ISSN 2278-3091

Volume 9, No.1.3, 2020

International Journal of Advanced Trends in Computer Science and Engineering

Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse6391.32020.pdf

https://doi.org/10.30534/ijatcse/2020/6391.32020

Classifying Violent Elements in Role-Playing Games Based on User Review using Naïve Bayes Technique

Nursyahirah Tengku Mohamad Hazmiza[1], Norizan Mat Diah*[2], Nasiroh Omar[3]

[1]Faculty of Computer and Mathematical Sciences, UiTM Shah Alam, Selangor, Malaysia, syahirahhazmiza@gmail.com

[2] Faculty of Computer and Mathematical Sciences, UiTM Shah Alam, Selangor, Malaysia, norizan@fskm.uitm.edu.my

[3] Faculty of Computer and Mathematical Sciences, UiTM Shah Alam, Selangor, Malaysia, nasiroh@fskm.uitm.edu.my

ABSTRACT

Role-Playing Game (RPG) is a game genre that provides a fantasy setting in which each player can choose a character according to the role that they like to play and ~~can~~ interact within the game's imaginary world. There exists time that the interactions inculcate implicit violence. This paper presents a classification of violent elements in RPGs via web-based visualization approach. The approach helps to users such as parents or gamers to identify whether the RPGs are appropriate to be played or contain any violent elements. The web-based system employs Naïve Bayes classifier to classify the violent elements in RPGs. The system uses exploratory testing to ensure the effectiveness, correctness, and reliability of the system. The results show that Naïve Bayes classifier is useful and efficient for the classification purposes. Despite the success of this project, there are some limitations in relation to dictionary of violent elements and ~~the~~ datasets. In the future, this research can be improved by adding more glossary of violent elements. Furthermore, dataset labels should be labelled correctly as violence instead of no violence to ensure to increase the correctness of percentage of violent elements found in the RPGs.

Key words: Classify, Naïve Bayes, RPG, Violence

1. INTRODUCTION

Video games are becoming an integral part of modern entertainment in the lives of human beings. Although video gaming has received widespread positive attention among gamers, it is said to be an addiction to them. There are a few researches focusing on the personality of the gamers that favour RPG [1],[2]. Previous research showed that violence in gaming has negative impact among teenagers [3]-[6]. The impact has raised concerns among parents of the consequence in decreasing time planned for homework or studies because of the extended time playing video games [7]-[10]. There are lots of violence can be found in RPG, such as brutal or aggressive behaviour, bad language, fear, sex, gambling, drugs, and discrimination. Nevertheless, RPG gamers deny that these violent elements can be found in the games that they are playing but there is no further research up to now [11].

Moreover, people currently play games to reduce their stress from problems related to social life, academic behaviour, or even work behaviour. Therefore, prior to starting to play the games, they need to find out whether the games contain any violent elements to avoid feeling more stressful while playing the games.

2. BACKGROUND OF RESEARCH

A video game is a digital game that displays visual and audio responses on a video device with audio capability and involves human interaction with a user interface [12]. There are several game genres preferred by people of all ages. These games can be played on any communication devices, such as smartphones, tablets, and computers. Some of the games can be played via the Internet. Players can interact with other players from all over the worldin a virtual game room, making it more interesting to play and can gain the most amazing gaming experience possible[13],[14].

Role-Playing Game (RPG) is a game that provides a fantasy setting in which each player can choose a character according to the role that they like to play and can interact within the game's imaginary world. Examples of RPGs that are popular among teenagers are Fortnite, Player Unknown's Battleground (PUBG), Grand Theft Auto (GTA), and Defense of the Ancients (DotA). The genre of these games helps to teach them how to solve problems, be creative, level up their social skills, encourage teamwork and cooperation, and have fun.

However, addiction and aggression are connected to playing video games[15]. Parents are concern about their children's social and academic performance that can be affected by video gaming [16],[17]. Currently, video games have evolved and raised concerns among parents that worry about their child behaviour. This is because video games do not contain elements that can help children to apply all the things they learn while playing the games. The social interaction within video games is also being concerned by the parents [18].

