A game analytics model to identify player profiles in singleplayer games

Luís Fernando Bicalho, Augusto Baffa and Bruno Feijó
ICAD/VisionLab - Departamento de Informática
Pontifícia Universidade Católica do Rio de Janeiro
Rio de Janeiro, Brazil
lfernando@icad.puc-rio.br, abaffa@inf.puc-rio.br, bfeijo@inf.puc-rio.br

***Abstract*—Data Science is a multidisciplinary area related to systems, methods, and processes to extract knowledge from a high volume of data. In this context, we use the term *game analytics* to designate the science of online analysis and metrics of games. Research works in this area have been focusing on the use of player behavior data to increase revenue and avoid users leaving the game too early. To help this behavior analysis, we created a classification method based on Richard Bartle's player types model, mixed with the definition of Casual and Hardcore players, resulting in eight archetypes: Casual and Hardcore Killers; Casual and Hardcore Achievers; Casual and Hardcore Socializers; and Casual and Hardcore Explorers. We used the first four types from this new model in a singleplayer *shoot'em up* game, which gathers players' behavior attributes during each match. The right profile is chosen using K-means and Decision Tree algorithms, based on data from previous gameplay sessions. This whole method was tested using two new questionnaires to match the player's profile evaluation with the game's final profile, revealing accuracy between 75% and 80%.**

***Keywords*-Telemetry; Game Analytics; Player Modeling; Machine Learning;**

I. Introduction

The game industry has been continuously growing during the last few years [1] [2], creating a highly competitive industry. This fact makes it difficult for new games, mainly independent ones, to stand out in comparison to the already established franchises. As stated in a study by Kyle Orland [3], 37% of the games registered on Steam, an online game store, have not even been uploaded once by registered users.

This very competitive scenario makes the game industry consumers seek for the most addictive, high quality, and innovative games [4]. Thus, game designers need better strategies to create an attractive gameplay, and a way of creating a game with high *replay value* (i.e., a form to create a compelling video game experience that keeps players coming back multiple times).

One of the well-known strategies is to use game data to analyze player behavior, focusing on improving specific gameplay characteristics [5], attracting older and newer player types. This research strategy is directly related to *game analytics*, which is an area focused on data analysis and metrics in games [6].

Most of the works on player behavior analysis use Bartle's Taxonomy [7], which is a system for classifying video game players in four groups: killers, achievers, socializers, and explorers [8] [9]. However, our work proposes a new approach to this taxonomy, defining a new axis that relates to player's dedication level. In this new approach, we add casual and hardcore types, which extend the model to eight types. For instance, in our system, a Socializer can be characterized as a Casual Socializer or a Hardcore Socializer.

Moreover, most of the works use unsupervised techniques, mainly K-means, to determine which player type fits better for a specific user, based on his/her mapped characteristics [10] [8] [11]. However, these approaches usually get the data from an already finished game, not having an algorithm ready for improvements in mechanics or difficulty.

That's why we decided to use a combination of K-means and Decision Tree algorithms, which would determine the player's classification during gameplay. At the end of each game session, we register this classification, using it as part of the training set for the next K-means executions. This approach allows us to change the game mechanics and difficulty, having new centroids and decision trees for each newly added session.

Our goal is to test this methodology in a *shoot'em up* singleplayer game, developed using *Unity Game Engine*, which we used because it is a popular engine in the industry and has extensive documentation [12]. Furthermore, we mapped the player's behavior through two questionnaires, letting us test the quality of our classification method.

In the present work, we are not interested in making comparisons with other player type models or fundamental game categories, such as the Keirsey model [13], DGD2 model [14], BrainHex model [15], and the Caillois' categories [16].

This paper is organized as follows. The section *Related Works* presents the works that needed to be studied or read, with special focus on three different groups, with all of them using game telemetry. The section *Main concepts* presents our algorithms and models, while the section *Methodology* shows how we used and tested our proposed method. The section *Results* summarizes our procedures and presents the final classification accuracy. Finally, *Conclusion* presents contributions and future works.

II. Related Works

In our literature review, we divided all related works into three main groups. The first group uses telemetry to analyze players behaviors, taking them as feedback for evaluating the current game design, which allows designers to make the gameplay better. The second group is very similar to the first but ends up using non-supervised algorithms to classify players in different clusters. This second approach allows the game designers to analyze these clusters and make changes to game mechanics to enhance the experience. Lastly, the third group uses adaptive algorithms, which changes the game difficulty or the design based on data gathered by telemetry. In this section, we present the works we considered most influential to our approach.

As a representative of the first group, Gagné et al.'s works [17] [18] present data analysis of a Real-Time Strategy (RTS) game, called *Pixel Legions* (Pixelante, 2011). To allow more precise visualization of the data for each game session, they created Pathways, a framework that shows player's squad movement trajectory and the moment of each death. This tool helped them answer questions related to the player's behavior between matches and within each match. However, the data collected was noisy and focused on each match instead of the player itself. Moreover, the framework, as said in their conclusion, "does not support selection, filtering, or remapping color to different variables within the data which makes the analysis of a single session's data require modifying the program code."

The works by Drachen et al. [10] [19] are representatives of the second group. The first one analyzed data from two AAA games: *Battlefield: Bad Company 2* (EA Dice, 2010), an action shooter, and *Tera Online* (Bluehold Studios, 2011), an MMORPG. They evaluated playing performance by gathering nine different types of information from each player, being also different for each game. This first work improves from Gagné et al. (op. cit.) not only on the volume and type of data but also on the machine learning algorithms used for clusterization (*K-means* and *SVM*).

