Human-centric Computing
and Information Sciences

RESEARCH

Open Access

CrossMark

Personality classification based on profiles of social networks' users and the five-factor model of personality

Alireza Souri[1*] , Shafigheh Hosseinpour[2] and Amir Masoud Rahmani[1,3]

*Correspondence:
a.souri@srbiau.ac.ir
[1] Department of Computer
Engineering, Science
and Research Branch, Islamic
Azad University, Tehran, Iran
Full list of author information
is available at the end of the
article

Abstract

Online social networks have become demanded ways for users to show themselves and connect and share information with each other among these social networks. Facebook is the most popular social network. Personality recognition is one of the new challenges between investigators in social networks. This paper presents a hypothesis that users by similar personality are expected to display mutual behavioral patterns when cooperating through social networks. With the goal of personality recognition in terms of analyzing user activity within Facebook, we collected information about the personality traits of users and their profiles on Facebook, hence we flourished an application using API Facebook. The participants of this study are 100 volunteers of Facebook users. We asked the participants to respond the NEO personality questionnaire in a period of 1 month in May 2012. At the end of this questionnaire, there was a link that asked the participants to permit the application to access their profiles. Based on all the collected data, classifiers were learned using different data mining techniques to recognize user personality by their profile and without filling out any questionnaire. With comparing classifiers' results, the boosting-decision tree was our proposed model with 82.2% accuracy was more accurate than previous studies that were able to foresee personality according to the variables in their profiles in five factors for using it as a model for recognizing personality.

Keywords: Facebook, Social networks, Personality, Data mining, NEO personality inventory

Introduction

Social Networking Sites (SNS) influence people's daily life with a remaining rapidity to become an important social stage for social interactions [1, 2]. Facebook, My Space, Twitter are successful considerable examples [3, 4]. Facebook, with more than one billion users as of October 2012, is the most prevalent social network among many Internet-based programs people use to communicate socially with others [5]. The main application of Facebook is to permit users to share their personal thoughts and stories to establish new relationships and preserve existing ones [6]. So Facebook gives users a great level of easiness in connecting and communicating with others than in the past. From this point of view, the relationship between using Facebook and psychosocial outputs attracts researchers' attention during the time [7, 8].

Souri *et al. Hum. Cent. Comput. Inf. Sci.* (2018) 8:24

Page 2 of 15

Personality is one of the interesting characteristics that can be considered for adaptation purposes. In the field of research, the personality of a person can be described as a set of specifications that compels a tendency on the behavior of the person; this tendency is unchanged through time and positions [9]. Having information about one's personality period gives hints about how he would react when encountering different situations. Detecting a user's personality can facilitate knowing his potential needs on different occasions [10]. Therefore, adaptive applications may take advantage of models of users' personality to adapt their behavior accordingly. There are a wide marketing, healthcare or recommender systems, among others.

On the other hand, we believe that the personality can be traced by investigating users' interaction in Online Social Networks (OSNs). In this context, an essential consideration would be whether virtual relationship and communication reflect user personality in the real world or offline life. For example, Barkhuus and Tashiro [11] and Cherubini et al. [12] state that Facebook which is currently the most frequently used OSN, is, apparently, a good similar example to the user's offline life. If a special user occasionally interacts with another user through the internet, it doesn't mean that they have many more real-life interactions. Additionally, it is proved that people use OSNs to maintain already existing relationships rather than establishing new ones (around 77% of Facebook users' social relationships in the real life are replicated in the virtual environment) [11, 12]. For example, these studies present that if a customer has many associations in a social network such as Facebook, he just cooperates by a minor percentage of them commonly [13].

In this context, we developed a movie recommender, a Facebook application intended to acquire evidence 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 with a hobby for users to display premier movies in Hollywood. This research efforts to discover guidelines for evaluating user personality in Facebook, without any requesting him to achieve exact personality assessments. This research uses machine learning methods for constructing user personality classifiers. This classification qualified based on 100 user's analysis for data collection of the application.

Most of the investigation is done to represent the existence of relationships between user personality and user interactions in social networks focus on investigating how single features correlate, on the average, with personal properties. In this approach, having data of a given user's interactions in social networks would make it possible to predict his personality, at least regarding some personality qualities.

The main contributions of the research are as follows:

- Developing a machine learning approach to predict the user personality through a personality test in the social network.
- Presenting a recommender system for friends relationships in social groups.
- Analyzing five kinds of personality features as their profile photos in the social groups.

The structure of this research is organized as follows: In "Related work" section, a brief literature review is presented. "Personality models and interaction data collection"

Souri *et al. Hum. Cent. Comput. Inf. Sci.* (2018) 8:24

Page 3 of 15

section shows data collection and personality models in the Facebook application. "Collected data description" section presents the data collection procedure from Facebook users. In "Data analysis" section data preprocessing is explained and "Building the personality classifier" section shows how personality classifiers are built. "Analysis of results" section provides the experimental results of personality prediction. Finally, "Conclusions and future works" section illustrates conclusions and future research.

Related work

This section presents a brief literature review for analyzing related studies in the personality analysis in social networks. For example, a personality trait that is called extroversion is positively correlated with both the size of a given individual's social networks and a number of social interactions that the individual is engaged in Asendorpf and Wilpers [14].

