between the same personality traits and the number of messages in the user wall.

However, the goal of this work is more ambitious. Our aim is to look for patterns of users' interactions that correspond to specific types of users' personality, by analyzing the collected data. Initially, different machine-learning algorithms were applied to produce classifiers of user personality. Techniques such as NaiveBayes [28], K-nearest neighbors [29], classification trees and association rules were used to analyze the dataset. Since the performance of the classifiers obtained by applying these techniques was similar, the final selection was made according to the simplicity and readability of the resulting model. In this sense, since classification trees provide a very good solution in terms of precision and readability, they were chosen. We worked with the Weka data mining toolbox, developed by the Waikato University, New Zealand. It consists of an open source set of tools and visualization algorithms for data analysis and predictive modeling. More specifically, the J48 implementation [30] of the C4.5 algorithm [31] was used. The discriminant analysis was performed using the lineal rule of Fisher [32] through the system R; this is another open source system focused on data and graphical analysis [33].

Data about user interaction and personality were utilized to train personality classifiers. The first step consisted on defining which type of prediction the classifiers were expected to do. Considering the output of ZKPQ-60, each user has six

Please cite this article in press as: A. Ortigosa et al., Predicting user personality by mining social interactions in Facebook, J. Comput. System Sci. (2013), http://dx.doi.org/10.1016/j.jcss.2013.03.008

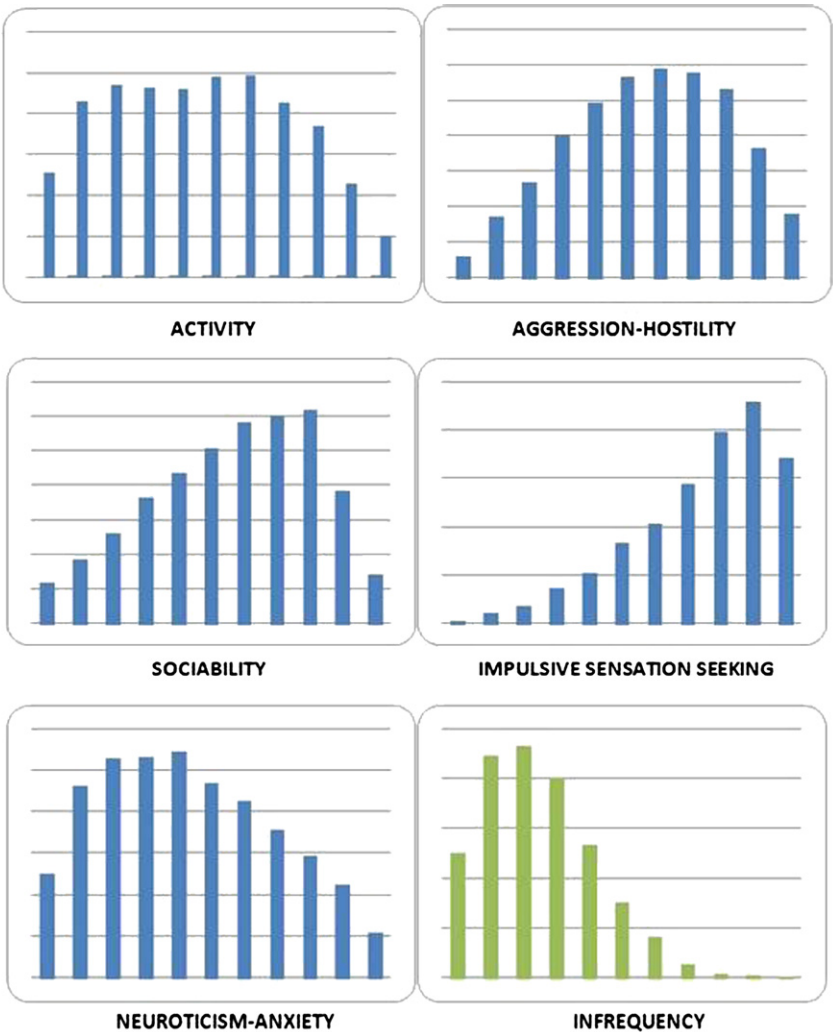

Fig. 6. Frequency of ZKPQ-50-cc traits and that of the infrequency dimension.