The second work [19] compared different methods for the clustering of nearly 70,000 *World of Warcraft* (Blizzard Entertainment, 2004) players, based on their leveling speed and online playtime data. They concluded that only Archetypal Analysis (AA) and k-means provided basis vectors that are intuitively interpretable in terms of the behavior of the players. Also, they showed that AA provided more varied basic vectors than the other methods. Nevertheless, in the present work, we decided to use the k-means method because it allocates players directly to groups (via cluster centroids), while AA requires a second analysis step. Furthermore, the superiority of AA would require more investigation with a variety of different situations.

Odierna and Silveira [8] also used the *World of Warcraft* game to classify players, based on the *World of Warcraft Avatar History* dataset, containing data from more than 90,000 game avatars. Their goal was to allow the game developers to create more exciting game features, based on player types that are related to Bartle's archetypes. Although their work contributes to making the game design better, it does not involve AI methods that process the data and avoid game designer's rework. They would also benefit from changing some of Bartle's archetypes to fit more specifically in their game genre, or just differentiating the casual players from the hardcore ones.

Other projects used telemetry allied with machine learning algorithms, like Eggert et al. [20], in which they presented a vast dataset from recorded Dota 2 (Valve Corporation, 2013) matches, to train their AI models. The goal was to compare which algorithm could best classify players by the genre's most common roles, with *logistic regression* being the overall most stable and efficient algorithm. However, as a case study, they did not use a more generic classification (like Bartle's), which narrowed their contribution and focusing on the MOBA game genre.

Finally, Baére and Feijó [21] is the work of the third group that influenced us the most. They measured players' satisfaction after playing two versions of a shoot'em up game, one with adaptive difficulty and the other without it. Their approach compared casual to hardcore players while using a simple algorithm, based on three player models: Easy, Medium, and Hard. Although each model had its limit values for players' accuracy, the number of deaths and the number of defeated enemies, their adaptive game had more effect on hardcore players than on casual ones.

III. Main Concepts

Some works in the second group of our literature review use unsupervised algorithms to classify each instance of the collected data in different groups [8] [11] [22]. Like other works of this second group, they choose centroids clustering algorithms, mainly *K-means*, because of its popularity and simplicity, and because it is considered a baseline technique [23]. Therefore, we chose *K-means* as the primary unsupervised method of our system. Other reasons that support our choice, as stated by C. Bauckhage et al. [23], are related to our data being numerical and not sparse, and our overall goal being to build a behavior model. This model is based on the centroids that result from the mentioned algorithm.

Nevertheless, we wanted to label each centroid, applying known archetypes to them. To make that possible, we based ourselves in the four Bartle's Archetypes [7], shown in Fig. 2, to create our behavior model. However, to make this classification update regularly during the gameplay, we needed a supervised algorithm, which would receive each instance of our data already labeled by the previous method.

For this purpose, we chose the *decision tree* algorithm, as it is the fastest compared to others like *KNN* and *Naïve Bayes* [24], and because of its high comprehensibility. It

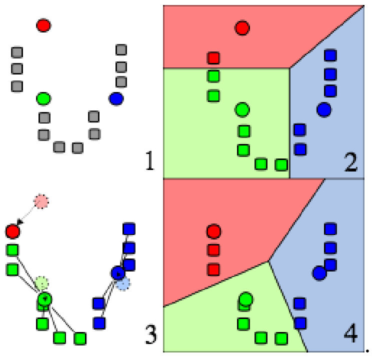

Figure 1. K-Means four main steps.

allows programmers and designers to understand the classification steps quickly, knowing, for instance, which attribute was more decisive for each group [25].

Thus, this section presents the characteristics of the algorithms of *decision tree* and *K-means* used in our system. Firstly, instead of implementing them, we decided to use *Accord.NET* framework, written in C# programming language [26], to execute these classification and clustering methods efficiently. Secondly, in this section, we explain each archetype of Bartle's model and how we used them to create our classification method, which considers the notions of hardcore and casual gamers.

A. K-means

K-means is an unsupervised technique that only receives unlabeled training sets, making predictions from the attributes of each point [27]. This algorithm places each point in one of the K clusters, or groups, according to specific criteria. It works in three main steps [28]:

1) Choose K random centroids (points in the given attributes' domain) to represent the median of each cluster (Fig. 1, step 1);
2) Place each data point in the cluster with the nearest median, resulting in K separated clusters (Fig. 1, step 2), on the form of a Vornonoi diagram [29];
3) Each algorithm iteration runs through the whole data set, re-positioning its median based on the cluster values stored (Fig. 1, step 3). This re-positioning is repeated a determined number of times, resulting in clusters, like in the fourth step on Fig. 1, containing the points that are similar to each other.

B. Decision Tree

Decision Tree is the most successful and one of the most straightforward learning algorithm, being easy to implement and serving as an excellent introduction to supervised learning [30]. We can consider it a supervised algorithm because it receives a set of labeled actions as training data and makes predictions for all unseen points [27]. In the context of our project, the labeled actions are the mapped player attributes during each gameplay session, while the prediction is the player classification based on the proposed taxonomy.

