# A GENERIC APPROACH FOR PLAYER MODELING USING EVENT-TRAIT MAPPING SUPPORTED BY PCA

M. Akif Gunes[0000-0002-4991-8248], M. Fatih Kavum[0000-0001-9223-2316], and Sanem Sariel[0000-0003-2993-6681]

Artificial Intelligence and Robotics Laboratory, Faculty of Computer and Informatics Engineering, Istanbul Technical University, Istanbul, Turkey
gunesme,kavum,sariel@itu.edu.tr
https://air.cs.itu.edu.tr/

**Abstract.** Modeling players based on their in-game events is essential for predicting their future behaviors. Player modeling studies mostly target a specific game or genre. This makes it difficult to transfer existing methods from one game to another. In this study, we propose a generic event-trait mapping and unsupervised learning approach for player modeling that extends our earlier modeling method with Principal Component Analysis (PCA). We present a case study of this new approach on the 10000 player data from World of Warcraft(WoW), a massive multiplayer online role-playing game (MMORPG). The base and the extended approaches are compared with an AutoEncoder (AE) based aproach on these data. The methods generate clusters as mixtures of different character traits. The results show that the extended approach is the best for player-modeling among these three methods.

**Keywords:** Player Modeling · Player Profiling · Event-Trait Mapping · PCA · World Of Warcraft

## 1 Introduction

The process of personality prediction in video games is referred to as player modeling. Player modeling not only helps with personality prediction but it is also an imperative tool in the game industry for monetary gain. It is key to adjust, improve, and develop products for different types of players to improve their satisfaction. The first and the most crucial step is to model players in a fast and reliable way either in real-time or during development. Player Modeling approaches may vary for games since one may wish to emphasize a distinctive trait of the game for a better result. Therefore, modeling different games with a common method is a challenge. Some generic models [16] use common aspects of games of different genre. They generalise game actions(events) to group them into smaller chunks and apply profiling for different genres, showing that a generic approach is plausile. However, for games with no common ground (no common group of events), this approach is not applicable.

In this study, we propose an unsupervised learning method which applicable to a variety of games genres. Our method builds on our earlier work, Event-Trait Mapping and Feature Weighting method (ET-FW) [10] and extends it (ET-PCA) with the use of Principal Component Analysis (PCA). ET-FW creates profiles as combinations of one or more personality traits in different proportions. We present a case study of ET-PCA on the World of Warcraft(WoW) [19], a massive multiplayer online role-playing game (MMORPG) that contains a wide spectrum of different in-game elements found in almost every game genre [6]. We compare the results with the base method and an AutoEncoder (AE) based learning method.

The rest of this paper is organized as follows. Section II presents a review of literature. The methodology is described in Section III. The case study on WoW data is presented in Section IV. This section includes the experimental results. Finally, concluding remarks are made in Section V.

## 2 Related Works

Halim et al. [11] explore the player profiles by using three different games with three different feature selection algorithms , four different clustering techniques and three different classifier methods. They use IPIP-NEO-120 personality tests on users to label them and select the traits as Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. In this work, manual labeling of players are needed by giving them personality tests in real life. Yee et al. suggest a clear traits list and pre-processed event data for WoW [19]. Their study includes forty-six events and the necessary calculation methods. This study has two objectives. The first of these is an investigation of events which can provide insights about a person's personality. The second objective is to probe the possibility of learning a person's personality just by studying their virtual behaviors. Worth and Book discover the relationships between the behaviors of humans in the real world and in the virtual world [18]. Brown and Mitchell used the NEO-PI-R questionnaire to measure personality tests based on five factors while measuring the relationship between personality and gambling style over poker games [13]. Wang et al.[17] present a behavior prediction method by using an auto-encoding neural network. They use traditional AE method as a pre-training step.Charles and Cowley use archetypal profiles other than trait profiles, while implementing their behavlet analytics method for profiling [5]. They also use domain-expert knowledge for creating the behavlets. New scenarios are created from the events, and these scenarios are used for profiling the player. We propose an Event-Trait mapping method that can cover these scenarios for mapping.

## 3 METHODOLOGY

The main goal in this research is to determine in-game personality traits from players' in-game events. Player behaviors are represented via traits, which can be thought of as classes. Players' trait scores are modeled using in-game events as features. Our work builds on our previous work, Event-Trait Mapping and Feature Weighting method (ET-FW) [10] which has shown to be successful in clustering users to traits in a casual mobile racing game called Dusk Racer. The first step of this clustering process is to determine event-trait relations. For this purpose, event-weight(EW), event-trait(ET), trait-event(TE) and user-event(UE) matrices are constructed to produce the final user-trait(UT) matrix.

