This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2017.2751602, IEEE Transactions on Affective Computing

Profiling Players Using Real-World Datasets: Clustering the Data and Correlating the Results with the Big-Five Personality Traits

Zahid Halim, Member, IEEE, Muhammad Atif, Ahmar Rashid, and Cedric A. Edwin

Abstract—Computer games provide an ideal test bed to collect and study data related to human behavior using a virtual environment having real-world-like features. Studies regarding individual players' actions in a gaming session and how this correlates with their real-life personality have the potential to reveal great insights in the field of affective computing. This study profiles players using data collected from strategy games. This is done by taking into account the gameplay and the associations between the personality traits and the subjects playing the game. This study uses two benchmark strategy game datasets, namely, *StarCraft* and *World of Warcraft*. In addition, the study also uses the *Age of Empire-II* game data, collected using 50 participants. The IPIP-NEO-120 personality test is conducted using these participants to evaluate them on the Big-Five personality traits. The three datasets are profiled using four clustering techniques. The results identify two clusters in each of these datasets. The quality of cluster formation is also evaluated through the cluster evaluation indices. Using the clustering results, the classifiers are then trained to classify a player, after a gameplay, into one of the two profiles. Results show that the gameplay can be used to predict various personality features using strategy game data.

Index Terms— Profiling players, personality prediction, cluster analysis, dispositional theory

1 INTRODUCTION

RESEARCH in the area of human behavior pertaining to human emotion recognition and other related affective phenomena is gaining momentum. Profiling of humans is a computational technique used to categorize individuals with common behavior in the same group and those having diverse traits in different groups [1] [2] [3]. Personality tests are an established mean of behavioral research in many fields [4]. They are also used in scientific settings to support analytic decisions. Minnesota Multiphasic Personality Inventory (MMPI) and the Eysenck Personality Questionnaire (EPQ) are the pioneer tools to study personality [5]. Before the development of an indigenous profession in a few countries, human behaviorists (also referred to as, psychological consultants) were often hired from abroad to serve the small clientele comprising mainly of large multinational companies. Later, due to their wide utility, direct personality assessment methods evolved. Today, companies use these methods to hire appropriate individuals, sports officials use them for recruiting team players, and educationists evaluate students' learning in the affective domain through these. The study of human personality finds its utility in a diverse set of fields including, but not limited to, employment test [6], individual and relationship counseling [7], trauma treatment [8] and career counseling [9]. However, there are two major criticism for deploying personality tests: criterion-related validity of personality measure and faking. Criterion-related validity of personality measures suggests that often personality tests are deployed to either measure the wrong outcome or the correlations developed are flawed [10]. Personality tests are also susceptible of being faked by respondents, because more than often the desirable answer is apparent and obvious. Personality tests mainly rely on the honesty of the respondent and her/his self-awareness. Moreover, faking, also known as Hawthorne effect [11], depends on the situation as well as the interaction of the respondent with the situation. For instance, the validity of personality tests conducted on employees will be significantly different from that of the conducted on potential candidates for the employment. Furthermore, the personality tests are more influenced by the motivational factors than the cognitive abilities of the participant [12]. However, the indirect personality assessment also suffers from issues like larger amount of time required and mapping of the assessment features to the personality traits [13]. Substantial amount of effort has been attributed to the development of various personality tests. These tests can be grouped into three strategies: inductive, deductive, and empirical [14]. In all these strategies, the subject taking the personality test is aware of the fact that her/his behavior is being monitored. This may cause the subject to fake all or few of the traits. Although some of these tests may have the

This work was supported by the GIK Institute graduate program research fund under GA1 scheme.

- *Z. Halim and A. Rashid are with the Faculty of Computer Science and Engineering, Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Topi, Pakistan, 23460. E-mail: {zahid.halim, ahmar.rashid}@giki.edu.pk.*
- *M. Atif is with the Department of Computer Science, National University of Computer and Emerging Sciences, Pakistan, 38000. E-mail: atif.muhammad@nu.edu.pk.*
- *Cedric A. Edwin is with the Department of Management Sciences, Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Topi, Pakistan, 23460. E-mail: aimal@giki.edu.pk.*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2017.2751602, IEEE Transactions on Affective Computing

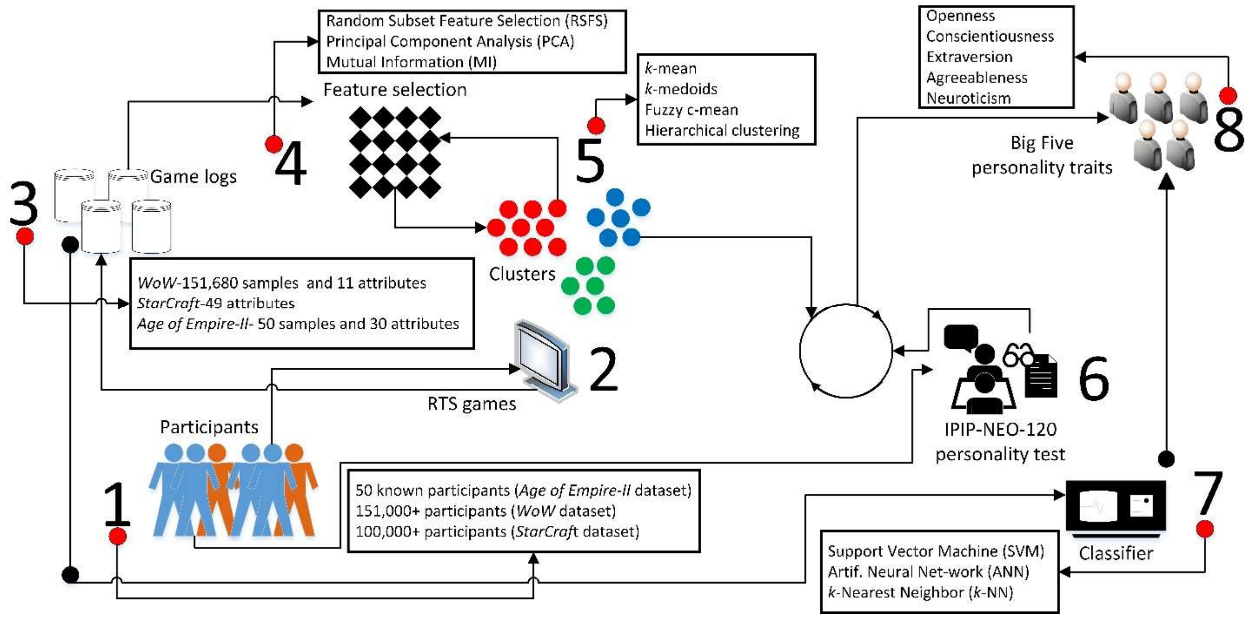


Fig. 1. Proposed system structure.

mechanism to identify the Hawthorne effect. Additionally, these personality tests are conducted under controlled conditions similar to a laboratory experiment. This has the disadvantage of not providing the subjects an environment where they could manipulate the environment's variables and based on this, make a choice using their preferences [15]. Subjects are asked to imagine a scenario, and based on the hypothetical situation or a previous similar experience, they opt for an answer. All this adds to the imprecision that these direct personality tests may have. There are many indirect personality assessment methods [16] but they also comprise of questions. However, there are few other factors that make these personality tests useful, such as the maturity of various models over the period of decades, time bounded, and convenience.

It will be interesting to see how a group of subjects is classified in one of the personality profiles using their actions in the real-world. This will require long durations for observation and making an assessment. For convenience, in such experiments a virtual environment may be useful, however, this will have the same disadvantage mentioned earlier, i.e., the subject being aware of the experiment. Computer games provide a rich mapping of the real-world scenarios to their virtual equivalent. This is evident from the fact that these have been used for in an assortment of fields to gain better performance. For example, Coller et al. [17] utilize video games to teach mechanical engineering students numerical methods, Koenig et al. [18] use gaming environment for patient rehabilitation, and Drettakis et al. [19] employ gaming environments for architecture and urban planning. There are many genres of games available, including predator/prey games, platform games, board-based games and real-time strategy games (RTS) [20]. Among these, RTS games provide a real-world-like environment and scenarios. This type of game is an effort to pretend the choices and procedures intrinsic to some real-world conditions. Mostly the rules are selected to reproduce what the real-world significances would be of each player's action and choice. Many strategy games provide multiple paths to play. Player behavior in reality is inclined by personality characteristics. Player's personality and behavior are correlated in the strategy games as if they are in the real-world [21]. One may be able to predict player's personality from her/his style of playing the game. Subjects may be asked to play RTS gaming session and their actions be recorded during this time. Later, these recorded features can be mapped to specific personality factors, ultimately classifying the subjects in various personality profiles. This will help in a real assessment of the subject's personality using a test bed that keeps the system peculiarities hidden from the subjects and allows them to exhibit their true behavior in real-world-like situations. However, mapping of the features recorded during the gaming session to the various personality traits will need an in-depth study with logical reasoning. This may be done by adding customized features to the RTS game.

1.1 Contribution

This work performs the profiling of human subjects using computer games. For recording various real-world-like features, two benchmark strategy game datasets, i.e., *StarCraft* and *World of Warcraft (WoW)*, are used. Since the information available in these datasets is of unknown subjects who cannot be engaged in further experiments, an additional dataset is also recorded using *Age of Empire-II* game utilizing 50 participants. For the profiling task, clustering is utilized. Different clustering techniques may produce dissimilar clustering formations for the same data. To avoid this and for generalization of the results, this work uses four clustering techniques, namely, *k*-means, *k*-medoids, fuzzy *c*-mean, and hierarchical clustering. Datasets from the three games, *StarCraft, WoW,* and *Age of Empire-II*, contain 57, 10, and 30 features respectively. To study the most and the least important features and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2017.2751602, IEEE Transactions on Affective Computing

their effect on the clustering formation, three feature selection techniques are used, which are Mutual Information (MI), Principal Component Analysis (PCA) and Random Subset Feature Selection (RSFS). The clustering results are then evaluated using three cluster evaluation indices: Davies–Bouldin Index (DBI), Dunn Index (DI), and Silhouette Coefficient (SC). This work later uses classifiers to predict an unknown player's profile/group, given her/his gameplay data. The classifiers used for this purpose are Support Vector Machine (SVM), Artificial Neural Network (ANN) and *k*-Nearest Neighbor (*k*-NN). For mapping the extracted profiles through clustering to the Big-Five personality traits, IPIP-NEO-120 personality test is taken by the same 50 participants who played *Age of Empire-II*, and their scores are recorded. The proposed system can be used to categorize people in various personality profiles using data recorded through an RTS game. Fig. 1 shows the complete system framework. The overall process to obtain the results include eight steps. During Step-1 and Step-2 human subjects are engaged for data collection pertaining their in-game behavior. For this, the RTS games are utilized. The next step is to preprocess this data and store it in a logfile. Step-4 covers studying the various features of the datasets and identify the least and the most important attributes. Step 5 involves profiling the data using a number of clustering techniques. Steps 4 and 5 are repeated multiple times to study the effect of various features on the resultant clustering. Step 6 creates the ground truth by engaging a set of known participants in the IPIP-NEO-120 personality test. The data from this step is then utilized in Step 7 to train the classifiers. Finally, Step-8 classifies the unknown participants, and maps each of them to one of the profiles created in Step 5. Novelty of this work is the proposed comprehensive framework that utilizes the in-game behavior of the participants using RTS games to assess personality. There are a few approaches proposed in the past (Section 2) that utilizes computer games to evaluate personality, however, these are limited in scope, lack generalization capability, and ground truth is not available. This proposal is extensively evaluated using three RTS games' data, a know set of participates are utilized for the ground truth creation and an extensive evaluation of the framework is carried out using ten machine learning approaches (four clustering techniques, three feature selection methods, and three classifiers).

