Profile Comparison Between Players Segmentation According to Bartle's Taxonomy and K-Means Algorithm

Daniel Calife* Ricardo Nakamura

Escola Politécnica da Universidade de São Paulo, Computer Engineering Department, Brazil

ABSTRACT

One of the main ways to understand players profile and behaviors is through their classification or clustering, both problems are widely discussed and can bring results to the most diverse areas applied to games, such as marketing, game design and game development. This segmentation can be accomplished by adopting some semantics of the game itself and its design, or else, by applying some statistical treatment.

In this work we classify the players from a social game called Vida Rock 2 according to a semantic model based on Bartle's Taxonomy and also cluster according to a statistical model of K-Means. The monetization profile of the players is analyzed and then compared according to these segmentations, analysing the ability to describe players who made some purchase in the game, not taking into account this feature in any of the models proposed.

Keywords: game analytics, data science, player modeling, classification.

1 INTRODUCTION

Players' behavior analysis and classification are frequent challenges in the field of games, which can bring results in its various areas, such as development, game design and marketing [1]. Just as in other online services, user actions can, and usually are, recorded, thus creating a large amount of data to be processed and analyzed. One of the major concerns regarding these data is how to understand them and generate actions that give concrete results.

One of the approaches to handle this data is segmentation of players by categories, classes or clusters. This separation can be accomplished following some semantics from the game itself and its design, or else, following some statistical treatment [2].

In this work, two types of segmentation will be applied. One following the Bartle´s Taxonomy [3], which separates players into four categories: Explorers, Socializers, Achievers and Killers, within a graph where the x-axis represents their interaction with Players or World and the y axis represents their preference between Act and Interact, as Figure 1 demonstrates.

*e-mail: calife@gmail.com

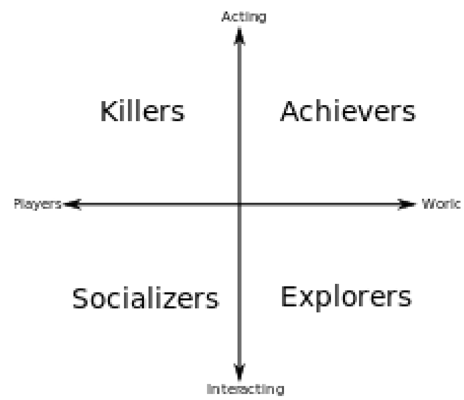

Figure 1: Bartle´s Taxonomy [2].

As we can see in figure 1 Explorers are players who prefer to interact with the world, know every aspect of levels and maps, create own maps, find hidden places, and Socializers prefer the more social aspect of the game, interacting and helping other players. On the acting side, we have players who prefer more concrete results like Achievers who like to have items, points, money and others success measurements and Killers who act on other players, enjoying competition.

Another approach of clustering applied in this work is the method of K-Means [4], whose objective is to group the elements into k groups according to their distance from a mean. The K-Means algorithm, along with other approaches, are widely used to cluster many aspects of games, like behavioral profiling, monetization and progression [1][2].

There are other approaches and methods to players´ segmentation taking into account some in game features and its meaning or purely based on statistics [5][6][7][8]. Therefore, we adopt these two approaches because of their simplicity of application and their use as base for more elaborated segmentations.

Thus, these two forms of segmentation will be compared taking into account their ability to describe players who have made purchases in a game. Leading to the objectives described in the next sub-section.

1.1 Objectives

The objective of this work is to compare a classification based on game design concepts (Bartle's taxonomy) with a statistical based clustering (K-Means). For that, it is necessary to describe a heuristic model according to Bartle's Taxonomy, and also, apply feature reduction techniques in order to use K-Means algorithm [9].

This comparison will test the ability of both techniques to describe players who have made some purchase in the game, and their conversion rate, but without using this feature in any of the proposed models.