- $EW$: Represents the weights of the preprocessed events. It is a diagonal square matrix, with the size of event $\times$ event. The weights can be either determined by a domain expert or generated with the extracted data. In this research, the inverse of frequencies of the events has been used for weighting.
- $ET$: It is a mapping table between in-game events and traits. This relation is determined by a domain expert. The scores are given from 0 to 5 where 0 means *not related* and 5 means *fully related*. One event can be mapped to more than one trait. As stated in the review study by Hooshyar et al. [12], mapping of in-game events to traits is carried out by experts in the majority of profiling processes. We also prefer this way when doing mapping.
- $TE$: Just like the event-trait matrix, this matrix is filled with the same intuition. This time, traits are mapped to events. This relation is determined by a domain expert. The size and dimensions of ET and TE are the same. Again, the scores are given from 0 to 5.
- $ET \circ TE$: Represents the final mapping. Note that the multiplication operation here is an element-wise operation called Hadamard product [7]. This gives a resulting matrix of the same size of ET and TE whose elements take on values from 0 and 25.
- $UE$: Represents the frequencies of events done by each user. Since the total number is different for each event, it is normalized and used.
- $UT$: This matrix shows the relation between users and traits. It is the final product, and computed by Equation 1.

$$UT = normalize(UE \times EW) \times (ET \circ TE) \tag{1}$$

After the computation of UT by Equation 1, the Expectation-Maximization (EM) method is applied to cluster users with this matrix. To see each persona trait ratio, mean trait values given by the EM method for each cluster is used.

Mapping traits to game events reveal different behavior groups of the players and their distributions, leading to a better understanding of the player-game interaction. In order to better distinguish these player groups, we further extend our previous work with Principal Component Analysis (PCA).

In traditional profiling methods, the game logs are reduced to a certain number of records using PCA, and then the players can be profiled by any clustering method [9], yet in complex games, these clusters are very difficult to interpret. Also, since PCA is an unsupervised learning method it determines dimensions of maximum variance without reference to class labels. To apply PCA, the sum-product of the trait and variance values of events in each PCA component is taken to create PCA component/Trait relation (Equation 2).

$$sp = \sum_{n=0}^{n} a_n * b_n \tag{2}$$

where $sp$ is the sumProduct value, $n$ is the index of events in the PCA component, $a_n$ is the variance of the particular event in that PCA component, and $b_n$ is the trait value taken from the ET matrix. Then, k-means clustering algorithm [3] is applied to the selected first-k PCA components to create clusters. Finally, cluster trait relations are obtained using this relation as the result of the ET-FW method by applying sum product operation between "PCA component/Trait" table and "PCA component/mean values of clusters" relations.

The following section presents a case study of this method on the WoW game.

# 4 A Case Study on Wow Data

## 4.1 Data Collection and Preprocessing

Player data used in this study are collected from a third party website, WoW-Progress [1]. More than 650000 guild information can be accessed here which represent in-game association of player characters in the game. Players collaborate to form teams, and they socialize and help each other in their guild. These guild data are categorized by the WoWProgress company according to language, realm and tier in their website. Blizzard, the original publisher of the World of Warcraft game, has also set up an API system to benefit the developers [2].

In order to access the detailed game logs of these players, 51 different API functions are called for each player, and more than 10000 events are obtained. The events which are not directly correlated to the traits in the raw data are eliminated, and the remaining events are preprocessed, yielding 175 events. In the preprocessing phase, some of the stats are rewritten as averaged percentages, or rates. [1] Because it is known that the player levels fluctuate greatly in terms of events, only 120 level players are taken into consideration for our experiments. Eventually, over 10000 players are registered in the database with 175 events.

## 4.2 Trait Selection

The selected traits in our study are mostly inspired from Bartle's [4] and Ferro et al.'s studies [8]. According to Bartle's study, personas are considered to have four main traits, namely; killers, achievers, socialisers and explorers. Ferro et al.'s paper illustrates the extended edition of Bartle's Player Type Graph. The selected traits from these studies for our case study are as follows: competitive, casual, explorer, grinder, social, craftsman, supportive, and DPS-player [2].

