Big Data Analytics in Cloud Gaming: Players' Patterns Recognition using Artificial Neural Networks

Victor Perazzolo Barros
Postgrad. Program in Electrical and Computer Engineering
Mackenzie Presbyterian University, UPM
Sao Paulo, Brazil
victor_ perazzolo@hotmail.com

Pollyana Notargiacomo
Postgrad. Program in Electrical and Computer Engineering
Mackenzie Presbyterian University, UPM
Sao Paulo, Brazil
pollyana.notargiacomo@mackenzie.br

***Abstract*—The Cloud Gaming model emerges with the evolution of the Cloud Computing and communication technologies. Through smartphones, PCs, tablets, consoles and other devices, people can access and use games on demand via data streaming, regardless the computing power of these devices. The Internet is the fundamental way of communication between the device and the game, which is hosted on a environment known as Cloud, enabling a large scale offer. The variety, volume, velocity, value and veracity (Big Data 5Vs) of data that is involved in these Cloud environments exceed the limits of analysis and manipulation of conventional tools, therefore, Big Data platforms are required to handle and interpret this data. The model known as Big Data Analytics is an effective and capable way to, not only work with these data, but understand its meaning, providing inputs for assertive analysis and predictive actions. A method is presented in this study to identify and analyze players' patterns in a virtual environment. With this information, it is possible to optimize user experience, revenue for developers and raise the level of control over the environment. Results are presented based on a dataset of the World of Warcraft game. By using a neural network, it was possible to identify with an average of 91% of accuracy the players' assiduity based on patterns in the game. Using the Hadoop technology and visualization tools on a Cloud based cluster, it was possible to map and identify the players' behaviors as well as their gameplay patterns.**

***Keywords*-Big Data Analytics; Cloud Gaming; Pattern Recognition; Artificial Neural Networks;**

I. Introduction

The growing demand for virtual entertainment on mobile devices such as tablets and smartphones is changing the way game developer companies are operating in the market in recent years. The consumption of games had a high growth as the adoption of new mobile technologies increased [8]. The variety of devices, including computers, consoles and even smart TVs, stimulates a transformation in the traditional distribution offer of these games. In this scenario a concept emerges, trying to fulfill this demand for a accessible, online, hardware power independent and multi-device game, these elements are part of the Cloud Gaming model. The model implies that any device can access, via a client, the server on a Cloud remotely through the Internet, the servers process the information sent by the device and give the actions back via streaming [27], [28].

Different Cloud Gaming platforms appeared in the last years and are facing the challenges of a new model in the market. One of the first platforms are OnLive [22] and Gaikai [9], both acquired by Sony in 2012 and 2015, respectively. Today Sony offers its own Cloud Gaming platform, the PlayStation Now, using the Gaikai and OnLive technology [24]. Another platform is GamingAnywhere [10], the first open-source Cloud Gaming platform [15], designed to allow scalability and portability through its open code [14]. It is possible for developers and game publishers to customize this platform for a better integration with their products and business models. It can be structured in combination with NVIDIA GRID, a solution by NVIDIA [21] that offers game streaming through the Internet. The graphic power of a GPU cluster can process and render high workloads of video and games [12].

All the data generated from theses systems can be properly analyzed to create resources for the developers to improve both games and platforms, but traditional systems may not support the demand that will be handled, that is where systems capable to manage this information, like Big Data systems, are essential [25], [19].

II. Big Data Analytics for Cloud Gaming

Great volumes of data are generated all the time in a Cloud Gaming environment. Each interaction made by a player creates data that are transferred and stored, and if properly analyzed, can contain valuable information. This information can be vital for the continuity and improvement of a game. Know the players' characteristics and behavior can lead to a better performance and enhancement of the game in terms of attractiveness and immersion, for example. When gathered together the behavior of multiple players around the world, patterns can be detected and even predictive analysis can be made to foresee the actions and intentions of these players inside the game [23].

Traditional Business Intelligence (BI) tools may not support working with the volumetry generated in the Cloud

978-1-4673-9005-7/16/$31.00 ©2016 IEEE

Gaming environments. To manipulate and understand the data, the industry is using Big Data platforms, which are capable to ingest great volumes of data and present accurate analysis, once all the data is considered [25], [19].

The statistical data collected in the most of the games are known as Game Telemetry, the game companies normally use this data to build the games statistic models and get some insights about the game. The difficulty faced in this model is to know what data will be analyzed and which information will be generated efficiently with them. Another problem is the volume of data that is stored, they are usually collected on a gross basis, and to treat and structure them in a way that they can be effective, an inefficient amount of time can be given to accomplish this task [7], [23].

It is known as the concept of Big Data, the manipulation of amounts of data that are not supported by traditional technologies and techniques, as well as the conciliation of databases with different architectures, structured and unstructured. In addition, innovative ways to process information for better insight and decision-making. The properties of the concept of Big Data are called the 5Vs: Volume, Velocity, Variety, Value and Veracity. These properties involve the entire model, known as Big Data Ecosystem (BDE), that involves Big Data Infrastructure, Big Data Analytics, Big Data Management and Big Data Security [6], [13].

The property of the concept known as Big Data 5V represent the characteristics expected from the system when handling with the data. It is expected the model to handle volumes of data that conventional systems do not support. Within velocity, ideally, the system should operate in real time, or very close to it. The difference between operating in (near) real time or not may cause direct impact on the results, it can be the difference between a game being discontinued or become viral [23].

From the moment that a game becomes viral, the flow of information tends to increase exponentially and then, four factors of the Big Data analysis help developers to proceed with the viral spread: increase the number of players; enlarge the contact of such players with other people, in order influence them in a given period of time; ensure that the players spend more time in the game; increase the probability that the players' contacts become new players [23].

This sequence creates a dissipation chain that offers greater chances of success for a game and, having an infrastructure capable of supporting this demand and prepared to work the data that is in it, another element is added to assist with the optimization of a game, both in terms of revenue and technically: the pattern recognition. Big Data Analytics uses analysis algorithms that run on robust platforms and can reveal patterns and correlations, it can also be used in predictive analysis [13].