This work contributes to the field of affective computing by presenting a machine learning-based framework that indirectly evaluates personality using a participant's gameplay data. Extensive experiments using benchmark datasets and well-known personality test make this framework robust. The proposed framework can be utilized to assess personalities using any RTS gameplay data. There are few existing approaches (mentioned in Section 2) that address a somewhat similar issue, however, they either focus on the influence of culture on personality or are dependent only on dataset from a single computer game. This work is evaluated on three RTS games that makes it generalizable.

The paper is organized as follows. Section 2 reviews the related work in the domain of profiling and direct/indirect personality assessment. Section 3 explains the dispositional theory. Section 4 explains the methodology. Section 5 lists the experiments and their results. Section 6 discusses the results and mentions some limitations of the proposed approach. Finally, Section 7 concludes the paper with suggestions on future research in this area.

2 Related Works

Profiling people and the personality assessment attracts interest from an assortment of fields. Previous work exists that utilizes diverse techniques to directly or indirectly understand the behavior of subjects under various conditions. This section covers such related literature.

Worth et al. [22] explore the relation between general people behavior in video games and their personality traits. A questionnaire is distributed among a group of university students to analyze their behavior in a video game environment. The same group also completes the measures of personality and psychopathic traits. Using feature selection, their work identifies four factors aggressing, winning, creating, and helping. These factors are then mapped to different items in the questionnaire to correlate these with the personality traits. The results suggest that the players' behavior in the video games resembles that of the real-world. The questionnaire used in [22] is general in nature and can be used for a variety of video games. Experiments in Bialas et al. [23] show that the culture influences a player's play style. Three play style groups are identified using the Hofstede's cultural dimension theory [24], including: competitiveness, cooperation, and tactical choices. Players are expected to show variation in these groups based on their cultural background. For extracting various statistical features *Battlefield 3* game is used. Analysis of Variance (ANOVA) test is used to categorize the player's nation-wise play style into cooperativeness and competitiveness. Results from Multivariate Analysis of Variance (MANOVA) test show that the national culture accounts for variance of 5.6% and 4.2% in competitive and cooperative play styles respectively. Comparison also shows that the German and Swedish players have more cooperative behavior than the players from the United Kingdom (UK) and the United States (US). Burris et al. [25] explore the relationship between current religious status, childhood play style and adult personality. The study is conducted using 431 undergraduate Canadian students. Five Factor Model (FFM) result indicates that lifelong spirituals (i.e., those individuals whose identities remained unchanged from childhood) are more agreeable and conscious and less open to experience. Furthermore, the study suggests that the childhood openness to practice is related to an inclination for imaginary play and the grownup openness to familiarity has been found to be related to apostasy. Brown et al. [26] study the relation between gambling play style and personality. A group of 64 players is used in the study to be classified in two main dimensions: tight or loose and aggressive or passive. For this, the study uses observations from on-table poker behavior. Personality traits of the players having superstitious beliefs are compared to those who think otherwise. The data is collected using a mobile phone, where the frequency of hands played pre-flop is used to classify a tight/loose play and the frequency of raises made pre-flop is used to classify a passive/aggressive play. The results show a higher mean age among loose compared to tight players. Symborski et al. [27] investigate the prediction of a player's real-world behavior using avatars data in the massively multiplayer online role-playing games (MMORPGs). The games *Guild Wars,* and *Aion* are used for collecting the participants' data. Ground truth is established using the demographic information. The experiment consists of three phases (1) an initial laboratory session by the participants to develop ground truth, (2) recording the quantitative data using MMORPG, and (3) the creation of a prediction system. The prediction model is created using the discriminant analysis. Overall, the results in [27] support the concept of identifying real-world features of the players using gameplay, like, gender, age, education level, extraversion level and submissive ideology with an accuracy of 83%, 70%, 66%, 68%, and 65%, respectively. Work in [28] investigates the relation between age and play style. A survey is conducted to collect data from 13376 *Battlefield 3* players. Around 60 play-style variables are studied to establish the relationship, where it is found that 46 play-style variables can explain 45.7% of variance in age. The data is collected over a period of six weeks through a website. It is found that older players kill/die less in a game session, and they score less and focus on winning. It is also found that the age influences class and vehicle preferences. The study suggests that the player models can use age as a predictive variable to help design a better game. Bean et al. [29] explore a link between Big-Five Inventory (BFI) and *WoW* players. In their study 1210 participants were engaged who aged above 19 years. Using MANOVA, a connection between BFI personality characteristics and play style is established. The game playing preferences and other demographic information about the players is collected using a questionnaire. The questions related to play style include: Player Verses Players (PVP), Role Playing (RP) or Player Verses Environment (PVE) and specialty of character. Results show that the players opting as PVE in the game are less agreeable, not open to experiences, and conscientious, compared with the players choosing other styles of play, i.e., PVP and RP. The results support the play style followed in *WoW* to be related to the real-life personality traits. Liu et al. [30] present a study comparing the inventory questions used in standard BFI and the open-ended questions to access a virtual agent's personality. The comparison is based on two personality traits, i.e., extraversion and neuroticism. Gesture and posture is used to convey clues about the agent's personality. Data is collected using 59 and 84 subjects in two experimental settings where they are shown a virtual agent's clip and asked questions related to its personality traits. Results show that inventory questions enable to differentiate between emotional stability and agreeableness. However, open-ended questions enable the participants to perceive the high stability agent as extraverted and the low stability agent as disagreeable. The work in [31] examines personality traits represented by the Big-Five model. The data is collected using a thorough analysis of the facial expressions of emotion in the vlog scenario (blog with postings in video form). Regression analysis is performed to predict the personality traits using facial expression cues. These cues include the percentage of time each facial expression of emotion is active, the frequency with which an active segment appears in the facial expression signal, the average duration of the facial expression segments and the percentage of time the signal is active in segments. Results show that these cures are only able to predict extraversion impressions. Bostan et al. [32] present a framework that predicts the behaviour of both player and non-player characters (NPC) in a computer game. They take into consideration 29 psychological needs proposed by Murray [33] in developing a unified personality model. The motivational variables are suitable for creating realistic NPCs in a computer game. Wright et al. [34] study a range of widely-studied models of behavioral game theory for evaluation and then suggest modifications to the best-performing model which is suitable for prediction of initial play by humans. Some work on profiling without the involvement of personality traits can be seen in [35] and [36].

The works closely relate to this proposal are [22], [23], [26], [27], [29], and [31]. The majority of the participants used in [22] are not frequent video game players. The data is gathered based on the in-game behavior, frequent game players as participants would have resulted in a more reliable data. The questionnaire used in [22] is of a general nature and is applicable to any category of the video games. However, in our case the subjects are chosen such that they are either frequent game players or have motivation towards the task. Additionally, the current study focuses only on the RTS category of games for data collection. The study in [23] investigates the cultural influence on play style using PsyOps database [37]. The uniform distribution of players from various cultural backgrounds is not guaranteed in the database, this may result in biasing the findings towards cultures having a larger number of samples. The work in [26] uses poker style to correlate it with the FFM that requires focused sessions for observing the in-game behavior. The work in [27] predicts players' characteristics using avatar observations. The ground truth for personality measurement is created using authoritarian ideology and demographics reported by the subjects. However, in our case the ground truth is generated using the IPIP-NEO-120 personality test. The approach in [29] uses the FFM to understand the personality assessment of *WoW* players. The study focuses on only one RTS game, i.e., *WoW* for drawing results. Additionally, the number of role-playing participants is also less which makes it difficult to correlate the personality traits of role-playing players with their gameplay. This

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2017.2751602, IEEE Transactions on Affective Computing

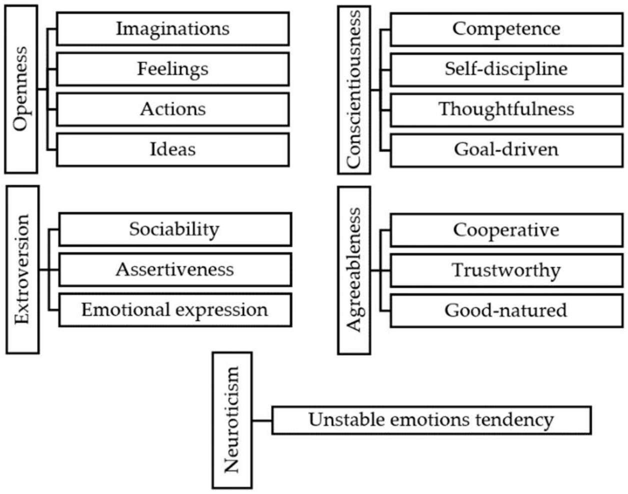


Fig. 2. The Big-Five personality traits and their details.

study has a balanced distribution of participants and three RTS games are used to extract results. The study in [31] uses one vlog per user for experiments. In our case, we use the average of multiple runs using the *Age of Empire-II* game. The studies in [22] and [29] depend on the questionnaire for correlating subjects' behavior with the personality traits. However, in this proposal real-life data is gathered using three RTS games and correlated with the personality traits extracted using the IPIP-NEO-120 personality test. Table 1 lists the key features of the current proposal and closely related previous works.

3 Dispositional Theory

Dispositional theory, also called the trait theory, is a domain in psychology used to study human personality. The theory is primarily interested in measuring traits that are defined as habitual patterns of behavior [38]. Dispositional theory assumes that there are a finite number of traits that can be measured and observed. The two key approaches used in the dispositional theory for measuring traits include Eysenck Personality Questionnaire, (EPQ) [39] and Big-Five personality traits (also known as the Five Factor Model) [40]. EPQ is a questionnaire to access a person's personality traits. The theory underlying EPQ primarily concerns physiology and genetics. The personality assessment through EPQ is based on temperament and has three aspects only. Temperament is a genetically-based personality aspect, which is developed and evolved right from the birth. Instead of focusing on temperament, a biochemical phenomenon, the FFM is based on assessing personality in a socio-cultural concept. Traits in the FFM have approximately equal influence from environment and hereditary circumstances [41]. The Big-Five personality traits are based on the common language description of personality. There has been critique on both of these models. However, the FFM is the widely-used approach for various purposes [40]. This work also utilizes the FFM for correlating the extracted clusters with the personality traits.

The traits comprising of the FFM include: openness, conscientiousness, extraversion, agreeableness, and neuroticism, often referred to as OCEAN. Extraversion explains if a person is talkative, assertive, and energetic. Agreeableness shows how much a person is trustworthy and cooperative. Conscientiousness explains weather a person is orderly, responsible, and dependable. Neuroticism displays properties like calmness, non-phobic, and not upsetting easily. Openness suggests a person's intellectual, creative, and sovereign mindedness. Fig. 2 lists the Big-Five personality traits and their sub-dimensions. Each of the Big-Five factors is quite broad and consists of a range of more specific traits. For further details on the FFM, the reader is referred to [42].