To explain the logic behind the algorithm, firstly we need to use the concept of information entropy, which is defined as the value that quantifies uncertainty, i.e. the value of a choice [31]. The entropy value varies from zero to one, and is calculated using (1), in which: $E(S)$ is the current's situation (S) entropy; p_i is the probability of case i; n is the total number of cases; and b is the logarithmic base, representing the number of results.

$$E(S) = -\sum_{i=1}^{n} p_i \times log_b(p_i), E(S) \in [0, 1] \tag{1}$$

To decide which of the attributes is the most relevant, we need to calculate their entropy gain towards the initial one, shown in (2). This calculation allows us to choose, comparatively, the attribute (A) that has the more significant entropy gain (G).

$$G(A) = E(I) - E(A) \tag{2}$$

C. Bartle's Taxonomy

Richard Bartle, a researcher in the massive multiplayer online (MMO) game industry, helped creating the first *Multi-User Dungeon* (MUD), also known as *Multi-User Dimension* or *Multi-User Domain* in a more generic approach [32]. It is a text-base virtual world environment, in which players could interact with each other and play different adventures together [33]. However, since 1991, the concept of MUD was expanding to non-game applications, which brought the question if sports, pastimes and other types of entertainment could be considered as MUDs.

The answer to that question was brought by Bartle [7], in 1996, through the definition of a taxonomy, allowing to classify players in four different player types: *Achievers*, *Killers*, *Socializers*, and *Explorers*. This helped defining MUD as a "game" in which different types of players could interact with each other, or with themselves. For instance, the act of cooking is considered a MUD, as its "user" fits on the *Explorer* classification [7].

MUDs were also the precursor of the MMO game genre, which represents, allied with the MOBA sub-genre, 25% of the game market revenue in 2017, with the perspective to grow almost 50% until 2021 [34]. These facts make the taxonomy defined by Bartle still relevant, for both multiplayer and singleplayer virtual games.

In the Bartle's model (Fig. 2), there are two central axes: Player-World, and Action-Interaction, representing the source of players' interest. The horizontal axis goes from the extremely player-oriented gameplay to the world-oriented one, while the vertical axis goes from the action-oriented gameplay to the interaction-oriented gameplay. To fully

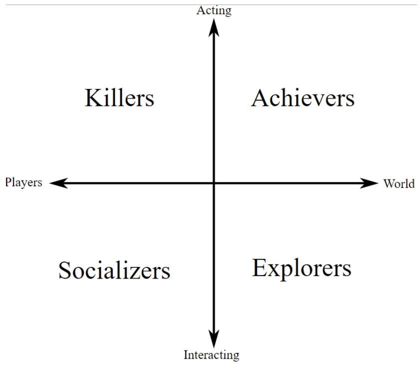

Figure 2. The Bartle Taxonomy.

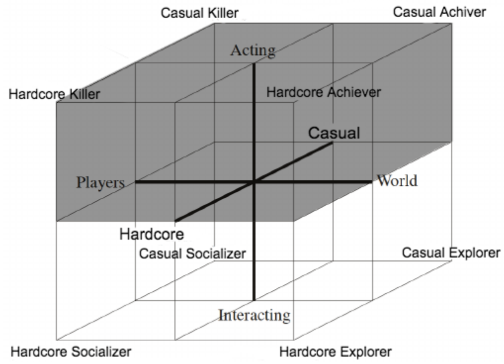

Figure 3. The proposed expansion of the Bartle Taxonomy to include casual/hardcore players. The shadowed region is the particular case of singleplayer shoot'em up games.

understand the difference between each orientation, we need to know the definition of each archetype, as shown below:

1) *Achievers* are focused on mastering the game, on the rewards it has to offer. They share the world with other players, or non-playable characters (NPCs), and add a competitive element to the environment. Therefore, they are proud of their status in the game hierarchy, and how fast they reached their current level.
2) *Killers* are focused on acting on other players, or NPCs, most of the time showing their superiority over them. They seek more power and abilities, that can help them affect others. Therefore, they are proud of their level of authority and their fighting skills.
3) *Socializers* are focused on interacting and talking with other players, or NPCs. Also, finding more about other people is more interesting for socializers than competing, or bossing them. Therefore, they are proud of the relationships and of their influence towards other players.
4) *Explorers* are focused on interacting with the world, the game environment. The sense of discovery or finding new areas and game elements fulfills them more than just achieving a great status in the game. Therefore, they are proud of their knowledge and of searching for new places and possibilities.

We quickly find four-fold models of personality and player types in gaming that seem to be compatible among them, such as Bartle (*op. cit.*), Keirsey [13], and DGD2 [14] models. Furthermore, there are pieces of evidence that these classifications are reduced to two or three categories depending on the game genre (see the insightful discussion on this subject in Gamasutra community of industry and game design experts [35]). Therefore, we propose to reduce the Bartle model to two-player types for the particular case of singleplayer shoot'em up games: *Achievers*, who focus on collecting items and coins, and *Killers*, who focus on killing enemies.

D. The Proposed Expansion

These two types do not seem enough to contemplate all player characteristics and behaviors. To solve this problem, we added a new axis representing the player's dedication to the game genre in question, as shown in Fig. 3. This axis allows us to create two more types: Casual Players and Hardcore Players.

Casual players are the ones who play video games for fun, or to take a quick break from the daily obligations. They usually play on mobile, or on web pages, during coffee break or work intervals [36] [2]. They also usually play casual games, defined as simple to play, easy to learn, and straightforward to get rewards, turning the gameplay into an enjoyable experience [37] [38]. But there is still a difference between playing casual games and playing games casually [38]. In this paper we will consider a casual gamer the ones who play games casually, independent from the game genre.

Meanwhile, the Hardcore players are the one's who dedicate a big part of their time to play different game genres, having a good knowledge of the industry, and spending more money than the average users [39] [2].