Within Big Data Analytics the data analysis can be classified into three classes, according to the depth of the analysis: Descriptive Analytics, Predictive Analytics and Prescriptive Analytics. In Descriptive Analytics, historical data are explored to present what occurred. Regression algorithms can be used to identify trends in the databases. This level of analysis is associated with Business Intelligence (BI) systems. The Predictive Analytics tries to predict probabilities and future trends, it uses statistical techniques such as linear regression and logistic regression to understand behaviors and predict actions. Through data mining, identifies and extracts patterns to provide information and forecasts. The Prescriptive Analytics is focused on efficiency and decision-making. It is possible to simulate analysis in complex systems in order to collect information about the behavior of this system, identify problems and, through certain techniques, find optimal solutions to these problems [13].

These analysis classes combined allows the management and manipulation of the data, know individual profiles and identifying groups with common or distinct characteristics, focused on mapping the games and players in the Cloud Gaming environments. This should be operated within a Big Data Analytics Infrastructure, that involves the Hadoop ecosystem and which will be in a parallel layer of the Cloud Gaming environment, however, interconnected to the applications. Following, an experiment with Neural Networks and Big Data Analytics tools will demonstrate a modeling and explore more of the concepts already shown.

III. Behavior Analysis using Neural Networks

To visualize and demonstrate a practical application of behavior analysis, an artificial neural network was developed to identify the assiduity of the players' based on some of their avatars' (fictional characters controlled by players) characteristics [4]. The dataset used in this study is the World of Warcraft Avatar History (WoWAH) [18], a dataset extracted from the Massively Multiplayer Online Role-Playing Game (MMORPG) World of Warcraft (WoW) [3].

Although the WoW game is not a pure Cloud Game, it has most of the elements defined by the Cloud Gaming concept, therefore enables the study based on these precepts. The game is accessed through a client installed in the player's device, which in turn, reaches the Cloud environment where the game is hosted via the Internet, the servers process part of the information and send the results to the client.

At first, the modeling defined for this study seeks to understand player's behavior through its avatar's characteristics and identify if it has low, medium or high assiduity in the game using a neural network. The main goal is to identify players' patterns and behaviors based on login/logout time records, length of time in the game (assiduity/frequency) and in the characteristics and attributes of their avatars using Big Data and data visualization tools.

With this modeling, many different analysis can be made, such as: analyze the assiduity of a group of players based on their avatars' characteristics, understand the influence of guilds on the players' engagement, observe the players' most used movement patterns, identify frauds (fraud detection) by revealing behavior outliers and see how innovations can impact on players' gameplay patterns, for example. Using this technic, other analysis models can be performed to support developers in the administration and improvement of the games. That is useful to understand the attractiveness of a game over the players and, if it is found that a certain group of players are not immersed, it is known that they can eventually leave the game. With the right analysis, the developers can take some actions to avoid the illegal activities and game churn.

A. *Defining the Dataset*

The WoWAH Dataset have the record of 91,064 avatars over 1,107 days, from January 2006 to January 2009. It has the total amount of 138,084 log (text) files divided in 1,095 folders, which represents more than 36 million record lines and the total size of 3.4 GB. The data includes the avatars' gameplay times and their attributes, such as their ID (Avatar ID), Guild, Level, Race, Class and Zone. It also have the time of the query (Query Time) and the number of the query sequence (Query Sequence Number). There are two factions in the game: Alliance and Horde. In this dataset, the Horde faction was observed in the game world server "TW-Light's Hope realm" [18].

B. *Preparing the Data*

The log file is comma separated and has 12 fields: Null, Query Time, Query Sequence Number, Avatar ID, Guild, Level, Race, Class, Zone, Null, Record Line Number and Null. The Null fields represents non-relevant information. A log containing 569 records was randomly selected to be a sample used to train the neural network, namely: 06-Aug-2007 00:05:17. A table was created containing all the records from this log, but some fields (in red) were discarded. For this modeling, only the fields: Guild, Level, Race and Class (in blue) were used, as seen in the Figure 1:

Null	Query Time	Query Seq. No.	Avatar ID	Guild	Level	Race	Class	Zone	Null	Record No.
"0	08/06/07 00:01:32	1	49719	79	51	Orc	Warlock	Thousand Needles	0"	-- [1]
"0	08/06/07 00:01:32	1	52518	161	45	Orc	Warrior	Tanaris	0"	-- [2]
"0	08/06/07 00:01:32	1	49944	167	47	Orc	Warrior	Tanaris	0"	-- [3]
"0	08/06/07 00:01:32	1	34216	53	53	Orc	Warrior	Un'Goro Crater	0"	-- [4]
"0	08/06/07 00:01:32	1	41004	115	30	Orc	Rogue	Undercity	0"	-- [5]

Figure 1. Example of the log file organized as a table with the fields and values.

The values of the fields were grouped and an integer numeric value was associated to each group in order to make the neural network understand these values. These groups and ranges were defined based on an estimated average of the values in the log files.

To the Guild field, where the value represents the number of the guild that the players are in, two groups were created and numbers assigned: "0", to represent an avatar that does not belong to a guild, and "1", to for those who belongs to a guild. The Level field was divided into three groups: "Low Level" (from 1 to 23), "Medium Level" (from 24 to 47) and "High Level" (from 48 and above). For the Race field, five numbers were associated to each of the values, namely: Blood Elf, Orc, Tauren, Troll and Undead. The Class field was divided into two groups: "Melee", which have the Dark Knight, Hunter, Paladin, Rogue and Warrior classes, and "Magic", with Druid, Mage, Priest, Shaman and Warlock classes.