4 Methodology

This section explains the methodology used for recording participants' in-game behavior, extracting profiles and predicting personality. Since this work presents an experimental research, a research hypothesis is formed here. The testing hypothesis is listed as follows:

If a gameplay data recorded from human subjects

TABLE 1
Features Summary of the Proposed and Closely Related Works

Works	Features					
	Real-world data/observations	Benchmark datasets	No. of participants	Data collection via survey forms	In-game data/features used	Results correlated with ground truth
Worth et al. [22]	✓		219	✓		
Bialas et al. [23]	✓	✓	9368	✓	✓	
Brown et al. [26]	✓		44		✓	
Symborski et al. [27]	✓		114	✓	✓	✓
Bean et al. [29]	✓	✓	1210	✓		
Teijeiro-Mosquera et al. [31]	✓		281			
Halim et al. (Current proposal)	✓	✓	155125	✓	✓	✓

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2017.2751602, IEEE Transactions on Affective Computing

through an RTS game is profiled using machine learning techniques, it can be utilized to map subjects to the Big-Five personality traits.

The data capturing, clustering, and classification is based on the above-mentioned initial hypothesis. This study uses three datasets. Two of these are the benchmark datasets extracted from *StarCraft* and *WoW* games. The third one is created using *Age of Empire-II* game and 50 participants. It is observed that more than 55% of the Internet consumers are online gamers, though the number of Massively Multiplayer Online Game (MMOG) subscribers universally rose in 2010 to 21 million [43]. Amongst numerous categories of MMOGs, massively multi-player online role-playing games (MMORPGs) is the most common category, with a 95% portion of the MMOGs marketplace [44]. The MMORPG marketplace is worth US $6 billion globally, with a forecasted worth of US $8 billion in the years to come [44]. These are the reasons of choosing *StarCraft*, *WoW*, and *Age of Empire-II* in this study. Details on the datasets, demographics of the subjects and the personality measurement follow in this section.

4.1 Datasets

The two benchmark RTS game datasets used in this study include: World of Warcraft history dataset (*WoW*)† and *StarCraft* dataset‡. Other than these two datasets, an additional dataset is created using 50 participants who played *Age of Empire-II*.

WoW is a massively multiplayer online game. The *WoW* history dataset contains 91065 avatars' data and their related attributes. The avatar data was recorded for three years. Each avatar status was recorded after 10 minutes and their respective attributes were stored in the dataset. There are 151,680 samples in the data. If an avatar does not appear in successive samples, it means that it has logged out. The directory structure of *WoW* archive encompasses two sub-directories. There are four first-level directories each containing avatar samples of three months. The second-level directory named as *day:month:year* contains 144 log files because in one day there are 1440 minutes and the samples are taken after 10 minutes. The log file is named as *HH: MM: SS* showing the time of sampling. Each log file contains two arrays, i.e., persistent storage and rounded information. The information collected about online avatars is stored in a persistent storage array. Each row stores information about avatar history perceived during sampling time. A row is a string that consists of 11 fields parted by comma. The 11 attributes that are recorded about each avatar online are: *dummy-item 1*, *query time*, *query sequence number*, *avatar id*, *guild*, *level*, *race*, *class*, *zone*, *dummy-item 2*, and *dummy-item 3*. The query time is between January 2006 and January 2009. The q*uery sequence number* and *avatar id* is a positive integer. *Guild* is a positive integer between 1 and 513. *Level* is also a positive integer between 1 and 80. There are five races in *WoW*, i.e., *Blood Elf*, *Orc*, *Tauren*, *Troll*, and *Un-dead*. Various classes in *WoW* are: *Death Knight*, *Druid, Hunter*, *Mage*, *Paladin*, *Priest*, *Rogue*, *Shaman*, *Warlock*, and *Warrior*. The zone is one of the 229 zones listed in *WoW.* They differ in terms of *territory, dungeons, arena,* etc. The *WoW* dataset has data of players from January 2006 to January 2009. However, this study uses two months' data, January 2006 and February 2006.

StarCraft is another RTS game used in this study. There are three types of fictions in *StarCraft*, varying in terms of units, build tree, gameplay styles etc. The three fictions are *Protoss*, *Zerg*, and *Terran*. There are 57 attributes in *Protoss vs. Protoss* (PVP), *Protoss vs. Terran (PVT)* and *Protoss vs. Zerg* (PVZ). In *Terran vs. Protoss* (TVP), *Terran vs. Terran* (TVT) and *Terran vs. Zerg* (TVZ) there are 52 attributes, while *Zerg vs. Protoss* (ZVP), *Zerg vs. Terran* (ZVT) and *Zerg vs. Zerg* (ZVZ) have 49 attributes. In the *StarCraft* dataset, all values are integers except *midbuilds* having seven classes in each fiction. For the current study, all the *midbuilds* are replaced with integer numbers from zero to six. Unknown *midbuilds* are replaced with zero. Attributes of one fiction are different from other fiction.

The third dataset is recorded from a group of participants using the *Age of Empire-II* game. In this study a multiplayer version of the *Age of Empire-II* was played by 50 subjects and at the end of the session, 30 gameplay attributes were captured for each of the subjects. These include, *military*, *economy*, *technology*, *society*, *total score*, *units killed*, *units lost*, *building razed*, *building lost*, *units converted*, *largest army*, *food collected*, *wood collected*, *stone collected*, *gold collected*, *trade profit*, *tribute sent*, *tribute received*, *feudal age*,

TABLE 2
AN OVERVIEW OF THE DATASETS' FEATURES

Dataset	Features (Excluding dummy items)
World of Warcraft history dataset (*WoW*)	(1) query time (2) query sequence number (3) avatar id (4) guild (5) level (6) race (7) class (8) zone
StarCraft	*Number of features depending on the race*
Age of Empire-II	(1) military (2) economy (3) technology (4) society (5) total score (6) units killed (7) units lost (8) building razed (9) building lost (10) units converted (11) largest army (12) food collected (13) wood collected (14) stone collected (15) gold collected (16) trade profit (17) tribute sent (18) tribute received (19) feudal age (20) castle age (21) imperial age (22) map explored (23) research count (24) research percent (25) total wonders (26) total castles (27) relics captured (28) relic gold (29) villager high (30) survival to finish

† http://mmnet.iis.sinica.edu.tw/dl/wowah/
‡ https://games.soe.ucsc.edu/project/starcraft-data-mining

castle age, *imperial age*, *map explored*, *research count*, *research percent*, *total wonders*, *total castles*, *relics captured*, *relic gold*, *villager high*, and *survival to finish*. The attribute *military* has five types, including: infantry, archers, cavalry, siege weaponry, and naval units. Feature *economy* indicates the financial health of the team, the higher the better. The feature *technology* has varying strengths and weaknesses depending on the civilizations. The features like, *units killed*, *units lost*, *building razed*, *building lost*, *units converted*, *largest army*, *food collected*, *wood collected*, *stone collected*, and *gold collected* represent the collection items and the attribute *total score* is mainly dependent on these. All tributes (*tribute sent/received*) made in *Age of Empires-II* have a 30% tribute tax. The features *feudal age*, *castle age*, and *imperial age* are utilized for advancement in civilizations. The attribute *villager high* is used to gather resources indicated by the remaining features. In addition to recording the game data for the 50 subjects using *Age of Empire-II*, demographic information was also collected before the gameplay. This included: gender, date of birth, qualification, devices familiar with, preferred strategy games, preferred gaming platform and the duration participant play games each day. Table 2 provides an overview of the features extracted from the three RTS games.

In addition to the abovementioned datasets, for the evaluation of the classifiers a new dataset is recorded (in Section 5.3) using 50 previously unknown participants. These were randomly picked volunteers having previous experience of playing RTS games.

4.2 Participants

For the creation of the third real-world gameplay dataset using *Age of Empire-II,* participants from a university environment were solicited. The subjects were randomly selected, provided they had previous experience of playing *Age of Empire-II* or some other RTS. The benefit of this dataset is to have known participant, who can be utilized in additional experiments and drawing correlations. Most of the participants of this activity are undergraduate/graduate students, with a few of them being research assistants and their friends. Participation in this experiment was voluntary and no compensation was made to the participants. Table 3 shows the participants' demographics.

4.3 Measuring Personality

To create the ground truth and for the purpose of evaluating the participants' personality, IPIP-NEO-120 questionnaire is used. After recording the *Age of Empire-II* data, the questionnaire is completed by the 50 subjects and their score on the Big-Five is recorded in terms of percentile. The IPIP-NEO-120 questionnaire comprises of 120 questions related to personality. The choice of the IPIP-NEO-120 questionnaire is made based on its wide use in the community to measure personality on the Big-Five model. These results are later used to correlate with the clustering formations.

TABLE 3
PARTICIPANTS' DEMOGRAPHICS

No. of participants	50
Male	70%
Female	30%
Player average age	22 years
Average RTS expertise	Medium

4.4 Profiling

For extracting analogous groups (profiles) from the datasets, clustering is employed. Clustering is an approach for statistical analysis that groups the data using a distance function. Good clusters are formed by means of decreasing the inter-cluster similarity and increasing intra-cluster similarity. Initially, all records in the dataset are clustered. This is done due to the absence of labels in the data. The data is preprocessed for categorical variables. All categorical variables are assigned an integer number. Later, to find the number of clusters in the data, four clustering techniques are used, including *k*-means, *k*-medoids, fuzzy c-mean, and hierarchical clustering. Since the problem at hand is of unsupervised learning, the use of the four clustering techniques helps in averaging-out the results and provides a bigger picture about the number of profiles in the underlying datasets. This also helps in generalizing the results.

k-means is an unsupervised partitioning-based clustering technique. *k*-means partitions *n* objects into *k* clusters where, each object is assigned to the cluster having nearest mean [45]. The *k*-means algorithm converges to a local minimum. Before it converges, calculations of cluster centers and distance are undertaken. The standard *k*-means starts with a random set of initial *k* points for clustering. This work has also used the same. The *k*-means is iterated until there is no change in the cluster representatives. This aspect is dependent on the initial selection of the cluster centroids. *k*-medoids is a classical unsupervised clustering technique. It clusters the data consisting of *n* objects into *k* clusters where the value of *k* is known a priori. For data with noise and outliers, *k*-medoids is more robust as compared to the *k*-means clustering [46]. It reduces the sum of pairwise variations. Hierarchical clustering is a method for cluster analysis that builds a hierarchy of clusters. It has two types, agglomerative and divisive. This work uses agglomerative hierarchical clustering. For *n* objects to be grouped, and an *nxn* distance matrix, the plain process of Johnson's hierarchical clustering is used. The fuzzy c-means clustering provides a degree of membership instead of hard clustering. It is a soft clustering technique [47], where an object may belong to more than one cluster. It is based on minimization of the objective function according to a threshold value. The membership degree shows how closely an object belongs to a cluster. The partitioning process is carried out through an iterative optimization of the objective function with an update of the membership U_{ij} and the cluster center c_j. The function is listed in (1) [48].

$$F_m = \sum_{i=1}^{N} \sum_{j=1}^{C} U_{ij}^{\ m} \left\|x_i - c_j\right\|^2 \qquad 1 \le m \le \infty \qquad (1)$$

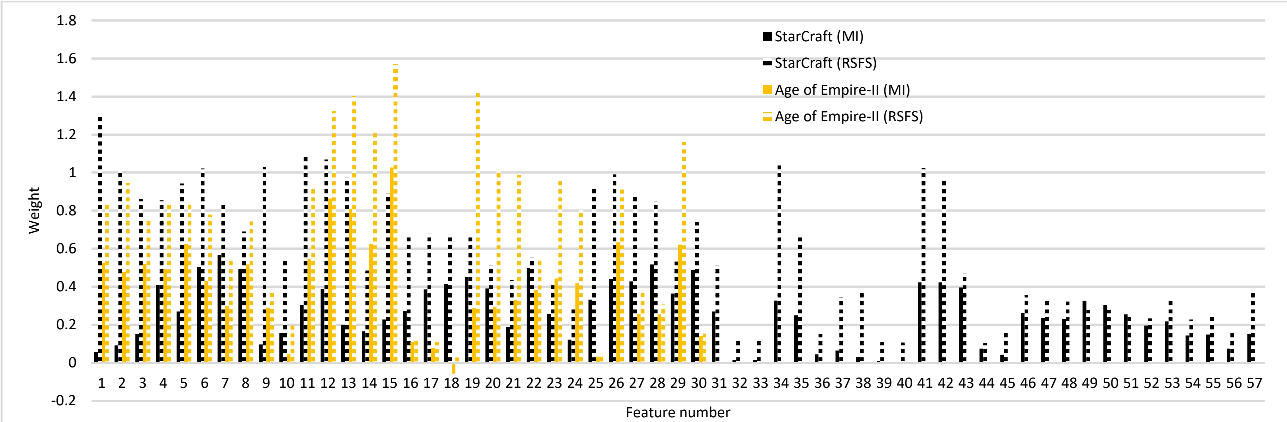


Fig. 3. Feature selection results for the *Age of Empire-II* and *StarCraft* datasets.

