The Thirty-First International Florida
Artificial Intelligence Research Society Conference (FLAIRS-31)

Identifying Niche Stage in MMORPGs Using Ensemble Classifier

Luiz Bernardo Martins Kummer, Júlio Cesar Nievola, Emerson Cabrera Paraiso
Graduate Program in Informatics, Pontifícia Universidade Católica do Paraná, Rua Imaculada Conceicao, 1155, Curitiba-PR, Brazil
{luiz.kummer, nievola, paraiso}@ppgia.pucpr.br

Abstract

In this paper, we present a Machine Learning based approach to identify the Niche stage in Massively Multiplayer Online Role-Playing Games during their usage lifecycle. The Niche stage is the last stage of the lifecycle and represents a risky situation, because its rentable usage may end soon, as the engagement of the new players are dropping. This paper is related to identifying risky situations based on the players' commitment. Commitment is a measure that illustrates the players' engagement to a game. Our approach consists in evaluating how different ensemble's configurations can affect the Niche identification. We support that the game industry may use the Niche identification to evaluate how attractive a game is. Four different ensemble's configurations are proposed, being each one associated to a research hypothesis that we aim to accept or reject. We applied this approach to a real dataset, which contains three years of usage data from World of Warcraft. We could identify some similarities and differences between all the configurations. It was also possible to identify an assumption that the players' behavior tends to be a Niche behavior over time. The results show that the reduced ensemble with entrance requirement is the best configuration, because it has less classifiers than the others and it can capture the changes on players' behaviors over time.

Introduction

Massively Multiplayer Online Role-Playing Game (MMORPG) is a game in which a very large number of players interact with one another within a virtual world. This kind of entertainment game is very popular. The revenue for the WoW (World of Warcraft), one of the most famous MMORPG (Lee et al. 2011), was more than one billion dollars in 2014 (Tassi 2016). To captivate new players, games become increasingly complex, opening an opportunity for using Artificial Intelligence (AI) (Silver et al. 2016). AI can provide solutions to different game needs, such as intelligent behavior (Galway, Charles and Black 2008). Among the different subareas of AI in games, Machine Learning is one of the most used (Galway, Charles and Black 2008).

In this work, we are using Machine Learning to identify the Niche stage in games, based on data gathered when they are played (usage data). The Niche stage is the last stage of a set of stages defined by Cook (2007) (more details in the "Game Usage Lifecycle" section) and represents a risky situation that may end the rentable usage of a game. It turns the Niche identification a valuable information to game producers.

The usage data applied to this work contains the players' commitment. Lee et al. initially collected the dataset in (2011) and Kummer and colleagues added the commitment attribute in (2017). The commitment is divided in three degrees, which means that players are labeled as low, average, or high on commitment. The dataset time-span starts in January 2006 and finishes in January 2009 (37 months).

In this work, we apply different configurations on the ensemble aiming at evaluating aspects related to the quantity and quality of the ensemble's classifiers. The analysis is formalized through four research hypotheses (more details in the "Identifying the Niche Stage Using Ensemble" section).

We advocate that in MMORPGs, the Niche stage is a key factor to follow. It can be used as one of the metrics a game producer should gauge over time, to decide when to release a new version, validate the acceptance of an upgrade, or even, to finish the operation of a game.

The paper is organized as follows: after the introduction, a game usage lifecycle overview is described, and then related works are presented. Next, the proposed approach based on ensemble is described, proposing four research hypotheses. Experiments that accept or refuse the hypotheses are presented. Finally, conclusions and future works are discussed.

Game Usage Lifecycle

The game lifecycle can be represented by the quantity of players playing a game over time as well as their behavior

when playing. Speller (2012) shows how game producers analyze the game lifecycle. Game producers have some concerns when a game is on the market (e.g., "*Are players still motivated to play*?"). To answer this kind of question, they use some metrics which can identify the usage of their games. The most used is called MAU (Monthly Active Users). MAU counts the number of unique users (players) during a specific time-span (usually the previous 30 days). Figure 1 illustrates the game lifecycle through the MAU's view for several games.

