Evaluating Video Game Acceptance in Game Reviews using Sentiment Analysis Techniques

Larissa F. S. Britto
Departamento de Computação
Universidade Federal Rural de Pernambuco
Recife, Brazil
larissa.feliciana@ufrpe.br

Luciano D. S. Pacífico
Departamento de Computação
Universidade Federal Rural de Pernambuco
Recife, Brazil
luciano.pacifico@ufrpe.br

***Abstract*—The game industry has become one of the most profitable markets in the entertainment industry. Knowing what customers and users think and how they feel about a game is a central piece to drive the decision-making process, of any game developer or game studio, towards the user satisfaction. Sentiment Analysis is being widely used by companies to discover what customers are saying about their products. In this paper, a complete process to evaluate video game acceptance using game user reviews is proposed, by means of the application of Sentiment Analysis techniques. Through a sentiment classification approach, we infer user acceptance. Also, a new dataset is proposed, composed of game user reviews written in Brazilian Portuguese language. Some classifiers applied for sentiment classification in literature are adopted.**

***Index Terms*—vieo game acceptance, sentiment analysis, natural language processing, artificial intelligence, machine learning**

I. Introduction

The advance in technologies and the emergence of social networks, smartphones and tablets, make video games even more popular and accessible to different audiences. The number of gamers worldwide is still on the rise and will exceed three billion by 2023, The global market for video games will generate revenues of about $159.3 billion in 2020, a +9.3% year-on-year increase [1].

To be successful in this competitive market, a game developer needs to know the reasons that lead customers to play a game, and how to keep players engage. With the Internet and social networks popularization, users can express their opinion freely, providing feedbacks and reviews that can be accessed through such networks and game platforms. Game reviews can be viewed as expert experience reports, giving players an idea on what to expect from the game, working as purchase guides to their readers [2]. But in fact, reviews are a rich source of user opinions and sentiments. As reported in literature, sentiments expressed by the users about different aspects of the game have strong correlation to the user's acceptance of the game [3], [4].

In this context, Sentiment Analysis approaches have been recently adopted to evaluate video game acceptance by users [4]–[6]. Sentiment Analysis (SA) is an automated process that uses Artificial Intelligence to analyze textual documents and identify sentiments and opinions. Through SA approaches, it is possible to explore game reviews automatically, identifying user acceptance rates contained in those reviews. When a review expresses positive sentiments and opinions, it means that the game is well-accepted. Negative reviews allow the game development team to know what are considered, by some users, the weakest points of a game, guiding future game design decisions and improvements.

SA is one of the most popular tasks in Natural Language Processing (NLP) due to its large number of applications. However, there is a lack of resources and frameworks in Brazilian Portuguese language, such as public datasets and precise NLP tools, making the evaluation of Brazilian players' opnions a hard task. Given this shortage of resources and the fact that Brazil is the 13th largest video game market in the world [7], in this work, we evaluate video game acceptance using sentiment analysis approaches. A new dataset, based on game reviews in Brazilian Portuguese, is proposed, and the dataset development process is described. The sentiments expressed in the reviews from the proposed dataset are analyzed and classified using three well-established classifiers from text classification literature: Logistic Regression, Random Forest and Support Vector Machines.

The remainder of this paper is organized as follows. The dataset development process, preprocessing approaches, feature extraction method and classifiers adopted in this work are described in Secion II. Our experimental setup and results are presented in Section III. Section IV concludes this paper.

II. Methodology

In this section, the dataset preparation process is described (Section II-A), as well as the preprocessing (Section II-B) and feature extraction method (Section II-C). The selected classifiers are also described (Section II-D1).

A. Dataset

The dataset proposed in this work is composed by game reviews extracted from Steam[1] using its Web API. Those reviews, all in Brazilian Portuguese language, express what players think and how they feel about the games and their features. The dataset will be available in a public repository

[1]https://store.steampowered.com/

[2], without any processing to give the researchers the freedom to choose which preprocessing steps are more suitable to their researches.