These definitions of casual and hardcore players, however, are getting older and obsolete, as stated by Satoru Iwata, ex-president of Nintendo, who thinks the meanings are wider [40]. Thereupon, Adams and Ip [41] stood away from the binary distinction between casual and hardcore players, proposing a gamer dedication evaluation method, based on fifteen qualities [42] [41], which are described below.

1) Technologically savvy - Dedicated gamers usually have more interest in new technologies, mainly the ones related to the game industry.
2) Have the latest high-end gear - Dedicated gamers will be up to the latest hardware news on the industry, acquiring the best consoles and computers. They usually own, or have owned, older game platforms.
3) Willingness to pay - Dedicated gamers usually don't wait for the promotions or special offers to get the

games they want. They need to play it, and even buy products related to their favorite franchises.

4) Prefer violent/action games - Dedicated gamers prefer games that are more violent and action-packed than the market average.
5) Prefer games that have depth and complexity - Dedicated gamers prefer games that challenge their knowledge, that make them spend more time trying to beat or master it.
6) Play games over many long sessions - Dedicated gamers play regularly and usually spend hours in a single session.
7) Hunger for gaming-related information - Dedicated gamers constantly search for the latest game news, previews and reviews, also looking forward to interviews with industry experts, game magazines, books and strategy guides.
8) Discuss games with friends online - Dedicated gamers love to discuss about the game industry trends and news, mostly through forums and social media groups.
9) Play for the exhilaration of defeating (or completing) the game - A dedicated gamer will play a game for the pleasure of beating it, defeating difficult enemies. They usually care more about good game mechanics than graphics, acting or even the story, forgiving the possible flaws from this game characteristics.
10) Much more tolerant of frustration - Dedicated gamers don't abandon games because of frustration. They are used to play difficult and challenging games, helping them to mitigate this possible frustration.
11) Engaged in competition with himself, the game, and other players - Hardcore gamers want to feel happy when rewarded after beating a difficult challenge, or by improving their skills. This also allows him/her to compete against other players and/or computer-controlled opponents. For instance, less dedicated gamers wouldn't spend time mastering a character, or learning all his combos, in a fighting game like Mortal Kombat (NetherRealm Studios, 2011).
12) Age at which first started playing games - If someone has aged constantly playing video games, since he/she was younger, he/she can be considered an experienced and dedicated gamer;
13) Comparative knowledge of the industry - Dedicated gamers are likely to have more knowledge of the game industry trends and new technologies, not just because of their will to search for this type of information, but also because they play various game genres;
14) Early adoption - Dedicated gamers seek for over midnight game releases, pre-orders and beta or alfa test events, being one of the first with access to new games;
15) Desire to modify or extend games in a creative way - Dedicated gamers sometimes feel the need to create mods, altering from graphic, character skins, to creating new game modes, when the game offers a customization level. For instance, *Defense of the Ancients* (DotA) started as a mod of Warcraft III (Blizzard Entertainment, 2002), and helped creating the MOBA game genre, widely know for games like DOTA2 (Valve Corporation, 2013) and League of Legends (Riot Games, 2009) [43] [44].

Figure 4. *Space Shooter* screenshot.

This continuum of player characteristics served as the base to create a new gamer dedication axis on the player types graphics, allowing every combination of casual and hardcore types with the original four archetypes of players. The result was a three dimensional graphic, as shown in Fig. 3, with: Hardcore Achiever; Casual Achiever; Hardcore Killer; Casual Killer; Hardcore Explorer; Casual Explorer; Hardcore Socializer; and Casual Socializer. Nevertheless, in this paper we are using just the superior part, containing the variants of Achievers and Killers.

IV. Methodology

A. *Space Shooter: A Case Study*

To test the proposed player type model and the machine learning algorithms, we created a *shoot'em up* game, called *Space Shooter*, developed using the Unity Game Engine [12]. This type of game is a sub-genre of two-dimension (2D) action games, which have, as their principal challenge, innumerable hordes of different enemies shooting at the players direction [42].

To fight back its foes, the player can choose from two different spaceships, focusing on the precision of its shots, on dodging bullets, and on collecting items and coins. The main goal is to get a good score, always trying to overcome the highest one registered. One screenshot of the game is shown in Fig. 4.

The game is singleplayer, i.e. it can be played by just one user at a time, and has only one stage, that must be completed until the end. This allowed us to compute each game session as a new value to be used in the algorithm's training set. Whereas, we define twelve attributes, listed below, that are updated every half a second (0.5 second is the time interval).

A0) Number of direction changes (Mean);
A1) Position in X axis (Mean);
A2) Position in Y axis (Mean);
A3) Total time in movement (Mean);
A4) Number of items collected (Total);
A5) Number of coins collected (Total);
A6) Number of destroyed enemies (Total);
A7) Percentage of game completed (Total);
A8) Number of shots (Mean);
A9) Number of shots on target/enemies (Mean);
A10) Number of shots without enemies (Mean);
A11) Number of shots taken or Number of lives lost (Total);

At the end of each session, these attributes are stored, separated by commas, as a new line on the training set file. This file list is read every time a new session starts, serving as a knowledge base for the K-means algorithm to calculate the four new centroids.

In our experiment, although all the attributes seemed valuable, the last generated decision tree did not use three of them: A3, A4, and A11. If we exam the tree nodes in Fig. 6, we shall notice this elimination.

B. Classification Sequence

Firstly, the K-means algorithm has $K = 4$, as each cluster represents one of the four archetypes previously proposed. After reading the training set file, it stores the four resulting centroids in a array, but without the archetypes labels yet.