Table 4
Interaction parameters collected by TP2010.

Parameter	Min	Max	Mean	Standard deviation
Friends	1	2915	186.03	202.22
Wall	0	3480	88.39	141.63
AfYear	0	360	38.55	33.59
AfMonth	0	217	18.04	17.63
PostsYear	1	501	160.97	138.43
PostsMonth	1	500	59.52	64.50
Months	0	35	11.77	6.82
PostsPerMonth	0	660	34.13	56.34
AfPerMonth	0	140	6.65	9.91

values associated: five of them correspond to the five traits of the Alternative Five model, while the sixth corresponds to the value of the *infrequency* dimension. This last dimension, however, is only intended to validate the results obtained for the other five dimensions through the questionnaire and, therefore, was not considered when building the classifiers. The remaining five traits were analyzed independently from the others, as corresponds, since, as it has been mentioned above:

- Traits are supposed to have little or no correlation among them (by definition).
- Considering the intended scenario of use, it is possible that not all the traits are relevant for a given adaptive application.

Please cite this article in press as: A. Ortigosa et al., Predicting user personality by mining social interactions in Facebook, J. Comput. System Sci. (2013), http://dx.doi.org/10.1016/j.jcss.2013.03.008

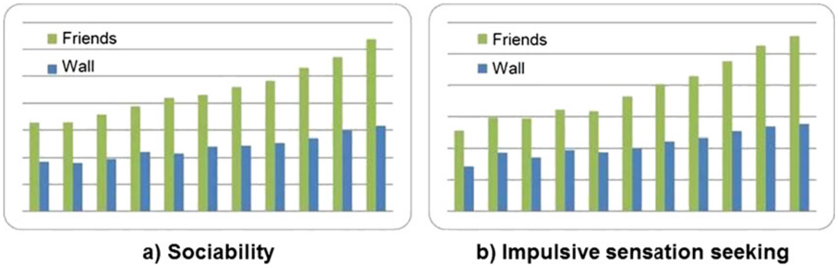

Fig. 7. Personality parameters vs. the average numbers of friends and wall messages.

- In no case would the value of a trait be inferred from the value of other(s) trait(s), as nothing is supposed to be known about the user personality.

Moreover, considering each trait separately simplifies the learning process and improves the accuracy of classifiers: it is easier to predict, for example, a user's score on the "Sociability" trait than to predict his score on the "Sociability", "Aggression-Hostility" and "Impulsive Sensation Seeking" traits at the same time, mainly considering that no relationship among traits exists.

Since the possible scores for each dimension of personality are eleven (from 0 to 10), we could think of eleven classes. However, for our prediction goal, it is not so relevant to know the exact score of one user in each dimension (e.g., whether he has scored either 9 or 10 in the sociability parameter), but whether he is "pretty sociable", "half sociable" or "little sociable". This is the reason why classifiers were trained to predict classes of personality. In other words, instead of predicting, for example, a value of 2 for the sociability trait, the classifier predicts that the user has a low sociability tendency. This is especially useful for adaptive systems, since they provide adaptation according to user profiles. In this direction, two alternatives were considered: dividing the scale either in three or in five classes, giving rise to 3-class and 5-class models, respectively.

In the 3-class model, each of the five traits is split up in three classes: *low, medium* and *high* (for example, *low activity*, *medium activity* and *high activity*). Because of this, five new synthetic attributes are added to each record with the labels of the corresponding classes. As each trait is analyzed independently, the dataset is divided into five datasets, each one containing one of the new label attributes.

In order to label each trait record, it is necessary to establish the relationship between the eleven possible outcomes for each trait and the three classes; that is, the limits for each class should be determined for each trait. The solution adopted was to approximate, for each trait, the values that produce a uniform distribution among the three classes. In order to build the classifiers, different subsets of the attributes shown in Table 2 were considered in order to find out which of them lead to the best results. Regarding the 5-class model, records were split up into five categories, trying that the number of records in each category were similar for each trait, as it was done for the 3-class model. In this case, the categories are: very high, high, medium, low and very low. Classifiers were trained for both options, and the results obtained are presented in the next section.