To determinate the assiduity of each group of characteristics, three ranges were created: "Low Frequency", with 1 to 14 repeated appearances, "Medium Frequency", with 15 to 49, and "High Frequency", with 50 or more of the same group appearances in the period of the log. The numbers associated with each group of values can be seen in Figure 2:

Attribute Value	Assigned Value
∉ to a Guild	0
∈ to a Guild	1
Low Level (1 - 23)	2
Medium Level (24 - 47)	3
High Level (48 - 80)	4
Blood Elf	5
Orc	6
Tauren	7
Troll	8
Undead	9
Melee: Dark Knight; Hunter; Paladin; Rogue; Warrior	10
Magic: Druid; Mage; Priest; Shaman; Warlock	11
Low Frequency of Patterns Appearance (1 - 14)	18
Medium Frequency of Patterns Appearance (15 - 49)	19
High Frequency of Patterns Appearance (50 - 100+)	20

Figure 2. List of numbers associated with each field group in order to make the neural network understand the values, as defined in the modeling.

An example of how the characteristics values was assigned to each of the groups, the Figure 3 was built:

Guild	Guild ∈ / ∉	Level	Level L/M/H	Race	Race	Class	Class Mel/Mag
161	1	48	4	Troll	8	Hunter	10
	0	1	2	Troll	8	Hunter	10
204	1	62	4	Blood Elf	5	Paladin	10
208	1	62	4	Undead	9	Priest	11
204	1	70	4	Tauren	7	Hunter	10
103	1	70	4	Orc	6	Rogue	10
21	1	70	4	Undead	9	Priest	11
53	1	61	4	Troll	8	Rogue	10
	0	64	4	Tauren	7	Hunter	10
	0	70	4	Tauren	7	Druid	11
19	1	70	4	Orc	6	Hunter	10

Figure 3. According to the number assigned to each field group as shown in Figure 2, the association was made for each value (rows) individually.

With the dataset sample structured in the defined modeling and in a way that the neural network can understand it, the patterns that appeared were grouped by the sequence of numbers that composes these patterns, and than, a single value was associated to represent this pattern. For example, the pattern "0,2,5,10", where "0" represents a non-guild avatar, "2" a low level player, "5" a Blood Elf Race and "10" that it belongs to a Melee Class, forms the "02510" pattern, which was assigned to "Pattern 1", once this was the first pattern to appear in the sample. The other patterns such as "02511", "02610" and so on, was associated to "Pattern 2", "Pattern 3" and further.

Besides the "Pattern Number", another value was associated with each pattern. This value represents the frequency of appearance of each pattern by the number of times the same pattern appeared in that sample. For example, if the same pattern appeared between 1 and 14 times, the number "18", representing "Low Frequency of Appearance", was associated with it. The value "19" for "Medium Frequency Appearance" means that the same pattern appeared between 15 to 49 times in that sample, and for "High Frequency Appearance", the number "20" was assigned, meaning that that pattern appeared 50 or more times, as shown in the example (Figure 4):

Guild ∉ or ∈	Level L/M/H	Race	Class Mel/Mag	Pattern Grouped	Pattern No.	Pattern Freq.
0	2	5	10	0,2,5,10	1	19
0	2	5	11	0,2,5,11	2	19
0	2	6	10	0,2,6,10	3	18
0	2	6	11	0,2,6,11	4	18
0	2	7	10	0,2,7,10	5	18
0	2	7	11	0,2,7,11	6	18
0	2	8	10	0,2,8,10	7	18
0	2	8	11	0,2,8,11	8	18
0	2	9	11	0,2,9,11	9	18
0	3	5	10	0,3,5,10	10	18

Figure 4. Sample with the values interpretable by the network with the format of the consolidate pattern (Pattern Grouped), number of the pattern (Pattern No.) and the assiduity of the pattern (Pattern Frequency).

It was noticed in previous tests that the neural network had difficulty in distinguish the "Pattern Grouped" values, since they have a narrow difference in terms of numbers, for example "02510" and "02511". This uncertainty caused considerable high rate of error, so the "Pattern Number" was added. The network had more accurate results with smaller and distinguished values.

After associating the "Pattern Number" and the "Pattern Frequency" to each of the patterns in the sample, it was possible to create two sets for the neural network to be trained. The first set was the Pattern Recognition (Figure 5), where the sequence of the four attributes (Guild, Level, Race and Class) is given as input to the network and, as the output, the unique number associated to that pattern (Pattern Number). The second set was the Assiduity Recognition (Figure 6), where the same input of the Pattern Recognition set is given but, as output, the assiduity number associated to each pattern (Pattern Frequency), since the neural network is trained using the supervised learning:

Input 1	Input 2	Input 3	Input 4	Output
1	4	9	11	49
1	4	7	10	44
0	2	8	10	7
1	4	9	10	48
0	2	5	10	1
1	4	9	11	49
1	4	9	11	49
0	2	6	11	4
1	2	7	11	28
1	4	8	11	47

Figure 5. Pattern Recognition training set example (only first 10 rows).

Input 1	Input 2	Input 3	Input 4	Output
1	4	9	11	20
1	4	7	10	19
0	2	8	10	18
1	4	9	10	19
0	2	5	10	19
1	4	9	11	20
1	4	9	11	20
0	2	6	11	18
1	2	7	11	18
1	4	8	11	19

Figure 6. Assiduity Recognition training set example (only first 10 rows).

C. Neural Network Architecture

The network used in this study was the Multi-Layer Feed-Forward with the Backpropagation learning algorithm. Through training and testing, the Backpropagation method is efficient on identifying patterns presented to the network [4]. The Multiple Back-Propagation v2.2.4 [20] software was used to simulate the neural network. The network have 4 neurons on the input layer, 7 neurons on the hidden layer and 1 neuron on the output layer, as seen in Figure 7:

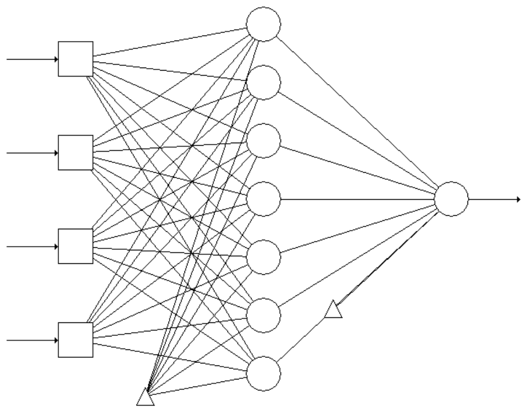

Figure 7. Developed Artificial Neural Network topology.