The highest value of the membership shows strong similarity to the cluster. The relationship between c_j and u_{ij} are calculated using (2) and (3).

$$c_j = \sum_{i=1}^{N} \frac{{U_{ij}}^m \; x_i}{\sum_{i=1}^{N} {U_{ij}}^m} \tag{2}$$

$$U_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|}\right)^{\frac{2}{m-1}}} \tag{3}$$

Iterations stop when || *uijk+1 –uijk* || is less than a threshold value, where the termination criterion is between 0 and 1, and *k* is the number of iterations.

4.5 Cluster validation

Once the clusters are formed, the results were evaluated through three cluster evaluation criteria, i.e., Davies–Bouldin index (DBI), Dunn index (DI), and Silhouette coefficient (SC).

Davies Bouldin Index (DBI):

The Davies Bouldin Index (DBI) computes the inter-cluster and intra-cluster relationship. DBI is calculated using (4), where, n corresponds to the number of clusters, σ_x denotes the average distance of all objects in a cluster to the midpoint of cluster c_i, and $d(c_i, c_j)$ is the distance among the midpoints of two different clusters c_i, and c_j.

$$DBI = \frac{1}{n} \sum_{i=1}^{n} \max_{j \neq i} \left(\frac{\sigma_i + \sigma_j}{d\,(c_i, c_j)}\right) \tag{4}$$

Dunn Index (DI):

Dunn Index (DI) is an internal cluster assessing metric grounded on the clustered data. DI can be computed using (5), where, $d(i, j)$ is the distance between two clusters. The value of $d'(k)$ signifies the distance among the objects inside the cluster. DI shows a cluster's density and how much it is separated from other clusters. For a clustering formation, higher value of DI shows good clustering.

$$DI = \min_{1 \le i \le n} \left\{ \min_{1 \le j \le n,\; i \neq j} \left\{ \frac{d\,(i,j)}{\min_{1 \le k \le n} d'(k)} \right\} \right\} \tag{5}$$

Silhouette Coefficient (SC):

Silhouette Coefficient (SC) checks each object's similarity with all other objects in their own cluster and their dissimilarity with objects belonging to other clusters. SC can visually show how well an object is placed in its cluster. SC value ranges between [-1, +1]. Where, +1 indicates good clustering while -1 indicates bad clustering. SC can be calculated using (6) and (7).

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{6}$$

$$SC = \frac{1}{N} \sum_{i=1}^{N} S(i) \tag{7}$$

Where, $a(i)$ shows the average dissimilarity of an object i within a cluster from all other objects in that cluster. Value $b(i)$ is the smallest average dissimilarity of an object i to the objects of all other clusters.

5 Experiments and Results

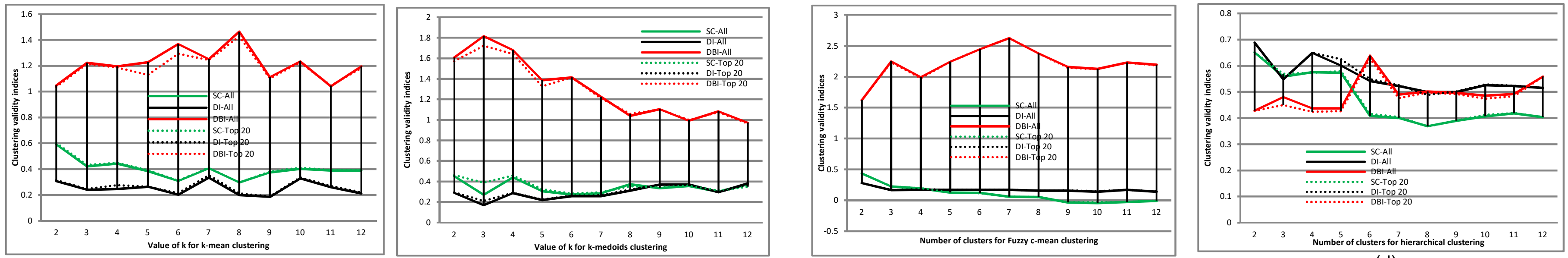
This section describes the experiments conducted and the results obtained. The feature selection is explained first followed by the clustering and classification experiments.

5.1 Feature Selection

To study the importance of various features (in the datasets) towards better clustering formation, various feature selection techniques are applied. Feature selection is important for predictive analytics. It serves to identify if two or more of the independent variables are correlated to the dependent (predicted) variable. In such cases, the estimates of the coefficients in a regression model tend to be unstable or counter intuitive. Therefore, only the key features are used for the predictive analysis. The feature selection techniques used in this study are Random Subset Feature Selection (RSFS), Principal Component Analysis (PCA), and Mutual Information (MI). PCA is chosen because it joins comparatively similar components to make new ones, better than the original one. In RSFS, the arbitrary subset classification is attained several times. This is essential to differentiate the good dimensions from a set of dimensions that simply look valuable, due to the arbitrary components of the method. Mutual information (MI) generally considers dependencies between features and class, as compared to statistical methods, like PCA, that do not consider class labels. It provides superior feature extraction than all other feature extraction techniques [49].

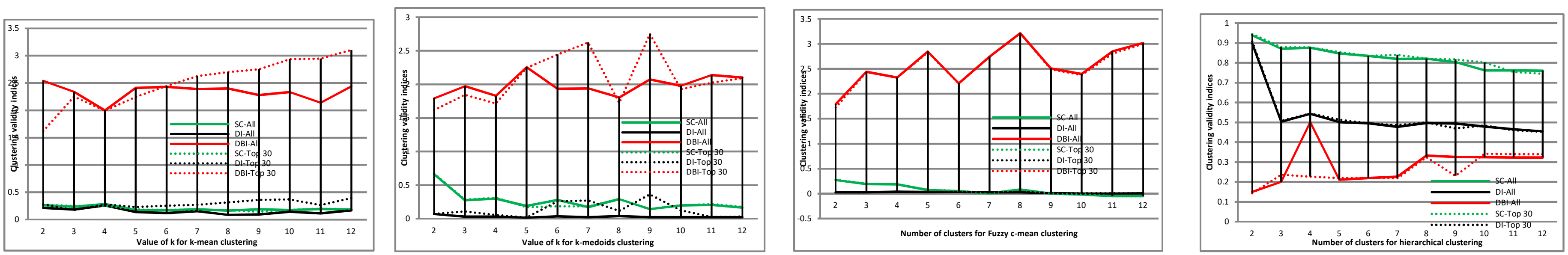
Fig. 3 shows game features' weights calculated through MI and RSFS for the 30 attributes in *Age of Empire-II* and 57 attributes of *StarCraft*. It can be interpreted from the graph that some features have lower weight.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2017.2751602, IEEE Transactions on Affective Computing



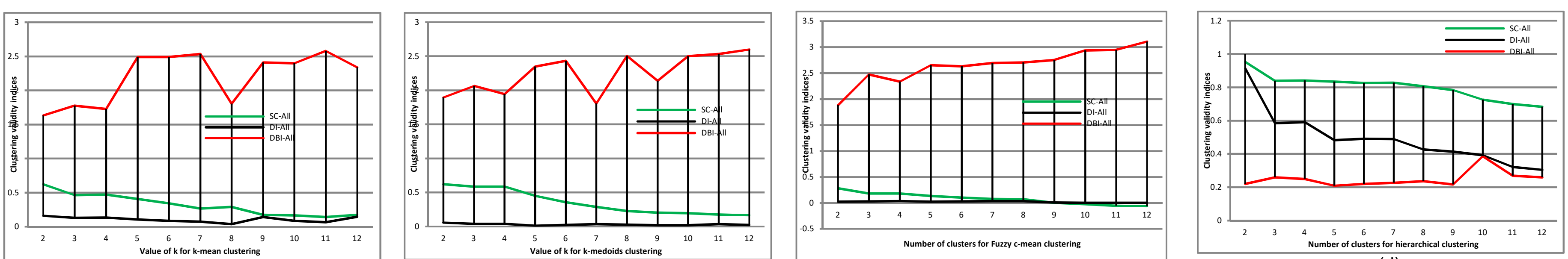
(a) (b) (c) (d)

Fig. 4. The DBI, DI, and SC values for the various clustering formations of *Age of Empire-II* dataset using *k*-means, *k*-medoids, fuzzy c-mean, and hierarchical clustering techniques.



(a) (b) (c) (d)

Fig. 5. The DBI, DI, and SC values for the various clustering formations of *StarCraft* dataset using *k*-means, *k*-medoids, fuzzy c-mean, and hierarchical clustering techniques.



(a) (b) (c) (d)

Fig. 6. The DBI, DI, and SC values for the various clustering formations of *WoW* dataset using *k*-means, *k*-medoids, fuzzy c-mean, and hierarchical clustering techniques.

