

Prediction Outcome for Massive Multiplayer Online Games Using Data Mining

Shazwani Samsurim, Nor Ashikin Mohamad Kamal, Marina Ismail, Norizan Mat Diah

Department of Computer Science, Faculty of Computer and Mathematical Sciences

Universiti Teknologi MARA

Shah Alam, Selangor, Malaysia

Article Info

Article history:

Received Jan 8, 2018

Revised Mar 15, 2018

Accepted Apr 21, 2018

Keywords:

Classification

Games

IBK

J48

MMOG

Naïve Bayes

ABSTRACT

Massive Multiplayer Online (MMO) game is one of the famous game genres among teenagers nowadays. MMO games allow gamers to interact and play with up to thousand players. Rainbow Six Siege (RSS) belongs to MMO type of game. However, due to many operators that are available in this game, the player needs to choose the right operator to counter the enemy operator. Therefore, based on the characteristic of the selected operator, this paper attempted to predict the outcomes of the game. In our prediction model, characteristics for these operators were extracted from 120 live stream replays. Three classification algorithms were utilized to predict the outcome of the game. Among these algorithms, IBK had obtained outstanding performance in the dataset. The accuracy of the model is 93.75%, applying 5-fold cross-validation test. The success rate reveals that our proposed model is suitable to predict the outcome of the game.

Copyright © 2018 Institute of Advanced Engineering and Science.

All rights reserved.

Corresponding Author:

Shazwani Samsurim,

Department of Computer Science, Faculty of Computer and Mathematical Sciences

Universiti Teknologi MARA

Shah Alam, Selangor, Malaysia.

Email: shazwanisamsurim@gmail.com

1. INTRODUCTION

The game industry is facing a surge of data which results from increasingly available highly detailed information about the behavior of software and users. Teenagers enjoy playing online games which are challenging but fair to all players. Before online games become popular, gamers usually play a single player game such as Mario and Hugo. After some time the games evolve and become massive multiplayer online. Massive Multiplayer Online (MMO) allows gamers to interact and play with other players up to thousand players. Gamers can stay at home and play with their friends as long they have an internet connection. The rise of online games become more popular when many platforms which provides various collection of online games and allows gamers to communicate with each other and create a community to discuss based the selected games such as STEAM, GARENA and ORIGIN. Game's genre like strategy and first-person shooter are the most well-known genres among gamers.

Data mining has been used to improve the strategy of the game [1]. One of the main challenges for game data mining is to predict strategies to make the game become more challenging and fair to all player. Formassive multiplayer game such as Rainbow Six Siege (RSS), data mining techniques can be use to predict the chosen operator for the user to counter enemy operator ability. RSS comes with multiple operators which have an individual ability to counter in the games thus, it is best to predict which operator is suitable to use for a user to challenge the opponent operator's ability. A large number of replays were collected from live streams from an international tournament. Each replay is divided into many rounds and is labelled based on player chosen operator and class.

This paper tests the accuracy of three classification algorithms to predict the output of massive multiplayer online game. The remainder of this paper is organized as follows. Section 2 discusses past studies of MMO and data mining in games. Section 3 expresses the methods involved in this research. The performance measurements used are discussed in Section 4. Section 5 presents the results, and the conclusion and future works are provided in Section 6.

2. BACKGROUND OF THE STUDY

This research focuses on the operator prediction based on the opponent operator that will be used by the user in RSS. There are several approaches in recognizing the pattern in operator selection. This section discusses the overview of massive multiplayer online game and previous study on games.

2.1. Massive Multiplayer Online Game (MMOG)

An online game is a game controlled by the computer where the player interacts using computer network [2]. Massive multiplayer online games (MMOG) have captured attention in the game industry since millions of people participate in the battle using the same server [3]. Table 1 shows the various MMOGs available on the internet. MMOG represent complex data that can be divided into sub problem thus it is necessary to do data mining in this game to study the pattern or strategy of the game to increase winning ratio of the game play. There are several papers regarding game data mining. Analyzing behavior data from massive multiplayer game data can be challenging since it involves huge numbers of active players across the world. This research paper suggested data mining to explore data and discover patterns that can reduce overall complexity of the data [4].