In a study that is done in 2010 [15], the participants are asked to report on their steady use of FB and also are asked to complete the NEO-PI-R as well as the Cooper Smith Self-Esteem Inventory [16]. The results manifest that extroverted people reported higher amounts of both SNS use and addiction to the internet.

In most of these research studies, generally the recent ones, information is collected by requesting the users to fulfill offline surveys. For example, Lampe et al. [17], Nosko et al. [18] and De Brabander and Boone [19] state that, while university students fill in most of the profile items (59%), a sample including university and non-university users only complete 25% of the information needed in profiles. More interestingly, they make it clear that if a user fills in his age and sentimental status can lead to inferring whether he keeps his profile public or private. Lampe et al. [17] proposed existing association rules between the number of groups and the total number of user's profile. Specifically, this correlation is bigger with reference data than others (place of birth, school, etc.), then comes contact data (sentimental state, address, etc.) and lastly favorable data (music, movies, books, etc.).

Another work [20] analyzes relationships between users from a different perspective: they try to investigate into who uses Facebook and the relationship between the Big Five, being shy, selfishness, being alone, and Facebook usage. The results indicate that the users of Facebook tend to be more extroverted and, but not permanent users and almost feeling socially alone, than nonusers. Also, the popularity of Facebook tendencies with specific structures are shown to vary that are results of certain characteristics, such as self-praise, feeling alone and shyness.

The other research study has used the data based on personality to examine the relationship between various types of Twitter users and personality, including popular users and influential ones [21]. This study has collected just 335 users that specify their Twitter accounts in their Facebook profiles.

However, the studies presented some interesting results: those users that are famous and influential are emotionally stable and have extrovert personalities (they got low score in the Neuroticism trait); famous users are highly 'imaginative' (and got high score in Openness), while influential users tend to be 'ordered' (and got high score in Conscientiousness).

Souri *et al. Hum. Cent. Comput. Inf. Sci.* (2018) 8:24

Page 4 of 15

A previously conducted study [22] indicate that individuals that are extrovert belong to more Facebook groups, but not necessarily in a relationship with more Facebook friends. They also come with the result that Neuroticism is not related to the posting of information that shows personality and those who are low in Neuroticism tend to put photos on their Facebook profiles. While Ross et al.'s study relies on self-reports by participants, in a follow-up study. Amichai-Hamburger and Vinitzky [23] illustrated that being extrovert have a positive effect on the number of friends, but not related to the use of Facebook groups, and individuals with high level of Neuroticism show more tendency to put their photos on Facebook than those individuals with low neuroticism.

Also, Nie et al. [24] presented a measurement approach to use personality visual attributes in the user's social features extracted in social media. Measuring the personality visual attributes are extracted into three challenges as follows: (1) feature selection (2) feature fusion and (3) feature absence. These challenges present a novel approach for evaluating personality distance between descriptive images in social media.

Huang et al. [25] proposed a personality characters analysis on effect online social associations. This research uses five personality theory to measure data collection of personality results. In addition, Bleidorn and Hopwood [26] presented a literature review for analyzing machine learning techniques for personality evaluation in social networks. In this review, some personality characteristics such as data collection, data extraction, and data prediction are analyzed. Finally, Lo Coco et al. [27] have presented a homogeneous classification for personality characteristics of a user's Facebook. This classification evaluates examined association rules between profiles of Facebook usage, relational characteristics and personality characteristics in online social interactions.

Personality models and interaction data collection

As it is mentioned above, the intention of this investigation is to acquire an approach to recognize user personality without asking them to response a specific questionnaire. Thus we use a model of personality that illustrates its structure. 117 volunteers from among Facebook users accepted to participate in this study. Their age ranged from 18 to 50. Putting incomplete data away, 100 users' profile information was the final data.

The traditional way of modeling personality structure is modeling factors. Three of the most famous models of structuring personality are the Eysenck three-factor model (that is known as the P.E.N. model, standing for Psychoticism, Extroversion, neuroticism) [28], Big Five model [29] and the Alternative Five [30]. There is no common consent about which model describes personality better. Nevertheless, it is usually accepted that their items or traits are frequently correspondents; the three of them present information about people's reactions to different situations, and they give information to decide which academic procedure is better considering different personalities.

In this work, the big five factors model is opted for measuring personality traits that classify personality of users into five agents: Conscientiousness, Extraversion, Agreeableness, Neuroticism, and Openness to Experience. Highly extraverted individuals are self-assured and warm, rather than calm and cautious. Agreeable individuals are coordinated and courteous. Conscientious individuals are organized and precise. Neurotic individuals are not prone to be emotionally resilient. Lastly, highly openness individuals are receptiveness and prefer innovation to the routine. The Big Five can state as

Souri *et al. Hum. Cent. Comput. Inf. Sci.* (2018) 8:24

Page 5 of 15

much of the variation in individuals' personalities as possible, using a small set of trait dimensions.