TABLE 4
BEST ANN ARCHITECTURE PERFORMANCE

Dataset	Accuracy	Sensitivity	Specificity	Precision	Recall	F-Measure	G-mean
Age of Empire-II	0.94	0.97	0.89	0.94	0.97	0.95	0.93
StarCraft	1.00	1.00	1.00	1.00	1.00	1.00	1.00
WoW	0.92	1.00	0.89	0.78	1.00	0.88	0.94

TABLE 5
SVM PERFORMANCE WITH THE GAUSSIAN RBF KERNEL

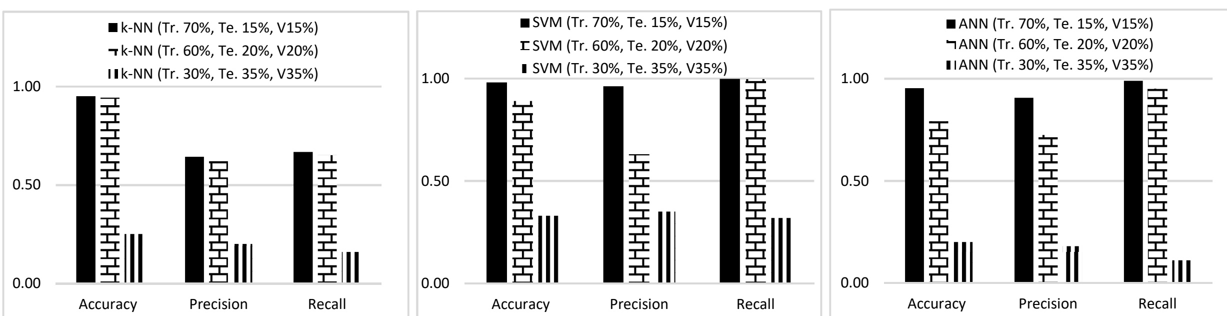
Dataset	Accuracy	Sensitivity	Specificity	Precision	Recall	F-Measure	G-mean
Age of Empire-II	1.00	1.00	1.00	1.00	1.00	1.00	1.00
StarCraft	0.96	0.99	0.91	0.95	0.99	0.97	0.95
WoW	0.98	1.00	0.97	0.93	1.00	0.97	0.99

TABLE 6
K-NN CLASSIFIER PERFORMANCE FOR *K*=3

Dataset	Accuracy	Sensitivity	Specificity	Precision	Recall	F-Measure	G-mean
Age of Empire-II	1.00	1.00	1.00	1.00	1.00	1.00	1.00
StarCraft	0.98	0.96	0.99	0.99	0.96	0.97	0.98
WoW	0.94	1.00	0.92	0.84	1.00	0.91	0.96

TABLE 7
CLASSIFIERS PERFORMANCE OVER ADDITIONAL DATASETS (AVERAGED OVER ALL DATASETS)

Dataset	Accuracy	Sensitivity	Specificity	Precision	Recall	F-Measure	G-mean
ANN	0.78	0.79	0.70	0.78	0.77	0.79	0.76
SVM	0.91	0.98	0.89	0.78	0.89	0.88	0.94
k-NN	0.88	0.87	0.75	0.75	0.84	0.84	0.90



(a) (b) (c)

Fig. 7. Classifiers performance over various training, testing, and validation rates (a) *k*-NN classifier (b) SVM classifier (c) ANN classifier.

The features having a higher weight play an important role in clustering. The features with a low weight do not have a major impact on clustering. MI and RSFS identify the same features as top 20 for *Age of Empire-II* dataset. These include: *military*, *economy*, *technology*, *society*, *total score*, *units killed*, *units lost*, *building razed*, *largest army*, *food collected*, *wood collected*, *stone collected*, *gold collected*, *castle age*, *imperial age*, *map explored*, *research count*, *research percent*, *total castles*, and *feudal age*. Feature selection when applied to the *StarCraft* dataset gives following 30 attributes as important ones using MI: *SecondGas*, *FirstExpansion*, *SecondExpansion*, *ThirdExpansion*, *FourthExpansion*, *ThirdGatway*, *FourthGatway*, *Range*, *Forge*, *Cannon*, *GroundWeapons1*, *FroundArmor1*, *GroundWeapons2*, *FroundArmor2*, *Citadel*, *Legs*, *Archives*, *Templar*, *Archon*, *Storm*, *DartTemplar*, *RoboBay*, *Shuttle*, *Observory*, *Obs*, *Stargate*, *FleetBeason*, *Tribunal*, *Arbitor*, and *Statis*. RSFS also gives almost similar set of features with the exception of four. It adds *FirstGas*, *Core*, *AirWeapons1*, and *AirWeapons2* by dropping out *ThirdGatway*, *Range*, *Stargate*, and *Tribunal*. Feature selection is not applied to the *WoW* dataset since it has only eight attributes (after excluding the three dummy columns).

5.2 Clustering

The datasets are clustered one by one using the four techniques. This results in four clustering formations for each of the dataset with twelve clustering results in total. This methodology enables to generalize the results. In addition to this, a clustering study using the most and the least significant features has also been carried out. To apply *k*-means on a dataset the value of *k*, i.e., the number of clusters, is prespecified. *Age of Empire-II* dataset is clustered using *k*-means for the values of *k* ranging from two to 12. Fig. 5 shows the DBI, DI, and SC values for the various clustering formations of the *Age of Empire-II* using *k*-means, *k*-medoids, fuzzy c-mean, and hierarchical clustering techniques. It is observed that for two clusters *k*-means provides optimum clustering formation. This is decided using the cluster evaluation indices of DBI, DI, and SC. For a given clustering formation, higher values of DI and SC indicate better clustering. In the case of DBI, the best clustering formation is indicated by the lowest index value. As shown in Fig. 4 (a) for the various values of *k*, SC has minimum values for *k*=2 using the clustering formation based on both, the complete datasets as well as

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2017.2751602, IEEE Transactions on Affective Computing

the top 20 features. For k=2, DI has the third lowest value and for DBI it is the second highest. It is also interesting to note that the cluster validity indices tend to have the same behavior for the complete dataset and top 20 features. The clustering results using k-medoids for *Age of Empire-II* datasets in Fig. 4 (b) show promising results for two clusters based on SC only. The fuzzy c-mean clustering shows an optimum clustering formation using two clusters based on all the cluster validity indices as shown in Fig. 4 (c). Same is the trend for the clustering results using hierarchical approach for the *Age of Empire-II* datasets (Fig. 4 (d)). Clustering results for the *StarCraft* dataset are shown in Fig. 5. As was the case for the *Age of Empire-II* dataset, the *StarCraft* data is also clustered for various values of k and then the optimum number of clusters is decided based on the cluster validity indices of DBI, DI, and SC. Fig. 5 (a) suggests optimum clustering for k=2 (using k-means clustering) based on the SC and DBI (using top 20 attributes only). However, clustering using the complete dataset yields optimum groups at k=4 for SC and DI indices. By clustering the same data using k-medoids optimum clusters are achieved for k=2 for all the cluster validity indices (using all the attributes as well as only top 20) with the exception of DI. Fig. 5 (c) show the clustering results for the *StarCraft* dataset using fuzzy c-mean technique, where all the indices suggest better clustering formation for categorization of the data into two groups with an exception of DI. Finally, clustering using hierarchical technique, as shown in Fig. 5 (d), suggests two clusters for all validity indices.

The third dataset, *WoW*, is also clustered to find the number of coherent groups. However, clustering analysis of the top features in not done in this case because of the fewer number of dataset dimensions. Fig. 6 lists the results, where the optimum clustering is found for two groups using all techniques and all validity indices with an exception of DBI in case of hierarchical clustering.

5.3 Profile Prediction

A classifier is used to predict a player's profile (i.e., cluster) based on her/his in-game behavior. Classification is a supervised learning technique. The classifiers used in this study are: ANN, SVM, and k-NN. These are tested on the three datasets separately. Two classes are used for this experiment as suggested by the clustering results. The ANN consists of the number of input layer neurons proportional to the feature count in the dataset. The output layer consists of two neurons representing one profile (cluster) each. The ANN is tested on various number of hidden layers and neurons in them. The best performance is observed using two hidden layers with ten neurons in each. For SVM, a set of kernels is used to identify better performance. They include: Gaussian radial basis function (RBF), linear kernel, and multi-layer perceptron (MPL). It is observed that the RBF kernel yields better results as compared to the other two kernels. For the k-NN classifier the values of k ranging from two to five are evaluated, where the best accuracy is achieved for k equals three. The three datasets consist of dissimilar number of items where 70% of the dataset is used for training and 30% for testing. However, in cases where a validation set is utilized as well, each dataset is randomly divided as 60% for training, 20% for testing, and 20% for validation. The datasets are shuffled before proceeding for the training. For training, input patterns and the desired output are presented to the classifier, whereas in the case of testing, only input patterns are presented. Table 4 lists the performance evaluation metrics for the best ANN architecture. Using ANN, the best accuracy of 99% is achieved for the *StarCraft* dataset while the second best is for the *Age of Empire-II* dataset which is 94%. Table 5 lists the performance measures for SVM using the RBF kernel since it produces better results as compared to the other two kernels. SVM gives the best accuracy for the *Age of Empire-II* dataset which is 100%. Its second-best accuracy is reported for the *WoW* dataset with a value of 98%. Table 6 contains the results for the k-NN classifier with its optimum settings. k-NN also returns the best accuracy for the *Age of Empire-II* dataset. However, its second-best accuracy is reported for the *StarCraft* dataset. Overall, SVM gives the highest average accuracy of 98% for the three datasets. k-NN has the second best average accuracy having a value of 97%, while ANN has the least average accuracy of 95%. To study the performance of the three classifiers over various training, testing, and validation rates an experiment is performed using three sets of configurations. They include (training=70%, validation=15%,

TABLE 8
CLUSTERING TECHNIQUES COMPARISON FOR THE THREE DATASETS

	Age of Empire-II				*StarCraft*				*WoW*			
Indices	k-means	k-medoid	Fuzzy c-mean	Hierarchal	k-means	k-medoid	Fuzzy c-mean	Hierarchal	k-means	k-medoid	Fuzzy c-mean	Hierarchal
SC-All	0.591	0.451	0.436	**0.651**	0.286	0.662	0.271	**0.938**	0.622	0.622	0.282	**0.952**
DI-All	0.334	0.382	0.278	**0.688**	0.257	0.070	0.037	**0.896**	0.159	0.059	0.039	**0.913**
DBI-All	1.041	0.972	1.620	**0.428**	1.998	1.792	1.791	**0.150**	1.632	1.805	1.881	**0.209**
SC-Top Features	0.602	0.461	0.436	**0.651**	0.275	0.672	0.281	**0.945**				
DI-Top Features	0.353	0.372	0.279	**0.688**	0.394	0.361	0.037	**0.907**				
DBI-Top Features	1.036	0.964	1.621	**0.424**	1.620	1.615	1.721	**0.143**				

testing=15%), (training=60%, validation=20%, testing=20%), and (training=30%, validation=35%, testing=35%). Fig. 7 lists the results for this, where the performance of all the classifiers decreases as the percentage of dataset instances in the training sets decrease. The average accuracy of all the classifies for (training=70%, validation=15%, testing=15%) is 96%, for (training=60%, validation=20%, testing=20%) it is 87%, and for (training=30%, validation=35%, testing=35%), the average accuracy falls below 47%. Thus, the best configuration for training the classifiers is (training=70%, validation=15% testing=15%). The training instances greater than 70% for the classifiers would result in overfitting. The results in Fig. 7 also indicates SVM to be more robust as the training, testing, and validation percentages change, since its accuracy deteriorates gradually as compared to ANN and *k*-NN. However, this trend cannot be generalized and depends on the problem at hand.

The three RTS game datasets utilized in this study have disjoint sets of features. A classifier trained on one dataset is not appropriate for the others. However, the data extracted from the same game, but different participants can be utilized for this. Further evaluation of the classifiers is done by engaging 50 additional randomly picked volunteers having previous experience of playing RTS games. The majority of these volunteers were university students. Initially, the dataset gathered using them is clustered using the *k*-medoids by setting *k* equals 2. This gives the ground truth to measure classifier accuracy. Unlike the previous experiments where an average of multiple runs was utilized to record the gameplay features, here the three games are played only once by each participant. The trained classifiers over the previously recorded datasets are used to predict the classes for this newly constructed data. The clustering results of the new dataset are utilized as the ground truth to compute accuracies. For each classifier, Table 7 list the average value of the performance measures over the newly created datasets using *Age of Empire-II*, *StarCraft*, and *WoW* datasets. SVM performs better here with an overall average accuracy of 91%.