1) Data Acquisition: The first step to dataset creation is the acquisition of the data. The acquisition can be made through two main methods: the application of Web Scraping techniques (where data can be gathered from websites mechanically, through properly automation softwares); or by the use of Application Programming Interfaces (APIs), generally made available by the websites themselves or by their users, enabling developers to interact with the systems easily and safety. APIs furnished by the websites allow the users to access the website content and data in a predefined manner, giving them the opportunity to avoid web scraping problems, such as complicated and changeable web page structures. In this work, Steam API [3] is used to collected the necessary information.

2) Polarity Annotation: After the data acquisition, the sentiment polarity annotation is performed. All the reviews were classified in two classes, according to the game acceptance (*positive* or *negative*). The polarity is also obtained through the review made by the user. In Steam, the player can evaluate the game by recommending it or not.

The "recommend" information is represented in the API by the *vote_up* variable, that is *true* when the player recommends the game, and *false*, otherwise. In the dataset, reviews in which the player recommends the game are considered with a positive polarity, and a negative polarity is attributed to reviews where the player does not recommend the game.

B. Preprocessing

One of the most fundamental steps in text classification is the preprocessing of the textual documents, which includes the cleaning of textual data, that will remove irrelevant information and any other noise that may worsen the classifiers performance. Documents are also standardized. In this work, the following steps were performed for data preprocessing.

1) Lowercase Conversion: A basic approach in text preprocessing, with great performance impact [8], was adopted in this work: the conversion of all document uppercase letters into lowercase letters. This process enables a union of words whose the only difference is whether the first letter is uppercase or lowercase. This conversion removes the problem of terms that, despite being the same, are considered different by algorithms, such as "jogo" ("game") and "Jogo" ("Game").

2) Special Characters Removal: Any special characters in the documents were removed, such as punctuation, symbols and digits. They have no meaning and do not indicate any sentiments polarity, being totally disposable.

C. Feature Extraction

Feature extraction intends to transform raw documents from the dataset into useful data supported by the classifiers. Bag-of-Word (BoW) was adopted in this work. In BoW model, a textual document is represented by its set of words. The text is converted into a matrix, where every column represents a word, each row represents a document, and each position contains the number of occurrences of a word in a document. The structure of the original document or the order of the words in that document are not taken into consideration by this representation [9]. Our dataset will finally be converted into a *Document-Term Matrix* (DTM), as represented in Fig. 1.

D. Classification

This section contains a brief description of all classifiers used in our experimental evaluation: Random Forest, Logistic Regression and Support Vector Machines.

1) Random Forest: Random Forest classifier (RF) [10], [11] is a method that combines a set of Decision Trees (hierarchical structures that represents a learning function [12]), in order to avoid the sensibility to noise and outliers that a single Decision Tree would represent, making the classifier more robust. The algorithm combines the results of several trees, aggregating the votes from those different estimators to make the prediction [13].

2) Support Vector Machines: Support Vector Machines (SVM) [14]–[16] are supervised learning algorithms based on Structural Risk Minimization principle [17]. SVM maps a set of data patterns from their original feature space into a new space, where their classes are linearly separable by an optimal hyperplane [18].

3) Logistic Regression: The discriminative model Logistic Regression (LR) [19], [20] is used to predict the probability of the possible outputs of a dependent variable, given a set of independent variables. LR assumes that the dependent variable (the class of the testing sample) can be predicted by a linear combination of of the feature set from the training samples (independent variables) and the model parameters.

III. Experimental Results

In this section, the experimental results obtained for the proposed dataset are presented. Aiming at evaluating the proposed dataset, we compare the performances of three different and well-established classifiers from SA literature when applied to Brazilian Portuguese documents. A ten-folds cross-validation framework was used in our evaluation, where the proposed dataset has been randomly split into ten parts to form the training and testing sets. Nine folds are used each time to compose the training set, and the remaining fold is used as the testing set. To generate a large variety of tests, the ten-folds cross-validation process has been executed ten times, and, for each execution, ten random distributions of the data have been obtained, in such a way that we had one hundred different tests evaluations (the ten-folds cross-validation method has been executed ten times, each time starting with a new random distribution of the data patterns into the folds). The adopted resampling process has been performed to avoid results obtained by chance. Four well-known classification metrics are adopted: Accuracy, Precision, Recall and F-Measure.

