Improving Players' Profiles Clustering from Game Data Through Feature Extraction

Luiz A. L. Rodrigues and Jacques D. Brancher
Department of Computer Science
Londrina State University - UEL
Londrina, Brazil
luiz_rodrigues17@hotmail.com, jacques@uel.br

***Abstract*—The study of players profiles is an emerging area which commonly uses demographic characteristics as determinant features. However, there is the need for a deeper understanding that these characteristics alone do not provide. Another problem is the use of predefined profiles that might be oversimplified or too embracing. This paper investigates players profiles based on their gameplay data, besides demographics, using an educational math game as a testbed. Thus, problems such as noise, mixed data types and high dimensionality must be tackled. To this end, we investigated two feature extraction methods to mitigate these difficulties, Principal Component Analysis (PCA) and Features Agglomeration (FA). Then, two unsupervised learning algorithms were used to find the profiles in our experiments, showing that both PCA and FA improved clustering performance, wherein the best results indicated four profiles: advanced, skilled, beginners and intermediated. Our findings provide game designers with insights about playing styles, can be used to adapt the game in real-time and to assess how distinct players profiles perform in the educational subject, as well as their playing performance.**

***Keywords*-player typologies; player types; player classes; educational game; math game;**

I. Introduction

The literature on player typologies (types, profiles, classes or categories) is yet an emerging field and criteria such as players' background (e.g. age and gender) are often used [1]. Players' age and gender have an important role in playing behavior and motivations, which have been highlighted in previous researches, and might also be used to define players' classes [2]. In spite of that, gender alone is not responsible for player enjoyment and there is the need for a deeper understanding of player profiles, especially for educational games [3].

Predefined categories, such as the famous taxonomies of Bartle [4] and Yee [5], have been used to stipulate players categories. However, Bartle's four types of players were designed to a specific game genre (Multi-Player Dungeons/Domains, MUD), not to be a general typology. Despite that, it has been often used out of its originally designed context, even though it suffers from a number of weakness [2]. Yee's model [5] was developed based on a long-term quantitative study, however, it is also based on a specific type of game, Massively Multiplayer Online Role-Playing Games (MMORPGs). It also defines players' classes through questionnaires, which in parts were based on Bartles, allowing it to be considered a refinement and empirical grounding of Bartle's types. This might be considered as valid as there is a historical link between MMORPGs and MUDs.

Nevertheless, these largely embracing, predefined, types might be oversimplified, missing important information, whereas it requires clarification and support by data to be useful [1]. In Drachen et al. [6], the limitations of predefined classes are also discussed, where it is stated that these methodologies for segmentation should be checked with actual game data, using unsupervised cluster analysis to validate the findings. Through this process, it is possible to define players types that might be viewed as *personas*, rather than defining them as an absolute playing preference.

The notion of finding players' types, specifically based on game data, has received significant attention by the community of designers, game developers and related fields [7], [8], [9], [10], [11], [12]. This type of research might be referred to as *in the wild*, considering that the behavioral data analyzed was gathered out of the researchers' control (e.g. game interaction or purchase behavior) [6]. Unsupervised machine learning techniques, often clustering algorithms, are methods applied to solve this problem. However, the use of these techniques requires expertise in both the game being evaluated and on the cluster analysis, since the selection of which technique to use involves goals of analysis, available data and context [13].

In addition, there are several challenges that need to be addressed [6], such: mixed data types in a single dataset (e.g. numeric, categorical and binomial values) that might be necessary features; noisy datasets [14]; algorithms that requires parameters selection (e.g. number of clusters to be found/used [15]); datasets of high dimensionality (e.g. number of features bigger than number of instances); and provide results that are easily interpretable to stakeholders (e.g. developers or designers), which might not be specialists on these analysis. Another problem is that, although the potentials of clustering algorithms to analyze game behavioral data, the field is still in its infancy [13].

Moreover, Bauckhage et al. [13] states that the definition of guidelines for which approach to use in this context is difficult, considering the variety of game designs, questions to be answered and hundreds of models available. They further claim that future works on game analytics are necessary. Based on the aforementioned context, the goal of this paper is to identify players profiles based on data from an educational math game, SpaceMath[1].

The use of in-game behavior can be viewed as problematic since it is usually based on a single game or genre, which limits the possible behaviors. It shows the behavior of its users that originated through the game mechanics available [1], which is of interest to this research. Thus, we adopted behavioral data, along to demographics, to identify players' profiles in an educational math game that lead its users to practice arithmetic operations. This is a key aspect of this research as this type of game have shown many benefits to its users [16], [17], [18], [19]. This will allow us to determine how different types of players play the game, providing insights to improve its design and may also be used to provide real-time feedback to improve a specific user's experience (e.g. adapt game content to promote experiences suitable to specific playing styles) and, thus math training.

A dataset with 199 rows (players) and 21 columns (features) was available to this research, where 9 of these columns were related to players' demographics (e.g. genre, age and school year) and 12 originated through players interactions with the game (e.g. average score, time and amount of levels played). Seeking to reduce data's noise, dimensionality and, thus improve clustering results, we experimented with two feature extraction methods: Principal Component Analysis (PCA) and Features Agglomeration (FA). Thus, three versions of our dataset were analyzed. To identify the profiles, two clustering algorithms were applied to each dataset version, K-means (KM) clustering and Hierarchical Clustering (HC). To validate the developed models and find the best setups, two metrics suitable for cases wherein the ground truth is unknown were used: Silhouette Coefficient (SC) and Calinski-Harabaz Index (CHI). Thus, the main contributions of this paper might be summarized as: (1) the identification of players' profiles, using both behavioral and demographic features, in an educational gaming context, (2) an analysis of the characteristics of these profiles considering both behavioral and demographic data and (3) clustering results improvement through feature extraction.