The configuration parameters were set as follows (Figure 8):

Learning method	Supervisioned
Number of neurons in the input layer	4
Number of neurons in the hidden layer	7
Number of neurons in the output layer	1
Number of hidden layers	1
Transfer function	Sigmoid
Number of epochs	Variable
Root Mean Square Error	0.1
Learning Rate	0.01
Momentum	0.9
Weights variation	Between -1 e 1
Number of training patterns	569
Number of testing patterns	Variable

Figure 8. Developed Artificial Neural Network parameters.

D. Tests and Results

The neural network was used to train and test two types of sets. The first, called Pattern Recognition, was modeled to understand if given the four inputs, separately, representing the characteristics of the players, the network would be capable to identify this profile as a single pattern. The second, called Assiduity Recognition, was modeled to understand if given the same inputs (Figure 6), the network would be capable to identify how is the assiduity of that pattern in the game.

For the Pattern Recognition, the sample: 06-Aug-2007 00:05:17 with 569 records was used as the training set, organized with 4 inputs and 1 output, as shown in Figure 5. For the testing set, the sample: 27-Jul-2007 23:05:19 with 709 records was randomly selected. The testing set sample was modeled as the training set and, for a better visualization of the results, only 100 records were (randomly) selected to be the full testing set.

Once defined both training and testing sets, the neural network was executed. As a result, it was observed that 49 distinct patterns were found in the training set. It means that, in the sample from 06-Aug-2007, 49 unique combinations of this four characteristics could be found from a total of 60 possible combinations, according to this modeling. From the 49 patterns found in the training set, which represent 100% of the patterns of this sample, the neural network recognized 41 singular patterns, which represent 84% of the accuracy.

The black line in the graphics refers to the desired output and the red line refers to the output given by the network. When the red line overlaps the black line, it means that the network could recognize the pattern in that point. The Y-axis represents the number of the patterns that the network is trying to recognize, while the X-axis, the number of entries, in this case, the number of records (rows) of the sample. Below are found the training (Figure 9) and the testing (Figure 10) results:

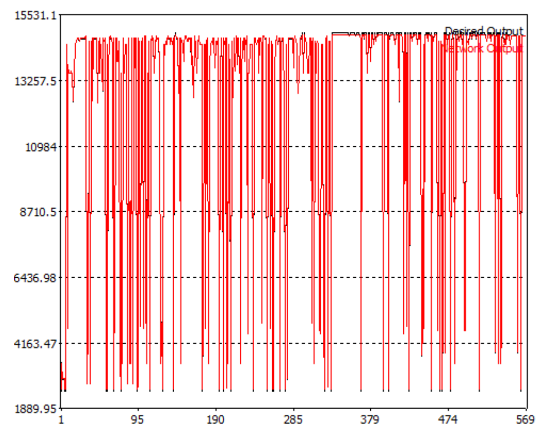

Figure 9. Neural Network Pattern Recognition training outputs. Desired output (black line) and Network output (red line).

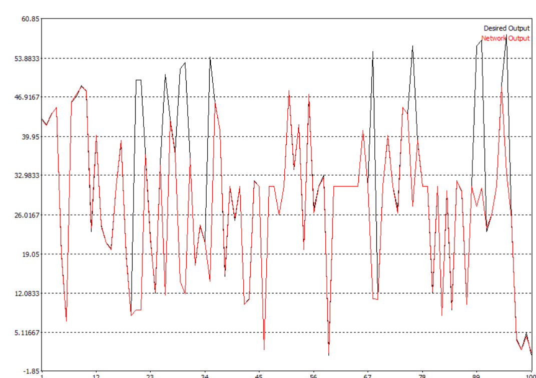

Figure 10. Neural Network Pattern Recognition testing outputs. Desired output (black line) and Network output (red line).

For the Assiduity Recognition test, the same sample was used for the training set, but with the output metrics representing the frequency of appearance of that pattern in the game. The testing set received the same modeling, but the sample was the log from one year further: 06-Aug-2008 00:00:08, with 399 records, which only 100 records were randomly selected to compose the assiduity testing set. As a result, the network could identify 91% of the patterns and outputs if these patterns have low, medium or high assiduity in the game, as shown in the Figure 11:

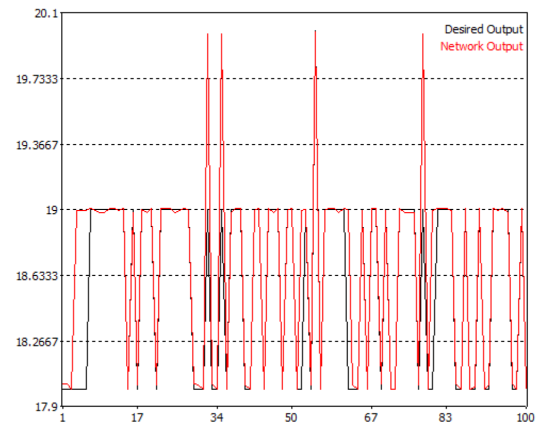

Figure 11. Neural Network Assiduity Recognition testing outputs. Desired output (black line) and Network output (red line).

This was a specific modeling designed to be used as an experiment to show how neural networks can be used to analyze and extract information from the players' data in Cloud Games. The results observed and different modeling can be used to build metrics for decision-making by the developers and game publishers. The volume of data used in this experiment to train and test the neural network was restrict, only small samples of data were used. It is known that the volume of data in the Cloud Gaming is superior to the supported by conventional analytic tools, therefore it is necessary the use of Big Data platforms to manipulate and analyze the data. An approach that involves the global information of the dataset, in order to represent a Big Data problem is presented below.