6. Results

Given the interaction data from a certain user, the personality classifiers predict to which class the user belongs, for each of the five personality traits. Classifiers were trained using a 3-class model and a 5-class model. The results obtained for both models are presented next.

6.1. Personality prediction with 3-class model

The first scenario corresponds to the 3-class model. The best performance was achieved when using all the attributes that appear in Table 4. The accuracy of the classifiers built for each trait is presented in Table 5. As it can be seen, the accuracy is higher than 70% for all traits, considering a 99% probability confidence interval. Sociability is the trait that can be predicted more exactly: near 72% accuracy.

Finally, the following rule is an example of knowledge represented by these classifiers: "*If a user started using Facebook less than* 3 *months ago, has more than* 1213 *friends, more than* 65 *'active friends' during the last month and a monthly average of more than* 16 *'active friends', then he has high activity*". Interestingly, these results did not change when the infrequency dimension was taken into account to remove suspect records.

6.2. Personality prediction with 5-class model

Regarding the 5-class model, the best performance was obtained for classifiers using the same variables than the previous case, and the accuracy results are shown in Table 6. Logically, the more classes the model has, the worse performance the

Table 5
Classification accuracy with 3 possible classes.

Scale	Accuracy	Confidence interval (99%)
Activity	70.44%	(69.31–71.55)
Aggression-Hostility	71.43%	(70.31–72.53)
Sociability	71.87%	(70.75–72.96)
Impulsive Sensation Seeking	71.45%	(70.33–72.55)
Neuroticism-Anxiety	70.05%	(68.91–71.16)

Table 6
Classification accuracy with 5 possible classes.

Scale	Accuracy	Confidence interval (99%)
Activity	62.53%	(61.34–63.71)
Aggression-Hostility	64.73%	(63.55–65.89)
Sociability	64.28%	(63.10–65.45)
Impulsive Sensation Seeking	63.18%	(61.99–64.35)
Neuroticism-Anxiety	63.39%	(62.20–64.56)

Table 7
Classification accuracy with 5 possible classes when errors between adjacent classes are not considered.

Scale	Accuracy	Confidence interval (99%)
Activity	79.98%	(78.98–80.94)
Aggression-Hostility	82.82%	(81.87–83.73)
Sociability	81.92%	(80.81–83.03)
Impulsive Sensation Seeking	79.87%	(78.87–80.84)
Neuroticism-Anxiety	80.87%	(78.89–81.82)

Table 8
Comparison of classification accuracies.

TP2010 ($n = 20{,}988$)	Weka 3 classes	Weka 5 classes	Weka 5 classes*	R 3 classes
Activity	70.44%	62.53%	79.98%	39.15%
Aggression-Hostility	71.43%	64.73%	82.82%	40.00%
Sociability	71.87%	64.28%	81.92%	40.97%
Impulsive Sensation Seeking	71.45%	63.18%	79.87%	41.13%
Neuroticism-Anxiety	70.05%	63.39%	80.87%	37.15%

corresponding classifiers would get. In this case, about 6–7% of precision is lost. Anyway, this 5-class model would be more beneficial for those needing a finer classification of the users.

In some applications it could be the case that, even if a finer classification is preferred, not all the classification errors have the same consequences. For example, it would be different to classify a user with *very low activity* as a user with *low activity* than to classify him as a user with *high activity* or *very high activity*. In the context of intelligent systems, it is probable that adaptation decisions for *very low activity* and *low activity* would not be too different; obviously, the user would benefit more from adaptation intended for *very low activity*, but would also benefit from adaptation for *low activity* (contrary to what would happen with adaptation for *high activity* or *very high activity*). If that were the case in the application context in which knowing the user personality is needed, it would be interesting to calculate the error of the classifier without considering "neighbor errors". Table 7 shows the classification accuracy for the 5-class model when errors between adjacent classes are not considered.