The remainder of this paper firstly discuss related researches (Section II) and presents the testbed game used (Section III). Then, we describe how players' profiles were identified and evaluated (Section IV), and show the results of performed experiments (Section V). Finally, we discuss our findings, draw conclusions and implications, as well as future directions of researches (Section VI).

II. Related Works

In Schneider et al. [20] an analysis of players' profiles according to Bartle's Taxonomy is presented. To accomplish this, players responded to the Bartle's Test and then played a general purpose game. A sample of 25 volunteers participated in the experiment, where the majority of them were classified as *achievers* and *explorers*. They found unexpected results from the identified profiles, which was argued to be due to the intersection of them. Also, they claim that, for the sake of game customization, analyze the behavior within the game is a more reliable approach than to take general profiles.

An investigation of players' profiles is presented by Calife and Nakamura [21]. Although they adopted the same taxonomy as Schneider et al. [20], they use a heuristic model based on player's behavior instead of the Bartle Test. In addition, they compare the identified profiles through this model against clusters found by the KM algorithm. Their goal with these segmentation procedures was to understand how players that invested real money in the game are distributed among these groups. While the Bartle's Taxonomy showed that a large amount of buying players are on the interaction side, KM found a cluster with the highest conversion rate. Thus, they claim that combining information from both approaches can help to find players that might increase the game revenue.

Drachen et al. [12] used a commercial game, *Tomb Raider: Underworld*, to find players' types on a large dataset (1365) of users that finished the game. Through an emerging self-organization map, four types were found and labeled as: *Veterans*, *Solvers*, *Pacifists* and *Runners*. Their findings showcase that the evaluated game is played with the flexibility provided and that their approach allows the identification of surprising or unwanted behavior. Thus, the results are valuable to game designs, enabling them to know how the game is being played, and to dynamically use it in order to adjust the game according to player's needs and skills.

Siqueira et al. [8] used another commercial game as testbed, *World of Warcraft* (WoW). Their goal was to extract players' profiles, besides a model to predict the probability of a player renewing its subscription. Extracted profiles were: *Beginners*, *Intermediate I*, *Intermediate II* and *Professional*. The developed model allowed them to find that *Professionals* and *Intermediate II* renewed their subscription, whereas *Beginners* and *Intermediate I* did not. Furthermore, their analysis revealed important players metrics to aid the game company in user-oriented testing and to detect which players are more likely to maintain their subscription.

Also using the game WoW, Thurau and Drachen [11] introduced the use of Archetypal Analysis (AA) to define classes of players, comparing it to other four algorithms. AA describe players in terms of their relation to basis vectors, whilst KM, for instance, does that according to centroids. Thus, while KM represents averages and assigns each player to a specific group, AA find extreme classes of players defining each player's label according to the nearest basis

[1] spacemath.rpbtecnologia.com.br

vector.

In Drachen et al. [6], an adaptation of the AA, Simplex Volume Maximization (SIVM), and KM clustering are applied on data from two different games, *Tera* and *Battlefield 2: Bad Company 2*. They demonstrate how these algorithms yield different but complementary results, where KM provides reliable insights on players' general distribution (e.g. cluster with low performance) and SIVM detects subversive players (e.g. gold farmers or cheaters). Furthermore, this approach also demonstrates how to incorporate design's knowledge of data exploration through features selection and on the analysis process. Also, it presents a description of profiles to aid in their interpretability.

The proposal of Francillette et al. [10], an ongoing research, is based on two types of data - information received by players and captured by their behavior inside the game - alike this research. They explore players' profiles based on their style and skills to produce a matchmaking system. Using DBSCAN, a density-based clustering algorithm, allowed them to stipulate the number of players (cluster size) and the similarity of them within each match through the algorithm's parameters, considering players' success rate and interaction frequency to game mechanics. However, the impact of their approach on players' experience was not evaluated.

Benmakrelouf et al. [9] explored players' characteristics and performance to identify their profiles using a combination of regression and clustering. A serious game was used as a testbed, where they found that features such as the number of access to the game and quests visited were the most influential to score and retention (duration of the game). Thereafter, they were able to find three distinct playing behaviors, which originated the following profiles of players: *Beginners*, *Intermediate* and *Advanced*. These insights are valuable to guide personalization in serious games, which is an important factor in the motivation and effectiveness that these games provide.

Halim et al. [7] uses four clustering techniques (KM, K-medoids, fuzzy c-mean, and HC) as part of a process of players' personality evaluation in three RTS games datasets. These techniques were evaluated using the Davies-Bouldin Index, Dunn Index, and SC. This was made after the features selection process was applied, which included the following techniques: PCA, Mutual Information and Random Sub-set Feature Selection. Thereafter, clustering showed that HC overcame the other algorithms in all evaluation metrics for every dataset. However, it presented unbalanced clusters (one cluster containing a single sample), which thus, led the researchers to present their descriptive analysis based on the second best algorithm, K-medoids. Furthermore, their findings provided reliable results to the personality prediction, however, the discussion of this process is not within the scope of this paper.

In this section, we identified works that are based on questionnaire data or fitting players clusters in predefined ones. These approaches suffer from the limitation of too embracing or oversimplified profiles [1]. This paper addresses these limitations by searching for profiles on behavioral and demographic data. On the other hand, there researches that also follows this approach, however, the majority of them investigate behavioral data only, ignoring players demographics. Also, they are mostly focused on general purpose games, wherein the main focus is to provide insights into playing styles from the entertainment perspective. Our research differs from these by investigating an educational math game, where profiles will aid not only from the entertainment perspective but also from the educational. Furthermore, we perform feature extraction previous to clustering players data to determine the profiles. This aims to improve clusters separation, a procedure which is scantily employed in these applications, as well as explicitly presenting clustering results according to their separation (e.g. SC measure). This is also addressed in this paper where two measures are used to justify parameters and final results selection.