IV. Big Data Analytics for Pattern Recognition

The platform used in this study was the Cloudera Distribution for Hadoop (CDH) [5], [11] with a cluster of four virtual machines (VMs) in a public Cloud environment. One of them, the master host, with one NameNode and one DataNode and the three others (slaves) with only one DataNode each. In addition to the nodes on master host, some other components were installed: HBase, HDFS (Hadoop Distributed File System), Hive, Impala, HUE (Hadoop User Experience), Pig, Spark, YARN (MR2) and ZooKeeper [1], [16], [17].

The entire WoWAH dataset was used to test how the modules of the Hadoop ecosystem embedded in the CDH would perform and which results would be obtained with all the data. As the dataset is not in a relational database, instead, it is divided in text files, the first step was to organize them in a table within a database, since it is already structured [5], [18].

The data was prepared in a way that it could be imported and analyzed. Using the Metastore Manager through HUE, a new database was created and also a table associated with this database. The table contains the same fields (columns) as the files in the WoWAH dataset, so it was imported to the database using the HDFS. The table "avthist" was formatted with 12 columns and 36,513,647 rows, containing the total of 438,163,764 values.

With the database ready, many different queries could be performed using Impala, an analytical SQL-like tool, which works similarly to Hive. The Tableau (v10) software was used as the data visualization tool for the results generated by the Impala queries. Connectors can be applied to integrate the CDH to the Tableau, in a way that the last can have the data ingested in real time [17], [26].

The first analysis made intended to understand the global information about the characteristics of the players within the database. A initial query was created using Impala to count the frequency of appearances by Race, Class, Level and if the avatars belongs to Guilds or not. Accessing the results using Tableau and performing a graphical visualization, the results showing the most assiduous combination of these characteristics can be seen in the Figure 12:

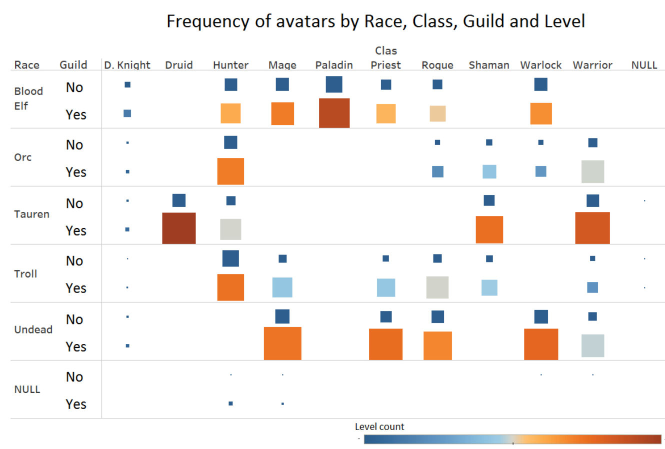

Figure 12. Comparative counting of avatars' characteristics combinations. The bigger the square sizes, the more incidences of specific combinations, as lower levels appearing in dark blue and higher in dark orange.

The "NULL" column and values represents non-identifiable characters or values with errors in the dataset.

It can be observed a relevant difference between the combinations with and without Guilds, both their Level and frequency in the game, which shows a relation between attractiveness and guild participation. It can also be seen that the two most assiduous combination are "Undead Mage" and "Tauren Druid" that belongs to Guilds.

Considering this results, a query was made to relate the amount of avatars with and without Guilds, the result can be seen in the Figure 13:

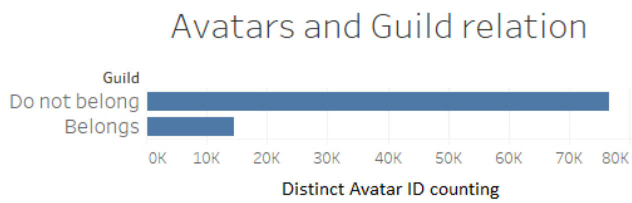

Figure 13. Relation between the number of avatars that belongs and do not belongs to guilds. The Y-axis shows the number of players.

An assiduity query by players was made and compared with the guild participation analysis. It was discovered that they are inversely proportional: 84% of the avatars that does not belongs to any Guild represents 17% of the total frequency in the game. The 16% of the avatars that belongs to Guilds are 83% of the time in the game.

In order to understand players' behavior, a heatmap was developed using queries that relates the avatars and their characteristics with the Zones that they were at some point in the game. With this approach, the most and less frequented Zones can be viewed in real time and be filtered by Race, Class and Level, which allows to get different results depending on the filter selected. The zones considered cold (less frequented) appear in dark blue, while areas considered hot (most frequented) in orange to red. The heatmap was plotted over the WoW continent maps, divided by Zones (Figure 14):

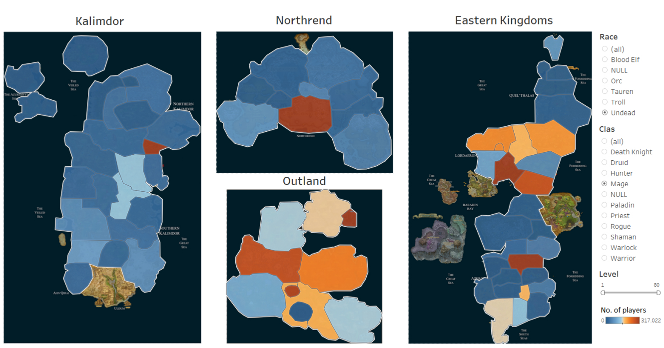

Figure 14. World of Warcraft's heatmap. On the left: the Kalimdor continent, on the right: Eastern Kingdoms, at the top: Northrend and at the bottom: Outland. The selected filter shows Race: "Undead" and Class: "Mage" of all Levels. The areas without colors did not have any incidence of players, most of them were built in patch/expansion released after the period of the dataset.