After the traits selected, Event-Trait mapping is performed. The mapping process is done by taking averages of the scores of 4 experts playing the game [3].

[1] Event explanations are available at: https://playerprofiling.github.io/WoWEvents/
[2] Trait explanations are available at: https://playerprofiling.github.io/WoWTraits/
[3] A sample Event-Trait table @see: https://playerprofiling.github.io/EventTrait/

### 4.3 Applying Profiling Steps

The Event-Trait Mapping and Feature Weighting method (ET-FW) is applied on the Event-Trait(ET) matrix and User-Event(UE) data. The resulting stacked graph for the trait clusters can be seen in Figure 1. The results show that there are 8 different clusters with 8 different trait ratios. The top three highest-valued traits are considered for cluster representation. For example, the first cluster in the Figure 1 involves the grinder trait as the most dominant trait amont other seven traits. Then, DPS-player and Crafter trait follows as the second and the third most dominant trait for that cluster.

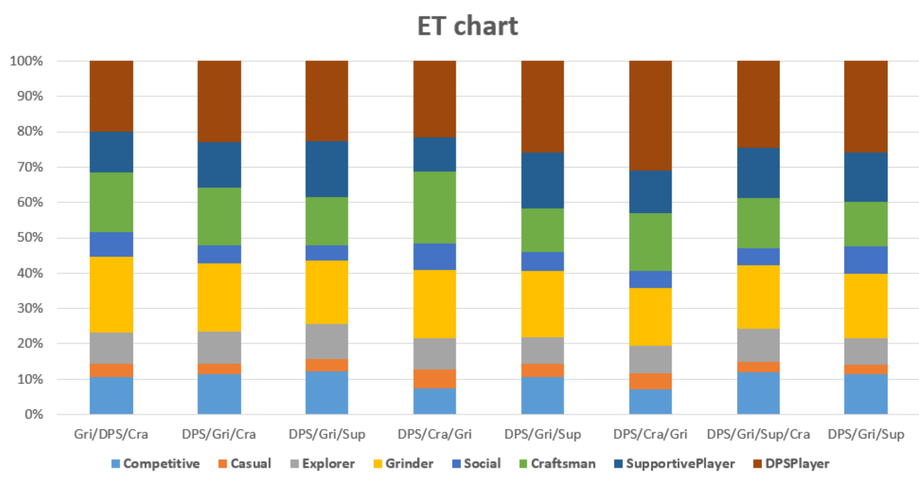


**Fig. 1.** The resulting stacked graph of the Trait-Cluster relation for ET.

| Weighted | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Event\Trait | Com | Cas | Exp | Gri | Soc | Craf | Sup | DPS |
| PCA Comp1 | 0.49 | -2.48 | 15.12 | 8.03 | 0.00 | 87.73 | 0.00 | 0.00 |
| PCA Comp2 | 21.12 | -4.95 | 20.55 | 16.28 | 0.00 | 31.66 | 10.60 | 5.51 |
| PCA Comp3 | -7.99 | -3.69 | 3.99 | -0.77 | 0.00 | 10.58 | 19.82 | 32.06 |
| PCA Comp4 | 4.56 | 3.99 | -3.01 | -1.13 | 0.00 | -8.46 | 28.61 | -2.94 |
| PCA Comp5 | 9.45 | 10.53 | 1.30 | 0.65 | 0.00 | 1.73 | -27.55 | -13.28 |
| PCA Comp6 | 17.29 | -0.12 | -8.33 | -2.60 | 0.00 | -8.67 | -6.12 | 12.55 |

**Table 1.** Trait values of PCA components.

Then, our new method, Event-Trait modeling supported with PCA (ET-PCA), is applied on the same $ET$ and $UE$ matrix. After that, a new matrix that shows the relationship between PCA components and clusters is created. Table 1 shows this relationship between the PCA components and the traits obtained by using PCA and k-means clustering methods. The sum product formula given in Equation 2 is used between these two relations to create the final stacked graph (Figure 2).