III. SpaceMath

This section presents a brief description of SpaceMath, a game which encourages its players to practice math, more specifically, the four arithmetic basic operations.

Inside the game, the user incorporates an astronaut, which is in a parallel dimension of space, and must solve arithmetic puzzles to escape from it. At each level, an arithmetic challenge is presented, which the player must correctly solve it (in up to 90 seconds) to advance to a next parallel dimension. To correctly solve the challenge, the astronaut must explore the scenario until it finds all the pieces that form the challenge's answer (numbers between 0 and 9). Once all pieces are collected in the correct order, the player must dive into the portal to advance to the next level. If the answer is incorrect, the astronaut is teleported back to the beginning.

Different types of boxes are distributed in the scenario. They might hide pieces of the challenge's answer; aliens that can not touch the player (or s/he is teleported to the begin again); or just occupy space to disturb players path. Initially, these boxes are also used to guide players towards the correct answer, using different colors to advise them. For example, one could identify that a green box is hiding the dozen and that the unit is behind a pink one if the answer has exactly two numbers. However, after reaching a 10 wins streak, it advances to the second difficulty level and, from there on, every box contain the same color (white). Furthermore, the astronaut has a portals device, which should be used to protect itself from aliens and boxes, teleporting them to another unknown parallel universe. However, this device is not able to create portals that can save itself.

Moreover, every scenario and math puzzle is procedurally generated in real-time. Thus, each time a challenge is answered and the astronaut enters the portal, s/he is teleported to a new automatically created parallel dimension. This leads to an endless game, requiring the player to explore several scenarios and, therefore, to practice several arithmetic problems. Note that describing these generation processes is not the scope of this paper.

Fig. 1 demonstrates a randomly selected level as example. In order to solve the math challenge/puzzle (top middle), the player needs to collect the second answer's piece (7) - the first (1) was collected already - and enter on the portal (bottom right) without being touched by the alien (middle right). Boxes at the figure's left are hazards which might be empty or hiding other aliens. As mentioned, considering that this puzzle's answer has two numbers (1 and 7), the first piece was below a green box and the second below a pink one. Furthermore, at the figure's top left there are information about the player's status (nickname, current score, shots and time available, and the current streak of explored universes without a loss), all in Portuguese. On the top right there are keys that can be used while playing: keyboard directional arrows or A, S, D and W to move the avatar; spacebar to activate the portal device; backspace to drop back a collected number; and return to confirm game's message (e.g. after a player's loss).

IV. Study

This section describes the research process performed in this paper. Firstly, we introduce the analyzed dataset, its preparation process, and the sample's characteristics. Thereafter, we present the conducted experiments, describing the clustering methods, parameters selection, validation metrics and, finally, the setup of the experiments.

A. Dataset

Before starting to play SpaceMath, users were required to create an account, using a unique login, in order to identify their demographic characteristics. This account's data, together with some performance statistics (in-game metrics), originated our dataset columns (features). Thus,

Figure 1. Randomly selected level of SpaceMath. The player already collected one piece of the answer (1) and, to advance to the next level, must to collect the other (7) and then enter on the portal (bottom right).

our original dataset contained 21 columns, representing each player's features. Also, the dataset was composed by 199 rows, one for each player, where all of them played at least 10 game levels. Table I presents a brief description of every dataset's feature.

As can be seen in the table above, nine features (rows one to nine) are demographic data, whilst the remaining 12 were originated by user interactions to the game. The demographic features seek to capture players personal characteristics (Age, Genre, HasNet), educational environment (SchoolType) and situation (SchoolYear), playing games habits (PlayingTime and Gamer), and affinity with math (LikesMath and KnowsMath). In-game metrics have the goal to point out users playing styles (i.e. scores, amount of shots fired, time and amount of levels played, sequence and total of wins, and winning percentage according to each type of arithmetic challenge).

Note that the testbed game is available online for free access from anyone with a computer having internet access. It has been online for approximately three months at the time of writing this paper. Its disclosure was made through email lists, social networks, and suggestion to colleagues, teachers, and professors. Thus, the analyzed dataset was constructed in no rigid way, as the game might be played at home, school activities or work intervals for example. This

Table I. Dataset's features description. Types of features are abbreviated to: Continuous = Cont.; Boolean = Bool; and Categorical = Catg.

Feature	Type	Description
		Demographics
Age	Cont.	User's age
Genre	Bool	Whether the player is male (1) or female (0)
HasNet	Bool	Internet access from home (1) or not (0)
SchoolType	Catg.	municipal (0), public (1), federal (2) or private (3)
SchoolYear	Catg.	Between 1st and 9th year, or finished (0)
PlayingTime	Cont.	Average gaming time per week in hours
Gamer	Bool	Considers itself a gamer (1) or not (0)
LikesMath	Catg.	How much enjoys math in a five-point scale
KnowsMath	Catg.	Knowledge in math in a five-point scale
		Behavioral
AvgScore	Cont.	Average score per level
MaxScore	Cont.	Maximum summed score achieved
AvgShots	Cont.	Average of shots fired per level
AvgTime	Cont.	Average of time spent to complete each level in seconds
SumTime	Cont.	Total time spent playing the game in seconds
MaxSeq	Cont.	Largest sequence of wins achieved
TotalPlayed	Cont.	Total of played levels
SumWins	Cont.	Total of wins in all levels
RatioSum	Cont.	Ratio of wins per played levels of summation challenges
RatioSub	Cont.	Ratio of wins per played levels of subtraction challenges
RatioMult	Cont.	Ratio of wins per played levels of multiplication challenges
RatioDiv	Cont.	Ratio of wins per played levels of division challenges

approach was adopted to capture players playing behavior in the environments that they will naturally interact with the game, being in their own houses our in school activities, for instance.