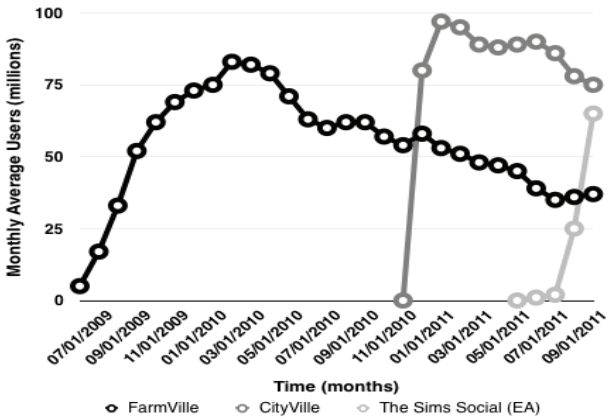

Figure 1. MAU of different games (adapted from (Speller 2012)).

The game genre lifecycle was defined by five stages (shown in Figure 2), representing each one specific player and game producer behaviors. According to Cook, when a new kind of game is released on the market, players will initially test the genre, if it is accepted, then the usage of that genre grows until it starts to fall. The decline is due to a lack of motivation, which can represent that the genre is not fun as it was before. The decline remains until the end of the genre.

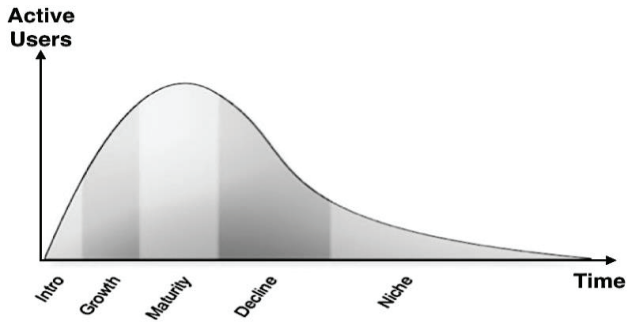

Figure 2. Game genre lifecycle (adapted from (Cook 2007)).

The Intro stage represents the curiosity stage. At this moment, players are very curious about the new genre and will test it to discover what is offered to them. If the genre is accepted, then the Growth stage begins. The number of active players grows and more games in this genre are developed. On the Maturity stage, the genre is consolidated, and great game producers start to develop games in this kind of genre. The players are very interested in the genre. The Decline stage is related to the beginning of demotivation, when players start to have a lack of motivation and less games are developed. The final and the worst stage is the Niche stage. In that stage, the remaining players are very attached to the game and hardly-ever leave it. This stage is characterized by having more players with high commitment than low commitment (Cook 2007). Low commitment players are players who started to play recently and are learning how to play the game.

Zhu, Li and Zhao in (2010) studied online game lifecycle, interested in identifying player behavior. They identified the following four stages interviewing players: Try, Tasting, Retention and Abandonment. The Try stage is characterized by a player who is curious about the game and intends to try the game for the first time. In the Tasting stage the player has already approved the game and starts to accumulate profit (e.g. levels, items, quests, and friends). The next stage is the Retention, where players know everything about the game. When the game does not have challenges anymore, it leads players to leave the game (Abandonment).

Comparing the studies of Cook (2007) and Zhu et al. (2010), it is possible to identify some similarities between player behaviors on both models. The Intro stage of Cook represents the first behavior stage, like the Try stage of Zhu and colleagues. The Growth and Maturity stages of Cook represent the acceptance of a genre, like the Tasting stage, which also represent an acceptance and a growth of players' abilities. The Decline stage of Cook is like the Retention stage, because it represents the beginning of a disgust behavior. At last, the Niche stage represents the same aspect of the Abandonment stage, because it is the stage where several players leave the game, remaining only very attached ones.

As suggested by Kummer and colleagues in (2017), the game genre behavior fits with the game usage behavior, because initially players are interested, spend a lot of time playing and improving their abilities and then leave the game when it is not fun anymore (lack of new game content).

Related Works

Machine Learning is widely used in games, ranging from virtual bot detection (Kang et al. 2013) to enemy behavior prediction (Weber and Mateas 2009). Usage data are any data which contain information about usage (e.g., login records or a list of online players). The data specificity can change from one game to another (e.g., all player action`s in a match, or just the final result). In this section, we present some related works that focused in one or both aspects: Machine Learning (game mining) and usage data.

Speller in (2012) studied the behavior of usage lifecycles especially in the MAU context. The author created a dynamic system based on usage metrics aiming at predicting future behavior of MAU in a bad, normal, or good situation.

Tarng and colleagues in (2009) created a model to predict the player departure (abandonment) based on the time spent playing (this time reduces over time until the abandonment). The classification method was based on the Support Vector Machine (SVM) algorithm. Castro and Tsuzuki in (2015) also studied about player departure and their object of study was the login rate.