2 THE GAME VIDA ROCK 2

Between 2007 and 2014 the company Colorcube Games created and operated four social games launched on Orkut and Facebook social networks. Its first game, Vida Rock, was a pioneer on Orkut social network for the Brazilian market, presenting social mechanics, creating bands with friends, and also implementing the free to play model of monetization, allowing micro payments within the game. Vida Rock had 2 million registered players and during its peak had around 600 thousand players per day.

After Vida Rock success, the games released later by Colorcube were more elaborated and refined. For this work it was chosen the game Vida Rock 2 [10]. The second version of Vida Rock follows the career of an aspiring rockstar who must train his musical skills, earn money and fight for success. Therefore, the main measure of evolution in the game is Popularity, which act as the player level. A screenshot from the game can be seen on figure 2.

Figure 2: Vida Rock 2 game on Orkut.

To earn popularity the player must attend to musical events and succeed in them, success is achieved according to their musical abilities, which can be in an instrument or as a vocalist. Therefore, another important attribute in the game is its musical ability, to develop these skills the player must train, which is one of the main actions in the game.

The player has two main attributes in the game: Energy and Money. With energy the player can train and participate in events. With money, the player can buy items to customize his character and his house.

As Vida Rock 2 is a social game, there are still other mechanics, such as forming bands with friends, challenging other players for duels, gifting, using songs created by friends, training together and poking. All these features makes Vida Rock 2 a good candidate to explore player´s segmentation and modeling.

In addition to these more game design specific attributes, other more generic attributes are stored, such as account creation date and last visit date, number of visits, numbers of friends etc.

Released in January of 2011 Vida Rock 2 on Orkut reached approximately 1 million registered players, in the chart presented on figure 3 we can see the evolution of daily player visits over time.

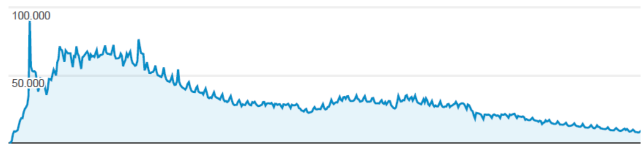

Figure 3: Evolution of players visits from 2011 until 2014.

3 DATA PREPARATION

Vida Rock 2 database has more than 66 tables, some exceeding 20 million records. To reduce the number of records in the database, were eliminated players considered inactive and that had never made a purchase with real money. With this reduction we reached the approximate number of 380 thousand players, from these players were extracted features that describe their evolution and style of play, 17 attributes were chosen:

- osid and playerid: primary identification keys for players;
- popularity: main attribute of player evolution;
- status: describes how many assets the player has;
- moneytotal: amount of virtual money owned;
- charismatotal: describes the level of interaction with other players;
- trainings: number of times the player has trained their skills;
- events: number of musical events the player participated in;
- friendvisits: how many friends visits the player received;
- duels: number of duels participated;
- duelvictories: number of duels won;
- bonuscredits: amount of real money player have won;
- cashcredits: amount of real money player have purchased;
- days: number of days played;
- ownervisits: how many times player accessed the game;
- smallenergydrink: consumption of spare energy given by friends;
- energydrink: consumption of spare energy purchased with real money;
- creationdt: account creation date;
- lastvisitdt: date of last player's visit.

3.1 Outliers removal

After data was collected, an exploratory analysis was performed to identify possible patterns and outliers. At first, it was soon noticed the large number of players with popularity values between 0 and 100, representing players who did not interact or interacted very little with the game. As we can see on the popularity chart presented by figure 4.

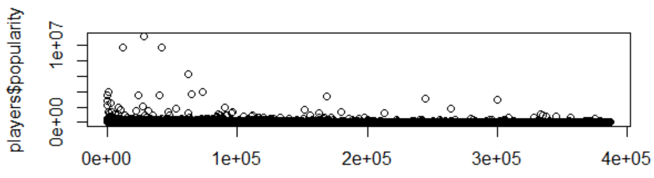

Figure 4: Players´ Popularity.