Players' movement patterns can be analyzed through this type of visualization. Since this information can be accessed in real time using specific avatar's characteristics, a detailed study can be developed to understand each race or class behavior. Using additional information, like the players' IP addresses, applied to this analysis model, the game developers can build a real geographical world heatmap that shows the hotter and coolers regions/counties on the planet. This information can help to decide where to build another Cloud server hub to improve players' gameplay performance, for example.

Understand the assiduity of the players can bring important information for pattern analysis. A query counting the avatars' gameplay time, with the number of hours spent by all players individually during the period of analysis of the dataset, as well as their respective Level, Race and Class was carried out and plotted, as can be seen in Figure 15:

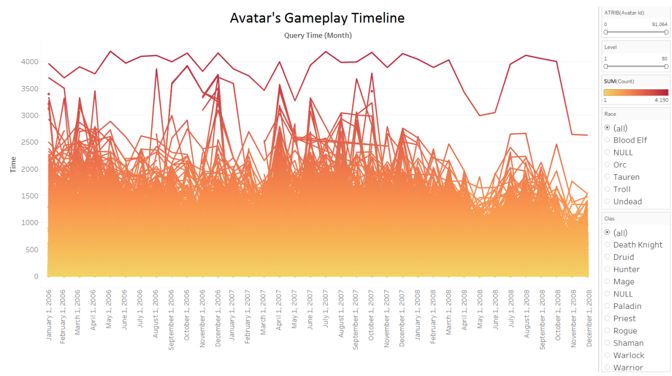

Figure 15. Avatars' gameplay timelines. Each point in the Y-axis represent the total amount of time spent by each player per month (X-axis). This chart can be filtered by Race, Class, Level and the avatars can be selected individually.

Each timeline refers to an avatar, all avatars are included in this plot (can be filtered). It is visible in this graph the presence of an outlier (the higher line), an avatar who played significantly longer than most of the other players during the whole period. The outlier detected is the avatar number (Avatar ID) 182. On average, it played 635 hours per month, which represents being online 88% of the time (Figure 16).

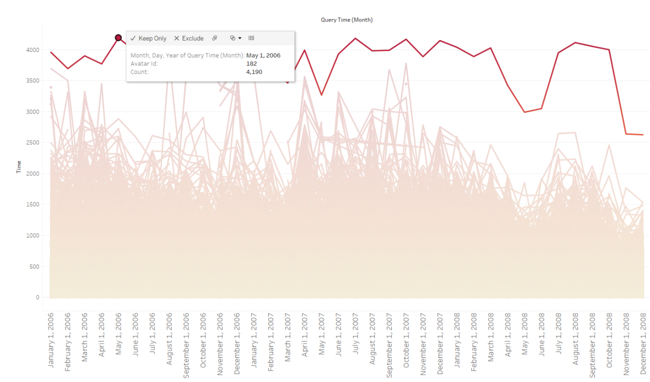

Figure 16. Avatar 182 gameplay timeline. Each point in the highlighted line (Y-axis) represent the total amount of time spent by avatar 182 per month (X-axis).

For a better understanding of its behavior, a specific analysis was made based on its gameplay patterns. A query combining Query Time, Level, Guild, Race, Class and Zone of the Avatar ID 182 was performed and a visualization chart was created (Figure 17):

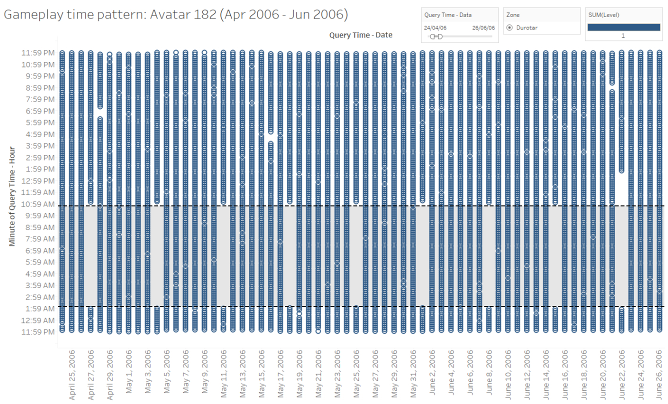

Figure 17. Avatar 182 gameplay time pattern from April 2006 to June 2006. The blanks spaces represent the avatar's inactivity in the game, the time it was offline. The circles represent the record of activity, the time the avatar was online. The Y-axis shows the hours and the X-axis, dates.

Each circle represents a record containing the minutes of the day that the player was online. The color represents the levels achieved by the avatar and the Zone tab in the upper right corner shows the areas that this avatar had been.

The result shows that despite being the most assiduous player of the database, it has not advanced any level, remaining at level 1. It did not belong to any guild and has not changed any zone, staying in "Durotar" for the entire period. In addition, a set of patterns based on logins and logouts time can be detected. As seen in the Figure 17, the avatar goes online directly for six days and logs out for a period of about eight hours every Thursday. The dashed line indicates the pattern.

A comparison of the three years gameplay time was plot and it is possible to identify the behavior and avatars pattern during the whole period, as shown in Figure 18:

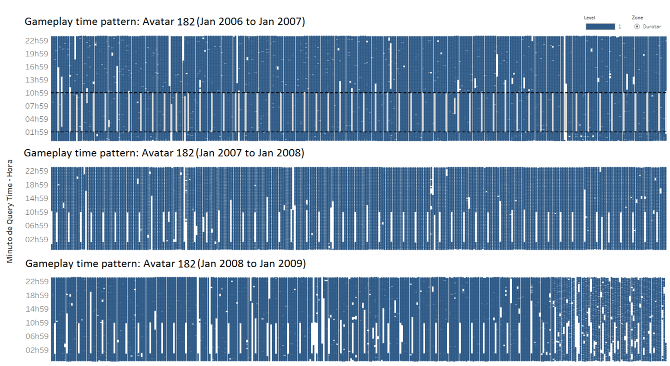

Figure 18. Avatar 182 gameplay time pattern from January 2006 to January 2009. The blanks spaces represent the avatar's inactivity in the game, the time it was offline. The circles represent the record of activity, the time the avatar was online. The Y-axis shows the hours and the X-axis, dates.