In order to evaluate the relationship between user behavior in social networks and personality, it needed to do classification technique so in order to classify personality to five classes that explained before its necessary to collect interaction data in users' profiles. To achieve this aim, data on profiles were collected through a Facebook application programmed using the Facebook API. This application distributed between contact which includes classmates and friends or coworkers and their friends. Participants were led to our application link, in which the aim of the research was explicated. Before the application could run, the application asked the user for permission to access their information in their profile like the number of friends or posts. The application involved the questionnaire NEO-FF-R-60 that they must answer them too. The application collects data on their profile until that moment which runs the application and stored them in a database. The collected data enumerated for each user that furnished to build the classifiers are Likes, Favorites, Language, Book, Job, Education, Sport, Activity, Game, Group, Cinema and movies, Music, Subscriber, Friends, Interests and hobbies, Links, TV shows, Question, Post, number of

Table 1 The components of Facebook used in research

Variable	Description	Type of variable
Likes	The number of posts which include videos or pages and etc. you like	Numeric
Language	The number of languages the person knows	Numeric
Book	The number of books that the person has verified	Numeric
Work	The number of jobs that a person has	Numeric
Education	The number of Level of education for example diploma, bachelor	Numeric
Sport	The number of teams and players of interest	Numeric
Activity	The number of latest added buddy, recent events, liked pages, will appear in this section	Numeric
Game	The number of memberships of the games available on Facebook	Numeric
Group	The number of memberships of the groups available on Facebook	Numeric
Movie	The number of favorite person's videos	Numeric
Music	The number of favorite person's music	Numeric
Friends	The number of the user's friends on Facebook	Numeric
Interested	It gives users the ability to create on Facebook a new list of their favorite pages or categories that work in a specific context and always follow them up	Numeric
Links	The number of links that his friends have put on the wall	Numeric
TV shows	The number of the person's favorite television series	Numeric
Note	The number of notes in the user's wall	Numeric
Question	The number of questions the user has placed in his wall	Numeric
Post	The total number of user posts since the moment of membership	Numeric
Number of only text	The number of posts that only use text	Numeric
Number of the photo in timeline	The number of photos the user posted in his profile	Numeric
Number of the photo without text	The number of posts that are only photos	Numeric
Newsfeed	The number of feedback from friends	Numeric
Show in timeline	The number of photos that have text	Numeric

Souri *et al. Hum. Cent. Comput. Inf. Sci.* (2018) 8:24

Page 6 of 15

Table 2 Distribution sex

Sex	Frequency	Frequency percentage
Female	47	47
Man	53	53
Total	100	100

only texts, amount of photos in timeline, number of photos without text, news feed, shown in timeline according to Table 1. To encourage users to take part in our research, we promised them to email them their personality test results. These data considered as input variables of a rapid miner in order to classification.

Modeling techniques were provided by the Rapid miner toolbox. Elements like age and gender were eliminated since they were not included to improve the classifier accuracy.

Collected data description

Among users who participated in the survey, with removing incomplete data, just 100 instances responded reliably and correctly to personality inventory and gave permission to access their profile. Elements like age and gender were shown just for giving general information. Otherwise, there was no use according to Table 2.

Data analysis

Today's real-world databases are highly sensitive to noisy, lost, and unsteady data due to their enormous size and their origins from different, heterogeneous resources [31]. Low-quality data leads to a low-quality outcome. There are numerous methods for preprocessing data. Data preprocessing can be used to eliminate outliers and noise, and solving unsteadiness [32].

To avoid missing information, the following mechanism is performed: in order to apply the suitable credible value to fill in the missing value, use the attribute mean for all samples belonging to the same class as the given tuple.

By observing the dataset, because of a low number of data set, we could find the data which their value was far from the average and known as noise and outlier data. Therefore ignoring method was applied because they might have an influence on the accuracy of the model and give inaccurate and unrealistic results.

Building the personality classifier

As it is declared above, most of the related works found in the psychology field struggle to find a correlation between the personality of users and their interactions in social networks through Statistical approaches. Whereas our probe focuses on seeking a criterion for forecasting users personality without asking them to fulfill the personality inventory.

Formerly, different machine-learning algorithms were used to establish classifiers of user personality. Techniques such as Naive Bayes [33], decision trees [34] and neural network [35], support vector machine (SVM) [36] were used to analyze the dataset [37, 38]. In this research, we applied some tricks for boosting classification accuracy. We focused on ensemble methods. A combination of classification was a composite model

Souri *et al. Hum. Cent. Comput. Inf. Sci.* (2018) 8:24

Page 7 of 15

Table 3 Adjusting modeling parameters

Setting parameters	Model	Abbreviations
–	Naïve Bayes	NB
Index split = 4	Decision tree	DT
Level of confidence for pruning = 0.25		
Minimum number of samples in each leaf = 2		
The maximum depth of the tree = 50		
The number of hidden layers = 1	Neural network	ANN
Learning rate = 0.3		
Momentum rate = 0.2		
Number of training cycles = 300		
Core rate = X*Y	Support vector machine	SVM
Constant = 0		
The number of repetitions = 10	Ada boost-Naïve Bayes	Ada-NB
The number of repetitions = 10	Ada boost-decision tree	Ada-DT
The number of repetitions = 10	Ada boost-neural network	Ada-NN
The number of repetitions = 10	Ada boost-support vector machine	Ada-SVM

that included a set of classifiers. After individual classifiers voted, a class label anticipator was returned by the combination based on the group of votes. Combinations were more accurate than their component classifiers. Ada Boost [39] was one of the famous combination methods which we used in the present study.