Another focus of research relates game usage data and the player behavior identification. Drachen et al. studied in (2012) ways to identify behavioral patterns based on player telemetry (usage data). In their research, two datasets were used, one of an online RPG (Role-playing game) and another of a FPS (First Person Shooter) game. The authors applied K-means and Simplex Volume Maximization algorithms to identify clusters. The profiles identified were very specific to each game (e.g., in the FPS game, the identified clusters were: Medic, Assault and Driver). Drachen et al. published another research on this topic in (2014). The goal was to evaluate different results from the same dataset according to a clustering algorithm chosen. The authors worked with Archetypal Analysis, K-means, C-means, non-negative Matrix Factorization and Principal Component Analysis. One of the conclusions was that K-means is a good algorithm to do this kind of task, because it is less susceptible to outliers (outliers exist with a certain frequency in games and they are represented by players who liked the game very much, playing more time than ordinary players). The identification of the ideal number of clusters was done changing the k value between two and 24 evaluating the mean squared error with the profiles identified.

Ben and Mateas (2009) were motivated to study the prediction of strategies in strategy games. They used Machine Learning algorithms to model players and then predict the enemy strategy before it is applied. The authors collected replays of professional matches and then applied algorithms to model the domain and the opponent. The task was classified as a classification problem and the following algorithms were applied: C4.5, KNN, NNge and LogitBoost.

Pingen and Geert in (2014) also studied the strategy prediction but using neural networks. They trained a model to identify a strategy and propose a counter-strategy. This approach was incorporated in a virtual player (bot).

Uysal in (2016) studied commitment in an investment model. The author based his model on a romantic relationship model, but with a new context (player and game). A questionnaire was applied to 176 people aiming at identifying degrees of satisfaction. This approach did not use Machine Learning techniques.

Kummer and colleagues proposed in (2017) a key risk indicator (KRI) based on commitment applied over the lifecycle. This indicator can illustrate how players' motivation changes compared to MAU. The KRI could identify risky situations where the MAU could not.

The next section presents the methodology used to identify the Niche stage based on ensembles.

Identifying the Niche Stage Using Ensemble

As described by Rokach (2010), "The idea of ensemble is to build a predictive model by integrating multiple models". It means that, given an instance to be predicted, the ensemble applies many classifiers and then through an internal configuration, the final class is labeled to the instance. One of the possible configurations can be the majority vote, where the class most returned from all classifiers is the class chosen. Kummer and colleagues used the ensemble in (2017) to identify commitment (applying the majority vote policy). In this paper, we propose the Niche identification method based on their method, as illustrated in Figure 3.

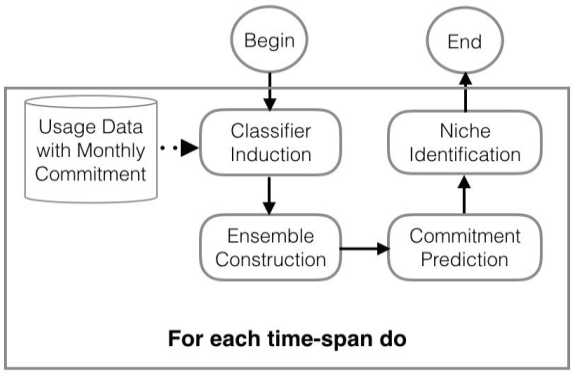

Figure 3. Niche identification method.

This paper focuses in proposing a way to identify the Niche ("Niche Identification" step), as well as changing the configurations on the "Ensemble Construction" step, because it affects the Niche identification. The proposed method works as follows: given a determined time-span (for instance a month), the usage data with commitment is collected. A classifier is induced based on commitment and added to an ensemble. The ensemble's prediction gives players' commitment from the usage lifecycle point of view (many months represented by many classifiers), which differs from the monthly classifier point of view (only one month with its classifier) (Kummer et al. 2017).

To make it clear let us describe a running example. In the first month of the series, there are 2,000 players. They are labeled as following: 1,200 as low, 500 as average and

300 as high committed players (according to the classification proposed by Kummer et al.). Then, a classifier is induced using that data. The classifier is added to the ensemble that gives for the players of the first month its commitment degrees. When a second month is over, the same occurs again, a classifier is induced based on the second month usage data and added to the ensemble. Now the ensemble has two classifiers (one for each month) and uses them to identify for both, the first and the second months, the commitment of their players. When the third month comes, the same occurs and so on. In this strategy, the ensemble will grow incrementally. After 50 months, there will be 50 classifiers in the ensemble.