Parents' concerns start to rise up when their children start to interact with other players that they do not know. Their children might be harassed, stalked, sexually used, deceived, mingled with the wrong group of people, or even exposed to undesirable advertising by online strangers [7],[19]. Gamers go to the virtual field to run away from the reality where they have that supposed 'freedom' to do anything [18]. At Iowa State University, there was a report about college students, who spent greater time video gaming during junior high and high schools, engaged in more self-reported aggressive behaviour as adults due to playing violent and non-violent video games [20],[21]. Video games are sometimes seen as destructive activities that can corrupt the moral values in the gamers themselves [22]. Around 64% of E-rated games contain intentional violence [23].

In order to help parents and communities to monitor their children's behaviour, they need an overall view of the gamesplayed by their children. This is the reason why parents should know the violent elements that might appear in the games. However, it is difficult to identify the violent elements in video games due to the absence of an efficient application or system to help parents and gamers. Furthermore, by using classification, it helps parents to identify whether the games their children are playing contain any violent elements, thus, they can further monitor their children's behaviour in the future.

2.1 Technique for Classification

There is research related to data classification that can be used as a guideline for this research.

- **Decision Tree:** A previous research focused on two different genres, which are a sports game in I Am Playr and a music game in Lyroke [24], aimingto analyse the accuracy of the training model classification. There were no violence elements detected in this research. The work used game telemetry to collect data from users. Telemetry refers to obtaining data through remote access to transmit the data collected from a game server to a collection server and formatted there to support further analysis [25]. The work analysed datasets from January to March 2014 and from March to April 2014for I Am Playr and Lyroke, respectively. In this research, a decision tree was used to demystify what events are likely to result in players disengage. There were various algorithms to developthe decision tree, however, the accuracy of the classification obtained was not high.

- **Gradient Boosting Machine (GBM):** Aprevious research focused on the tile-based game [26] with the aim is to analyse the accuracy of the classification of the system. The violent elements found in this research included abusive acts, cheating, and false complaints. A total of 907 complaints were selected from CCSoft Okey Player Abuse (COPA) Database. In this research, Gradient Boosting Machine (GBM) was used as a formalism to classify individual complaints. It was based on ensembles of sequential weak learners and it helped to minimise an error term based on classification error using gradient descent method. Overall, the accuracy of classification using GBM was high.

- **Support Vector Machine (SVM):**This technique focused on neurogaming platforms [27]. The aim was to analyse the accuracy of classification on EEG signals. There were no violent elements detected in this study. The EEG data was collected from 30 college-aged students. Support Vector Machine (SVM) classifier was used to classify the EEG data. SVM could be used to classify two classes by finding the optimal hyperlane in which the expected classification error of test samples is minimised. Overall, the accuracy of classification using SVM was high.

- **Naïve Bayes:** Aprevious work focused on the multiplayer online battle arena game in Dota 2 [28]. The aim was to analyse the accuracy of classification of the classifier. There were no violent elements detected in this research. The work chose nine player roles that strike the balance between coverings common play styles in details. The researcher used the replay files to analyse the behaviour of each role and manually labelled 708 replays of highly professional players. Naïve Bayes classifier was used to classify the players' behaviour in terms of specific roles within a team of players of the game. The classifier was trained and evaluated using 10-fold cross-validation. Overall, the accuracy of the classification using Naïve Bayes was high after reducing the set of classes.

Table 1 describes the summary of four data classification techniques conducted in the research.

Table 1: Summary of four data classification techniques

Game genre	Sports game and music game	Tile-based game	Neurogaming platform	Multiplayer online battle arena
Aim	To analyse the accuracy of the training model	To analyse the accuracy of the system	To analyse the accuracy of classification on EEG signals	To analyse the accuracy of classification of player roles
Violence	No violence	Have violence	No violence	No violence
Types of Data	Dataset from game telemetry	Dataset from CCSoft Okey Player Abuse (COPA) Database	Survey Interview	Replay files
Classification Technique	Decision Tree: C4.5 algorithm	Gradient Boosting Machine	Support Vector Machine	Naïve Bayes
Accuracy	Not high	High	High	High

As shown in Table 1, the research focuses on a variety of game genres. There are various techniques used in the research and the final results are varied. All research is conducted with the same purpose, which is to check the accuracy of classifying the various types of data. Since there are only a few researches conducted on violence, this project will focus on discovering violent elements in RPGs by using Naïve Bayes.