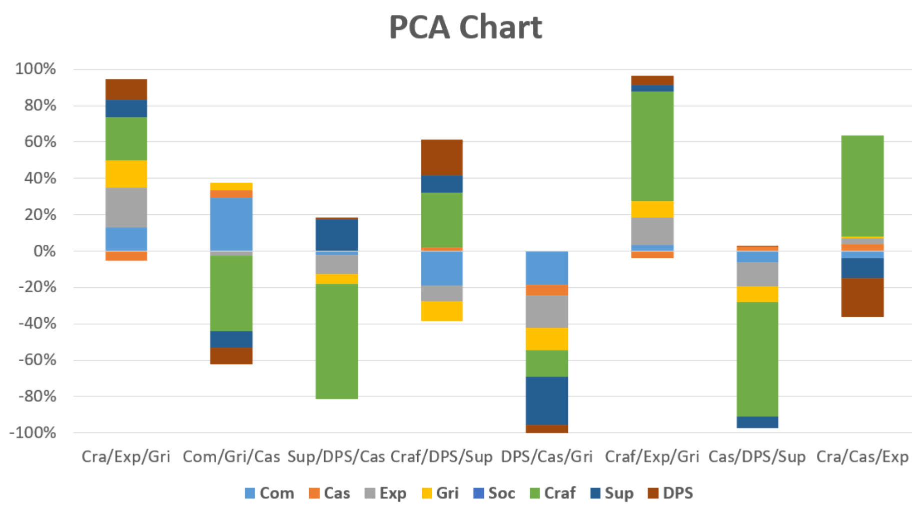


**Fig. 2.** Stacked Graph of Trait-Cluster Table for our new extended method

Finally, the AE method is applied to the WoW dataset and 175 preprocessed events reduced to eight encoded events. Creating a Trait/Cluster table similar to our methods is not possible with AE since existed AE methods are reconstructing the input nodes in the output so, we don't have intepretable relationships between original nodes and encoded nodes. For this reason, at the end of the profiling process, the players are clustered without having any information about their mixing rate of character traits, but only their cluster numbers. Leaky Relu [14] is used as the activation function. The Adam optimizer [20] with a 0.0005 learning rate is used. Mean Square Error algorithm is selected as the Loss function, and the program was run with 200 epochs and 256 batch sizes. The resulting loss value is 0.0184 after training.

### 4.4 Analysis of The Results

First, the personas (cluster results) are compared between the three methods according to the selected traits. Then, the quality of the clusters are compared by analyzing their trait ratios and using silhouette analysis.

**Evaluating methods with two different personas** In order to better explain this section, cluster IDs are assigned to the outcomes of the profiling methods.

| Competitor Trait | | | | | |
|---|---|---|---|---|---|
| **UserId** | **ET-FW** | | **ET-PCA** | | **AE** |
| | Cluster ID | Competitor Rank | Cluster ID | Competitor Rank | Cluster ID |
| 778 | 2 | 1 | 1 | 1 | 2 |
| 1448 | 2 | 1 | 0 | 2 | 2 |
| 888 | 2 | 1 | 0 | 2 | 2 |
| 4786 | 2 | 1 | 0 | 2 | 2 |
| 473 | 6 | 2 | 1 | 1 | 2 |
| 3873 | 1 | 3 | 1 | 1 | 4 |
| 5433 | 6 | 2 | 1 | 1 | 1 |
| 6273 | 6 | 2 | 0 | 2 | 2 |
| 2008 | 1 | 3 | 1 | 1 | 4 |
| 5240 | 6 | 2 | 0 | 2 | 5 |

**Table 2.** Top 10 competitor players and their assigned clusters for each method. The order of displaying the character trait relative to other clusters shown as rank. We can't report the rank for AE method since it reconstructs the input nodes in its output. So it is not possible to have a proper relationship between original and encoded nodes.

For ET-FW, ET-PCA and AE methods these clusterIDs are simple index numbers starting from 0 to 7. Also they are in same order as the clusters in the stacked chats given in Figure 1 and Figure 2.

Firstly, top ten players that have competitive character trait more than other ten-thousand players are selected by taking the players with the highest number of "ArenasPlayed", "DuelsWon" and "WorldHonorableKills" events. Table 2 reports the results for these players.

The top three clusters which have the highest competitive trait values for the ET-FW method are Cluster 2 (trait value: 130.53), Cluster 6 (trait value: 102.38) and Cluster 1 (trait value: 80.13). The order of displaying the competitor character trait relative to other clusters shown as rank. Hence, the competitor ranks for these clusters are 1, 2 and 3, respetively. The distribution is the same in Table 2 for the ET-FW method. So, for the competitive trait, the ET-FW method puts the selected players into the most related clusters.