In all built models split validation method was used that 70% of data was used to train data and 30% of data was used to test the model. By comparing F-measure and accuracy of classifiers obtained by applying these techniques one of the classifiers was elected as a proposed model for each personality factor.

Data about user profiles and personality inventory were utilized to train personality classifiers. The first step included defining the kind of prediction and anticipation the classifiers were expected to do.

In order to model all parameters that must adjust in rapid miner software, they are demonstrated in Table 3.

Analysis of results

In this step, after running rapid miner the results analyzed to find out which classifiers for each personality traits is more accurate than others. After setting the parameters for each of the target variables, in order to find suitable classifier for each five-factor we run all eight classification technique for every five factors on the same testing data, therefore, a total of 40 models was run. The reason for this repetition was to find the most appropriate classifier for each of personality factors since it was possible that a classifier would respond better in one personality factor than other classifiers in another factor. For instance, if boosting-decision tree is selected as an appropriate classifier for extraversion, it would not be granted that is proper for consciousness factor.

The experimental results were examined and analyzed using typical procedures of F-measure and accuracy according to Tables 4, 5, 6, 7 and 8. For this purpose, the

Souri *et al. Hum. Cent. Comput. Inf. Sci.* (2018) 8:24

Page 8 of 15

Table 4 Accuracy, F-measure in extraversion factor

Model	Accuracy	F-measure
Naïve Bayesian	50	63.41
Neural network	76.67	85.71
Decision tree	93.33	96
SVM	88.89	80
Boosting-Naïve Bayesian	46.67	40
Boosting-neural network	80	88
Boosting-decision tree	93.33	96.15
Boosting-SVM	86.67	92.59

Table 5 Accuracy, F-measure in openness factor

Model	Accuracy	F-measure
Naïve Bayesian	60	25
Neural network	86.67	66.67
Decision tree	66.67	66.67
SVM	83.33	28.57
Boosting-Naïve Bayesian	66.67	44.44
Boosting-neural network	90	40
Boosting-decision tree	66.67	66.67
Boosting-SVM	86.67	50

Table 6 Accuracy, F-measure in consciousness factor

Model	Accuracy	F-measure
Naïve Bayesian	97.83	97.14
Neural network	96.67	Unknown
Decision tree	95.56	96.33
SVM	95.56	96.33
Boosting-Naïve Bayesian	97.83	97.14
Boosting-neural network	97.87	96.55
Boosting-decision tree	97.83	97.14
Boosting-SVM	97.78	96.55

Table 7 Accuracy, F-measure in agreeableness factor

Model	Accuracy	F-measure
Naïve Bayesian	66.67	79.17
Neural network	93.33	96.55
Decision tree	86.67	77.80
SVM	93.33	96.55
Boosting-Naïve Bayesian	86.67	92.86
Boosting-neural network	93.33	96.55
Boosting-decision tree	96.67	98.31
Boosting-SVM	93.33	96.55

Souri *et al. Hum. Cent. Comput. Inf. Sci.* (2018) 8:24

Page 9 of 15

Table 8 Accuracy, F-measure in neurotic factor

Model	Accuracy	F-measure
Naïve Bayesian	83.33	70.59
Neural network	80	62.5
Decision tree	76.67	36.36
SVM	76.67	58.82
Boosting-Naïve Bayesian	76.67	58.82
Boosting-neural network	86.67	60
Boosting-decision tree	86.67	77.78
Boosting-SVM	93.33	50

Fig. 1 Comparing the accuracy and F-measure of extraversion
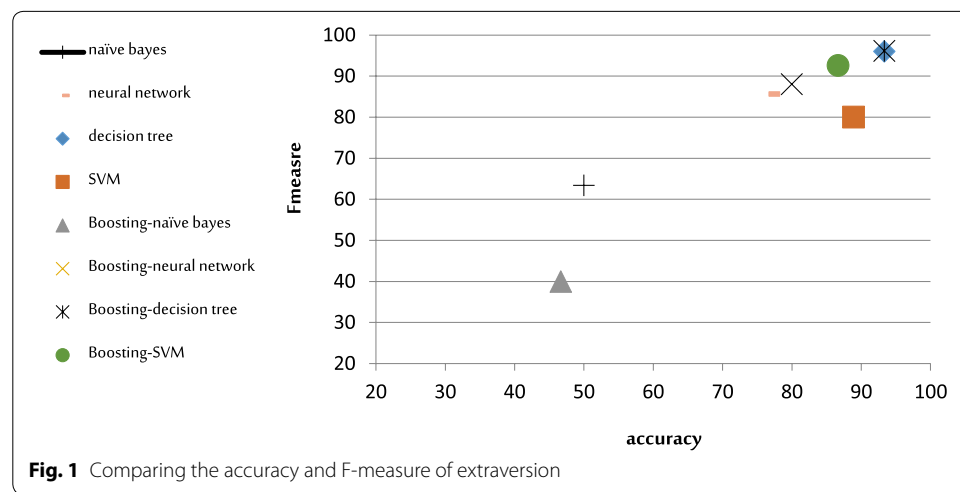
F-measure which is the result of combining two indexes of precision and recall is selected according to Eqs. (1) and (2).

$$
precision = \frac{TP}{TP + FP}
$$
$$
recall = \frac{TP}{TP + FN} = \frac{TP}{P}
$$
(1)

$$
F = \frac{2 \times precision \times recall}{precision + recall}
$$
(2)

As a result, in comparison with models, it is suitable to have a better accuracy and F-measure. For example, as shown in Table 4, the decision tree boosting with a precision of 93.33%. And the F error is equal to 96.15% as the best model for extraversion prediction, and Boosting-Naïve Bayesian model with a precision of 46.67% and the F measure of 40% is not recommended.