B. Sample

Here we describe the demographics and behavioral metrics of the analyzed dataset. Standard Deviation is referred to as ±.

Players were mostly males (82.91%), already finished middle school (89.95%) and had an average age of 20.43 years (± 8.9). They were mostly students from public (57.28%) and private (31.65%) schools and had internet access at home (94.47%). In addition, they indicated their affinity regarding math. When asked if they like the subject, where 1 means *I don't even want to hear about it* and 5 means *Yes, I consider it the most important subject of school*, the average response was 3.53 (± 1.09). When questioned how themselves considers their knowledge in the subject, the average was 3.23 (± 0.76), considering *Very low* as 1 and *Very high* as 5. Also, 49.25% of them considered themselves as gamers and players weekly playing hours was 12.25 (± 21.71) on average. Thus, it is expected that these players did not face difficulty on the math challenges and that it served as a training activity for them.

Table II displays statistics of the dataset's behavioral features. For each one, it shows the minimum, average, standard deviation and maximum value. As can be seen in the table, players' average ratio was close to 90% for puzzles of all four operations. We argue that this originates from the game's mechanics, which lead players to win more than losing. One factor is that the game aid players to solve the challenges by only presenting the numbers that correctly solve them. Thus, remaining to the player the need to just collect the numbers in the correct order. Another reason is the guidance the game provides through boxes' colors, allowing players to identify that collecting green, pink and blue boxes, if necessary, correctly solves all challenges. Also, it demonstrates that the total time playing and the total of played levels had a large standard deviation, indicating that these characteristics had a big difference between players. In contrast, while the average score, time and shots per level had small variation, the maximum score variation was close to 300 points. This corroborates to the variation on maximum level reached near to five and average score per level of approximately 56 (roughly 280 points in five levels).

Table II. DISTRIBUTION OF DATASET'S BEHAVIORAL FEATURES.

Feature	Min	Mean	SD	Max
score_mean	34.54	55.86	3.76	64.11
time_mean	10.79	24.2	6.27	55.92
shots_mean	3.2	10.5	4.64	39.78
max_level	4	12.16	4.93	35
total_played	10	52.33	38.74	231
sum_wins	7	46.41	34.96	208
max_score	173	746.92	283.31	1967
sum_time	134	1253.12	924.84	4675
sum_ratio	0.0	0.84	0.15	1.0
sub_ratio	0.25	0.89	0.12	1.0
mul_ratio	0.5	0.87	0.12	1.0
div_ratio	0.6	0.92	0.09	1.0

C. Data Preparation

The first preprocessing step was to transform the feature *SchoolType* to avoid the indication that one type is higher (or better) than others. Therefore, this column was replaced by three new boolean features, each one representing one school type. Note that, despite the four types, this process resulted in three columns since the fourth case is represented when they all have zero value. At the second step, min-max normalization was applied due to mixed types of features, avoiding that e.g. score dominates or influence the clustering process more than SchoolYear.

After these initial procedures, the selected feature extraction methods were applied. PCA [22] usage is based on its characteristic of joining components that are comparatively similar to each other, creating the Principal Components (PC), that are expected to be better than the originals [7]. FA is a procedure that uses agglomerative HC [23], however, instead of agglomerating data instances, it recursively merges features based on their similarity, resulting in the average value of merged ones. Here, Euclidean distance was employed to determine which features should be merged. We selected FA as an alternative to dimensionality reduction, investigating if its results could overcome the widely used PCA.

In sum, at the end of preparation, the dataset was composed of 199 rows (players) and 23 columns (features). Also, two alternative versions were available to be clustered as well. They originated after the application of PCA and FA that aims to reduce the dataset's dimensionality (number of features) and improve clustering results. Thus, the presented experiments are conducted in a total of three datasets. Note that all parameters settings (e.g. number of PC and number of agglomerated features) are evaluated for both feature extraction methods parameters, in order to find the best one.

D. Clustering

Considering that (1) we want to discover the different users' profiles through data and (2) we do not know how many of them exist, our approach was to use an unsupervised learning procedure, clustering. It requires the comparison of algorithms and parameters selection, given that there is no correct choice [13]. Thus, the experimented methods were KM [15] and HC [23], since they have been used in the literature (e.g. [7], [11]), providing interesting results, and operates in different ways.

KM is a widely used technique, which might be applied as a baseline algorithm [13]. It is based on centroids that are determined by the average of all data points assigned to it. Iteratively, the algorithm optimizes these centroids, updating them based on the nearest neighbors, until points belonging to each cluster does not change any more or a limit of iterations is reached. HC is a method that repetitively combines data instances (players) into clusters, and then those clusters into larger clusters, forming a hierarchy [23]. In contrast to KM, the HC employed is accomplished in a *bottom-up* approach, which is called *agglomerative* [24]. It starts joining the most similar objects and recursively merge the resulting clusters to other objects and/or clusters according to their similarity. Euclidean distance was adopted as the similarity measure for KM and HC.

The number of clusters (k) to be found is a required parameter for both algorithms, which might be defined through different procedures for each one. To determine it, we need to select the model that fits better our dataset. Common approaches to this end are internal and external evaluation, however, we have no prior knowledge of classes - often the case of clustering - [23] which led us to the internal method.

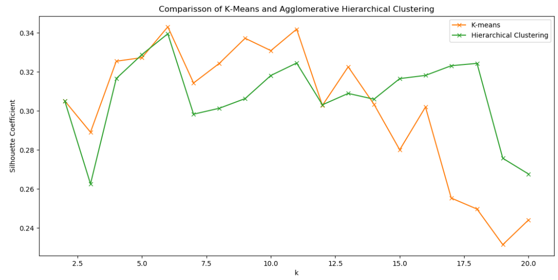
To accomplish this, two measures were used: SC [25] and CHI [26]. The former encompasses intra-cluster and nearest-cluster distances. It is bounded between -1 and +1, where values close to +1 indicate better clustering, around 0 suggest overlapping, and values near to -1 points incorrect cluster assignment. While SC is bounded to a predefined range, CHI does not have this property. It is determined by the ratio of between-clusters dispersion mean and within-cluster dispersion. A higher CHI score indicates a model with clusters better defined.