Data mining algorithm was used to identify the difference between human and AI bots in [9]. In the study, 142 replays were collected and trained using machine learning. The result shows that it is possible to increase the win ratio of the game if the bots able to recognize the opponent's strategy and successful build the order.

Several approaches have been used to model the opponent [10, 11, 18]. The approaches were applied on the Real Time Strategy game to predict the strategy of the opponents so that player can respond accordingly to the game. The predictions had applied KNN, decision tree and logistic regression classifiers to predict the accuracy of the strategy [10]. A Bayesian model was used to predict the result or outcomes of isolates battle and also predict what unit will be used to defeat a given army [11]. The goal of the opponent model approach is to learn and select which combination of unit is more effective against others.

Table 1. Massive Multiplayer Online Game

Title	Graphics	Setting	Subscription Model	Release Date
Star Wars Galaxies	3D	Science Fiction	Pay to Play	2003
World of Warcraft	3D	Fantasy	Pay to Play	2004
Rainbow Six Siege (RSS)	3D	Science Fiction	Buy to Play	2015
Tibia	2D	Fantasy	Free to Play	1997
Xyson	3D	Apocalyptic Fantasy	Buy to Play	2011

Rainbow Six Siege (RSS) has been one of the successful MMOG with huge number of players over the world. Many websites provide the gameplay that stores all the gaming events and to replay the game of good players from that. As a result, the RSS community has accumulated their experience in the form of replays for several years. Recently, RSS game has gained a lot of interest in the gaming industry. Beside Dota2 and Counter Strike, RSS placed third for the most popular game in the world. Figure 1 shows the most popular MMOG in the world in December 2017.

Current Players	Peak Today	Game
307,644	719,771	Dota 2
180,592	515,266	Counter-Strike: Global Offensive
49,879	70,620	Tom Clancy's Rainbow Six Siege
41,985	50,534	Team Fortress 2

Figure 1. Example of Massive Multiplayer Game Online

2.2. Overview of Rainbow Six Siege

Rainbow Six Siege (RSS) is one of the popular First-Person Shooter with real-time strategy which uses current technology and is divided into two teams which is attack and defend. Players can choose 38 operator which is 18 per team with different abilities to win the games. The operator will be classified into 4 different classes in attack and defend. There are 10 players connected either random or in a party which is 5 persons per team. Both teams have different task at the early stage of the game (first 30 seconds) which on attack they need to find the location of the objective and identify the operator taken by the defend using drone. On the defend side they need to protect the objective from attacker thus they need to barricade or reinforce the wall in order to slowdown or prevent the attacker to secure the objective. There are several gameplay modes which is plant the bomb, rescue hostage and secure the area with seventeen different maps. RSS makes the team win by killing all enemy and secure the objective for attackers while on defend either kill all enemy or defend the objective within a time limit. A study has been conducted to predict opponent positioning using opponent modeling in first person shooter [5, 6]. This research focuses on one map which is presidential plane since the map is popular and have many replays on the web.

3. METHODOLOGY

Methodology is the specific steps taken during the project investigation Figure 2 shows the phases involved in this research. It starts with dataset development, extraction of features, classifications by three classifiers Naïve bayes, IBK and J48. It is then followed by results and evaluation.