Souri *et al. Hum. Cent. Comput. Inf. Sci.* (2018) 8:24

Page 10 of 15

Fig. 2 Comparing the accuracy and F-measure of openness

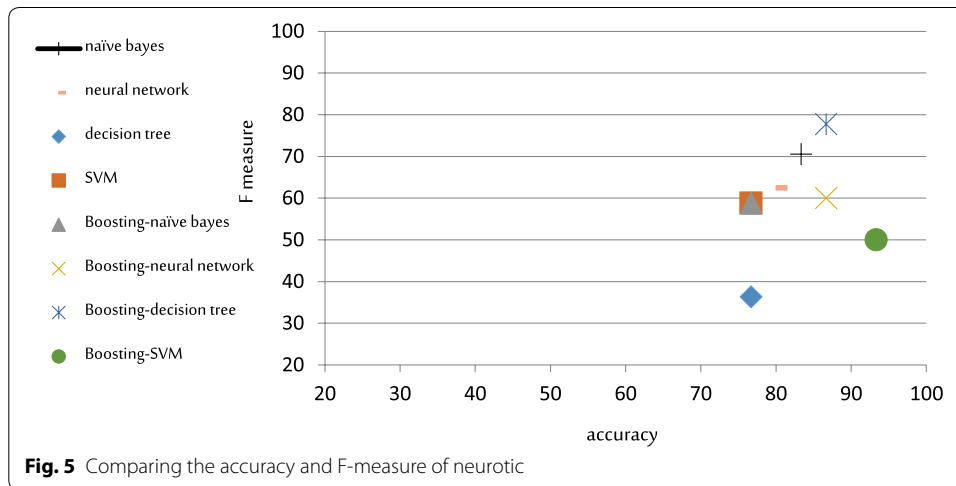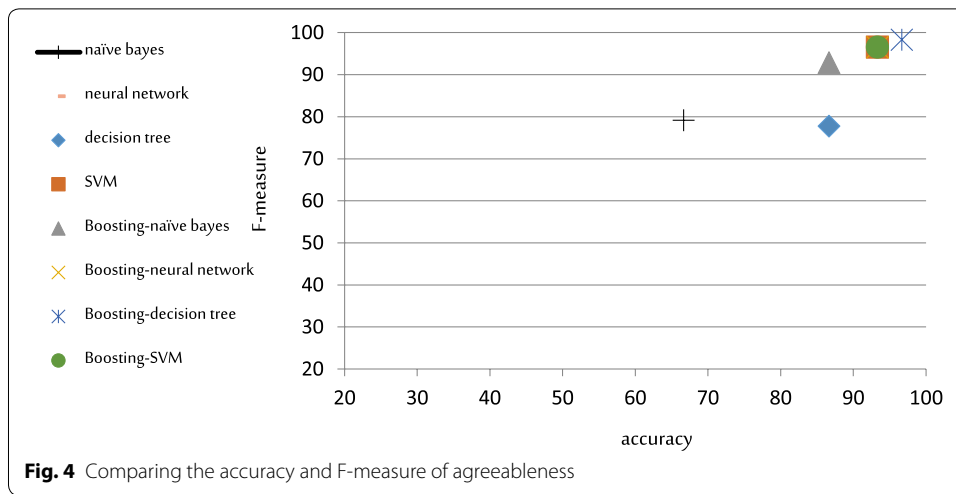**Fig. 3** Comparing the accuracy and F-measure of consciousness
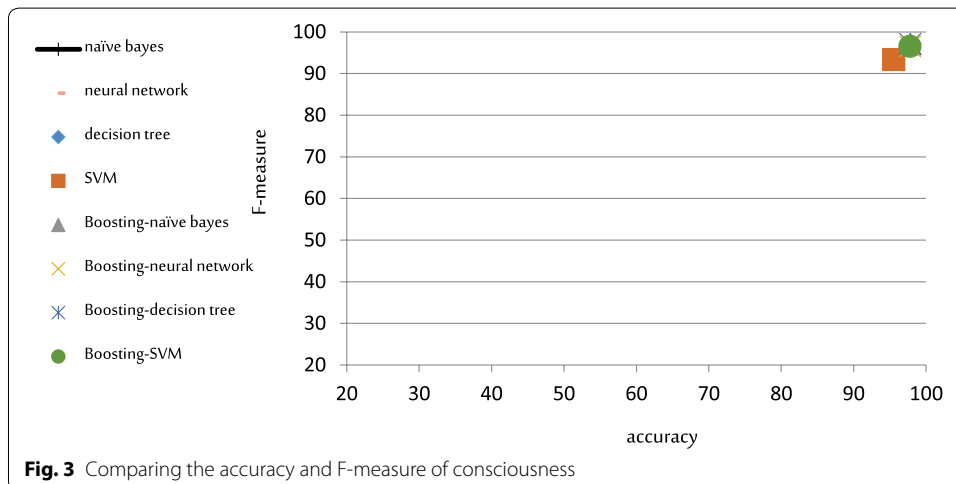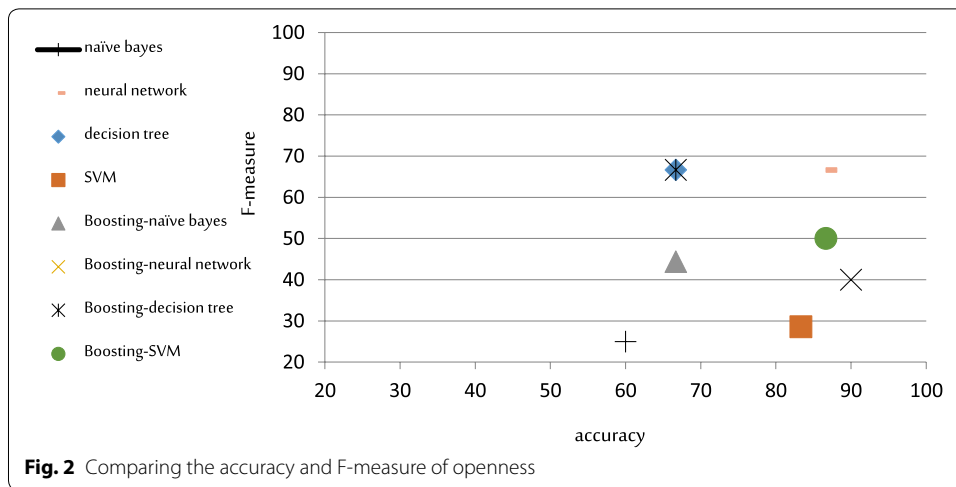
By transmitting the interaction data from the profile of a certain user, the personality classifiers would have the ability to predict which class of users belong to each of the five personality factors.