To associate each archetype to its correspondent cluster, we decided to “look” at these points in the attributes space, searching for characteristics that would differentiate one from another. Therefore, we chose the attributes A5, total number of coins collected, and A6, total number of enemies defeated, as the most relevant.

This decision was not made arbitrarily. We first tested the game using fixed generic names for each cluster (cluster 0, cluster 1, cluster 2 and cluster 3), associating them as labels for each training set line. This allowed us to pass this labeled list to the Decision Tree algorithm, which generated a new tree.

In the latest tests, this tree was assuming characteristics similar to the the ones seen in Fig. 6, which was based on the last test, with a training set of 138 instances. Thus, we can notice that its more decisive attributes are A6, A5 and A7. As A7 did not seem relevant to decide an archetype, we chose to focus on A5 and A6, being the most important, respectively, to determine an Achiever and a Killer.

Now we could iterate in the centroids array, searching which one of them had the bigger A5 value. This first one would be the “Hardcore Achiever” centroid. Then we would search, among those remained, which one had the bigger A6 value. This second one would be the “Hardcore Killer” centroid. This same process was repeated for the rest of the array, choosing, respectively, the “Casual Achiever” and the “Casual Killer” centroids.

Figure 5. Example of *Space Shooter’s* final screen.

Having all the training dataset labeled with the archetype names, the decision tree algorithm will always generate a new tree to choose between them. Then, the current player attributes will be passed to this renewed tree, resulting in the current classification. This one will change every period of time, depending on the current attribute values, and would result in a final classification at the end of the game session, as shown in Fig. 5.

C. Questionnaire Design

To test if the classification shown in the end of the completed game is compatible with the player’s profile, we designed two questionnaires. The first verifies if the player is classified as an Achiever or a Killer. It was based on the work by Schneider et al. (2016) [45], which presents a questionnaire containing twenty questions, resulting in a percentage for each player type.

Their approach differs from the usual *Bartle Test of Gamer Psychology* [46] [47], as it does not have binary questions forcing the player to fit in a profile (e.g. one answer indicates an achiever profile and the other a socializer one). They use, instead, the same five answers for every question:

- “I do not understand/I do not identify myself” (0 points);
- “I identify myself a little” (1 point);
- “I identify myself partially” (2 points);
- “I identify myself” (3 points);
- “I identify myself totally” (4 points).

Each answer has a weight related to it, making it more difficult to have different people choosing the same one. This approach is very similar to the Likert scale, as shown by Joshi et al. [48]. Moreover, the player who does not identify him/herself with any answer scores 0% in every profile, which is a more honest and precise evaluation [45].

The questions are also different from the usual Bartle’s Test, having five questions to identify each player type (total of 20). We only use ten of them, as we considered Achievers and Killers only. The following list shows the proposed questions [45]:

- Achiever

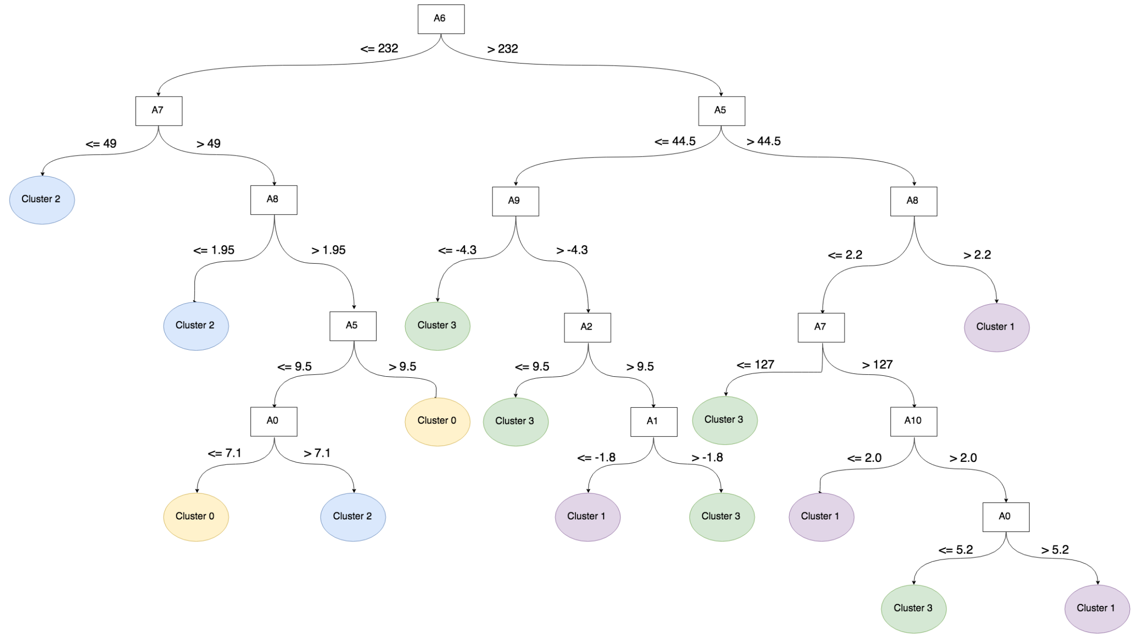

Figure 6. Decision Tree generated by *Accord.NET* algorithm.

  - "I like to conquer new badges in games";
  - "I get impressed with players that conquered high rewards";
  - "I play electronic games until the end with 100% of achievements";
  - "I love new items and medals";
  - "I like exposing my achievements (for example, on Facebook)".
- Killer
  - "I am very competitive in games";
  - "I like exploding things in games";
  - "My favorite games are first person shooters";
  - "I am known for my aggressiveness in games";
  - "I do not like talking in games, what I really like is shooting".