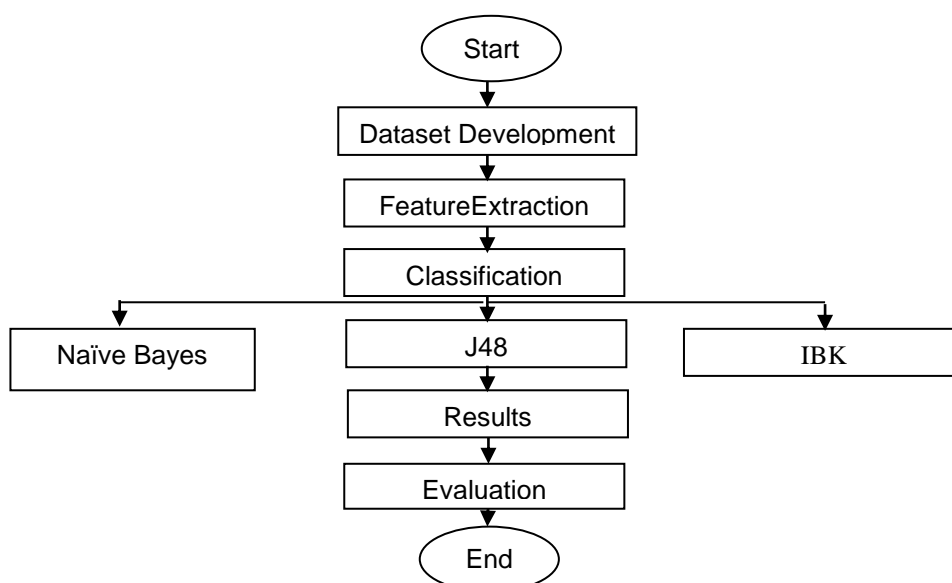


Figure 2. Research Flowchart

3.1. Dataset Development

In the gaming industry, several websites are dedicated to collect and sharing RSS replays. These websites allow sharing a large portion of replays of game play from professional and high-ranked amateur matches. Therefore, using all these data collected from those websites, it is possible to mine these websites in order to build a collection of all possibilities of strategies that user can use to win the game play in RSS. Due to a large number of replays available, it is possible to learn a variety of operator ability and also the strategies on several maps against different play styles to help improve user skills in order to win the game. The data can help improve online matchmaking, modelling player behavior and predicting strategy [10-12]. The data of the replays were collected from www.reddit.com and <http://steamcommunity.com>, two popular RSS website. The collections of the replays from the professional tournament were downloaded from these websites. The scope of this research is limited to one map only. Table 2 shows the list of operators based on class in the RSS.

Table 2. List of RSS Operators based on Class

Attack		Defender	
Class	Operator	Class	Operator
Breacher	Ash, Thermite, Hibana, Sledge, Zofia	Trap	Frost, Kapkan, Ela, Lesion
Shield	Montage, Blitz	Core Defense	Castle, Mute, Rook, Doc, Bandit, Jaeger
Damage	Glaz, Fuze, Black beard, Capitaio, Buck	Damage	Vigil, Tachankan, Caveira, Smoke
Utility	Twitch, Ying, Thatcher, IQ, Dokkaebi, Jackal	Information	Valkyrie, Mira, Echo, Pulse

The remainder of this section will describe data mining approach to automatically learn strategies from collection of replays and subsequently will predict win or lose of the RSS game.

3.2. Feature Extraction

RSS replays are collected based on secure area game modes and presidential plane map. A dataset containing 32 games are created based on the operator’s class taken and result of the operator ability counter. Table 3 represents the characteristic of operator’s class for the attacker and the defender team in RSS where the info is extracted from *rainbowsixsiege.fandom*.

Table 3. Characteristic of Operator Class

Class	Armor (Low, Medium, High)	Movement (Speed 1,2,3)	Flexibility (Low, Medium, High)
Breacher	Low	2	Medium
Shield	High	1	Medium
Damage (Attack)	Low	3	Medium
Utility	High	2	High
Trap	Medium	2	High
Core	High	1	Low
Damage (Defend)	Low	3	High
Information	Medium	3	High

The goal of this research is to build a general model of RSS that can predict the outcome of the game play based on the operator selection. This differs from the previous study which focused on player’s action. By applying data mining on game replays, the algorithm can develop a set of rules that can predict the result of the game play hence the user can learn the operator’s ability through the algorithm suggested then they will improve the strategy to win the game. Table 4 shows the data used in this study based on the matchmaking that had been recorded from the replay. Every attack operator will be matched on every defend operator and the result of win and loss will based on the replay.