For the ET-PCA method, the top three clusters that have the highest competitive trait values in the result are Cluster 1 (trait value: 33.63), Cluster 6 (trait value: 14.92), and Cluster 5 (trait value: 8.5). These three clusters are the only clusters that has the positive value for the competitor trait. Thus, it can be said that the other clusters are not related to the competitor trait. Non-related cluster information cannot be provided, since there is no negative value for ET-FW traits. By observing Table 2, it can be said that ET-PCA offers slightly better results while selecting players' clusters since it does not generate rank-3 clusters as does ET-FW.

For AE, six out of ten players are clustered in the same group (Cluster2), and others distributed separately in Clusters 4, 1, and 5. A cluster/trait relationship cannot be obtained for AE; therefore, this method cannot be evaluated like the other two.

| Grinder Trait (roll, acquired items) | | | | | |
|---|---|---|---|---|---|
| **UserId** | **ET-FW** | | **ET-PCA** | | **AE** |
| | Cluster ID | Grinder Rank | Cluster ID | Grinder Rank | Cluster ID |
| 9132 | 2 | 1 | 5 | 1 | 2 |
| 425 | 2 | 1 | 5 | 1 | 0 |
| 7069 | 2 | 1 | 5 | 1 | 0 |
| 3764 | 2 | 1 | 0 | 2 | 2 |
| 5423 | 6 | 2 | 5 | 1 | 0 |
| 6850 | 2 | 1 | 5 | 1 | 2 |
| 225 | 2 | 1 | 5 | 1 | 2 |
| 1448 | 2 | 1 | 0 | 2 | 2 |
| 480 | 2 | 1 | 5 | 1 | 2 |
| 3886 | 2 | 1 | 5 | 1 | 2 |

**Table 3.** Top 10 grinder players and their assigned clusters for each method. The order of displaying the character trait relative to other clusters shown as rank. We can't report the rank in AE due to its constraints (Table 2).

Users with grinder personas are selected by taking the players that have the highest number of "EpicItemsLooted" and "NeedRollsMadeOnLoot" events from the remaining ten-thousand people. Table 3 reports the results for these players.

The top three clusters that have the highest grinder trait values for ET-FW method are Cluster 2 (trait value: 190.62), Cluster 6 (trait value: 154.4) and Cluster 1 (trait value: 134.00). From Table 3, it can be said 9/10 of the selected personas are in the first 2 top ranked grinder clusters.

For the ET-PCA method, the top three clusters with the highest grinder trait values in the results are Cluster 5 (trait value: 22.07), Cluster 0 (trait value: 17.15) and Cluster 1 (trait value: 5.01). In addition, there is also cluster 7 which has the positive number for Grinder trait, too. The rest of the traits have negative values. From Table 3, it might be seen that the all of the selected players are in the first 2 top ranked grinder clusters.

For AE, seven out of ten players are in Cluster 2, and the other players are in Cluster 0. Again, the quality of clusters with respect to traits cannot be assessed, but it can be seen that AE could not cluster the players into top ranked grinder traits in the same group.

**Comparing Cluster Quality** Silhouette analysis is applied to the generated clusters for statistical analysis. The Silhouette Coefficient is a useful metric for evaluating clustering performance. The metric is computed using Equation 3:

$$sc = \frac{b - a}{max(a, b)} \tag{3}$$

where $a$ is the mean intra-cluster distance and $b$ is the mean inter-cluster distance to the closest cluster. The score ranges from -1.0 to 1.0, where higher the score, the better is the outcome. Table 4 reports this analysis. Overall results shows that, ET-PCA has better clustering results then that of ET-FW.

| Cluster | MSC (ET-FW) | MSC (ET-PCA) |
|---|---|---|
| cluster0 | 0.13 | 0.12 |
| cluster1 | -0.11 | 0.06 |
| cluster2 | 0.19 | 0.07 |
| cluster3 | 0.18 | 0.02 |
| cluster4 | 0.07 | 0.07 |
| cluster5 | 0.19 | 0.12 |
| cluster6 | 0.06 | 0.13 |
| cluster7 | -0.05 | 0.12 |
| Overall | 0.05 | 0.09 |

**Table 4.** Mean Silhouette Coefficient analysis results for ET-FW and ET-PCA.

## 5 CONCLUSION

In this paper, we present an Event-Trait Mapping based method extended with the use of PCA and a case study of this method on the WoW data. Persona clusters are created with this method as combinations and mixtures of different character traits. Two personas are selected for evaluating the resulting clusters on the WoW data. The selected top ten players' clusters are analyzed. The results indicate that the extended method (ET-PCA) groups players to the most related clusters. In addition, with our new extended method, persona's traits can have negative values. Having negative trait values in a cluster means that players does not show that speciality for a particular trait which is a good indicator. Furthermore, when the trait ratios of the clusters are observed on a stacked graph, it is seen that the Et-PCA modeling method gives more visible results.