As can be seen also a large concentration of players who played 0 or 1 day, according to the histogram on figure 5.

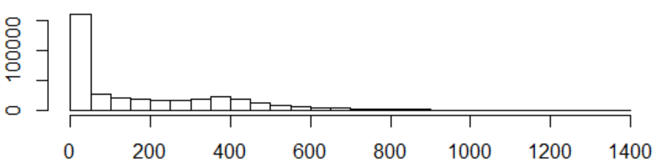

Figure 5: Number of days played histogram.

Thus, players with less than 100 popularity and less than 2 days played were removed, resulting in approximately 260,000 players records.

After this first removal of players who had little engagement with the game, the popularity relation was assessed with each other attributes, finding players who are very far from the values in which most other players meet. As an example, we can see the chart of popularity x status. Figure 6 shows the graph before the removal of outliers and Figure 7 shows the same graph after the elimination.

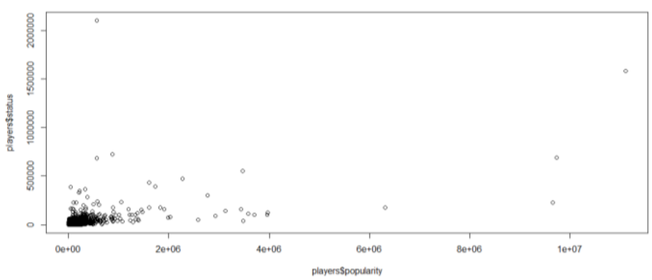

Figure 6: Popularity x Status, before outliers removal.

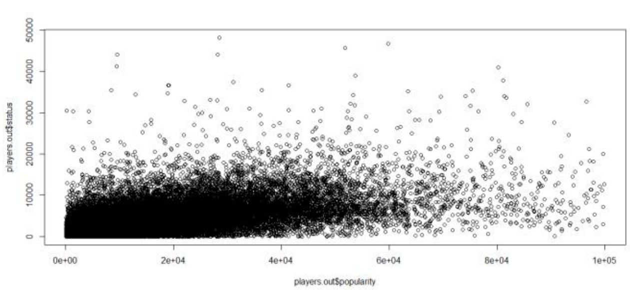

Figure 7: Popularity x Status, after outliers removal.

After an analysis of each feature, only the players with values below those determined below were left:

- popularity <100,000;
- status <50,000;
- moneytotal <500,000;
- charismatotal <5,000,000;
- trainings <3,000;
- events <1,000;
- friendvisits <500;
- duels <150;
- smallenergydrink <50;
- ownervisits <650;
- bonuscredits <160.

3.2 Correlation Matrix

One of the objectives of the experiment is to apply a clustering algorithm on players, to identify different profiles and how they behave in the game and how they monetize. To perform these type of algorithms with a higher quality it is necessary to reduce the number of features that describe the players.

At first we can identify the features that have a very strong correlation, which in this case can be considered redundant.

The figure 8 shows the correlation matrix between the players´ attributes, excluding the identification and date columns, after the normalization of their values.

Analysing the matrix we can notice that Popularity is the feature with higher correlation value with other features, thereafter, which makes sense, as the popularity reflects the evolution of the players through all of theirs actions, this feature was removed from the model.

The features that have lower correlation values are smallenergydrink, days and duel, indicating that these features are very representative to create a profiling model.

Even with one of the lowest correlation values, the feature days was removed from models, since it is very difficult to fit its semantic inside the Bartle´s Taxonomy, as every profile enjoy playing and visiting the game.