Table 4. Characteristics of selected RSS class operator

Class 1	Class 2	Armor		Movement		Flexibility		Result
		Class 1	Class 2	Class 1	Class 2	Class 1	Class 2	
Breacher	Trap	Low	Medium	Speed 2	Speed 2	Medium	High	Lose
Damage	Trap	Low	Medium	Speed 3	Speed 2	Medium	High	Lose
...
...
Trap	Utility	Medium	High	Speed 2	Speed 3	High	High	Lose
Core Defense	Breacher	High	Low	Speed 1	Speed 2	Low	Medium	Lose

3.3. Classification Algorithms

Three classifiers are used in this study. There are Naïve Bayes, (K-Nearest Neighbor), IBK and decision tree (J48).

3.3.1. Naïve Bayes

Naïve Bayes is used to predict the class depending on the probability of belonging to the class [13]. The probability is calculated based on Equation 1.

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \quad (1)$$

where $P(c/x)$ is the posterior probability of class (target) given predictor (attribute), $P(c)$ is the prior probability of class. $P(x/c)$ is the likelihood which is the probability of predictor given class and $P(x)$ is the prior probability of predictor.

3.3.2. IBK (k-Nearest Neighbor)

Lazy learning is a learning method in which the system tries to generalize the training data before receiving queries is delayed until a query is made to the system [14]. Lazy can deal with changes of problem area and solve multiple problem [15]. IBk classifier is like k-Nearest Neighbor (KNN) classifier where by k is a used defined constant [19]. The similarities of different instances are calculated using Euclidean distance as in Equation 2.

$$d_{i,j} = \sqrt{\sum_{s=1}^W (V_i^s - V_j^s)^2} \quad (2)$$

where V_i is the vector of the i-th instance, while V_j is the s-th elements of vector V_i , and d_{ij} is the distance between V_i and V_j .

3.3.3. J48 (Decision Tree)

The J48 classifier (Weka [17] implementation of the well-known C4.5 decision tree algorithm) will generate some rules in order to predict the output variable and also help in describing the critical contributions to become easily understandable [16]. The additional features of J48 were used to find missing values, decision trees pruning, continuous attribute value ranges and derivation of rules.

4. PERFORMANCE MEASUREMENT

The results of the classifiers are evaluated by using accuracy, true positive (TP) rate, precision, F-measure and mean absolute error (MAE). Accuracy assesses the overall effectiveness of the algorithm. It is given by Equation 3.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} * 100 \quad (3)$$

where TP (true positive) and TN (true negative) are the numbers of correctly predicted positive and negative samples respectively. FP (false positive) and FN (false negative) are the numbers of incorrectly predicted positive and negative samples, respectively.

The TP Rate determines the play case ratio for predicted correctly cases to the total of positive cases. It is a probability corrected measure of agreement between the classifications and the true classes. In this TP evaluation, TP rate will determine number of examples predicted positive that are actually positive for the result of the game.

Precision define as positive predictive value that will calculate how many positive predictions are correct for the result of the game. Equation 4 shows the formula to calculate precision.

$$Precision = TP / (TP + FP) \tag{4}$$

where TP refer to true positive and FP refer to false positive. F-Measure is a combination of recall and precision which is in a single performance. Equation 5 shows the formula to calculate F-measure.

$$F-Measure = 2 * Precision * Recall / (Precision + Recall) \tag{5}$$

The mean absolute error (MAE) is used to measure how far the predictions differ from the actual values [20]. The formula for MAE is given in Equation 6.

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - y_i| \tag{6}$$

where n=the number of errors and $|f_i - y_i|$ = the absolute errors.