To decide whether the player is an Achiever or a Killer, we decided to sum the points related to the questions of each archetype, and get the maximum value from their result, as shown in (3). If the sum result is equal for both types, the player is classified as both, lowing the chances of the game classification being wrong. This also happens, for instance, if the player is defined as 55% Killer and 45% Achiever, i.e. he/she is classified as both if the distance between both Killer and Achiever percentage is below or equal to 10 percentage points.

$$PT = \max\left(\sum_{i=1}^{5} A_i, \sum_{j=1}^{5} A_j\right), (A_i, A_j) \in [0, 4] \qquad (3)$$

The second questionnaire focuses on identifying if the game user is a Casual or a Hardcore player. To measure his/her dedication, we used the previous cited definition of a hardcore player, on the fifteen characteristics presented in Section III-D. Thus, we created the following questions (associated with each characteristic, respectively):

- "I always deal with technology and seek for new releases and trends" (7 points);
- "I like to have the latest high-end computers/consoles" (7 points);
- "I'm willing to pay anything for a game" (5 points);
- "I prefer violent/action games" (1 points);
- "I prefer games that have depth and complexity" (3 points);
- "I play games over many long sessions" (10 points);
- "I always search for the game industry latest information" (6 points);
- "I frequently talk about games, both via social media and with people" (10 points);
- "I always feel happy when completing (or defeating) a game" (7 points);
- "I don't get easily frustrated while playing a game" (9 points);
- "I am usually engaged in competition with myself, the game, and other players" (6 points);
- "I started playing games when I was little" (2 points);
- "I have played all the types of game genres, and I constantly compare one game to another" (10 points);
- "I buy games and consoles on their pre-release, or import them from other countries to be one of the first

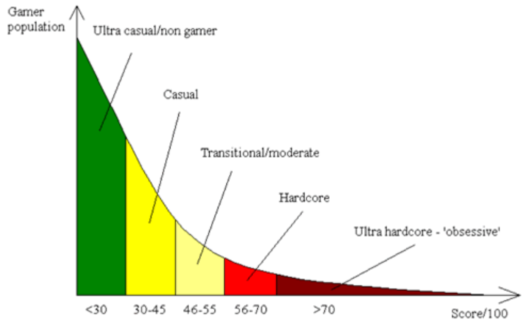

Figure 7. Casual and Core by gamer dedication.

to play” (9 points);

- “I think of modifying and extending some of the games I play” (8 points);

To answer each question, we repeated the same method used in the first questionnaire, with those five weighted responses. Besides that, we can notice that each question is also weighted, as we based ourselves on the work by Adams and Ip [41]. This method allows us to give more importance to some questions, when compared to others.

To quantify the player dedication, we used (4), in which: A_i represents the answer weight for question i; Q_i represents the weight for question i; and GD is the gamer dedication factor, which is represented by the sum of the multiplication of both question and answer weights, divided by 4 multiplied by the weights, representing the maximum points the user can make. This results in a percentage, that is interpreted according to Fig. 7, as shown in the list below, considering Non-gamers as Casuals, and Ultra Hardcore gamers as Hardcores.

1) Casual gamer - Has GD factor below or equal to 45%;
2) Moderate gamer - Has GD factor between 45% and 55%, with these limits included;
3) Hardcore gamer - Has GD factor above 55%.

$$GD = \frac{\sum_{i=1}^{15} A_i \times Q^i}{\sum_{i=1}^{15} 5 \times Q^i} \tag{4}$$

V. Results

The training data collected came from 138 game sessions, but the questionnaires to test the algorithms classification accuracy, were answered by 43 people, with various positioning towards the game industry.

The data was manipulated with Python language, more specifically using Pandas library, generating three graphics. The first one (Fig. 8) shows that our game was able to classify using all four archetypes, with more Hardcore than Casual players, maybe because of the overall game difficulty not being so high, compared to other *shoot'em ups*. Moreover, when considering Achievers and Killers, we had almost 58% of the testing group classified as the first, and 42% classified as the second archetype.

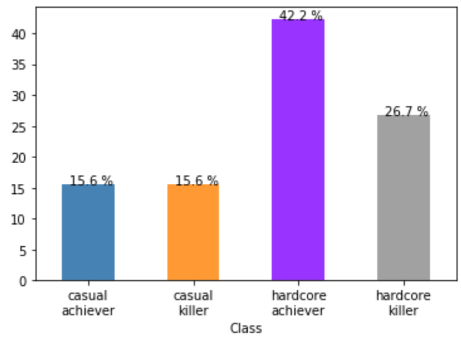

Figure 8. Total of each archetype found on game sessions.

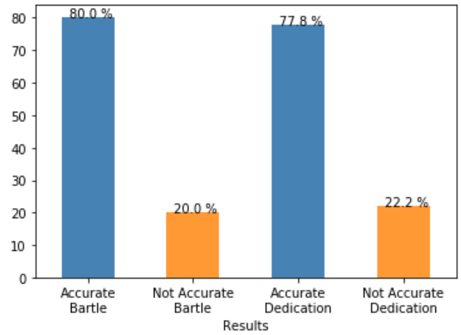

Figure 9. Accuracy results for both dedication and Bartle archetypes.

Even the classification having a good variation, we needed to test its accuracy towards the players' real profile. So, we compared the game results with the ones based on the questionnaires criteria. The only reservation we had was if the player was classified as Moderate by the gamer dedication questionnaire. In this case, any game classification, Casual or Hardcore, was considered accurate, as the player is in the range between both types.