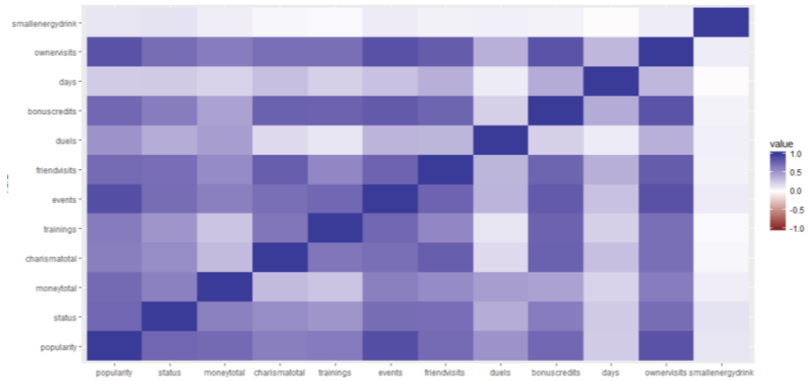

Figure 8: Correlation matrix.

$$(events{+}status{+}ownervisits{+}moneytotal){-}\\(charismatotal{+}friendvisits{+}duels{+}smallenergydrink) \quad (1)$$

$$(duels{+}smallenergydrink{+}events{+}moneytotal){-}\\(charismatotal{+}friendvisits{+}status{+}ownervisits) \quad (2)$$

4 Players Segmentation

In this section, the models of both segmentations methods are described and their results are presented and discussed.

4.1 Bartle´s Taxonomy Classification

To perform the classification by the taxonomy of Bartle it is necessary to define a model that calculates the location of each player according to the axes player / world and action / interaction. For this purpose, some features were selected according to their semantic representation of these behaviors in the game, creating in this way, two new attributes for the players, one representing the acting/interacting axis and other representing the player/world axis.

The features selected according to their meaning, and excluding monetization features, in the game were:

- Player interaction: charismatotal, friendvisits;
- Interaction with World: status and ownervisits;
- Player Action: duels, smallenergydrink;
- Action with World: events, moneytotal.

To calculate the value in the player/world axis it was applied the equation on (1).

To calculate the value in the axis action/interaction it was applied the equation on (2).

The figure 9 shows the distribution of players according to the proposed classification:

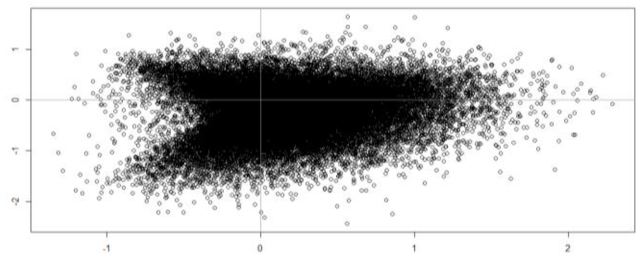

Figure 9: Players distribution according to Bartle´s Taxonomy model.

Although the position of the players in the taxonomy is continuous, it is possible to consider a predominant behavior for each player, leading to the following distribution of players:

- Killers: 12%
- Achievers: 12%
- Socializers: 40%
- Explorers: 36%

These results shows that the majority of players are more likely to interact with others players or the world, than to act on those subjects.

4.2 K-Means Clustering

In order to perform the K-Means clustering algorithm, the same features used on the previous classification were considered, but this type of segmentation has better results and is more easily analyzed when there are only two features used.

To reduce the features, we applied the PCA (Principal Component Analysis) algorithm [4], extracting the first two main components. using the elbow method [11].

In figure 10 we have the graph that shows the distribution of players according to their main components PC1 and PC2:

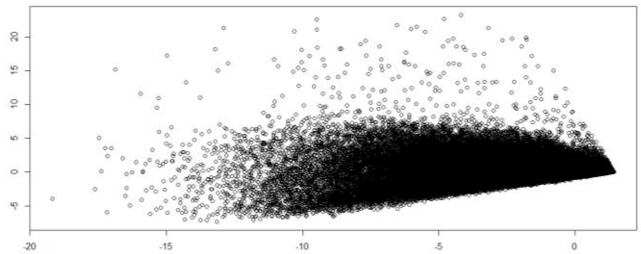

Figure 10: Players distribution according to main components from PCA algorithm.