5. RESULTS AND EVALUATION

This section focuses on the evaluation of the selected classification algorithm. Table 5 shows the accuracy for every classification algorithms, Naïve Bayes, IBK and J48. The results are obtained by using 5-fold cross validation. These measurements are obtained by using Weka [17] toolkit.

Table 5. Classification algorithm and overall accuracy

Classifier	Accuracy (%)
Naïve Bayes	0.75
IBK	0.94
J48	0.66

All the classifier predicts the outcome of the game according to the operator class selected. From Table 5, it can be concluded that IBK algorithm has the highest accuracy compared to Naïve Bayes and J48 which reaches accuracy of 93.75%. While Naïve Bayes reaches 75% of accuracy and J48 have the lowest accuracy value 65%. Figure 3 shows the performance evaluation of Naïve Bayes, IBK and J48 classification algorithms. It compares three types of performance evaluation which are True Positive (TP) rate, precision and F-Measure.

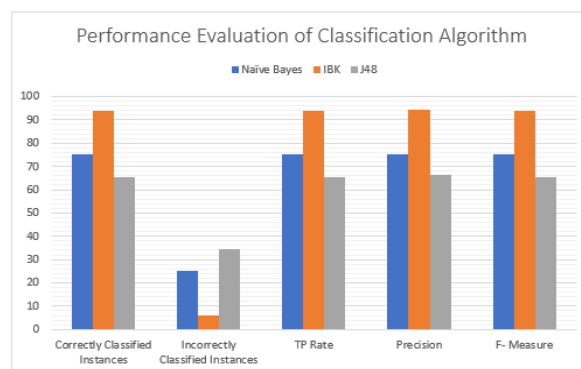


Figure 3. Performance measurement for classification algorithm

From Figure 4, it is observed that J48 algorithm obtain the highest error rate while IBK obtain the lowest error rate. Therefore, IBK classification algorithm perform well since it has the highest accuracy rate.

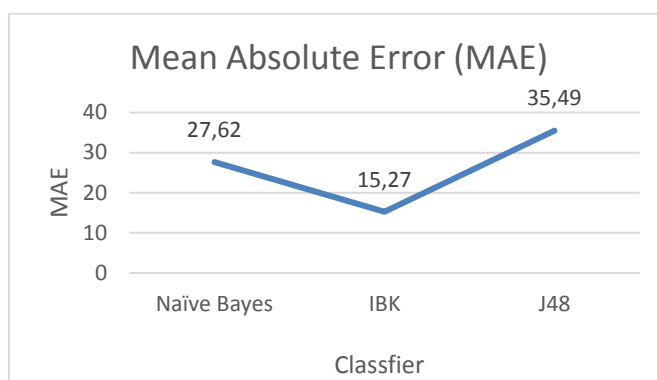


Figure 4. Error rate for classification algorithm

6. CONCLUSION AND FUTURE WORKS

This paper determines the best algorithm by comparing three classification techniques to predict the output of the RSS. This study mainly focus on three classification techniques, Naïve Bayes, J48 and IBK. Experiments have been conducted to determine the performance for each of the algorithm. Among the three classification algorithms, it can be concluded that IBK is the best algorithm to predict the outcome of RSS game compared to Naïve Bayes and J48. Therefore, IBK model is suitable to be applied for prediction of MMO games since IBK algorithm perform better than Naïve Bayes and J48. As for future works, the author plans to investigate the effect of feature selection to the game outcome prediction. Furthermore, using other classifiers should also be used in future experiments.

ACKNOWLEDGMENTS

This work is supported by UiTM Internal Research Grants Scheme (LESTARI) (600-IRMI/DANA 5/3/LESTARI (118/2016)) from the research Management Centre (RMC), Universiti Teknologi MARA, Shah Alam, Selangor. The appreciation also goes to Faculty of Computer and Mathematical Sciences (FSKM), Universiti Teknologi MARA, Shah Alam, Selangor for giving a moral support in the production of this paper.