The same analysis was performed for the avatar number (Avatar ID) 57, the second most assiduous avatar in the database, and its game pattern can be seen in the Figure 19:

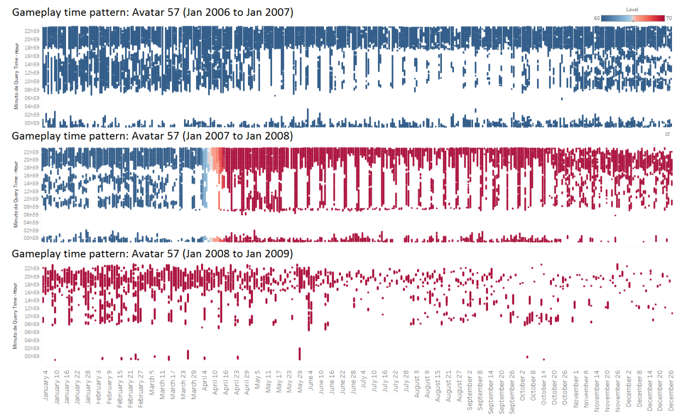

Figure 19. Avatar 57 gameplay time pattern from January 2006 to January 2009. The blanks spaces represent the avatar's inactivity in the game, the time it was offline. The points represent the record of activity, the time the avatar was online. The Y-axis shows the hours and the X-axis, dates.

A comparison can be made between this two gameplay time patterns. Related to the avatar 182, a very linear time pattern is observed and a strange behavior can be identified, since it is the most assiduous player but did not upgrade its level or move to any other zone. According to this particularities mentioned, based on its evolution/movement behavior, it is believed that the avatar is administered by a robot (bot). The usage of bots are considered an illegal activity in most of online games, including WoW [2].

Bots are generally used in online games to perform repetitive activities or those which a human player does not want to perform. Frauds are common in online games, but complex to be detect given the significant number of players in the virtual world. Using this model of analysis, there is the possibility of detection with higher accuracy.

As for the avatar 57, it can be observed that the gameplay time is not as linear as the avatar 182, but still shows some patterns. The assiduity of the player was more intense by the end of 2006 and 2007 compared to the end of 2008. The gameplay time pattern of player relatively changed around November 2007, seven months after the release of the "Burning Crusade" expansion. The color transition section refers to the time of release of this expansion (03-April-2007, in Taiwan), which allowed the player to increase its level from 60 to 70. At the end of 2008, the player becomes less assiduous, with a possible tendency to leave the game.

By analyzing the six most assiduous avatars, not including the first (Avatar 182), once inferred that it represents a non-human avatar, is possible to notice certain similarities between their patterns. The avatars 57 (2nd most assiduous), 388 (3rd) and 271 (5th) had higher assiduity at the beginning of the period (Jan-2006) and at the time of the first expansion release, and lower assiduity in the second expansion "Wrath of the Lich King" period, released in November 2008. They did not go beyond the level 70, so it is likely that they have not adhered to the second expansion. It is noticeable that the assiduity of these players had decreased significantly at the end of 2008, as seen in the Figure 20:

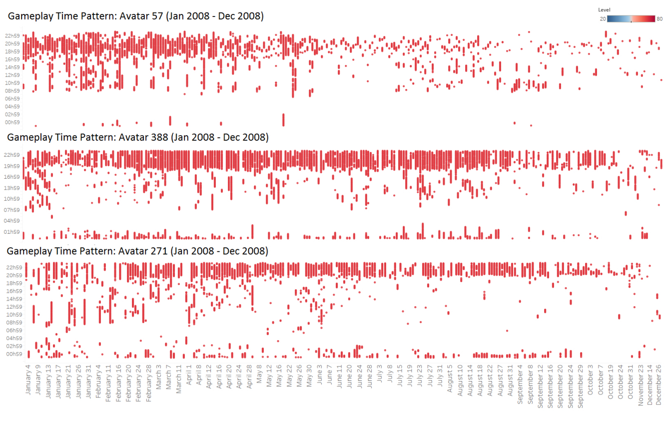

Figure 20. Avatars 57, 388 and 271 gameplay time patterns from January 2008 to December 2008. The blanks spaces represent the avatar's inactivity in the game, the time they were offline. The points represent the record of activity, the time they were online. The Y-axis shows the hours and the X-axis, dates.

In the opposite side, some assiduous players who have joined the second expansion (since they upgraded their level beyond 70), like the avatars 1003 (4th most assiduous), 1450 (6th) and 162 (7th), also have similar patterns between them, but in contrast to the assiduity from the other players mentioned previously, as shown in Figure 21:

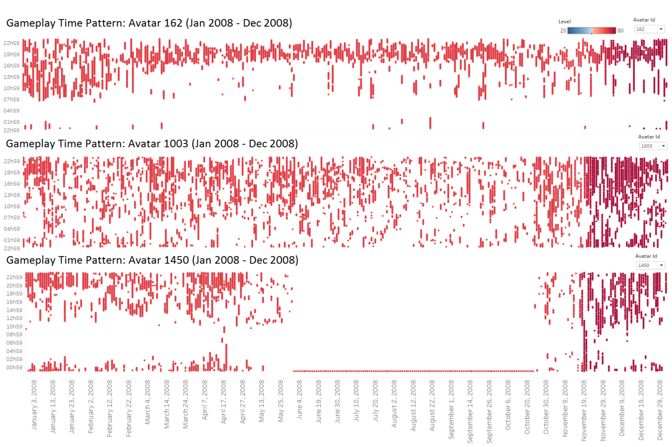

Figure 21. Avatars 162, 1003 and 1450 gameplay time patterns from January 2008 to December 2008. The blanks spaces represent the avatar's inactivity in the game, the time they were offline. The points represent the record of activity, the time they were online. The Y-axis shows the hours and the X-axis, dates.