3. METHODOLOGY OF RESEARCH

The method used in this study consists of fivephases, which are selection, extraction, pre-processing, classification, and calculation. Figure 1 shows the illustration of every phase involved.

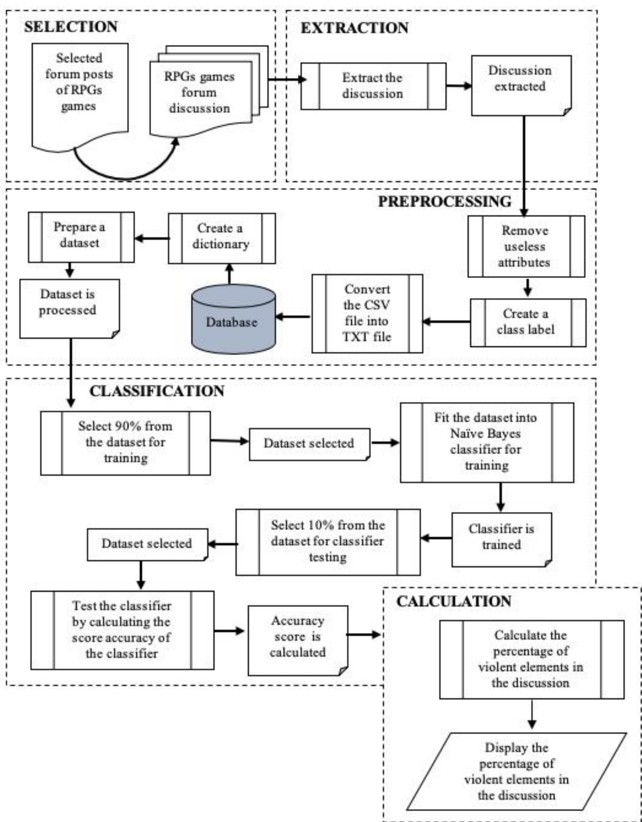

Figure 1: Phases involved in the methodology of research

3.1Selection

In this phase, the important data to be used for this research is identified first, such as the RPG's name and discussion of the RPG. In this project, five RPGs were selected. These five RPGs were selected from the list of popular games listed in a video game digital distribution platform called Steam. Subsequently, the verified website was selected to be extracted to ensure the validity of the data to be used in this project.

3.2 Extraction

Data extraction is a process of retrieving data from data sources for further data processing or data storage. For this research, data extraction was done by using a technique called data scraping. Also known as web scraping, it is a technique in which a computer programme extracts data from human-readable output coming from another programme.

3.3 Pre-processing

Pre-processing consists of a number of steps that need to be performed as a preparation for text classification. In this phase, there are two main parts of pre-processing, which are data pre-processing and text pre-processing. In data pre-processing, the extracted data discussions contain useless attributes that are not needed in the text classification process. This is because the text classification process to be conducted after pre-processing only involves text datasets.

Next, a class label needs to be created in order to apply it in the text classification process. The goal is to learn a rule that computes the label from other attribute values. In thisphase, there are two labels, which are 0 and 1. The instances labelled as 0 indicate that they do not contain any violent elements, while the instances labelled as 1 indicate that they contain violent elements. Figure 2 shows a list of violent elements [29].

violence, aggressive, carnage, attack, mild, conflict, unsafe, gore, weapon, injury, maiming, death, harm, torture, dismemberment, horrific, kill, cruel, abhorrent, murder

Figure 2: A list of violence elements

Labelling the class label is conducted by finding each word from the list of violent elements and searching through the attribute called discussion in the dataset. If all the words from the list of violent elements are found in each instance of discussion attribute, the class label is labelled as 1, while if it is not found, the class label is labelled as 0.