In Figure 11 we can see the same graph of figure 10, but now separated into clusters after the application of the K-Means algorithm creating 4 groups:

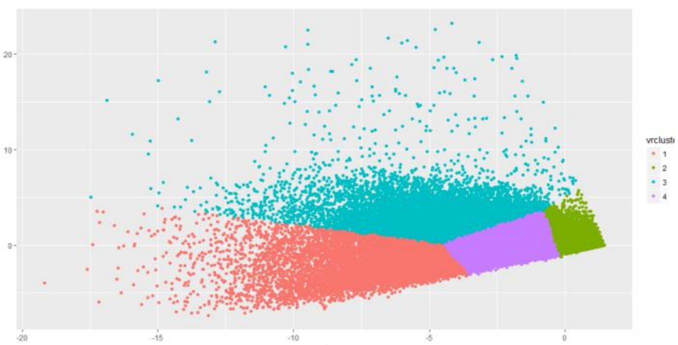

Figure 11: Players distribution according to main components from PCA algorithm colored by their clusters.

The distribution of the players among the 4 clusters are as follows:

- Cluster 1: 4%
- Cluster 2: 75%
- Cluster 3: 4%
- Cluster 4: 17%

4.3 Joining K-Means and Bartle´s Taxonomy

In Figure 12 we can see the players distributed according to the Player/World and Action/Interaction axis, defined in the Bartle´s classification and separated by their clusters defined by K-Means algorithm:

We can see that clusters 2 and 4 are more associated with players with more centralized and less defined profiles, while cluster 3 is more associated with players who prefer to act and cluster 1 are players who prefer to interact.

5 MONETIZATION PROFILE

From the total of players used in the sample to analyze its segmentations and behaviors, 11,807 players made at least one purchase with real money in the game.

With the segmentations, we can understand how these players are distributed among the groups. According to the classification of Bartle we have the following distribution of players who have bought something in the game:

- Killers: 3%
- Achievers: 11%
- Socializers: 27%
- Explorers: 59%

And taking into consideration the player´s total it describe the conversion percentage of:

- Killers: 1%
- Achievers: 4%
- Socializers: 3%
- Explorers: 7%

Although there is a higher conversion rate among Explorers, on overall all conversions are very close to the overall conversion average of ~4%.

Analyzing the same statistics through K-Means clusters we have the following percentage of players who have made a purchase:

- Cluster 1: 21%
- Cluster 2: 31%
- Cluster 3: 7%
- Cluster 4: 41%

And with a conversion percentage of:

- Cluster 1: 25%
- Cluster 2: 2%
- Cluster 3: 8%
- Cluster 4: 11%

In this case, we can clearly notice that there is a fairly high conversion rate from the players present in cluster 1. It would be quite difficult to identify the characteristics of the players from this particular cluster, since the PCs (Principal Components) have no semantic value inside the game or player. But when we come across Bartle's classification information, we realize that these players are defined as those who are more interested in interacting than acting in a more predominant way.

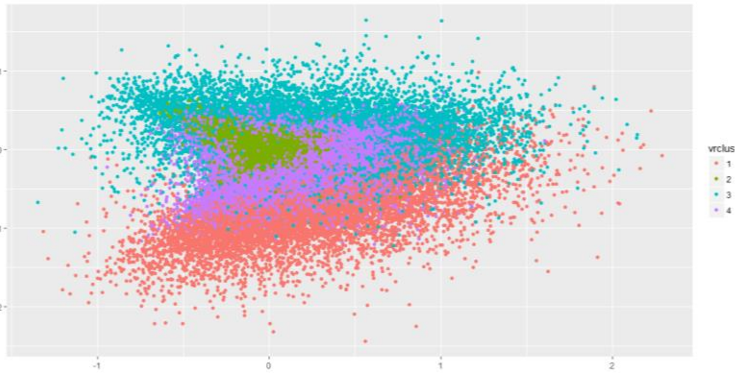

Figure 12: Combination of Bartle´s and K-Means

6 CONCLUSION

The objective of this work was to compare the monetization profile of players, based on two different forms of segmentation. We could note that both techniques can describe some aspects of the monetization profile of players and conversion rates.