Obviously, accuracy in this case (around 80% in all the traits) is higher than that in which all the errors are considered. More interestingly, this accuracy is also higher than that in the 3-class model. For this reason, we think that this classifier represents a good trade-off between accuracy and fine-grained information. Table 8 presents a summary of the results obtained for each classifier, including Fisher discriminant. As it can be seen, the results when using Fishing discriminant are quite worse. We tried to use the same relevant variables used in the best results with classification trees, but only around 40% precision is achieved.

7. Conclusions and future work

In this paper, we have shown that data from online social networks profiles, specifically Facebook profiles, can be mined in order to build classifiers able to predict user personality. On one hand, we have developed TP2010, a Facebook application based on the ZKPQ-50-cc questionnaire to get the user personality (extended with questions related to infrequency). On the other hand, interaction data from 20,988 users, with known personality, have been analyzed to try to infer personality from these interactions. Different machine-learning techniques have been used, especially classification trees. Classifiers were trained using a 3-class model and a 5-class model. The results for each of these models have been presented. When using a 3-class model, the best performance was achieved by using all the attributes appearing in Table 4, such as number of friends, number of posts in his wall per month, or number of "active friends", among others. The accuracy of the classifiers built for each personality trait in this model is higher than 70% for all personality traits. Regarding the 5-class model, the accuracy of the classifiers decreases a bit: it is higher than 62% for all traits. This model would be useful if a finer user classification were needed. However, when considering five classes but ignoring errors between adjacent classes, the results are better: the accuracy ranges from 79.87% for *activity* to 82.82% for *Aggression-Hostility*. This latter model is a good trade-off between accuracy and fine-grained information. More sophisticated machine-learning techniques were also tried, but not significant improvement was obtained.

All these results show that it is possible to obtain information about the users' personality rather precisely starting from information about their interactions in OSNs, in a quite unobtrusive way, with no need of asking them to fulfill the corresponding questionnaires. Another advantage of this approach is the capability of inferring the user personality by analyzing his previous interactions with the social network offline; that is, there is no need to monitor users dynamically while interacting with the network. Regarding works with similar goals, our approach is based on analyzing data from actual Facebook profile and activity information rather than self-reports, with the aim of avoiding possible sources of inaccuracy. More important, most of related works are based on very limited (and often homogeneous) samples, and lead to some contradictory findings. Our main goal was to use a large and representative sample of Facebook users to be able to extract general conclusions. At the end, we were able to use a sample of 20,988 users, which is larger than those used in most previous works (several orders of magnitude): the most similar dataset size found is the one used in [21], where data from 5000 users were used to train automatic classifiers.

At the time of developing this work, the API of Facebook did not allow developers to extract the information about one user directly, but reading the walls of the user's friends, looking for that user's messages. This is a very time-consuming method, since some users have more than a thousand of friends, and the Facebook API limited the number of queries to its database to 100 every 10 minutes (at the time when collecting the data for this experiment). Nevertheless, being able to obtain data from more than 20,000 heterogeneous users (different countries, levels of study and so on), only by inviting to our friends to join TP2010, has been an unexpected success, being the classification accuracy obtained with this large dataset surprisingly high.

The approach presented here has some limitations. The data collected may suffer from a self-selection bias, as we only got data from users that are active on Facebook and have decided to join TP2010. However, other works found that, despite some trends regarding extraversion and openness to experience, personality factors were not so influential in terms of Facebook usage [34]. With this consideration, it is probably that our dataset were heterogeneous and unbiased regarding user personality. The distributions shown in Fig. 6, except that of *Impulsive Sensation Seeking*, supports that assumption as well.