Using these metrics, the three dataset versions were assessed with $k = 2$ to $k = 20$, for both algorithms, in order to determine the model which provides the best separation. Note that, to mitigate the KM drawback of centroids initialization, it was executed 100 times for every k. The scikit-learn toolkit was used to conduct all experiments presented in this paper [27].

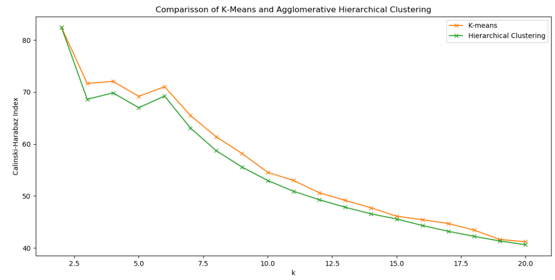
V. Results

This section begins presenting the clustering results for the dataset version with original dimensionality after its preparation (23 features). Fig. 2 demonstrates that the two algorithms indicated an optimal k of six according to SC (0.3431), whereas for CHI (82.47) they also agreed, but with $k = 2$. As suggested by the poor results, some features are probably acting as noise or interfering in the profiling procedure, showing the importance of our approach of extracting new features to improve clusters separation.

Results of using PCA before the clustering algorithms are presented in Fig. 3. It presents the highest score, for both metrics, according to the number of PC used. Every setup of PC - from one to 23 - was tested with k varying from two to 20. The highest score for each setup is presented in



(a) Silhouette Coefficient for each k.



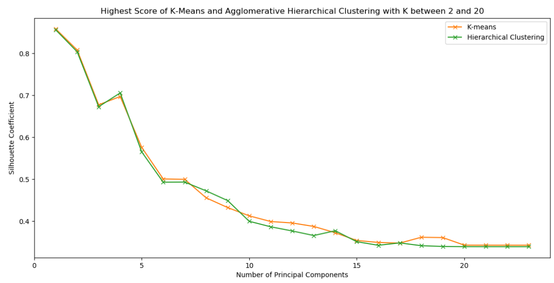
(b) Calinski-Harabaz Index for each k.

Figure 2. Clustering performance of prepared dataset according to both metrics.

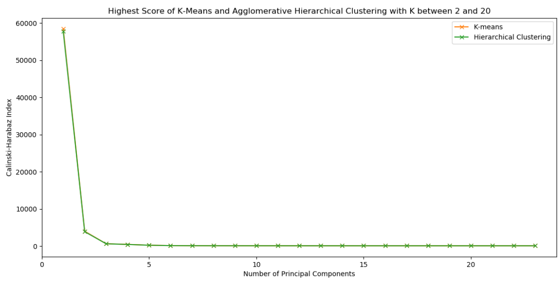
the figure. As can be seen, using only the first PC led to the highest score for both SC (0.8586) and CHI (58499.80). Thus, we further investigated this setup, which is displayed in Fig. 4. The figure shows that, according to SC and CHI, the best results were achieved with $k = 9$ and $k = 20$, respectively.

Finally, the results of the third dataset version are displayed in Fig. 5. Similar to the assessment of PCA usage (Fig. 3), Fig. 5 demonstrates the highest score, for both metrics, according to the number of agglomerated features used as input to the clustering algorithms. The prepared dataset had its features agglomerated from 1 to 22, where each setup was tested with k ranging from two to 20. Only the highest score of each setup is presented in the figure. Note that, unlike PCA, the setting with 23 as parameter was not performed since it would not change the prepared dataset in this case. Fig. 5 shows that agglomerating the 23 features to 4 led to the highest scores for both SC (0.8749) and CHI (3022.26). Then, our further investigations of this version found that these scores were achieved with $k = 4$, as can be viewed in Fig. 6. Similar to the use of PCA projected features, Fig. 6 also shows KM and HC mostly yielding results alike to each other, where the only setup with a substantial difference between them originated using the prepared dataset measured through SC (Fig. 2(a)).

To summarize these findings, Table III presents the highest scores for every dataset version. It demonstrates that both feature extraction methods improved clustering results. Also, it shows that using FA to reduce the dataset's dimension to

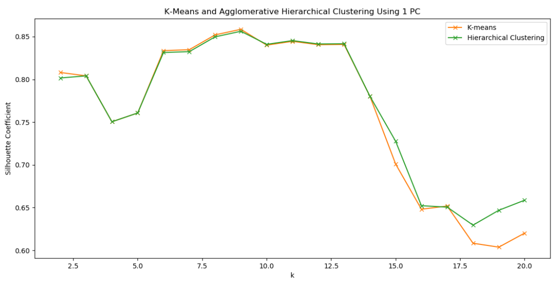


(a) Highest Silhouette Coefficient for each PCA setup.

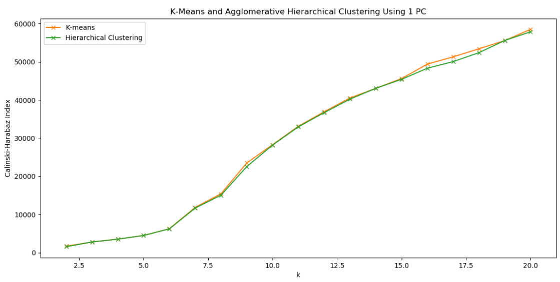


(b) Highest Calinski-Harabaz Index for each PCA setup.