It is noticed the level transition, from 70 to 80, by the shade of red in the middle of November of 2008. The avatar 162 barely changed its gameplay time pattern the entire year, while the 1003 increased the amount of time spent in the game by October 2008, a month before the expansion release, until the end of the year. It is noticeable that the avatar 1450 not only increased its time online, but returned to the game also a month before the second expansion release, once it has not played in the period from June to October 2008.

The offer of new features, scenarios and other innovative resources that are implemented in patches and expansions draws the players' attention and curiosity for a game. In the researched case, the release of the expansions and its adherence or not by these players can be one of the reasons for the increase/decrease in assiduity and the change in the gameplay patterns. According to the obtained data, it can be concluded that this is one of the major factors, but others can not be discarded.

The game developers can control all the data generated by the players on their games and platforms. Data lakes can be built using different sources to generate more accurate information and results which will provide necessary inputs for assertive and detailed analysis about their environments. The methodology and tools used in this research can be applied to support game developers on their search for data science technics and optimize relations between them, their games and their public.

V. Conclusion

The Cloud Gaming is changing the way the industry relates with its audience and the game market itself. The challenges are not only in the infrastructure of the model, but how to extract useful information of all structured and unstructured data that will be generated, in order to drive the improvements of the games and the understanding of the players. The knowledge of the games must be entirely in developers' control. As for the management and enhancement of the Cloud Games, they should understand how users interacts with the continuous changing virtual environment.

The objective of this research was to show how to extract relevant information from data based on players' characteristics, actions and behaviors using Big Data and neural networks of a online game environment. The results of the tests with the neural networks shows how it can be trained to recognize groups of players by their characteristics and learn the frequency of these groups in the game. Using the Big Data platform CDH, it was possible to map their gameplay patterns by game usage time and relate it to game modifications, like expansion releases. Using virtual geographic coordinates, a heatmap was created to identify frequented zones according to the characteristics of each avatar or groups of avatars. In addition, illegal activities of the players were detected.

Different analysis combining the use of this two methods can be made depending on the modeling, leading to inputs and insights that empowers and supports the developers' decision-making.

References

[1] Apache Software Foundation Oficial Website, 2016. http://www.apache.org

[2] Blizzard Entertainment, Batlle.net Oficial Policy Website, 2016. https://us.battle.net/support/en/article/policy

[3] Blizzard Entertainment, Inc. Oficial Website, 2016. http://us.blizzard.com/en-us/

[4] Cenggoro, Tjeng Wawan, et al. "Recognition of a human behavior pattern in paper rock scissor game using backpropagation artificial neural network method." Information and Communication Technology (ICoICT), 2014 2nd International Conference on. IEEE, 2014.

[5] Cloudera Oficial Website, 2016. http://www.cloudera.com/

[6] Demchenko, Yuri, Cees De Laat, and Peter Membrey. "Defining architecture components of the Big Data Ecosystem." Collaboration Technologies and Systems (CTS), 2014 International Conference on. IEEE, 2014.

[7] Drachen, Anders, et al. "Guns, swords and data: Clustering of player behavior in computer games in the wild." 2012 IEEE conference on Computational Intelligence and Games (CIG). IEEE, 2012.

[8] Fleury, Afonso, Davi Nakano, and J. H. D. O. Cordeiro. "Mapeamento da Indstria Brasileira e Global de Jogos Digitais." So Paulo: GEDIGames/USP, 2014.

[9] Gaikai Oficial Website, 2016. https://www.gaikai.com/

[10] GamingAnywhere Oficial Website, 2016. http://gaminganywhere.org/

[11] Gualtieri, Mike, et al. "The Forrester Wave: Big Data Hadoop Distributions, Q1 2016." (2016).

[12] Hou, Qingdong, et al. "A Cloud Gaming System Based on NVIDIA GRID GPU." Distributed Computing and Applications to Business, Engineering and Science (DCABES), 2014 13th International Symposium on. IEEE, 2014.

[13] Hu, Han, et al. "Toward scalable systems for big data analytics: A technology tutorial." IEEE Access 2: 652-687, 2014.

[14] Huang, Chun-Ying, et al. "GamingAnywhere: an open cloud gaming system." Proceedings of the 4th ACM multimedia systems conference. ACM, 2013.

[15] Huang, Chun-Ying, et al. "GamingAnywhere: The first open source cloud gaming system." ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 10.1s (2014): 10.

[16] Apache Hadoop User Experience Oficial Website, 2016. http://gethue.com/

[17] Apache Impala Oficial Website, 2016. http://impala.apache.org/

[18] Yeng-Ting Lee, Kuan-Ta Chen, Yun-Maw Cheng, and Chin-Laung Lei, "World of Warcraft Avatar History Dataset" In Proceedings of ACM Multimedia Systems, 2011.

[19] Mayer-Schnberger, Viktor, and Kenneth Cukier. Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt, 2013.

[20] Multiple Back-Propagation Oficial Website, 2016. http://mbp.sourceforge.net/

[21] NVIDIA Oficial Website, 2016. http://www.nvidia.com.br/object/nvidia-grid-br.html

[22] OnLive Oficial Website, 2016. http://onlive.com

[23] Perrow, Mike. "Knowing your game: How electronic game companies use Big Data for retention and monetization." Technical White Paper, 2014.

[24] Playstation Now Oficial Website, 2016. https://www.playstation.com/en-us/explore/playstationnow/

[25] Shea, Ryan, et al. "Cloud gaming: architecture and performance." IEEE Network 27.4: 16-21, 2013.

[26] Tableau Software Oficial Website, 2016. http://www.tableau.com/

[27] Xue, Zheng, et al. "Playing high-end video games in the cloud: a measurement study." IEEE Transactions on Circuits and Systems for Video Technology, v. 25, n. 12, p. 2013-2025, 2015.

[28] Wallis, Thomas. "The challenges and opportunities for cloud in gaming." Cloud Gaming USA, 2015

[29] World of Warcraft Avatar History Dataset available at http://mmnet.iis.sinica.edu.tw/dl/wowah/