Identifying the user personality and incorporating it in the user model can be very useful for intelligent and adaptive applications, in general [35]. Knowing the user personality can contribute to identify, for example, his potential preferences or needs in different contexts. This can be helpful, for example, in the domains of assistive technologies, recommender systems, or adaptive environments, among others. In the context of assistive environments, we have some experience on developing tools to guide people with cognitive limitations in the framework of the project ASIES ("Adapting Social & Intelligent Environments to Support people with special needs"). This project aims to enhance user independent living and focuses on providing training and supporting daily activities of people with special needs by means of adaptive systems. Young people with cognitive limitations come to the Universidad Autonoma de Madrid (UAM) to be trained for two years to achieve labor and social integration [36]. According to the expert trainers of this group, one of their main difficulties relates to emotion control. In fact, this is one of the main competences to acquire during their training. In this direction, knowing the user personality could help to better understand their emotional reactions in different situations. Moreover, adaptive feedback could be given to them according to their personality, taking special care of their potential emotion reactions in particularly difficult situations. For example, we have built a mobile assistant to guide users with special needs outdoors [37]. If they do not follow the instructions properly, the assistant gives them instructions to go back to their path. These instructions could be given in different ways depending on the user personality, with the goal to fit their potential personality and emotional needs at this situation.

In the educational area, it has been proved that the student personality, as well as the composition of workgroups regarding their members' personality (homogeneous vs. heterogeneous), affects to the group performance [38]. In this direction, [3] shows that student performance correlates with the personality dimensions, and [39] goes further, establishing relationships among personality, intelligence and the possibilities for good academic performance. Therefore, in adaptive training systems such as [40] the tasks to be accomplished at each time, as well as the feedback to be provided or even group formation, could vary for each user/group depending on the user personality, to better fulfill the user needs to this respect.

We are planning to combine the personality extraction as implemented in TP2010, with other techniques that also analyze different type of information to extract user personality, such as weblog text analysis [41]. We have some experience on analyzing natural language to extract information about the user emotions, both from the essays they elaborate [42] and from what they write in Facebook [43]. In this last work we also detect emotional changes. Knowing the user personality might help to understand these emotional changes better. Finally, another possibility would be to analyze other sources of user behavior, such as mouse movement patterns, which have shown to relate to some user features such as learning styles [44]. In summary, mining several sets of heterogeneous information together would be expected to lead to more accurate user personality classifications.

Acknowledgments

This work has been funded by the Spanish Ministry of Science and Education, project ASIES (TIN2010-17344) and by Comunidad Autonoma de Madrid, project E-Madrid (S2009/TIC-1650). Special thanks to all the users that have interacted with TP2010, who have made this study possible.