Figure 3. Clustering performance according to the number of PC used.

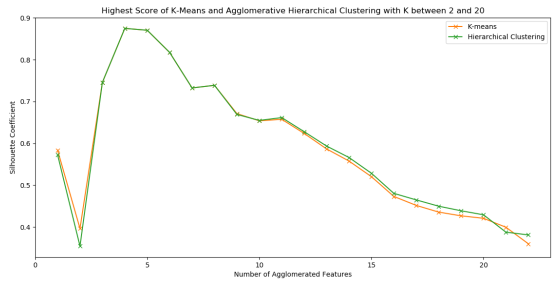


(a) Silhouette Coefficient for each k

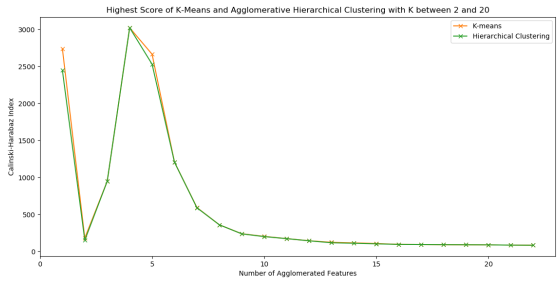


(b) Calinski-Harabaz Index for each k

Figure 4. Clustering performance using the dataset projected using one PC.

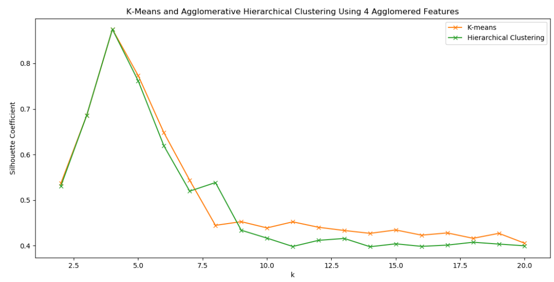


(a) Highest Silhouette Coefficient for each FA setup

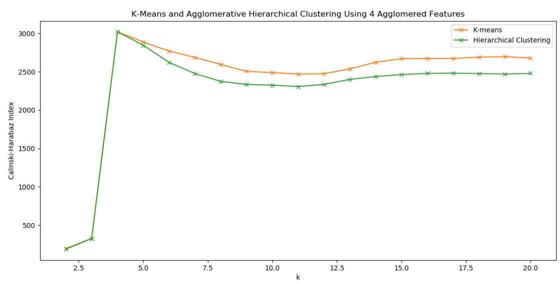


(b) Highest Calinski-Harabaz Index for each FA setup

Figure 5. Clustering performance according to the number of agglomerated features.



(a) Silhouette Coefficient for each k



(b) Calinski-Harabaz Index for each k

Figure 6. Clustering performance using the dataset with features agglomerated to four.

4 provided the best results according to SC, excelling the best PCA setup. In contrast, the PCA method overcame FA when measured through CHI. Our further investigations are based on results yielded by FA, since (1) its results were close to the best possible according to SC (0.8749/1), while CHI does not have a boundary, and (2) the best result of CHI found 20 clusters which is inappropriate to our goals.

Since both KM and HC achieved comparable results according to the evaluation metrics, we analyzed players' distribution. Once again, the algorithms produced the same outcome, with the same number of players per cluster.

Table III. SUMMARY OF CLUSTERING RESULTS. IT PRESENTS, FOR EVERY DATASET'S VERSION, THE HIGHEST SCORE OF SC AND CHI, AND THE NUMBER OF CLUSTERS (k) AND ALGORITHM THAT YIELDED IT. PREPARED = AFTER DATA PREPARATION; FIRST PC = PROJECTED PCA RESULT USING THE FIRST PC; AND FA TO 4 = FEATURES AGGLOMERATED TO 4.

	Silhouette Coefficient			Calinski-Harabaz Index		
Dataset	Score	k	Algorithm	Score	k	Algorithm
Prepared	0.3431	6	KM	82.47	2	KM
First PC	0.8586	9	HC	58499.80	20	KM
FA to 4	0.8749	4	Both	3022.26	4	Both

Therefore, we further investigated KM results due to: it provides a way to predict new players' groups naturally; its clusters can be assessed through its centroids; and it is most used in the literature than HC. Table IV demonstrates groups characteristics, where boolean features (i.e. *ismale* and *isgamer*) are displayed as percentages and *schooltype* shows the distribution between players categories within the group, for simplicity, wherein 0, 1, 2 and 3 represent municipal, public, federal and private schools, respectively. Thus, 1 represents that the average players were from public school and 2.6 that most of them were from federal and/or private institutions. However, the table displays only the characteristics that had a statistically significant difference between groups. This condition was checked through the Dunn's Post hoc test after the Kruskal-Wallis test.

A. Profiles Interpretation

Produce interpretable results is one of the most relevant characteristics of this type of analysis [23]. Thus, we analyzed the information in Table IV in order to interpret the groups' differences.

G0 is fully composed of gamers who weekly play the most in comparison to the other groups. In addition, players from this group were mostly males from public institutions. These are the players who played SpaceMath the most and achieved the highest performance in terms of maximum score and level reached. Thus, we considered them as **advanced** players due to their strong background with games and the best performance in our testbed. We did not use the nomenclature hardcore since the identified characteristics do not necessarily imply that they are, for instance, extremely competitive, as stated in [28].

Table IV. PROFILES CHARACTERISTICS.