During the text pre-processing, there are two main processes involved, which are tokenisation and normalisation. Tokenisation is a process of splitting all the texts into a list of tokens or words. This process is conducted to ensure that the programme can have a glossary that stores each word encountered in the *TXT* file. Next, removing non-alphabetic is done to ensure that the glossary stores only words without symbols and numbers. Lastly, removing duplicate words in the glossary is done by calculating the total frequency of each word in the glossary. Therefore, the glossary stores a list of words with the frequency of each word.

In the normalisation process, the feature vector is done in preparing a dataset for text classification. It is a process of converting the texts into numbers. In this project, the programme split the text in words and stored into a word list. Then, it converted each word in the word list into numbers according to the glossary created earlier. Once converted, theywere stored in a feature set list. Next, the programme labelled each *TXT* file with 0 and 1 according to the class label that was labelled earlier and stored in a label list.

3.4 Classification

Text classification is a process of classifying text into one or more classes. In this project, the text used was the discussions of each RPG. The purpose is to classify the discussions either it is violence or not, according to the labelled dataset. Multinomial Naïve Bayes is the classifier used in this process. In the text classification, the data is split into training and testing datasets with 90% of the data are for training and another 10% of the data is for testing purposes. In this programme, there weretwo training sets and two testing sets that trained and tested feature set list and label list separately. Both of the training sets for feature set list and label list were trained in a Multinomial Naïve Bayes classifier. This classifier used a multinomial distribution for each of the features. Then, the classifier predicted the testing set for feature set list based on the testing set of the label list. The score accuracy of the prediction wascalculated, and the confusion matrix of the predictions was generated.

3.5 Calculation

In this project, the calculation of the percentage of violent elements in eachRPG was done manually. The percentage of violent elements was calculated to measure whether the selected RPGswere violence or not violence. The calculation was done by calculating the total number of class label 1, which is the discussion that contains violent elements, dividing by the total number of the instances in the discussions, and multiplying by 100% to get the percentage.

4. RESULTS AND FINDINGS

Classification testing of the Naïve Bayes classifier was done to analyse the results of the classification. Table 2 summarises the results of the score accuracy of the classification for each RPG.

Table 2: Results of the accuracy of the classification for each RPG

RPGs	Score Accuracy (%)	Total number of datasets
RPG 1	95.45	216
RPG 2	80.95	204
RPG 3	81.82	161
RPG 4	66.67	26
RPG 5	75	38

Based on the results of the score accuracy of the classification in Table 2, the score accuracy of the classifier is different for each RPG. With the accuracies around 75%, the results are not perfect but quite promising with respect to the complex classification tasks and limited datasets. From the analysis, it reveals that more datasetsare needed to achieve a better result in the classification process.

Table 3 shows the results of the confusion matrix for each RPG. In the confusion matrix, the predicted classes are represented in columns, while the actual classes are represented in rows. There were four cases been measured, which are True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). TP measured the number of cases for which the classifier predicted 'violence' and the datasets were actually violence. TN measured the number of cases for which the classifier predicted 'no violence' but the datasets were actually had no violence. FP measured the number of cases for which the classifier predicted 'violence', but the datasets were actually had no violence. FN measured the number of cases for which the classifier predicted 'no violence' but the datasets were actually violence.

Table 3: Results of the confusion matrix for each RPG

RPGs	Confusion Matrix
RPG 1	$\begin{bmatrix} 18 & 0 \\ 1 & 3 \end{bmatrix}$
RPG 2	$\begin{bmatrix} 13 & 3 \\ 1 & 4 \end{bmatrix}$
RPG 3	$\begin{bmatrix} 7 & 0 \\ 2 & 2 \end{bmatrix}$
RPG 4	$\begin{bmatrix} 2 & 0 \\ 1 & 0 \end{bmatrix}$
RPG 5	$\begin{bmatrix} 3 & 0 \\ 1 & 0 \end{bmatrix}$

Based on the results of confusion matrix for RPG 1 in Table 3, out of the 18 actual instances of 'no violence' (first row), the classifier predicted all 18 instances correctly as no violence, while out of the 4 actual instances of 'violence' (second row), the classifier predicted 3 instances correctly as violence but only 1 instance predicted incorrectly as no violence. Overall, out of 22 actual datasets, the classifier predicted 21 datasets correctly and only 1 dataset was predicted incorrectly.