According to Table 4 and Fig. 1, boosting-decision tree with an accuracy of 93.33% and the F-measure to 96.15% could be selected as the proposed model to predict extroverts and Boosting Naïve Bayesian model with an accuracy of 67.46% and the F-measure to 40% was not proposed as an appropriate model.

According to Table 5 and Fig. 2, a neural network with an accuracy of 86.67% and the F-measure to 66.67% could be selected as the proposed model to predict openness and Naïve Bayesian model with an accuracy of 60% and the F-measure to 25% was not proposed as an appropriate model.

According to Table 6 and Fig. 3, boosting-decision tree and boosting-Naïve with an accuracy of 97.83% and the F-measure to 97.14% could be selected as the proposed model to predict in consciousness people and decision tree and SVM models with an

Souri *et al. Hum. Cent. Comput. Inf. Sci.* (2018) 8:24

Page 11 of 15

Fig. 4 Comparing the accuracy and F-measure of agreeableness

Fig. 5 Comparing the accuracy and F-measure of neurotic

accuracy of 95.56% and the F-measure to 93.33% was not proposed as an appropriate model.

According to Table 7 and Fig. 4, boosting-decision tree with an accuracy of 96.67% and the F-measure to 98.31% as the Boosting-Naïve Bayesian model to predict agreeableness people could be selected and Naïve Bayesian model with an accuracy of 66.67% and the F-measure was 79.17% which was not proposed as an appropriate model.

According to the Table 8 and Fig. 5, boosting-decision tree with an accuracy of 86.67% and the F measure to 77.78% as the Boosting-Naïve Bayesian model to predict neurotics people could be selected and decision tree model with an accuracy of 76.67% and the F-measure is 36.36% which was not proposed as an appropriate model.

According to the limited number of samples, these results are obtained, which may achieve other results with other examples and even more examples or by repeating

Souri *et al. Hum. Cent. Comput. Inf. Sci. (2018) 8:24*

Page 12 of 15

Table 9 Description of personality traits in two class

Personality traits	Class-one (low)	Class-two (high)
Extraversion	Dissociable, solitary	Warm, sociable
Openness to experience	Conventional	Strong, tenacious
Agreeableness	Sentimental and merciful	Reliable and reasonable
Neuroticism	Vulnerable	Tendency to stress and worry
Conscientiousness	Selfish and cautious	Organized and regular

these conditions for other people achieve a different result. After finding the right classifier for each personality trait, personality can be predicted with five factors. There is no correlation between personality traits so each one is predicted independently. It should be noted that each result for each trait can't have an influence on other results too. For instance, a person with low extraversion can't decide that he is in a low or high group of Conscientiousness. The other traits are also like that so each trait act autonomous. Considering the output of NEO-60, each user had five values associated with the five traits of the big five model.

It is worth mentioning, for our prediction goal, that it was not so significant to know the exact score of one user in each factor hence possible scores for each trait of personality categorized them into two classes: low and high. For example, a value of 2 for the extroversion trait, the classifier forecasted that the user had a low extroversion tendency. As each trait was interpreted independently, the dataset entered to software five times, each time containing one of the five factors as a new label attribute according to Table 9.