5.4 Descriptive Analysis

The clusters extracted from the underlying datasets in Section 5.2 are based on either a distance measure or correlation. This results in coherent groups, where individuals in the same group have alike features. The three cluster validity indices, i.e., DBI, DI, and SC, give the measure of how well separated these clusters are from each other, in addition to the internal cluster cohesion. However, in order to study the cluster properties, a descriptive analysis is required. Usually, the descriptive analysis is based on the feature set that is used for clustering [50]. In this proposal, the results of the IPIP-NEO-120 personality test are also used for the descriptive analysis in addition to the dataset features. The experiments in Section 5.2 concludes that there exist two clusters in each of the three datasets. The four clustering techniques may result in varying cluster formations. Out of these, the hierarchical clustering gives the optimum clustering validity values for all the indices over the three datasets (Table 8). The reason being splitting and merging of clusters in the hierarchical clustering based on the cluster validity indices. However, the visual inspection of the clusters formed with the hierarchical clustering identifies that one of the clusters contains single sample. Thus, indicating an imbalanced clustering formation due to one versus all samples. Therefore, the next best clustering technique is the *k*-medoids with overall eight best clustering validity indices (excluding the hierarchical clustering results) out of the 14 possibilities. This descriptive analysis is based on the clustering formation using the *k*-medoids results. The cluster-wise maximum, minimum, and average values of the *Age of Empire-II* dataset attributes are listed in the appendix. Cluster-2 has average *military* five times more than cluster-1, similarly cluster-2 has an average value of *society* six times more than cluster-1. For game attributes, like *food collected*, *wood collected*, *stone collected*, *gold collected*, *total castles*, and *relic gold* cluster-2 has average values four times higher than cluster-1. For other attributes, like *total score*, *units killed*, *units lost*, *building razed*, and *map explored* cluster-2 has average values three times larger than cluster-1. Features having almost double average values for the players in cluster-2 than those in cluster-1 include: *technology*, *largest army*, *research count*, *research percent*, *relics captured*, and *villager high*. For the game feature *feudal age*, cluster-1 has double value than cluster-2. The scores of game attributes *military*, *economy*, *technology*, and *society* are computed based on the other game features having varying weight. The *total score* is the overall total using the aforementioned four attributes. Thus, the higher values of cluster-2 for the attributes *military*, *economy*, *technology*, and *society* are influenced by their higher value in other game attributes, like *total score*, *units killed*, *units lost* etc. The players in cluster-2 have been more inclined towards constructing castles and wonders, resulting in a much higher score for *society* in comparison to the players in cluster-1. An interesting pattern worth noticing is that the *units lost*, and *units killed* for cluster-2 are both three times higher than the cluster-1. This may indicate an aggressive attitude of the players in cluster-2 in comparison to those in cluster-1. The individuals in cluster-2 have been interested in collecting game items for increasing their *gold*. As compared to the individuals in cluster-1 there is an inclination towards *technology* and its use in the cluster-2's individuals. As the game *Age of Empire-II* starts there is a dark age with less access to technology. For access to some of the technologies an update to *feudal age* is required. Access to all technologies requires an update to *imperial age*. Interestingly, the values of the *feudal age* and *castle age* are relatively higher for cluster-1, however, the participants in cluster-1 have not utilized these higher values as indicated by its two times lesser value of *technology* in comparison to cluster-2. Further inspection identifies relatively higher values of *imperial age* for cluster-2. Thus, justifying cluster-1 for lesser *technology* due to being in *feudal age* and *castle age* most of the times.

As mentioned earlier, the participants also take IPIP-NEO-120 personality test. The results obtained for this test are grouped cluster-wise and maximum, minimum, average, and standard deviation of the Big-Five personality traits are listed in the appendix. Based on the cluster's average values for the Big-Five personality traits, the two groups differ in neuroticism, extraversion, and agreeableness. The participants in cluster-2 score approximately 15% higher in neuroticism and 10% more in extraversion as compared to those in cluster-1. The participants in cluster-1 have 11% more agreeableness as compared to those in cluster-2. Fig. 8 shows a bar chart using the gameplay features and the Big-Five personality traits for the two clusters. Since the descriptive analysis is based on the *k*-medoids results, Fig. 8 also shows the same using *Age of Empire-II* dataset. There are 30 features in *Age of Empire-II* dataset. However, not all of these are used for plotting the bar chart. Only those features are utilized that have significant difference. Each gameplay feature value of cluster-1 is subtracted from its corresponding feature in cluster-2 and an overall difference is obtained by averaging the difference of the 30 features. Features having difference greater than the overall average difference are plotted. The gameplay features' values in Fig. 8 are scaled in the range between 0-100 for better visualization. Cluster-2 individuals are on an average 15% higher in neuroticism indicating experiencing feelings like worry, fear, and anger more as compared to those in cluster-1. They also have higher scores in gameplay features of *food collected* and *wood collected*. Thus, we can correlate higher scores in the RTS collection features with the higher neuroticism, because fear of losing or anger in the RTS environment would cause the player to collect more of such items. A 10% more extraversion is observed in the cluster-2 individuals, whereas they score 19% less in the technology-based features, such as: *feudal age*, *castle age*, and *imperial age*. This indicates a negative correlation between score in technology-based features for RTS games and extraversion. Finally, cluster-1 individuals have 11% more agreeableness and they also score 29% higher in the gameplay features of *economy* and *total score*. Thus, the RTS game features referring to the monetary aspects and score have positive correlation with the player's agreeableness. The personality traits of openness and conscientiousness are not correlated with any of the gameplay features because cluster-1 and cluster-2 exhibit similar behavior for these. Fig. 9 shows the visual representation of personality demographics and the RTS games' features preferences of the two clusters using various man icons. The 22 icons used in Fig. 9 are created using Flaticon[§].

The discussion above is based on the gameplay features' values and the percentile obtained by the participants on the Big-Five personality traits. In order to be statistically correct and to evaluate the hypothesis formed in Section 4, Pearson correlation coefficient (r) is also computed. This is done using the *k*-medoids clustering results, these being the optimum. The value of r=1 indicates perfect positive correlation, whereas, -1 shows perfect negative correlation. The value of r when computed for neuroticism and *food collected* is 0.472 and for neuroticism and *wood collected* it is 0.6127. This indicates a positive correlation between the collection features of an RTS game and neuroticism. The Pearson correlation coefficients (r) for extraversion/*feudal age*, extraversion/*castle age*, and extraversion/*imperial age* are -0.3482, -0.0492, and -0.5872 respectively. Indicating an overall negative correlation between extraversion and technology oriented features of an RTS game. The value of r for agreeableness/*economy* and agreeableness/*total score* is 0.7521 and 0.6124 respectively. This shows a strong positive correlation between agreeableness and RTS game features like economy and score. The correlation between other Big-Five personality traits was also calculated, however, these are not listed here because either a weak positive or a weak negative correlation was observed in them. The results mentioned in this section and the abovementioned correlations enable to accept the hypothesis set in Section 4.

6 Discussion

Finding personality traits using an indirect method is vital to numerous fields, including, but not limited to, psychology, affective computing, human resource development etc. However, they pose certain limitations. Falsifying or manipulating all or some of the personality traits by the subject is one of these. An indirect assessment tries to find personality traits using data collected from the participants through observations or other activities. This work presented an approach to find personality traits using in-game behavior of the participants extracted from three RTS games. Initially, the data is clustered using four clustering techniques. Later, a classifier is trained to predict the participants' profile using the in-game data. Correlation is established between the extracted profiles and the personality traits through IPIP-NEO-120 personality test's results taken by 50 participants.

For the initial phase of profiling, use of four clustering techniques has an advantage of generalizing the results. The data used in the experiments is collected from three different games, each representing a disjoint set of participants. For this reason, the four clustering techniques are applied separately over these datasets. Two of the clustering techniques, i.e., *k*-means and *k*-medoids are partitioning-based approaches for clustering. Fuzzy c-mean clustering is a soft computing approach where one sample may reside in more than one clusters against the hard clustering techniques, like *k*-means and *k*-medoids. While the hierarchical clustering is a tiered approach towards clustering based on splitting and merging of the partitions. All this enables to retrieve the profiles from the underlying data through various views. Further, to evaluate the quality of extracted clusters three benchmark cluster validity indices; DBI, DI, and SC are used. Although only one cluster validity index may have been fine, however, use of three enables to later opt for the best results based on majority votes. This is done for the descriptive analysis presented in Section 5.4 using the results in Table 8. The clustering results identify two disjoint groups in the three datasets. However, in some of the cases optimum clustering is obtained for either 7, 10, 11 etc. number of clusters.

[§] http://www.flaticon.com/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2017.2751602, IEEE Transactions on Affective Computing

Fig. 8. Cluster wise gameplay features and the Big-Five personality traits.

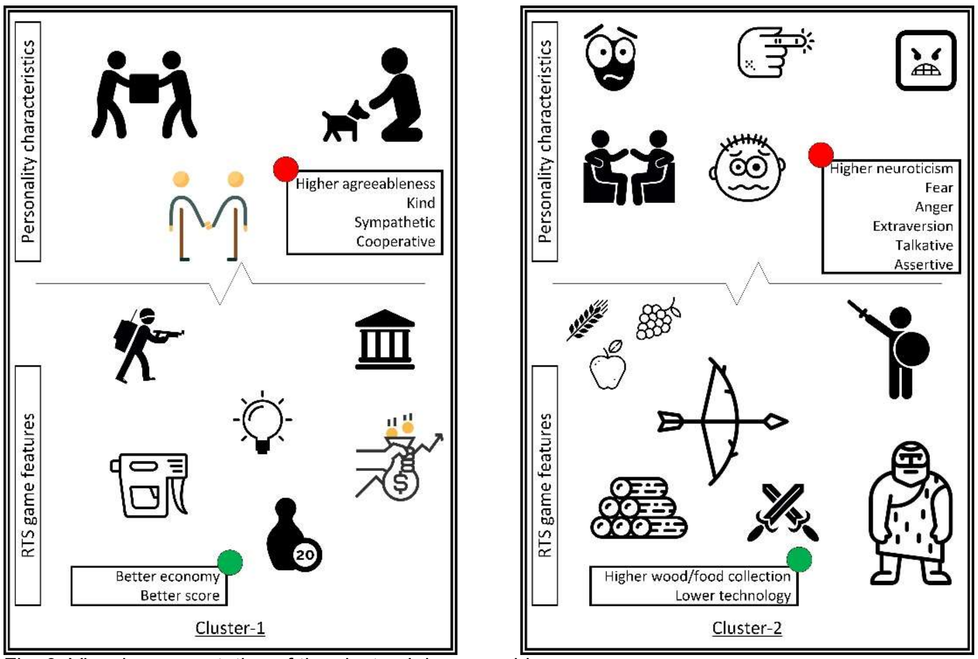


Fig. 9. Visual representation of the clusters' demographics.

For instance, using k-means clustering over the *Age of Empire-II* dataset DI suggest optimum clustering for k equals 7.