Following this criteria, the results were very similar for both gamer dedication and Bartle's archetypes, with accuracy between 75% and 80%, as shown by Fig. 9. This 20% not accurate can be explained by the generic results presented by the questionnaires, which maps the player profile in a general context, not considering his/her inexperience towards the *shoot'em up* game genre, for example.

VI. Conclusion

Our research work belongs to the second group of game analytics works we identify in section II of this paper, which uses telemetry strategies allied with unsupervised algorithms to classify the players behavior. Nevertheless, we present some valuable contributions to that group. Firstly, we decided not only to integrate this classification into the game itself (saving the player attributes as part of the training set) but also to use a supervised algorithm to classify the player while he/she plays the game. This strategy resulted in a more precise final classification. Secondly, we proposed an extension to the original Bartles model, adding a new axis representing the gamer dedication level of the player. This extension enhanced the interpretability of our classification. Thirdly, we designed simple questionnaires that facilitate the survey work. As an overall contribution, our approach revealed a classification algorithm with accuracy between 75% and 80%. What is even a more valuable contribution, we presented a complete strategy (from telemetry to testing procedures) which helps game designers to analyze clusters of users and make changes to game mechanics to enhance the players experience.

Even if the game content has been changed or updated, our algorithm is capable of adapting itself in the long run, as the training set continues to grow. For instance, if the difficulty of an action game is changed, the future number of hardcore players reduces in comparison to the previous classifications.

Moreover, our approach allows the creation of an adaptive model, which would base itself on the player classification to vary the difficulty of the game, changing parameters like enemy speed, the number of enemies on screen, enemy fire rate, player fire rate, and other relevant gameplay characteristics. In the end, this paper presents a combination of different approaches, showing that they work well together.

There are many future works to do. We should investigate if player type scores do significantly predict player experience. Also, future work is needed to use a larger and more representative sample, representing a broader range of ethnic and socioeconomic groups of players. Another required investigation is to test our approach in other game genres to cover other portions of the classification graphics in Fig. 3. Also, we should investigate some questions towards the use of different supervised and unsupervised algorithms, the creation of other questionnaires, and better game design guidelines based on telemetry.

Acknowledgments

This work was partially supported by CNPq (National Council for Scientific and Technological Development, linked to the Ministry of Science, Technology, and Innovation), CAPES (Coordination for the Improvement of Higher Education Personnel, linked to the Ministry of Education), FINEP (Brazilian Innovation Agency), and ICAD/VisionLab (PUC-Rio).

References

[1] T. Wijman, "Newzoo global games market report 2019," Newzoo, Tech. Rep., 2019. [Online]. Available: https://newzoo.com/solutions/standard/market-forecasts/global-games-market-report/

[2] D. Middelhoff and D. Schunk, "Gaming industry - facts, figures and trends," Clairfield International, Tech. Rep., 2018.

[3] K. Orland, "Introducing steam gauge: Ars reveals steams most popular games," https://arstechnica.com/gaming/2014/04/introducing-steam-gauge-ars-reveals-steams-most-popular-games/, 2014, acessed: 03/07/2019.

[4] S. Wright, "There are too many video games. what now?" https://www.polygon.com/2018/9/28/17911372/there-are-too-many-video-games-what-now-indiepocalypse, 2018, acessed: 10/07/2019.

[5] N. Statt, "How artificial intelligence will revolutionize the way video games are developed and played," https://www.theverge.com/2019/3/6/18222203/video-game-ai-future-procedural-generation-deep-learning, 2019, acessed: 24/05/2019.

[6] M. S. El-Nasr, A. Drachen, and A. Canossa, *Game Analytics: Maximizing the Value of Player Data*. Springer Publishing Company, Incorporated, 2013.

[7] R. Bartle, "Hearts, clubs, diamonds, spades: Players who suit muds," 06 1996.

[8] B. Odierna and I. Frango Silveira, *Player Game Data Mining for Player Classification*, 05 2019, pp. 52–61.

[9] T. Hynninen, "Identifying player profiles in video games," Ph.D. dissertation, Lappeenranta University of Technology, 2016.

[10] A. Drachen, R. Sifa, C. Bauckhage, and C. Thurau, "Guns, swords and data: Clustering of player behavior in computer games in the wild (pre-print)," 09 2012.

[11] T. Mahlmann, A. Drachen, J. Togelius, A. Canossa, and G. Yannakakis, "Predicting player behavior in tomb raider: Underworld," 09 2010, pp. 178 – 185.

[12] T. Deals, "This engine is dominating the gaming industry right now," https://thenextweb.com/gaming/2016/03/24/engine-dominating-gaming-industry-right-now/, 2016, acessed: 15/05/2019.

[13] R. Caillois, *Please Understand Me II: Temperament, Character, Intelligence*. Prometheus Nemesis Book Co., 1998.

[14] C. Bateman, R. Lowenhaupt, and L. Nacke, "Player typology in theory and practice," *Proceedings of DiGRA 2011 Conference: Think Design Play*, 05 2012.

[15] L. Nacke, C. Bateman, and R. Mandryk, "Brainhex: Preliminary results from a neurobiological gamer typology survey," vol. 6972, 10 2011, pp. 288–293.

[16] R. Caillois, *Man, Play, and Games*. University of Illinois Press, 2001.

[17] A. R. Gagné, M. El-Nasr, and C. Shaw, "A deeper look at the use of telemetry for analysis of player behavior in rts games," 10 2011, pp. 247–257.