REFERENCES

- [1] Miljkovic D, Gajic L, Kovacevic A, Konjovic Z. *The use of data mining for basketball matches outcomes prediction*. IEEE 8th International Symposium on intelligent and informatics. Subotica, Serbia. 2010: 309-312.
- [2] Hladky S, Bulitko V. *An evaluation of models for predicting opponent positions in first-person shooter video games*. Proceedings of the IEEE Symposium on Computational Intelligence in Games, Perth, Australia. 2008: 39-46.
- [3] Daneva, M. *How practitioners approach gameplay requirements? An exploration into the context of massive multiplayer online role-playing games*. 22nd IEEE Int. Requir. Eng. Conf. 2014: 3-12.
- [4] Bauckhage C, Drachen A and Sifa R. *Clustering game behavior data*. IEEE Trans. Comput. Intell. AI Games. 2015; 7(3): 266-278.
- [5] Jansz J and Tanis M. *Appeal of Playing Online First Person Shooter Games*. *Cyber Psychology Behav.* 2007; 10(1): 133-136.
- [6] Weber BG, Mateas M. *A data mining approach to strategy prediction*. IEEE Symp. Comput. Intell. Games. 2009: 140-147.
- [7] Dereszynski E, Hostetler J, Fern A, Dietterich T, Hoang T, Udarbe M. *Learning probabilistic behavior models in real-time strategy games*. 7th Artificial Intelligence and Interactive Digital Entertainment Conference. 2011: 20-25.
- [8] Park HS, Cho HC, Lee KY, Kim KJ. *Prediction of early stage opponents strategy for StarCraft AI using scouting and machine learning*. In Proceedings of the Workshop at SIGGRAPH Asia (WASA 2012). 2012: 7-12.
- [9] Chiu KSY, Chan KCC. *Game engine design using data mining*. Proceedings of the 26th IASTED International Conference on Artificial Intelligence and Applications. 2008: 352-357.

- [10] Hostetler J, Dereszynski E, Diettrich T, Fern A. *Inferring strategies from limited reconnaissance in real-time strategy games*. Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence. 2012: 367-376.
- [11] Hsieh JL and Sun CT. *Building a player strategy model by analyzing replays of real time strategy games*. *International Joint Conference on Neural Networks*. part of the IEEE World Congress on Computational Intelligence, Hong Kong, China. 2008: 3106-3111.
- [12] Synnaeve G, Bessière P. *A dataset for Starcraft AI & an example of armies clustering*. *Artificial Intelligence in Adversarial Real-Time Games*. 2012: 25-30.
- [13] Tveit A. *Game usage mining: Information gathering for knowledge discovery in massivement multiplayer games*. Proceedings of the International Conference on Internet Computing, 2002: 636-642.
- [14] Esmita AP, Gupta ME. *Process Mining: A Comparative Study*. *Int. J. Adv.* 2014; 3(11): 17-23.
- [15] Othman MF, Yau TMS. *Comparison of Different Classification Techniques Using WEKA for Breast Cancer*. 3rd Kuala Lumpur International Conference on Biomedical Engineering. 2006: 520-523.
- [16] Kaur G, Chhabra A. *Improved J48 Classification Algorithm for the Prediction of Diabetes*. *Int. J. Comput. Appl.* 2014; 98(22): 13-17.
- [17] Witten IH and Frank E. *Data Mining: Practical machine learning tools and techniques*. San Francisco, California: Morgan Kaufman. 2005.
- [18] Jiajia Z, Xuan W, Lin Y, Song W. *Building Opponent Model in Imperfect Information Board Games*. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2014; 12(3): 1975-1986.
- [19] Rashed M, Dingju Z. *An Efficient Lip-reading Method Using K-nearest Neighbor Algorithm*. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2015; 13(1): 180-186.
- [20] Sadiq H, Neama AD, Fadl MB-A, Najoua R. *Educational Data Mining and Analysis of Students' Academic Performance using WEKA*. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2015; 13(1): 180-186.