Bartle´s Taxonomy model of classification describe very well player´s profiles and its monetization's characteristics, with this model we can show that large amount of paying players are on the interaction side, but we could not spot a profile that stands out from the rest with a better conversion rate.

Applying K-Means algorithm we could find a cluster with a very high conversion rate of 25%, a conversion that is very far from market standards. The cluster which showed this conversion, was cluster 1, by itself the cluster has no meaning inside the game, specially after the principal components extraction.

Combining information from both models, we can define in a more objective way the target player that we have to seek to increase the game revenue.

7 FUTURE WORK

Based on the results brought by this work we can affirm that combining methods of classification can lead us to a more specific players´ profiles definition, so we can do a more in depth exploration and comparison by applying other classification and clustering methods, such as presented in [1][2].

The way we created our classification model for Bartle´s taxonomy, it was essential to have the in game knowledge of features meaning and application, making difficult to generalize it to work with other online games as well. One possible approach, is try to extrapolate the classification to other games using only common data, as player´s level, number of visits, session time, number of friends etc. In order to do that we can apply some machine learning techniques as seen in [8] a work that deals with the problem of not having access to specific games features and data.

ACKNOWLEDGEMENTS

The authors would like to thank Colorcube Games who kindly gave access to Vida Rock 2 and Foots database, including game, user and monetization data.

REFERENCES

[1] C. Bauckhage, A. Drachen and R. Sifa, "Clustering Game Behavior Data," in *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 7, no. 3, pp. 266-278, Sept. doi: 10.1109/TCIAIG.2014.2376982. 2015.

[2] A. Drachen, C. Thurau, J. Togelius, G. Yannakakis, Game Data Mining in Game Analytics: Maximizing the value of player data, editors: M. El-Nasr, A. Drachen, A. Canossa. Springer-Verlag London, 2013

[3] R.,Bartle. Hearts, Clubs, Diamonds, Spades: Players Who Suit MUDs. http://mud.co.uk/richard/hcds.htm last visit 04/08/2017. last visit 04/08/2017.

[4] J. ,MacQueen, Some methods for classification and analysis of multivariate observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics, 281--297, University of California Press, Berkeley, Calif., 1967.

[5] A. M.,Smith, C., Lewis, K., Hullett, G., Smith, G., and A. Sullivan. An Inclusive Taxonomy of Player Modeling. Tech. Rep. UCSC-SOE-11-13, University of California, Santa Cruz. 2013

[6] B.Manero, J. Torrente, M. Freire, B. Fernández-Manjón, An instrument to build a gamer clustering framework according to gaming preferences and habits, Computers in Human Behavior, Volume 62, 2016, Pages 353-363. 2016

[7] T. Hynninen. Identifying Player Profiles in Video Games. 2016, Master´s Thesis. 2016

[8] E. G. Castro and M. S. G. Tsuzuki, "Churn Prediction in Online Games Using Players' Login Records: A Frequency Analysis Approach," in *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 7, no. 3, pp. 255-265, Sept. 2015.

[9] K., Pearson. "On Lines and Planes of Closest Fit to Systems of Points in Space" (PDF). Philosophical Magazine. 2 (11): 559–572. 1901.

[10] Colorcube Vida Rock 2 Game. https://apps.facebook.com/vidarockii/ last visit 04/08/2017.

[11] P. R, Peres-Neto,D.A., Jackson & K.M., Somers, K. (2005). How many principal components? Stopping rules for determining the number of non-trivial axes revisited. Computational Statistics & Data Analysis, 49(4), 974–997.2005.