Feature	G0 (n=53)	G1 (n=61)	G2 (n=45)	G3 (n=40)
score_mean	56.55	54.48	56.68	56.13
time_mean	23.98	26.42	21.3	24.37
shots_mean	12.0	9.13	10.55	10.55
max_level	14.11	11.23	11.53	11.7
max_score	852.81	687.77	724.71	721.82
sum_time	1556.68	1176.52	1027.02	1222.08
ismale	0.94	0.67	0.96	0.78
gaminghours	18.85	5.97	18.47	6.1
schooltype	1.0	1.0	2.67	2.62
isgamer	1.0	0.0	1.0	0.0
sum_ratio	0.85	0.8	0.83	0.9

In contrast to G0, all players from G1 considered themselves as non-gamers and weekly play the less in comparison to all other groups. Also, this group had the highest percentage of females and the smallest average of both wins and maximum score achieved. They had the highest average of time per level between all groups, but only achieved the same maximum level. This is also reflected by the smallest average score. Hence, we considered this groups as **beginners** due to their lack of affinity with games and low performance in our testbed.

Similar to G0, G2 is also fully composed by gamers who weekly play substantially more than G1 and G3 and as much as G0. Also, the majority of its players are males, but, rather than being from public institutions, these players belong to federal and private institutions. They also differ from G0 in terms of total played time, being the group that players the less and had the smallest average time. In spite of that, their average score was similar to the highest ones. Therefore, we called them as **skilled** players, considering their previous familiarity with games and comparable performance despite less experience with the testbed game.

Finally, G3 is similar to G1 in terms of having less female players, weekly playing less than G0 and G2 and, in addition, being composed by non-gamers. Although, they mainly differ from G1 by school type, belonging to federal/private instead of public institutions, achieving a slightly higher average and maximum scores, and performing better in summation problems. Based on these characteristics, we named this group as **intermediate** players, considering their low affinity with games, persistence in playing SpaceMath and moderate performance.

Mainly, the most discriminating demographic characteristics were players school type, being a gamer, weekly playing time and genre. In terms of in-game behavior, summed playing time, maximum score and level, and average score and time were the most discriminating. This demonstrates that adding demographic features to player profiling through data is relevant, where these features alone could, possibly, provide similar clusters. However, as can be seen in Table IV, behavioral data could work similarly to distinct these players types. Thus, considering both demographic and behavioral data allows for a higher degree of differentiation when analyzing profiles, enabling the identification of their differences from both perspectives.

VI. DISCUSSION AND FINAL CONSIDERATIONS

This paper presented an analysis of players' profiles through an educational math game data. To accomplish it, a dataset of 199 players including both behavioral and demographic features was clustered. To identify the optimal number of profiles, two metrics for unknown ground truth cases were applied. Experiments demonstrated that HC yielded the best separation according to CHI when the original dataset was transformed with PCA. In contrast, the best performance according to SC was achieved by summarizing the dataset to four dimensions through FA. Considering that SC is bounded between -1 and 1, whilst CHI does not have such a property, we analyzed players' profiles based on clustering results from FA. Groups' characteristics suggested both demographic and behavioral features were discriminating between them, showcasing that using both perspectives together can improve the degree of differentiation in profiling. Based on these, we were able to identify the following profiles: **advanced**, **skilled**, **beginners** and **intermediate**.

These results suggest that **advanced**, which achieved the highest in-game performance, are the ones who had improved their knowledge the most. This is based on literature's evidence that these metrics might be used as indicators of improvements on the game's educational subject [29]. **Skilled** players are experienced players who perform moderately well. Identifying this type of players is important to employ some strategy able to prevent them to get bored and leave the game. Possible interventions might be to provide special rewards or adapt the game seeking to mitigate lack of interest or challenge [30]. Both **beginners** and **intermediate** players are inexperienced, suggesting that they might face difficulties while playing the game. An early identification of these players is relevant to enable the game to promote challenges consistent with their ability level in order to avoid negative experiences such as frustration [31].

Moreover, identifying players' profiles is relevant to game design, as developers have been making significant effort to create games that are interesting to a broad range of users from varied ages, genres and skill levels. This research provides valuable insights into how to know playing styles better, and thus, be able to drive the development process towards these differences. However, these profiles are also based on behavioral features, which would require testing phases during the design and development of the games. As previously mentioned, a reliable alternative to this problem is adaptation [30], which can be accomplished through Procedural Content Generation [32], [33]. Using this approach, the game would be able to adapt its contents (e.g. levels, educational challenges, difficulty) as it is played, aiming to promote personalized experiences for each profile. Furthermore, to the best of the author's knowledge, this is the first approach to use FA as part of the pipeline for finding players' profiles, which showed to be a valuable technique by overcoming the improvements provided by PCA on clustering results.

Nonetheless, our experiments were conducted with a modest sample size, where these volunteers were mostly males. Evaluating a larger sample with different subjects would possibly yield different results. In spite of that, the employed methodology could be reproduced in order to solve these limitations. In addition, we argue that this approach can be generalized to other games. Even though some in-game features are specific to our testbed game, they could be replaced by features that reflect other games' relevant characteristics. Furthermore, this methodology can be used for more than game design ends. For instance, when evaluating players' improvement after using some educational tool, knowing how the different profiles of players are affected by this tool can reveal valuable insights about their learning and experience.

Acknowledgements

We are thankful to Capes for the financial support and all volunteers who collaborated with this research.

References

[1] S. K. Holm, "Player types: What, why and how," in *Extended Abstracts Publication of the Annual Symposium on Computer-Human Interaction in Play*, ser. CHI PLAY '17 Extended Abstracts. New York, NY, USA: ACM, 2017, pp. 707–710. [Online]. Available: http://doi.acm.org/10.1145/3130859.3133220

[2] D. A. Dixon, "Player types and gamification," in *Proceeding of CHI2011*, 2011.