For future work, automated methods are planned to be develoed for crating Event-Trait matrix. In the short term, a validation procedure is planned to correct human errors in creating matrices. Also, game data that can be shown as a provenance graph might be useful to apply these methods to get rid of some of the handcrafted works as mentioned in Melo et. al.'s [15] paper. Another research direction deals with which traits should be chosen.

## References

1. https://www.wowprogress.com/
2. https://develop.battle.net/
3. Ahmad, A., Dey, L.: A k-mean clustering algorithm for mixed numeric and categorical data. Data & Knowledge Engineering **63**(2), 503–527 , https://www.sciencedirect.com/science/article/pii/S0169023X0700050X
4. Bartle, R.: Hearts, clubs, diamonds, spades: Players who suit muds
5. Charles, D., Cowley, B.: Behavlet analytics for player profiling and churn prediction
6. De Simone, L., Gadia, D., Maggiorini, D., Ripamonti, L.A.: Design of a recommender system for video games based on in-game player profiling and activi-

ties. In: CHItaly 2021: 14th Biannual Conference of the Italian SIGCHI Chapter. CHItaly '21, Association for Computing Machinery, New York, NY, USA , https://doi.org/10.1145/3464385.3464742
7. Elrawy, A., Kishka, p., Saleem, M., Abul-Dahab, M.: On hadamard and kronecker products over matrix of matrices. General Letters in Mathematics **4**, 13–22
8. Ferro, L.S., Walz, S.P., Greuter, S.: Towards personalised, gamified systems: An investigation into game design, personality and player typologies. In: Proceedings of The 9th Australasian Conference on Interactive Entertainment: Matters of Life and Death. IE '13, Association for Computing Machinery, New York, NY, USA , https://doi.org/10.1145/2513002.2513024
9. Gow, J., Baumgarten, R., Cairns, P., Colton, S., Miller, P.: Unsupervised modeling of player style with lda. IEEE Transactions on Computational Intelligence and AI in Games **4**(3), 152–166
10. Gunes, M.A., Solak, G., Akin, U., Erden, O., Sariel, S.: A generic approach for player modeling using event-trait mapping and feature weighting. In: AIIDE
11. Halim, Z., Atif, M., Rashid, A., Edwin, C.A.: Profiling players using real-world datasets: Clustering the data and correlating the results with the big-five personality traits. IEEE Transactions on Affective Computing **10**(4), 568–584
12. Hooshyar, D., Yousefi, M., Lim, H.: Data-driven approaches to game player modeling: A systematic literature review. ACM Comput. Surv. **50**(6) , https://doi.org/10.1145/3145814
13. Karlsson-Brown, S., Mitchell, L.: An observational investigation of poker style and the five-factor personality model. Journal of gambling studies / co-sponsored by the National Council on Problem Gambling and Institute for the Study of Gambling and Commercial Gaming **26**, 229–34
14. Li, Y., Yuan, Y.: Convergence analysis of two-layer neural networks with relu activation
15. Melo, S.A., Kohwalter, T.C., Clua, E., Paes, A., Murta, L.: Player Behavior Profiling through Provenance Graphs and Representation Learning. Association for Computing Machinery, New York, NY, USA , https://doi.org/10.1145/3402942.3402961
16. Shaker, M., Shaker, N., Togelius, J., Abou-Zleikha, M.: A progressive approach to content generation. pp. 381–393
17. Wang, S., Wang, H., Gao, Q., Hao, L.: Auto-encoder neural network based prediction of texas poker opponent's behavior. Entertainment Computing **40**, 100446 , https://www.sciencedirect.com/science/article/pii/S1875952121000434
18. Worth, N.C., Book, A.S.: Dimensions of video game behavior and their relationships with personality. Computers in Human Behavior **50**, 132–140 , https://www.sciencedirect.com/science/article/pii/S0747563215002460
19. Yee, N., Ducheneaut, N., Nelson, L., Likarish, P.: Introverted elves & conscientious gnomes: The expression of personality in world of warcraft. vol. 2011, pp. 753–762
20. Zhang, Z.: Improved adam optimizer for deep neural networks. In: 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS). pp. 1–2