Concerning the Niche identification, we propose an equation based on the work of Cook (2007) to identify when a game is in the Niche stage. That equation operates on the number of players on each commitment degree for a time-span. If the number of high committed players is higher than the number of low committed ones, then the game is in the Niche stage (Equation 1).

$$NICHE = \{x \in t | \, x_{low} < x_{high}\} \quad (1)$$

where x is a game, t is a distinct time-span and x_{low} and x_{high} are the numbers of players with low and high commitment degrees respectively.

To better explore the use of ensemble in the Niche identification method, we want to investigate some aspects around the relevance of the quantity of classifiers into the ensemble. Our motivation is the identification of how the commitment measurement changes according to different configurations in the observation window. We propose to use the Niche identification through the incremental configuration as a baseline (naming it as baseline ensemble), because it has the "big picture" of the players' behavior over time (this baseline is detailed in the next section). Considering that baseline, we proposed the following research hypotheses:

- **RH1**: An ensemble with an entrance requirement can get the same result as the baseline.
- **RH2**: An ensemble with the last six months can get the same result as the baseline.
- **RH3**: An ensemble with the last 12 months can get the same result as the baseline.
- **RH4**: The monthly classifier can identify the Niche occurrence as the baseline does.

The RH1 can be analyzed through an application of some requirement that controls if a new classifier can be added to the ensemble or not. It enables an evaluation about the necessity to have or not all the monthly classifiers inside the ensemble. The RH2 and RH3 aim at analyzing how the last months influence in the Niche identification. It has the idea of verifying if those observations windows contemplate changes on players' behavior sufficient to detect the Niche. The last hypotheses (RH4) aims at verifying if the ensemble is necessary. The idea consists in comparing the baseline ensemble prediction of a month *m* with its specific classifier (the month *m* classifier).

Experimental Results and Analysis

The experiments described in this section used the dataset presented by (Kummer et al. 2017). The baseline got the following results presented in Table 1. The "month count" is the sequence number of the months, the month 2008-03 is the 27 th one of the usage data. The "stage" column refers to the stage identified to the game in the month. As we can only identify when a game is in Niche, the other stages of Cook (2007) (Figure 2) were not explored, so when the Niche is not identified the "Other" label is applied. The Table 1 starts from the 27th month because the Niche was first detected in the 28th month and ends in the 37th because it is the last month of the series.

Table 1. Niche identification through the Baseline ensemble.

Month Count	Month	Stage
27	2008-03	Other
28	2008-04	Niche
29	2008-05	Niche
30	2008-06	Niche
31	2008-07	Other
32	2008-08	Niche
33	2008-09	Niche
34	2008-10	Other
35	2008-11	Niche
36	2008-12	Niche
37	2009-01	Niche

It is possible to notice that the Niche does not occur continuously after its first identification (2008-07 and 2008-10). In those cases, it happened due to a vacation month and a game upgrade. Table 1 clarifies the fact that the players' behavior in a game tends to be a Niche behavior during time. To check the RH1, we propose the use of a metric called Q Statistics (Yule 1900) to compute the diversity between the candidate new classifier and the ensemble based on the Q value. As described by Yule, positive Q values represent classifiers that tend to recognize the same objects correctly. On the other hand, negative Q values represents classifiers that make mistakes differently. This application allows the identification of classifiers that explore differently the instances space. In our case, we want to maintain in the ensemble only classifiers with

different behaviors (negative Q values), as classifiers with similar behaviors tend to predict similar results, there is not a new information, so removing them may not change the predicted values. After applying this policy, only three classifiers (from January 2006, January 2007, and November 2007) were maintained in the ensemble (named as reduced ensemble). We computed the commitment based on this reduced ensemble and applied a *t* test with $p < 0.05$ to validate this result with the baseline ensemble result. No statistical differences were found for all degrees of commitment. We also computed the Niche for the new result, and it was the same as the baseline. To check the RH4, we applied each specific classifier to its base month data (the first classifier applied only in the first month, the second only in the second month and so on). Table 2 illustrates the comparison between the monthly classifier, the baseline ensemble and the reduced ensemble perspectives. The columns low, average and high reference the quantity of players on each commitment degree. To save some space, we purposely suppressed the values between 2008-04 and 2008-10. For these months, only in 2008-07 and 2008-10 the Niche were not identified, for all the others it was.