Similarly, k-medoids clustering for the *Age of Empire-II* dataset gives optimum clustering for k equals 12 based on DI. The reason for this can be the optimum formation of the clusters for these particular cases of DI using the values of k as 7 or 12 combined with the randomness. However, when it comes to generalizing the results using the 432 options provided through the 4 clustering techniques X 12 values of k X 3 datasets X 3 cluster validity indices, optimum clustering is obtained for two clusters. The results in Fig. 4 through Fig. 6 supports this, where 96.99% of the cases suggest the existence of two clusters in each of the datasets. There are only 13 such instances where a particular cluster validity index on a specific dataset has suggested optimum clustering at a value of k other than 2. Thus, helping to opt for two clusters with a confidence of 96.99%. The results in Fig. 4d, Fig. 5d, and Fig. 6d show the optimum performance of hierarchical clustering for all the datasets and all cluster validity indices. However, later it is identified that this clustering technique leaves only one item in cluster-1 and all of the remaining items in cluster-2. This results in imbalance clustering formation not rich enough for results generalization. Based on the optimum results for the cluster validity indices, k-medoids performs the second best. The same is used for descriptive analysis of the clusters. The pairwise distance measure in the k-medoid helps in reducing noise and outliners in the final clustering formation. This is also supported by the fact that the k-medoids clustering for k equals 2 gets best cluster validity index values for 8 cases out of the 14 options. Once the data is clustered, it is used to train a classifier to predict the profile of a subject using her/his in-game behavior. For this, three classifiers are used, SVM, ANN, and k-NN. The data profiled in two clusters is used to train these classifiers. The results in Table 4, Table 5, and Table 6 suggest better performance of SVM as compared to ANN and k-NN classifiers with an average accuracy of 98% for the three datasets. SVM gives this accuracy with the RBF kernel. Table 5 lists the dataset wise accuracy, where 100% accuracy is achieved for the *Age of Empire-II* dataset, 96% for *StarCraft* dataset and 98% for the *WoW* dataset. The reason for SVM achieving better accuracy for *Age of Empire-II* dataset is due to the fact that SVM performs well on small datasets and *Age of Empire-II* has only 50 samples, whereas *StarCraft* and *WoW* have a much higher number of samples. SVM comparatively performs well for all datasets since the problem at hand is of binary classification at which SVMs are good, though for a given classifier a dataset can be devised that may defeat it. The trained SVM can later be utilized to classify an unknown person into one of the two clusters (i.e., profiles) provided the in-game behavior of the person is available. Thus, the person classified in a particular class is assumed to have similar personality traits as that of the other individuals who participated in this study for dataset creation.

To measure the personality traits of the 50 participants utilized to create *Age of Empire-II* dataset, the IPIP-NEO-120 personality test is taken by each individual. Later, using the formed clusters' descriptive analysis and the test results, a correlation is developed. The results suggest that participants in a cluster with high score in certain game attributes, such as: *technology*, *military*, and *society* have high neuroticism, extraversion, and openness but have low agreeableness and conscientiousness. It is also found that participants in a cluster with higher scores in game features, like *trade profit*, *feudal age* and *castle age* are high in conscientiousness and agreeableness. It is also observed that by combining the results with participants' demographics, the individuals who have been playing computer games since their childhood are higher in neuroticism and openness. Such individuals also have low extraversion, conscientiousness, and agreeableness as compared to novice players. The individuals in cluster-2 scores higher in neuroticism and such individuals experience more anger and greed [51]. By inspecting the in-game behavior of these participants it is found that they score four times more than the other cluster's members in the collection features, like *food collected*, *wood collected*, *stone collected*, *gold collected*, *total castles*, and *relic gold*. This finding of the proposal correlates with previous such studies on neuroticism and human behavior in [52] and [53]. Individuals high in neuroticism are time and space efficient [51], this also correlates with the low score of cluster-2 individuals for the *units converted* attribute and yet on average they score higher than cluster-1. The participants in cluster-2 have low agreeableness as compared to those in cluster-1. Studies show that individuals low in agreeableness are less cooperative and may not be trustworthy [54]. The RTS games being multi-player team-based experience requires collective behavior towards winning. The fact that the individual in cluster-1 are less cooperative correlates with their low game scores. Thus, supporting the theories presented in [55] and [56].

Other than the novelty and strengths of this proposal, there are a few limitations. The present study has a gender imbalance in participants of the *Age of Empire-II* dataset. Only 30% female participants were engaged in recording *Age of Empire-II* dataset and the IPIP-NEO-120 personality test. Demographics of the two benchmark

TABLE 9
KEY BENEFITS AND LIMITATIONS

Merits	Limitations
Enables indirect personality assessment	
	Gender imbalance
Good prediction accuracy	
Generalization using four clustering techniques, three classifiers and three datasets	Majority teenagers

datasets (i.e., *StarCraft* and *WoW*) are unavailable, therefore their gender distribution cannot be computed. Another limitation of this proposal is that the most of the participants were teenagers. This is evident from the participants' average age mentioned 22 years in Table 3. This study does not address the influence of various cultures on participants' personality traits and the information for this is neither recorded in the dataset nor is it sought in the demographics form. Table 9 lists the key strengths and limitations of this proposal.

7 CONCLUSION

This work presented the problem of profiling human subjects for accessing their personality traits. The personality assessment was conducted using the indirect method, instead of a direct questionnaire-based method. For the purpose of data collection, in-game behavior of various participants was recorded using *Age of Empire-II* game. In addition, two benchmark game datasets, *StarCraft* and World of WarCraft (*WoW*), were also used in the experiment. This study utilized the Real-Time Strategy (RTS) games for data collection these being reflective of many real-world scenarios. Once the datasets were recorded they were preprocessed to remove any noise and inconsistencies. Afterward, these datasets were profiled using four clustering techniques, including *k*-means, *k*-medoids, fuzzy *c*-mean, and hierarchical clustering. Various feature selection techniques were also applied to study the effect of most important dataset features on the cluster formations. Once the clusters were formed, they were evaluated for quality using three cluster validity indices, namely, Davies–Bouldin Index (DBI), Dunn Index (DI), and Silhouette Coefficient (SC). Using the four clustering techniques, three cluster validity indices, twelve values of *k* (a numeral indicating the number of clusters to be extracted), and three datasets, 432 options for the number of profiles in the underlying datasets were identified. Other than thirteen instances, remaining options had the optimum values for two profiles in the datasets. A benchmark personality test, IPIP-NEO-120, was taken by the participants and its results were used to correlate the clusters' features with the participants' personality traits. Later, three classifiers, namely, Support Vector Machine (SVM), Artificial Neural Network (ANN), and *k*-Nearest Neighbor (*k*-NN) were trained to predict profiles for any unknown subject. Where, SVM achieved the best accuracy, i.e., 98%. The results suggested participants' high in neuroticism, extraversion, and openness but low in agreeableness and conscientiousness falling in a cluster with high score in certain game attributes, such as: *technology*, *military*, and *society*. It was also observed that participants in a cluster with higher scores in game features, like *trade profit*, *feudal age* and *castle age* are high in conscientiousness and agreeableness. By combining the profiling results with participants' demographics, it was found that the individuals who have been playing computer games since their childhood are higher in neuroticism and openness. Such individuals also had low extraversion, conscientiousness, and agreeableness as compared to novice players. The proposed system can be used in evaluating personality using an indirect method based on RTS games' in-game features. This will help in finding the true personality traits that can be faked in the direct questionnaire-based personality assessment procedures.

This proposal can be extended in multiple ways in the future. The present proposal has used three RTS games' datasets having thousands of participants. However, 50 participants were engaged in the IPIP-NEO-120 personality test. In the future additional participants may be engaged in taking the personality test to see if the ground truth changes. The current form of IPIP-NEO-120 personality test consists of 120 questions; this causes a hurdle when it comes to asking participants for the personality test in addition to the game behavior recording. Some other brief personality tests may also be explored for this purpose. This study has collected in-game behavior of players using RTS game for personality assessment, collecting human behavior related features from other activities can be a future direction. Instead of using computer games, players' behavior in the real-world games, like football, cricket, etc. can be observed for profiling purpose. However, observation-based feature extraction may consume time and introduce inconsistencies. Computer vision can be a suitable option to extract various gestures and combine these with players' statistics, game score and previous history for creating a dataset (Big Data). This dataset can then be used for profiling and personality assessment. Another future direction to extend this work is to study the relation between the real-world/synthetic-environment behavior of participants with the sub-dimensions of the Big-Five personality traits. On the applicative side, this proposal can be used in customizing visual environment of a software to suit various personalities. However, for this task the in-game data of the subject is required. This can be done conveniently for customizing the gaming environment itself. For other software tools, an application can be developed to extract the in-game behavior of a player and export it to the software tool for personality assessment and customizing software's visual interfaces.

ACKNOWLEDGMENT

The authors would like to thank GIK Institute for providing research facilities.

REFERENCES

[1] G. Sporiš, D. Vuleta, D. Vuleta, and D. Milanović, "Fitness profiling in handball: physical and physiological characteristics of elite players," *Collegium antropologicum*, vol. 34, no. 3, pp. 1009-1014, 2010.

[2] L. Lorusso and G. Boniolo, "Clustering humans: on biological boundaries," *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, vol. 39, no. 1, pp. 163-170, 2008.

[3] Z. Halim, R. Kalsoom, and A.R. Baig, "Profiling drivers based on driver dependent vehicle driving features," *Applied Intelligence*, vol. 44, no. 03, pp. 645-664, 2016.

[4] J.J. Ryan, X.Y. Dai, L. Zheng, "Psychological test usage in the People's Republic of China," *Journal of Psychoeducational Assessment*, vol. 12, no. 4, pp. 324–330, 1994.

[5] P.T. Barrett, "Rejoinder to: The Eysenckian personality structure: A "Giant Three" or "Big Five" model in Hong Kong?," *Personality and Individual Differences*, vol. 26, no. 1 , pp. 175–186, 1999.

[6] J.G. Rosse, M.D. Stecher, J.L. Miller, and R.A. Levin, "The impact of response distortion on preemployment personality testing and hiring decisions," *Journal of Applied Psychology*, vol. 83, no. 4, pp. 634, 1998.

[7] A.B. Bakker, K. I. Van Der Zee, K.A. Lewig, and M.F. Dollard, "The relationship between the big five personality factors and burnout: A study among volunteer counsellors," *The Journal of social psychology*, vol. 146, no. 1, pp. 31-50, 2006.

[8] G.A. Bonanno, "Loss, trauma, and human resilience: have we underestimated the human capacity to thrive after extremely aversive events?," *American psychologist*, vol. 59, no. 1, pp. 20, 2004.

[9] J.W. Lounsbury, J.M. Loveland, E.D. Sundstrom, L.W. Gibson, A.W. Drost, and F.L. Hamrick, "An investigation of personality traits in relation to career satisfaction," *Journal of Career Assessment*, vol. 11, no. 3, pp. 287-307, 2003.

[10] M.R. Barrick and A.M. Ryan, *Personality and work: Reconsidering the role of personality in organizations*, vol. 20, John Wiley & Sons, 2003.

[11] E. Paradis and G. Sutkin, "Beyond a good story: from Hawthorne Effect to reactivity in health professions education research," *Medical Education*, vo. 51, no. 1, pp. 31-39, 2017.

[12] R.M. Guion and C.J. Cranny, "A note on concurrent and predictive validity designs: A critical reanalysis," *Journal of Applied Psychology*, vol. 67, pp. 239–244, 1982.

[13] J. R. Meloy, "Indirect personality assessment of the violent true believer," *Journal of personality assessment*, vol. 82, no. 2, pp. 138-146, 2004.

[14] F.P. Morgeson, M.A. Campion, R.L. Dipboye, J.R. Hollenbeck, K. Murphy, and N. Schmitt, "Reconsidering the use of personality tests in personnel selection contexts," *Personnel psychology*, vol. 60, no. 3, pp. 683-729, 2007.