For example, to find the personality of the person in five factors, at first, variables of Facebook profile' user enter to the model that selected as a Better predictor. On the other hand, the person fills the NEO questionnaire and for each factor receive a score so based on the score earned, they divide to one of the classes high or low described in the previous section and as the target variable entered into the model. After running the modeling five times for each personality traits; finally, it can be seen that with a few percent accuracies, the model can predict the personality of the person correct.

So key novelty of this journal is that with the help of modeling and classifying individuals, we can predict the personality of users on social networks without having any history of them or even filling the questions of psychology. The following example can help to better understand according to Table 10. A user with this condition entered the model and it is expected that for instance, being extraversion with using boosting decision tree classifier predict this trait correctly up to 90%. The others in the same way.

Table 11 shows comparison results of our work to other similar samples.

Conclusion and future works

Ultimately, we tried to identify the personality of users indirectly without the use of traditional methods so user personality was not just recognized based on a questionnaire that gave expression whereas we could predict personality of users by their profile that was displayed during the time. Therefore, in this work, we got assistance

Souri *et al. Hum. Cent. Comput. Inf. Sci.* (2018) 8:24

Page 13 of 15

Table 10 Example of the variable of one user

Language	Book	Works	Education	Friend	Likes	Activities	Games	Groups	Interests
0	1	1	3	187	10	0	0	5	0
Movies	Music	Sub-scrib-ers	Links	TV	Note	Question	Posts	Show in time-line	Number of the photo in time-line
0	1	0	25	1	0	0	181	7	20
Extraver-sion	Agree-able-ness	Open-ness	Neuroti-cism	Conscien-tious-ness	News-feed	Sport	Number of the photo without text		Number of only text
1	1	2	1	2	1	1	20		6

Table 11 Comparing with related work

Ortigosa et al. [40]	Our work	Factors
Divide to 3-class	2-class divide to	Classes
Weka	Rapid miner	Software
Nearly = 70%	Nearly = 90%	Accuracy
Naive Bayes, K-nearest neighbors, classification trees and association rules	Neural networks, naïve Bayes, decision tree, SVM, boosting	Methods

from data mining which discovered fruitful information from a series of irrelevant data. Within results achieved, the boosting-decision tree was our proposed model that with 82.2% accuracy was more accurate than previous studies that were able to foresee personality according to the variables in their profiles in five factors. Furthermore, we intend to do for more examples, different nationalities and different conditions and compare them with our results. By knowing the personality, this model can be used for other purposes, such as the recommender system of friends and social groups, and even can be used for promotional purposes.

The proposed method can be amended in some facets so we are planning to predict personality with other techniques such as text mining via utilizing words in posts and comments in user's timelines to predict personality. Moreover, the other proposal is researching what kind of photos will be used by each of the five kinds of personalities as their profile photos. Ultimately for increasing accuracy of classifiers, we are eager to use fuzzy classification in future works.

Author details
[1] Department of Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran. [2] Department of Computer Engineering, Islamic Azad University, Tabriz Branch, Tabriz, Iran. [3] Computer Science, University of Human Development, Sulaymaniyah, Iraq.