Table 2. Baseline Ensemble, Reduced Ensemble, and Monthly Classifier Comparison.

Month	Classifier Approach	Low	Average	High	Niche
2008-11	Baseline Ensemble	3,851	1,166	4,648	X
	Reduced Ensemble	3,886	1,055	4,724	X
	Monthly Classifier	4,586	3,598	1,481	
2008-12	Baseline Ensemble	2,910	977	4,230	X
	Reduced Ensemble	2,913	903	4,301	X
	Monthly Classifier	3,591	3,138	1,388	
2009-01	Baseline Ensemble	1,054	474	2,887	X
	Reduced Ensemble	1,066	420	2,929	X
	Monthly Classifier	1,369	1,743	1,303	

We can see that the reduced ensemble identifies when the Niche occurs as the baseline does, validating the RH1. All the Niche months identified from the monthly perspective are in the baseline, but it could not identify all the Niche months as the baseline does (e.g., in the last three months), rejecting the RH4. For all models, the Niche was found for the first time in 2008-04.

To verify the RH2 and RH3, two experiments were done. They started in the first month. For the first experiment (six-month size), a total of 31 iterations were applied (from the range 2006-01/2006-06 to 2008-08/2009-01). For the second experiment (12-month size), a total of 25 iterations were applied (from the range 2006-01/2006-12 to 2008-02/2009-01). Figures 4 and 5 illustrate for each iteration the quantities of Niche months identified.

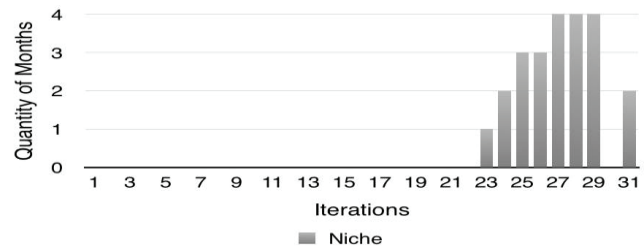

Figure 4. Six-month size niche identification.

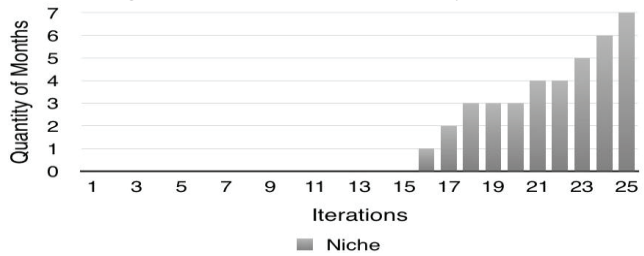

Figure 5. 12-month size niche identification.

The behavior of the six-month observation differs from the 12-month one. From the 12-month view, the Niche occurrence is always stable or growing from one iteration to another. It does not happen in the six-month view, where there is a growth in the number of Niche months and then it drops to zero. It means that looking at just six months, it is not possible to identify the Niche according to the baseline, because this window size does not contemplate the changes on player behavior sufficiently to identify the Niche. As Niche is an uncommon behavior in the lifecycle (the end of it), if only Niche months are observed, there are not any "uncommon" month, so the Niche cannot be identified.

We compared the last six months from all perspectives (baseline, reduced, six-months and 12-months) (illustrated in Table 3). The 12-month experiment is very similar to the baseline and reduced ones. The only disagreement was in the 2008-08, where it was a Niche for the baseline and reduced perspectives and "other" for the 12-month one. However, some similarities also exist in all experiments: the last two months were pointed as Niche; it means that even in a six-month observation window, these months showed a very distinct behavior that enable Niche to be identified. Another situation was the upgrade (2008-10), where the Niche was not identified in any perspective. As the six-month view and 12-month view did not fit into the baseline, RH2 and RH3 were rejected.

Conclusion and Future Works

In this paper, we proposed a method to identify the Niche and some experiments over the ensemble management. A baseline was defined, and four research hypotheses were described and tested on WoW usage data. Only one of four

Table 3. Observation Window Comparison.