However, for RPG 4, out of the 2 actual instances of 'no violence' (first row), the classifier predicted all 2 instances correctly as no violence, while out of the 1 actual instance of 'violence' (second row), the classifier predicted the instance incorrectly as no violence. Overall, out of 3 actual datasets, the classifier predicted 2 datasets correctly and only 1 dataset was predicted incorrectly. This shows that the Naïve Bayes classifier needs more datasets in order to achieve a better accuracy result in a classification.

5. CONCLUSION

Five RPGs from Steam website are selected, and the discussion of each RPG is extracted by using Google Chrome Web Scrapper Extension tool. After the dataset of the discussions is extracted, the pre-processing of the dataset takes place to prepare the data prior to the application of the Naïve Bayes algorithm. The pre-processing of removing useless attributes, creating a class attribute, tokenisation, removing non-alphabetic, removing duplicated words, and creating feature sets and labels are applied to prepare the datasets for classification. Next, the feature datasets of five RPGs are classified separately either it is violence or not. A

Naïve Bayes classifier is created and trained by using 90% of the dataset, while another 10% of the dataset are tested according to the labelled dataset. The score accuracy of classification is calculated and displayed. The results prove that the Naïve Bayes classifier is useful and efficient for classifying violent elements in RPGs.

ACKNOWLEDGEMENT

The authors would like to thank Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor, for sponsoring this research.

REFERENCES

1. L. T. Graham and S. D. Gosling, **Personality profiles associated with different motivations for playing world of Warcraft**, *Cyberpsychology, Behavior, and Social Networking*,vol. 16 (3), pp189-193, 2013. https://doi.org/10.1089/cyber.2012.0090
2. A. Simon, **Emotional stability pertaining to the game of dungeons & dragons**,*Psychology in the Schools*, vol. 24, pp 329-332, 1987.
3. S. Lemona, S. Brand, N. Vogler, N. Perkinson-Gloor, M. Allemand and A. Grob, **Habitual computer game playing at night is related to depressive symptoms**,*Personality and Individual Differences*, vol. 51, pp117-122, 2011.
4. V. Anand, **A study of time management: the correlation between video game usage and academic performance markers**, *CyberPsychology & Behavior*, vol. 10, pp 552-559, 2007. https://doi.org/10.1089/cpb.2007.9991
5. D. A. Gentile, P. J. Lynch, J. R. Linder and D. A. Walsh. **The effects of violent video game habits on adolescent hostility, aggressive behaviors, and school performance**, *Journal of Adolescence*, vol. 27, pp 5-22, 2004.
6. P. Holtz and M. Appel. **Internet use and video gaming predict problem behaviour in early adolescence**, *Journal of Adolescence*, vol. 34, pp 49-58, 2011.
7. H. J. Jiow, **Parental mediation of video games in Singapore**, Doctoral dissertion, National University of Singapore, Singapore, 2014.
8. A. Khoo, **Video games as moral educators?***Asia Pacific Journal of Education*, vol. 32(4), pp 416-429, 2012.
9. L. A. Kutner, C.K. Olson, D.E. Warner and S.M. Hertzog, **Parents' and Son's Perspectives on Video Game Play: A Qualitative Study**, *Journal of Adolescent Research*, vol. 23(1), pp 76-96, 2008. https://doi.org/10.1177/0743558407310721
10. M. M. Skoric, L. L Ching Teo and R. L. Neo. **Children and Video Games: Addiction, Engagement, and Scholastic Achievement**, *CyberPsychology & Behavior*, vol. 12(5), pp 567-572, 2009. https://doi.org/10.1089/cpb.2009.0079
11. B. Braun, J. M. Stopfer, K. W. Muller, M. E. Beutel and B. Egloff, **Personality and video gaming: comparing regular gamers, non-gamers, and gaming addicts and differentiating between game genres**, *Computers in Human Behaviour*, vol. 55, pp 406-412, 2016.
12. A. Qurotul, H. Taqwa, H. P. PancaOktavia and R. Untung. **Understanding How Gamification Influences Behaviour in Education**, *International Journal of Advanced Trends in Computer Science and Engineering (IJATCSE)*, vol. 8 (1.5), pp. 269-274, 2019.
13. M.S.Saleem Basha, A. Mohamed Abbas, Y. Gazala and Rajbunisa, **Augmented Reality Based Education for The Improvement of Sustainable Learning Ability in Oman Educational System**, *International Journal of Advanced Trends in Computer Science and Engineering (IJATCSE)*,vol. 8(1.2), pp.110-115, 2019.
14. N. H. M. Khir and M. Ismail, **Review on gamification in children computer interaction (CCI) for persona modelling**, *Bulletin of Electrical Engineering and Informatics*, vol. 8(4), pp. 1411-1417, 2019. https://doi.org/10.11591/eei.v8i4.1622
15. D. J. Kuss, **Internet gaming addiction: current perspectives**,*Psychol Res Behav Manag*, vol. 14(6), pp 125-37, 2013.
16. T. Sim, D. A. Gentile, F. Bricolo, G. Serpelloni and F. Gulamoydeen, **A conceptual review of research on the pathological use of computers, video games, and the Internet**, *International Journal of Mental Health and Addiction*, vol. 10(5), pp 748–769, 2012. https://doi.org/10.1007/s11469-011-9369-7
17. M. R. Mohamad Razali, Ahmad, S. and N. M. Diah, **Collaborative filtering content for parental control in mobile application chatting**, *Bulletin of Electrical Engineering and Informatics*, vol. 8(4), pp 1517–1524, 2019.
18. H. J. Jiow, R. Athwa, L.L Chew, M. H. Elias, N. Lim and K. Woo, **Revisiting video game ratings: shift from content-centric to parent-centric approach**,*SHS Web of Conferences*, vol. 33, 00065, 2017.
19. Lenhart, L. Rainie and O. Lewis. **Teenage life online: The rise of the instant-message generation and the Internet's impact on friendships and family relationships**, *Washington, DC: Pew Internet and American Life Project*, 2001.
20. C. Anderson and K. Dill, **Video games and aggressive thoughts, feelings, and behaviour in the laboratory and in life**, *Journal of Personality and Social Psychology*, vol. 78(4), pp 772-790, 2000.
21. M. S. Simatupang and R. Peter, **The Ambivalence of Electronic Games to Character and English Performance**, *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 8 (5C), pp.492-497, 2019.
22. R. Halverson, **what can K-12 school leaders learn from video games and gaming?***Innovate*, vol. 1(6), 2005.
23. J. J. Thompson, B. H. M. Leung, M. R. Blair and M. Taboada, **Sentiment analysis of player chat messaging in the video game Starcraft 2: extending a lexicon-based model**, *Knowledge-Based Systems*, pp 149-162, 2017. https://doi.org/10.1016/j.knosys.2017.09.022