[15] K.C. Parker, R.K. Hanson, and J. Hunsley, "MMPI, Rorschach, and WAIS: A meta-analytic comparison of reliability, stability, and validity," *Psychological Bulletin*, vol. 103, no. 3, pp. 367, 1988.

[16] T. Hamby, W. Taylor, A.K. Snowden, and R.A. Peterson, "A meta-analysis of the reliability of free and for-pay Big Five scales," *The Journal of psychology*, vol. 150, no. 4, pp. 422-430, 2016.

[17] B.D. Coller, and M.J. Scott, "Effectiveness of using a video game to teach a course in mechanical engineering," *Computers and Education*, vol. 53, no. 3, pp. 900-912, 2009.

[18] S.T. Koenig, A. Dünser, C. Bartneck, J. C., Dalrymple-Alford, and G.P. Crucian, "Development of virtual environments for patient-centered rehabilitation," *Proc. of* International Conference on Virtual Rehabilitation (ICVR), pp. 1-7, 2011.

[19] G. Drettakis, M. Roussou, A. Reche, and N. Tsingos, "Design and evaluation of a real-world virtual environment for architecture and urban planning," *Presence: Teleoperators and Virtual Environments*, vol. 16, no. 3, pp. 318-332, 2007.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2017.2751602, IEEE Transactions on Affective Computing

[20] Z. Halim, A. R. Baig, and K. Zafar, "Evolutionary Search in the Space of Rules for Creation of New Two-Player Board Games," *International Journal on Artificial Intelligence Tools*, vol. 23, no. 2, pp. 1-26, 2014.

[21] D. Williams, N. Yee, and S.E. Caplan, "Who plays, how much, and why? Debunking the stereotypical gamer profile," *Journal of Computer-Mediated Communication*, vol. 13, no. 4, pp. 993-1018, 2008.

[22] N.C. Worth and A.S. Book, "Dimensions of video game behavior and their relationships with personality," *Computers in Human Behavior*, vol. 50, pp. 132-140, 2015.

[23] M. Bialas, S. Tekofsky, P. Spronck, "Cultural influences on play style," *Proc. of IEEE Conference on Computational Intelligence and Games*, pp. 1-7, 2014.

[24] G. Hofstede, G. Hofstede, and M. Minkov, *Cultures and Organizations: Software of the Mind*, 3rd ed. New York: McGraw-HilI, 2010.

[25] C. T. Burris, Christopher and K. Raif, "Make Believe Unmakes Belief?: Childhood Play Style and Adult Personality as Predictors of Religious Identity Change," *The International Journal for the Psychology of Religion*, vol. 25, no. 2, pp. 91-106, 2015.

[26] S.C. Brown and L.A. Mitchell, "An observational investigation of poker style and the five-factor personality model," *Journal of Gambling Studies*, vol. 26, no. 2, pp. 229-234, 2010.

[27] C. Symborski, G.M. Jackson, M. Barton, G. Cranmer, B. Raines, and M.M. Quinn, "The Use of Social Science Methods to Predict Player Characteristics from Avatar Observations," *Predicting Real World Behaviors from Virtual World Data,* pp. 19-37, 2014.

[28] S. Tekofsky, P. Spronck, A. Plaat, J. Van Den Herik, and J. Broersen, "Play style: Showing your age," *Proc. IEEE Conference on Computational Intelligence in Games*, pp. 1-8, IEEE, 2013.

[29] A. Bean and G. Groth-Marnat, "Video gamers and personality: A five-factor model to understand game playing style," *Psychology of Popular Media Culture*, vol. 5, no. 1, pp. 27, 2016.

[30] K. Liu, J. Tolins, J.E.F. Tree, M. Neff, and M.A. Walker, "Two techniques for assessing virtual agent personality," *IEEE Transactions on Affective Computing*, vol. 7, no. 1, pp. 94-105, 2016.

[31] L. Teijeiro-Mosquera, J.I. Biel, J.L. Alba-Castro, and D. Gatica-Perez, "What your face vlogs about: expressions of emotion and big-five traits impressions in YouTube," *IEEE Transactions on Affective Computing*, vol. 6, no. 2, pp. 193-205, 2015.

[32] B. Bostan, "A motivational framework for analyzing player and virtual agent behaviour," *Entertainment Computing*, vol. 1, no. 3, pp. 139-146, 2010.

[33] H.A. Murray, *Explorations in personality*, New York: Oxford University Press, 1938.

[34] J.R. Wright and K.Leyton-Brown, "Beyond Equilibrium: Predicting Human Behavior in Normal-Form Games," *Proc. of AAAI*, 2010.

[35] M.G. Lee, "Profiling students' adaptation styles in Web-based learning," *Computers and Education*, vol. 36, no. 2, pp. 121-132, 2001.

[36] T. Xiang and S. Gong, "Video behavior profiling for anomaly detection," *IEEE transactions on pattern analysis and machine intelligence*, vol. 30, no. 5, pp. 893-908, 2008.

[37] S. Tekofsky, P. Spronck, A. Plaat, H. van den Herik, and J. Broersen, "Psyops: Personality assessment through gaming behavior," *Proc. of the 25th Benelux Conference on Artificial Intelligence*, 2013.

[38] L. Wacquant, "A concise genealogy and anatomy of habitus," *The Sociological Review*, vol. 64, no. 1, pp. 64-72, 2016.

[39] P.A. Almiro, O. Moura, and M.R. Simões, "Psychometric properties of the European Portuguese version of the Eysenck Personality Questionnaire—Revised (EPQ-R)," *Personality and Individual Differences*, vol. 88, pp. 88-93, 2016.

[40] C.L. Keyes, K.S. Kendler, J.M. Myers, and C.C. Martin, "The genetic overlap and distinctiveness of flourishing and the big five personality traits," *Journal of Happiness Studies*, vol. 16, no. 3, pp. 655-668, 2015.

[41] T.W. Smith, and P.G. Williams, "Personality and Health: Advantages and Limitations of the Five-Factor Model," Journal of Personality, vol. 60, no. 2, pp. 395-425, 1992.

[42] P.J. Howard and J.M. Howard, *The Big Five Quickstart: An Introduction to the Five-Factor Model of Personality for Human Resource Professionals*. 1995.

[43] Z.Z.E. Wan and X. Xu, online game report, 2006, Pacific Epoch Red Innovation Report Series, 2006.

[44] B. S. Woodcock, MMOG subscriptions market share by genre, 2008.

[45] A. Ahmad and L. Dey, "A k-mean clustering algorithm for mixed numeric and categorical data," *Data and Knowledge Engineering*, vol. 63, no. 2, pp. 503-527, 2007.

[46] X. Jin and J. Han, K-medoids clustering, Encyclopedia of Machine Learning, pp. 564-565, 2011.

[47] W. Cai, S. Chen, and D. Zhang, "Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation," *Pattern Recognition*, vol. 40, no. 3, pp. 825-838, 2007.

[48] T. Loster and J. Langhamrova, "Disparities between regions of the Czech Republic for non-business aspects of labour market," *Proc. 6th International Days of Statistics and Economics*, pp. 689-702, 2012.

[49] K.D. Bollacker and J. Ghosh, "Mutual information feature extractors for neural classifiers," *Proc. of IEEE International Conference on Neural Networks*, 1996, vol. 3, 1996.

[50] T. Deng, Y. Huang, S. Yu, J Gu, C. Huang, G. Xiao, and Y. Hao, "Spatial-temporal clusters and risk factors of hand, foot, and mouth disease at the district level in Guangdong Province, China," *PloS one*, vol. 8, no. 2, pp. 1-9, 2013.

[51] M. Hoerger, M. Coletta, S. Sörensen, B.P. Chapman, B. K. Kaukeinen, X. Tu, and P.R. Duberstein, "Personality and Perceived Health in Spousal Caregivers of Patients with Lung Cancer: The Roles of Neuroticism and Extraversion," *Journal of aging research*, vol. 2016, pp.1-7, 2016.

[52] M. Michikyan, K. Subrahmanyam, and J. Dennis, "Can you tell who I am? Neuroticism, extraversion, and online self-presentation among young adults," *Computers in Human Behavior*, vol. 33, pp. 179-183, 2014.

[53] N.L. Muscanell and R.E. Guadagno, "Make new friends or keep the old: Gender and personality differences in social networking use," *Computers in Human Behavior*, vol. 28, no. 1, pp. 107-112, 2012.

[54] L.A. Jensen-Campbell, K. A., Gleason, R. Adams, and K.T. Malcolm, "Interpersonal conflict, agreeableness, and personality development," *Journal of Personality*, vol. 71, no. 6, pp. 1059-1086, 2003.

[55] H.J. Bernardin, D.K. Cooke, and P. Villanova, "Conscientiousness and agreeableness as predictors of rating leniency," *Journal of Applied Psychology*, vol. 85, no. 2, pp. 232, 2000.

[56] L.A. Witt, L.A. Burke, M.R. Barrick, and M.K. Mount, "The interactive effects of conscientiousness and agreeableness on job performance," *Journal of Applied Psychology*, vol. 87, no. 1, pp. 164, 2002.

Zahid Halim (M'2014) received the B.S. degree in Computer Science from the University of Peshawar, Pakistan, in 2004, M.S. degree in Computer Science from the National University of Computer and Emerging Sciences, Pakistan, in 2007, and also the Ph.D. degree in Computer Science from the National University of Computer and Emerging Sciences, Pakistan, in 2010. He was with the National University of Computer and Emerging Sciences, Islamabad, Pakistan, as a Faculty Member (Lecturer and then Assistant Professor) from 2007 to 2010. Currently he is an Associate Professor with Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Pakistan. His current research interests include machine learning and data mining, probabilistic/uncertain data mining, and human factors in computing. Dr. Halim is a member of the IEEE Computational Intelligence Society.

Muhammad Atif received the BE degree in Information Technology from A.Q Khan Institute of Computer Science and Information Technology, KRL Kahuta, Pakistan, in 2014 and the MS degree in Computer System Engineering from Ghulam Ishaq Khan (GIK) Institute of Engineering Sciences and Technology, Pakistan, in 2016. He is currently working as a Lecturer at the department of computer science, National University of Computer and Emerging Sciences, Pakistan. Atif's research interests include: data mining, computer vision and computational intelligence.

Ahmar Rashid received the B.S. degree in computer systems engineering from GIK Institute of Engineering Sciences and Technology, Pakistan, in 1998, the M.S degree in computer software engineering from National University of Sciences and Technology, Pakistan, in 2007 and the Ph.D. Degree in Electronic Engineering from Jeju National University, Republic of Korea, in 2011. After his PhD in 2011, he joined the faculty of Computer Sciences and Engineering, GIK Institute of Engineering Sciences Technology, Pakistan, and is currently working as an Associate Professor. His research interests include evolutionary algorithms to solve static and dynamic optimization problems, computer vision and localization algorithms.

Cedric A. Edwin is the Head of Incubation Centre and Assistant Professor of Business Management at Ghulam Ishaq Khan Institute of Engineering Sciences and Technology since 2014. He did his PhD from Liverpool Hope University, UK. He also holds BA (hons) in Business Management with IT from Liverpool Hope University. Prior to joining GIK, he was a Lecturer in Marketing at Liverpool Hope Business School and taught marketing, strategic management and international business modules to undergraduate and postgraduate students and supervised MS students. His current research interests include: entrepreneurship and innovation, technology management, leadership and personality traits and corporate social responsibility. He is a member of European Business Ethics Network.