Last six month	Baseline/ Reduced	Six-month	12-month
Month	Niche	Niche	Niche
2008-08	X		
2008-09	X		X
2008-10			
2008-11	X		X
2008-12	X	X	X
2009-01	X	X	X

research hypothesis was confirmed. The reduced ensemble with entrance requirement could keep the baseline with only three classifiers, different from the baseline ensemble, which has 37 (RH1 acceptance). Observation windows of six and 12 months (RH2 and RH3) were not sufficient to identify the Niche as the baseline does. The experiments showed that even if a reduced ensemble can have few months (three in our case), it can have better results than ensembles with more classifiers, even if they refer to the last months of the series, because the reduced ensemble has only classifiers with different behavior. The rejection of RH4 (monthly classifier approach) symbolizes that it is necessary to look from a wider perspective to identify the Niche, justifying the use of an ensemble. It clarifies the fact that, for WOWAH dataset, the better strategy consists in looking at all the classifiers and accepting the different ones to enter the ensemble. That strategy enables the detection of changes on players' behavior over time, illustrating the tendency of the Niche behavior.

As future work, we intend to evaluate ways to identify the other stages of the lifecycle. We also intend to apply this approach to other games and revalidate the conclusions. Other possible work consists in evaluating how different time-spans of the same data can affect the Niche identification.

We hope that this paper helps researchers to interpret the game usage "on-the-fly", fomenting some new explorations, experiments and interesting results.

Acknowledgments

The authors would like to thank CAPES-Brazil (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) for its support in this research.

References

Weber, B. G. and Mateas, M. 2009. A data mining approach to strategy prediction. In *Computational Intelligence and Games*, IEEE Symposium on. IEEE, pp. 140-147.

Castro, E. G. and Tsuzuki, M. SG. 2015. Churn prediction in online games using players' login records: A frequency analysis approach. In *IEEE Transactions on Computational Intelligence and AI in Games* 7.3. pp. 255-265.

Cook, D. 2007. The Circle of Life: An Analysis of the Game Product Lifecycle. In *Gamasutra*. [Visited in 2017: http://www.gamasutra.com/view/feature/129880/the_circle_of_life_an_a nalysis_of_.php]

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

Drachen, A., Thurau, C., Sifa, R., & Bauckhage, C. 2014. A comparison of methods for player clustering via behavioral telemetry. In *arXiv preprint arXiv:1407.3950*.

Galway, L., Charles, D., Black, M.. 2008. Machine learning in digital games: a survey. In *Artificial Intelligence Review*, Volume 29, Issue 2, pp 123161.

Kang, A. R., Woo, J., Park, J., and Kim, H. K. 2013. Online game bot detection based on party-play log analysis. In *Computers & Mathematics with Applications*, 65(9), 1384-1395.

Lee, Y., Chen, K., Cheng, Y. and Lei, C. 2011. World of Warcraft Avatar History Dataset, In *Proceedings of ACM Multimedia Systems,* p. 123-128.

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

Pingen and Geert L.J. 2014. An implementation of a data-mining approach to strategy selection in Starcraft: Brood War. Dissertation presented at Radboud University Nijmegen.

Rokach, L. 2010. Ensemble-based classifiers. In *Artificial Intelligence Review*, v. 33, n. 1, p. 1-39.

Speller, T.H., 2012. The business and dynamics of free-to-play social casual game apps. Master thesis, MIT.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., ... and Dieleman, S. 2016. Mastering the game of Go with deep neural networks and tree search. In *nature*, 529(7587), 484-489.

Tassi, P. 2016. World of Warcraft Still A $1B Powerhouse. Forbes. [Visited in 16th, November 2016: Forbes.com]

Tarng, P., Kuan-Ta, C., and Huang, P. 2009. On prophesying online gamer departure. In *Proceedings of the 8th Annual Workshop on Network and Systems Support for Games*. IEEE Press, p. 16.

Uysal, A. 2016. Commitment to multiplayer online games: An investment model approach. In *Computers in Human Behavior*, 61, 357-363.

Zhu, L., Li, Y., Zhao, G. 2010. Exploring the Online-Game Life Cycle Stages. In *IEEE International Conference on E-Business and EGovernment (ICEE)*, 2436-2438.

Kummer, L.B.M., Nievola, J.C., and Paraiso E.C. 2017. A Key Risk Indicator for the Game Usage Lifecycle. In *Proceedings of the Thirtieth International Florida Artificial Intelligence Research Society Conference*, pp. 394-399.

Yule, G. 1900. On the association of attributes in statistics. Phil. Trans., A, 194, 257–319. 1900.