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

References

1. Correa T, Hinsley AW, De Zuniga HG (2010) Who interacts on the Web?: the intersection of users' personality and social media use. Comput Hum Behav 26(2):247–253
2. Rathore S, Sharma PK, Park JH (2017) XSSClassifier: an efficient XSS attack detection approach based on machine learning classifier on SNSs. J Inf Process Syst 13(4):1014–1028
3. Kang YS, Lee H (2010) Understanding the role of an IT artifact in online service continuance: an extended perspective of user satisfaction. Comput Hum Behav 26(3):353–364
4. Souri A, Asghari P, Rezaei R (2017) Software as a service based CRM providers in the cloud computing: challenges and technical issues. J Serv Sci Res 9(2):219–237
5. Bonds-Raacke J, Raacke J (2010) MySpace and Facebook: identifying dimensions of uses and gratifications for friend networking sites. Individ Differ Res 8(1):27–33
6. Buettner R (2017) Predicting user behavior in electronic markets based on personality-mining in large online social networks. Electron Mark 27(3):247–265. https://doi.org/10.1007/s12525-016-0228-z
7. Lee JY, Kim HS, Choi EJ, Choi SJ (2013) Exploratory study on online social networks user from SASANG constitution-focused on Korean Facebook Users. Paper presented at the online communities and social computing, Berlin, Heidelberg
8. Song H, Zmyslinski-Seelig A, Kim J, Drent A, Victor A, Omori K, Allen M (2014) Does Facebook make you lonely?: a meta analysis. Comput Hum Behav 36:446–452
9. Kumar U, Reganti AN, Maheshwari T, Chakroborty T, Gambäck B, Das A (2017) Inducing personalities and values from language use in social network communities. Inf Syst Front. https://doi.org/10.1007/s10796-017-9793-8
10. Tama BA, Rhee K-H (2017) A detailed analysis of classifier ensembles for intrusion detection in wireless network. J Inf Process Syst 13(5):1203–1212
11. Barkhuus L, Tashiro J (2010) Student socialization in the age of facebook. Paper presented at the proceedings of the SIGCHI conference on human factors in computing systems
12. Cherubini M, Gutierrez A, De Oliveira R, Oliver N (2010) Social tagging revamped: supporting the users' need of self-promotion through persuasive techniques. Paper presented at the proceedings of the SIGCHI conference on human factors in computing systems
13. Wilson C, Boe B, Sala A, Puttaswamy KP, Zhao BY (2009) User interactions in social networks and their implications. Paper presented at the proceedings of the 4th ACM European conference on computer systems
14. Asendorpf JB, Wilpers S (1998) Personality effects on social relationships. J Pers Soc Psychol 74(6):1531
15. Wilson K, Fornasier S, White KM (2010) Psychological predictors of young adults' use of social networking sites. Cyberpsychol Behav Soc Netw 13(2):173–177
16. Coppersmith A (1984) Self-esteem inventories. Consulting Psychologists Press, Palo Alto
17. Lampe CA, Ellison N, Steinfield C (2007) A familiar face (book): profile elements as signals in an online social network. Paper presented at the proceedings of the SIGCHI conference on human factors in computing systems
18. Nosko A, Wood E, Molema S (2010) All about me: disclosure in online social networking profiles: the case of FACE-BOOK. Comput Hum Behav 26(3):406–418
19. De Brabander B, Boone C (1990) Sex differences in perceived locus of control. J Soc Psychol 49:311–320
20. Ryan T, Xenos S (2011) Who uses Facebook? An investigation into the relationship between the Big Five, shyness, narcissism, loneliness, and Facebook usage. Comput Hum Behav 27(5):1658–1664
21. Quercia D, Kosinski M, Stillwell D, Crowcroft J (2011) Our twitter profiles, our selves: predicting personality with twitter. Paper presented at the privacy, security, risk and trust (PASSAT) and 2011 IEEE third international conference on social computing (SocialCom)
22. Ross C, Orr ES, Sisic M, Arseneault JM, Simmering MG, Orr RR (2009) Personality and motivations associated with Facebook use. Comput Hum Behav 25(2):578–586
23. Amichai-Hamburger Y, Vinitzky G (2010) Social network use and personality. Comput Hum Behav 26(6):1289–1295
24. Nie J, Wei Z, Li Z, Yan Y, Huang L (2018) Understanding personality of portrait by social embedding visual features. Multimed Tools Appl. https://doi.org/10.1007/s11042-017-5577-x
25. Huang H-C, Cheng TCE, Huang W-F, Teng C-I (2018) Who are likely to build strong online social networks? The perspectives of relational cohesion theory and personality theory. Comput Hum Behav 82:111–123. https://doi.org/10.1016/j.chb.2018.01.004
26. Bleidorn W, Hopwood CJ (2018) Using machine learning to advance personality assessment and theory. Pers Soc Psychol Rev. https://doi.org/10.1177/1088868318772990
27. Lo Coco G, Maiorana A, Mirisola A, Salerno L, Boca S, Profita G (2018) Empirically-derived subgroups of Facebook users and their association with personality characteristics: a Latent Class Analysis. Comput Hum Behav 86:190–198. https://doi.org/10.1016/j.chb.2018.04.044
28. Cattell RB (1946) Personality structure and measurement. Br J Psychol 36(2):88–103
29. Halverson CF Jr, Kohnstamm GA, Martin RP, Halverson CF, Kohnstamm GA (2014) The developing structure of temperament and personality from infancy to adulthood. Psychology Press, New york

Souri *et al. Hum. Cent. Comput. Inf. Sci.* (2018) 8:24

Page 15 of 15

30. Zuckerman M, Kuhlman DM, Joireman J, Teta P, Kraft M (1993) A comparison of three structural models for personality: the Big Three, the Big Five, and the Alternative Five. J Pers Soc Psychol 65(4):757
31. Norouzi M, Souri A, Samad Zamini M (2016) A data mining classification approach for behavioral malware detection. J Comput Netw Commun 2016:1
32. Yi G, Kim H-W, Park JH, Jeong Y-S (2018) Job allocation mechanism for battery consumption minimization of cyber-physical-social Big Data processing based on mobile cloud computing. IEEE Access 6:21769–21777. https://doi.org/10.1109/ACCESS.2018.2803730
33. Su J, Zhang H (2006) Full Bayesian network classifiers. Paper presented at the proceedings of the 23rd international conference on machine learning
34. Ma L, Destercke S, Wang Y (2016) Online active learning of decision trees with evidential data. Pattern Recognit 52:33–45
35. Liu C, Shu T, Chen S, Wang S, Lai KK, Gan L (2016) An improved grey neural network model for predicting transportation disruptions. Expert Syst Appl 45:331–340
36. Sady CC, Ribeiro ALP (2016) Symbolic features and classification via support vector machine for predicting death in patients with Chagas disease. Comput Biol Med 70:220–227
37. Park JH (2018) Practical approaches based on deep learning and social computing. JIPS 14(1):1–5. https://doi.org/10.3745/JIPS.00.0009
38. Souri A, Hosseini R (2018) A state-of-the-art survey of malware detection approaches using data mining techniques. Hum Centric Comput Inf Sci 8(1):3
39. Wang R (2012) AdaBoost for feature selection, classification and its relation with SVM, a review. Phys Procedia 25:800–807
40. Ortigosa A, Carro RM, Quiroga JI (2014) Predicting user personality by mining social interactions in Facebook. J Comput Syst Sci 80(1):57–71