[18] ——, "Analysis of telemetry data from a real-time strategy game: A case study," *Computers in Entertainment (CIE)*, vol. 10, 10 2012.

[19] A. Drachen, C. Thurau, R. Sifa, and C. Bauckhage, "A comparison of methods for player clustering via behavioral telemetry," *Proceedings of the 8th International Conference on the Foundations of Digital Games*, 07 2014.

[20] C. Eggert, M. Herrlich, J. Smeddinck, and R. Malaka, "Classification of player roles in the team-based multi-player game dota 2," 09 2015, pp. 112–125.

[21] B. Baère Pederassi Lomba de Araujo and B. Feijó, "Evaluating dynamic difficulty adaptivity in shoot'em up games," 10

2013.

[22] A. Drachen, A. Canossa, and G. Yannakakis, "Player modeling using self-organization in tomb raider: Underworld," 10 2009, pp. 1 – 8.

[23] C. Bauckhage, A. Drachen, and R. Sifa, "Clustering game behavior data," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 7, pp. 266–278, 09 2015.

[24] A. Ashari, I. Paryudi, and A. Min, "Performance comparison between naïve bayes, decision tree and k-nearest neighbor in searching alternative design in an energy simulation tool," *International Journal of Advanced Computer Science and Applications*, vol. 4, 12 2013.

[25] S. Kotsiantis, "Supervised machine learning: A review of classification techniques," *Informatica (Ljubljana)*, vol. 31, 10 2007.

[26] "Accord.net framework," http://accord-framework.net/, 2017, acessed: 15/05/2019.

[27] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning*. The MIT Press, 2012.

[28] D. J. C. MacKay, *Information Theory, Inference & Learning Algorithms*. New York, NY, USA: Cambridge University Press, 2002.

[29] F. Aurenhammer, "Voronoi diagrams - a survey of a fundamental geometric data structure," *ACM Comput. Surv.*, vol. 23, no. 3, pp. 345–405, Sep. 1991. [Online]. Available: http://doi.acm.org/10.1145/116873.116880

[30] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed. Upper Saddle River, NJ, USA: Prentice Hall Press, 2009.

[31] C. E. Shannon, "A mathematical theory of communication," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 5, no. 1, pp. 3–55, Jan. 2001. [Online]. Available: http://doi.acm.org/10.1145/584091.584093

[32] H. Hahn, *The Internet Complete Reference (2Nd Ed.)*. New York, NY, USA: McGraw-Hill, Inc., 1996.

[33] K. Stuart, "Richard bartle: we invented multiplayer games as a political gesture," https://www.theguardian.com/technology/2014/nov/17/richard-bartle-multiplayer-games-political-gesture, 2014, acessed: 09/06/2019.

[34] "Massively multiplayer online (mmo) and multiplayer online battle arena (moba) games market size worldwide in 2017 and 2021 (in billion u.s. dollars)," https://www.statista.com/statistics/830090/mmo-moba-market-revenue/, 2017, acessed: 09/06/2019.

[35] B. Stewart, "Personality and play styles: A unified model," https://www.gamasutra.com/view/feature/134842/personality_and_play_styles_a_.php?page=1, 2011, acessed: 23/05/2019.

[36] N. Taskiran and R. Yilmaz, *Handbook of Research on Effective Advertising Strategies in the Social Media Age*, 02 2015.

[37] E.-J. Waugh, "Gdc: Casual games summit 2006: An introduction to casual games," https://www.gamasutra.com/view/feature/131030/gdc_casual_games_summit_2006_an_.php, 2006, acessed: 29/06/2019.

[38] J. Kuittinen, K. Annakaisa, J. Niemel, and J. Paavilainen, "Casual games discussion," pp. 105–112, 11 2007.

[39] BITMOB, "Core and casual: Whats the difference?" https://venturebeat.com/2011/04/30/core-and-casual-whats-the-difference/, 2011, acessed: 30/06/2019.

[40] S. Iwata and S. Miyamoto, "Nintendo e3 2011: Deeper and wider," http://iwataasks.nintendo.com/interviews/index.html?disableNav=true/#/e32011/newhw/0/6, 2011, acessed: 30/06/2019.

[41] E. Adams and B. Ip, "From casual to core: A statistical mechanism for studying gamer dedication," https://www.gamasutra.com/view/feature/131397/from_casual_to_core_a_statistical_.php, 2002, acessed: 20/05/2019.

[42] E. Adams, *Fundamentals of Game Design*, 3rd ed. Thousand Oaks, CA, USA: New Riders Publishing, 2014.

[43] M. Walbridge, "Analysis: Defense of the ancients - an underground revolution," https://www.gamasutra.com/php-bin/news_index.php?story=18863, 2008, acessed: 03/07/2019.

[44] M. Fahey, "From warcraft obsession to game creation," https://www.kotaku.com.au/2009/02/from_warcraft_obsession_to_game_creation-2/, 2009, acessed: 03/07/2019.

[45] M. O. Schneider, . T. U. Moriya, A. Vieira da Silva, and J. C. Nto, "Analysis of player profiles in electronic games applying bartles taxonomy," 09 2016.

[46] J. Mulligan, B. Patrovsky, and R. Koster, "Developing online games: An insider's guide," 01 2003.

[47] M. Barr, "The bartle test of gamer psychology," https://matthewbarr.co.uk/bartle/, 2017, acessed: 23/06/2019.

[48] A. Joshi, S. Kale, S. Chandel, and D. Pal, "Likert scale: Explored and explained," *British Journal of Applied Science & Technology*, vol. 7, pp. 396–403, 01 2015.