24. H. Xie, D. Kudenko, S. Devlin and P. Cowling, **Predicting player disengagement in online games**,*Workshop on Computer Games*, pp 133-149, 2014.
25. M. El-Nasr, A. Drachen, and A. Canossa, **Game analytics: maximizing the value of player data**,*Springer*, 2013.
26. K. Balci and A. A. Salah, **Automatic classification of player complaints in social games**,*IEEE Transactions on Computational Intelligence and AI in Games*, vol. 9(1), pp103-108, 2017.
27. T. D. Parsons, T. McMahan and I. Parberry, **Neurogaming-based classification of player experience using consumer-grade electroencephalography**, *Technical Report LARC-2015-04*, Texas, USA,2015.
28. C. Eggert, M. Herrlich, J. Smeddinck and R. Malaka, **Classification of Player Roles in the Team-Based Multi-player Game Dota 2**. In: M. S. Simatupang and R. Peter, **The Ambivalence of Electronic Games to Character and English Performance**, *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 8 (5C), pp.492-497, 2019. https://doi.org/10.35940/ijeat.E1070.0585C19
29. R. S. Hamid and N. Shiratuddin, **Filtering violence and sexual content: the necessities of digital games content rating system for Malaysia's environment**, *Journal of Telecommunication, Electronic and Computer Engineering*, vol. 8 (8), pp. 83-89, 2016. https://doi.org/10.1063/1.4960881