[3] B. Horn, C. Clark, O. Strom, H. Chao, A. J. Stahl, C. Harteveld, and G. Smith, "Design insights into the creation and evaluation of a computer science educational game," in *Proceedings of the 47th ACM Technical Symposium on Computing Science Education*, ser. SIGCSE '16. New York, NY, USA: ACM, 2016, pp. 576–581. [Online]. Available: http://doi.acm.org/10.1145/2839509.2844656

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

[5] N. Yee. The daedalus project. [Online]. Available: http://www.nickyee.com/daedalus/

[6] A. Drachen, R. Sifa, C. Bauckhage, and C. Thurau, "Guns, swords and data: Clustering of player behavior in computer games in the wild," in *2012 IEEE Conference on Computational Intelligence and Games (CIG)*, Sept 2012, pp. 163–170.

[7] Z. Halim, M. Atif, A. Rashid, and C. A. Edwin, "Profiling players using real-world datasets: Clustering the data and correlating the results with the big-five personality traits," *IEEE Transactions on Affective Computing*, pp. 1–1, 2017.

[8] E. S. Siqueira, G. N. Rodrigues, C. D. Castanho, and R. P. Jacobi, "A data analysis of player in world of warcraft using game data mining," in *Proceedings of SBGames 2017*, 2017.

[9] S. Benmakrelouf, N. Mezghani, and N. Kara, "Towards the identification of players' profiles using game's data analysis based on regression model and clustering," in *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, Aug 2015, pp. 1403–1410.

[10] Y. Francillette, L. Abrouk, and A. Gouaïch, "A players clustering method to enhance the players' experience in multi-player games," in *Computer Games: AI, Animation, Mobile, Interactive Multimedia, Educational Serious Games (CGAMES), 2013 18th International Conference on*, July 2013, pp. 229–234.

[11] C. Thurau and A. Drachen, "Introducing archetypal analysis for player classification in games," in *Proceedings of the international workshop on evaluating player experience in games (EPEX'11) hosted at the 6th international conference on the foundations of digital games (FDG2011)*, 2011.

[12] A. Drachen, A. Canossa, and G. N. Yannakakis, "Player modeling using self-organization in tomb raider: Underworld," in *2009 IEEE Symposium on Computational Intelligence and Games*, Sept 2009, pp. 1–8.

[13] C. Bauckhage, A. Drachen, and R. Sifa, "Clustering game behavior data," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 7, no. 3, pp. 266–278, Sept 2015.

[14] A. Drachen and A. Canossa, "Evaluating motion: Spatial user behaviour in virtual environments," *International Journal of Arts and Technology*, vol. 4, pp. 294–314, 2011.

[15] J. A. Hartigan and M. A. Wong, "Algorithm as 136: A k-means clustering algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100–108, 1979. [Online]. Available: http://www.jstor.org/stable/2346830

[16] K. Kiili and H. Ketamo, "Evaluating cognitive and affective outcomes of a digital game-based math test," *IEEE Transactions on Learning Technologies*, vol. PP, no. 99, pp. 1–1, 2017.

[17] B. M. McLaren, D. Adams, R. E. Mayer, and J. Forlizzi, "A computer-based game that promotes mathematics learning more than a conventional approach," *IJGBL*, vol. 7, pp. 36–56, 2017.

[18] L. Rodrigues, R. P. Bonidia, and J. D. Brancher, "A math educational computer game using procedural content generation," in *SBIE 2017*, 10 2017.

[19] F. Ke, "A case study of computer gaming for math: Engaged learning from gameplay?" *Computers & Education*, vol. 51, no. 4, pp. 1609 – 1620, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360131508000523

[20] M. O. Schneider, Érika Tiemi Uehara Moriya, A. V. da Silva, and J. C. Néto, "Analysis of player profiles in electronic games applying bartle's taxonomy," in *Proceedings of SBGames 2016*, 2016.

[21] D. Calife and R. Nakamura, "Profile comparison between players segmentation according to bartle's taxonomy and k-means algorithm," in *Proceedings of SBGames 2017*, 2017.

[22] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37–52, 1987.

[23] P. Berkhin, *A Survey of Clustering Data Mining Techniques*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 25–71. [Online]. Available: https://doi.org/10.1007/3-540-28349-8_2

[24] A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1988.

[25] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53 – 65, 1987. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0377042787901257

[26] T. Caliński and J. Harabasz, "A dendrite method for cluster analysis," *Communications in Statistics*, vol. 3, no. 1, pp. 1–27, 1974. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/03610927408827101

[27] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[28] J. Kuittinen, A. Kultima, J. Niemelä, and J. Paavilainen, "Casual games discussion," in *Proceedings of the 2007 Conference on Future Play*, ser. Future Play '07. New York, NY, USA: ACM, 2007, pp. 105–112. [Online]. Available: http://doi.acm.org/10.1145/1328202.1328221

[29] K. Kiili, K. Moeller, and M. Ninaus, "Evaluating the effectiveness of a game-based rational number training - in-game metrics as learning indicators," *Computers & Education*, vol. 120, pp. 13 – 28, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360131518300125

[30] M. P. Silva, V. d. N. Silva, and L. Chaimowicz, "Dynamic difficulty adjustment through an adaptive ai," in *2015 14th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames)*, Nov 2015, pp. 173–182.

[31] C. Bauckhage, K. Kersting, R. Sifa, C. Thurau, A. Drachen, and A. Canossa, "How players lose interest in playing a game: An empirical study based on distributions of total playing times," in *2012 IEEE Conference on Computational Intelligence and Games (CIG)*, Sept 2012, pp. 139–146.

[32] R. Lopes and R. Bidarra, "Adaptivity challenges in games and simulations: A survey," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 3, no. 2, pp. 85–99, June 2011.

[33] M. Hendrix, T. Bellamy-Wood, S. McKay, V. Bloom, and I. Dunwell, "Implementing adaptive game difficulty balancing in serious games," *IEEE Transactions on Games*, pp. 1–1, 2018.