References

[1] Alfred Kobsa, Generic user modeling systems, in: The Adaptive Web: Methods and Strategies of Web Personalization, in: Lecture Notes in Comput. Sci., vol. 4321, Springer, Berlin, Heidelberg, 2007, pp. 136–154.
[2] Manuel de Juan Espinosa, Luis Fernando García Rodríguez, Nuestra personalidad: en qué y por qué somos diferentes, Biblioteca Nueva Universidad, Manuales y Obras de Referencia, Biblioteca Nueva, 2004.
[3] Tomas Chamorro-Premuzic, Adrian Furnham, Personality predicts academic performance: Evidence from two longitudinal university samples, J. Res. Pers. 37 (4) (2003) 319–338.
[4] Víctor Sánchez Hórreo, Rosa M. Carro, Studying the impact of personality and group formation on learner performance, in: JörgM Haake, Sergio F. Ochoa, Alejandra Cechich (Eds.), Groupware: Design, Implementation, and Use, in: Lecture Notes in Comput. Sci., vol. 4715, Springer, Berlin, Heidelberg, 2007, pp. 287–294.
[5] Anton Aluja, Jérome Rossier, Luis F. García, Alois Angleitner, Michael Kuhlman, Marvin Zuckerman, A cross-cultural shortened form of the ZKPQ (ZKPQ-50-cc) adapted to English, French, German, and Spanish languages, Pers. Individ. Differ. 41 (4) (2006) 619–628.
[6] L.R. Goldberg, T.K. Rosolack, The Developing Structure of Temperament and Personality from Infancy to Adulthood, Chapter The Big Five Factor Structure as an Integrative Framework: An Empirical Comparison with Eysenck's P-E N Model, Erlbaum, New York, 1994.
[7] Louise Barkhuus, Juliana Tashiro, Student socialization in the age of facebook, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10, ACM, New York, NY, USA, 2010, pp. 133–142.
[8] Mauro Cherubini, Alejandro Gutierrez, Rodrigo de Oliveira, Nuria Oliver, Social tagging revamped: Supporting the users' need of self-promotion through persuasive techniques, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI2010), Atlanta, GA, USA, April 2010.
[9] Christo Wilson, Bryce Boe, Alessandra Sala, Krishna P.N. Puttaswamy, Y. Ben Zhao, User interactions in social networks and their implications, in: Proceedings of the 4th ACM European Conference on Computer Systems, EuroSys '09, ACM, New York, NY, USA, 2009, pp. 205–218.
[10] Mitja D. Back, Juliane M. Stopfer, Simine Vazire, Sam Gaddis, Stefan C. Schmukle, Boris Egloff, Samuel D. Gosling, Facebook profiles reflect actual personality, not self-idealization, Psychol. Sci. 21 (3) (2010) 372–374.
[11] Samuel D. Gosling, Sam Gaddis, Simine Vazire, Personality impressions based on facebook profiles, Psychology (2007) 1–4.
[12] Inside facebook, available at http://www.insidefacebook.com, July 2012.
[13] Facebook newsroom, available at http://newsroom.fb.com/, July 2012.
[14] Tp2010, available at http://apps.facebook.com/test_de_personalidad, December 2009.
[15] Robert E. Wilson, Samuel D. Gosling, Lindsay T. Graham, A review of facebook research in the social sciences, Perspect. Psychol. Sci. 7 (3) (2012) 203–220.
[16] Fred Stutzman, Jacob Kramer-Duffield, Friends only: examining a privacy-enhancing behavior in facebook, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10, ACM, New York, NY, USA, 2010, pp. 1553–1562.
[17] Cliff A.C. Lampe, Nicole Ellison, Charles Steinfield, A familiar face(book): profile elements as signals in an online social network, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '07, ACM, New York, NY, USA, 2007, pp. 435–444.
[18] Amanda Nosko, Eileen Wood, Seija Molema, All about me: Disclosure in online social networking profiles: The case of facebook, Comput. Hum. Behav. 26 (3) (2010) 406–418.
[19] Rongjing Xiang, Jennifer Neville, Monica Rogati, Modeling relationship strength in online social networks, in: Proceedings of the 19th International Conference on World Wide Web, WWW '10, ACM, New York, NY, USA, 2010, pp. 981–990.
[20] myPersonality, http://mypersonality.org, July 2012.
[21] Yoram Bachrach, Michal Kosinski, Thore Graepel, Pushmeet Kohli, David Stillwell, Personality and patterns of facebook usage, in: Proceedings of the 3rd Annual ACM Web Science Conference, WebSci '12, ACM, New York, NY, USA, 2012, pp. 24–32.
[22] Jennifer Golbeck, Cristina Robles, Karen Turner, Predicting personality with social media, in: CHI '11 Extended Abstracts on Human Factors in Computing Systems, CHI EA '11, ACM, New York, NY, USA, 2011, pp. 253–262.
[23] Daniele Quercia, Michal Kosinski, David Stillwell, Jon Crowcroft, Our Twitter profiles, our selves: Predicting personality with Twitter, in: Proceedings of the IEEE Third International Conference on Social Computing (SocialCom), IEEE, October 2011, pp. 180–185.
[24] H.J. Eysenck, S.B.G. Eysenck, Personality Structure and Measurement, Routledge & K. Paul, 1969.
[25] Marvin Zuckerman, D. Michael Kuhlman, Jeffrey Joireman, Paul Teta, A comparison of three structural models for personality: The big three, the big five, and the alternative five, J. Pers. Soc. Psychol. 65 (4) (1993) 757–768.
[26] Marvin Zuckerman, Zuckerman–Kuhlman Personality Questionnaire (ZKPQ): An Alternative Five-Factorial Model, vol. 103, Hogrefe & Huber Publishers, 2002, pp. 377–396.
[27] Luis Fernando García Rodríguez, Baremos de la población española zkpq-50-cc, personal communication, Department of psicología biológica y de la salud, Universidad Autónoma de Madrid, 2010.
[28] Jiang Su, Harry Zhang, Full bayesian network classifiers, in: Proceedings of the 23rd International Conference on Machine Learning, ACM Press, 2006, pp. 897–904.

Please cite this article in press as: A. Ortigosa et al., Predicting user personality by mining social interactions in Facebook, J. Comput. System Sci. (2013), http://dx.doi.org/10.1016/j.jcss.2013.03.008

[29] Li Xiong, Subramanyam Chitti, Ling Liu, Mining multiple private databases using a kNN classifier, in: Proceedings of the 2007 ACM Symposium on Applied Computing, SAC '07, ACM, New York, NY, USA, 2007, pp. 435–440.
[30] Ian H. Witten, Eibe Frank, Data Mining: Practical Machine Learning Tools and Techniques, 2nd edition, Morgan Kaufmann Ser. Data Manage. Syst., Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2005.
[31] J. Ross Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.
[32] Ronald A. Fisher, The use of multiple measurements in taxonomic problems, Annu. Eugen. 7 (1936) 179–188.
[33] Michael J. Crawley, The R Book, 1st edition, Wiley Publishing, 2007.
[34] Craig Ross, Emily S. Orr, Mia Sisic, Jaime M. Arseneault, Mary G. Simmering, R. Robert Orr, Personality and motivations associated with facebook use, Comput. Hum. Behav. 25 (2) (2009) 578–586, Including the Special Issue: State of the art research into cognitive load theory.
[35] Peter Brusilovsky, Adaptive hypermedia, User Model. User-Adapt. Interact. 11 (1–2) (2011) 87–110.
[36] Prodis foundation, http://www.fundacionprodis.org/.
[37] Alberto G. García de Marina, Rosa M. Carro, Pablo Haya, Where should I go?: guiding users with cognitive limitations through mobile devices outdoors, in: Proceedings of the 13th International Conference on Interaccion Persona-Ordenador, INTERACCION '12, ACM, New York, NY, USA, 2012, pp. 46:1–46:8.
[38] Víctor Sánchez Hórreo, Rosa M. Carro, Studying the impact of personality and group formation on learner performance, in: Proceedings of the 13th International Conference on Groupware: Design Implementation, and Use, CRIWG'07, Springer-Verlag, Berlin, Heidelberg, 2007, pp. 287–294.
[39] Tomas Chamorro-Premuzic, Adrian Furnham, Personality, intelligence and approaches to learning as predictors of academic performance, Pers. Individ. Differ. 44 (7) (2008) 1596–1603.
[40] Estefania Martin, Rosa M. Carro, Supporting the development of mobile adaptive learning environments: A case study, IEEE Trans. Learning Technol. 2 (1) (2009) 23–36.
[41] Jon Oberlander, Scott Nowson, Whose thumb is it anyway?: classifying author personality from weblog text, in: Proceedings of the COLING/ACL on Main Conference Poster Sessions, COLING-ACL '06, Association for Computational Linguistics, Stroudsburg, PA, USA, 2006, pp. 627–634.
[42] Pilar Rodríguez, Alvaro Ortigosa, Rosa M. Carro, Extracting emotions from texts in e-learning environments, in: Leonard Barolli, Fatos Xhafa, Salvatore Vitabile, Minoru Uehara (Eds.), CISIS, IEEE, 2012, pp. 887–892.
[43] José M. Martin, Alvaro Ortigosa, Rosa M. Carro, Sentbuk: Sentiment analysis for e-learning environments, in: Computers in Education (SIIE), 2012 International Symposium, October 2012, pp. 1–6.
[44] Danilo Spada, Manuel Sánchez-Montañés, Pedro Paredes, Rosa M. Carro, Towards inferring sequential-global dimension of learning styles from mouse movement patterns, in: Wolfgang Nejdl, Judy Kay, Pearl Pu, Eelco Herder (Eds.), Adaptive Hypermedia and Adaptive Web-Based Systems, in: Lecture Notes in Comput. Sci., vol. 5149, Springer, Berlin, Heidelberg, 2008, pp. 337–340.

Please cite this article in press as: A. Ortigosa et al., Predicting user personality by mining social interactions in Facebook, J. Comput. System Sci. (2013), http://dx.doi.org/10.1016/j.jcss.2013.03.008