[2]https://github.com/larifeliciana/steam-reviews-portuguese
[3]https://steamcommunity.com/dev

	bom	jogo	amei	péssimo	eu
Document 1	1	1	0	0	0
Document 2	0	1	1	0	1
Document 3	0	1	0	1	0
...	...	...	...	...	...
Document *n*	0	0	1	0	1

Fig. 1: Bag-of-Words Document-Term Matrix.

The classifiers methods are evaluated through an empirical analysis concerning the testing set. A qualitative analysis is also performed in the proposed dataset, and the final statistics for this dataset are presented in Table I. The experimental results are shown in Table II.

SVM is able to obtain the best performances concerning all four selected classification metrics, in comparison to the other selected classifiers from literature, with an accuracy of 82.54%, with no significant difference to the second best approach, the LR, that was able to achieve an average accuracy of 82.40%. In turn, LR obtained the best average execution time (less than one second), while SVM reached 2.4 seconds, in average. The worst overall performance has been achieved by Random Forest, that achieved an average accuracy of 79.89%, and reached an average execution time of 2.3 seconds.

Another way to understand what users (players) think about a product (games) is to analyse what they are talking about the product. In Figures 2 and 3, the most frequent unigrams and bigrams from the positive and negative reviews from the dataset are presented, respectively. By removing the stop words, we are able to note a lot of important topics and features from the games that are approved and admired by the players, such as graphics ("gráficos"), gameplay ("jogabilidade", "jogabilidade boa"), soundtrack ("trilha", "trilha sonora") and storytelling ("história", "boa história"). We can also notice negative points that annoy users by analyzing Fig. 3, such as problems in game optimization ("mal otimização"), bugs ("bugs"), graphics ("gráficos"), gameplay ("jogabilidade"), downloadable content ("dlc"), and difficulty in playing online with friends ("multiplayer", "jogar online").

IV. Conclusion

In this work, video game acceptance was evaluate through a sentiment polarity classification approach, based on the fact that sentiments expressed in reviews about game aspects are strongly correlated to the user acceptance of the game [3], [4]. By classifying the game reviews, we can infer the user acceptance about a game.

A sentiment dataset in the domain of games, composed by reviews in Brazilian Portuguese, was proposed. The proposed dataset was analyzed in an attempt to understand what topics and elements from the games are more relevant to the players, positively and negatively, and which ones are not. We evaluate the performance of three well-established classifiers from Sentiment Analysis and Machine Learning literature: Random Forest, Logistic Regression and Support Vector Machines.

The experimental results pointed out that Support Vector Machines and Logistic Regression are able to achieve the best performances in relation to the four selected classification metrics (Accuracy, Precision, Recall and F-Measure). Logistic Regression has also achieved the best average execution time.

By analyzing the most frequent words, we could have a better understanding on how players feel about a game, as much as what they are discussing, what are their favorite game elements, and how good have been their experiences with that game. In positive reviews, we can observe some well-accepted elements, such as graphics, gameplay, soundtrack and storytelling. On the other hand, problems concerning online gaming, DLCs and bugs are the most frequently reported in negative comments.

The use of SA techniques allows the identification of problems in the game, areas of improvement and the understanding on what to change to increase video game acceptance. In the future, we intend to extend this work by analyzing how the game acceptance varies with time, as an attempt to understand how game changes and updates can influence the game acceptance by its users. Also, we intend to adopt quantitative analysis provided by Game Analytics to complement the qualitative analysis of SA, so we could have a better understanding on player's interaction and engagement with the games.

References

[1] NewZoo, "Global games market report," 2020.
[2] I. Livingston, L. Nacke, and R. Mandryk, "The impact of negative game reviews and user comments on player experience," *Proceedings - Sandbox 2011: ACM SIGGRAPH Video Game*, 08 2011.
[3] B. Strååt and H. Verhagen, "Using user created game reviews for sentiment analysis: A method for researching user attitudes." in *GHITALY@CHItaly*, 2017.
[4] A. Vieira and W. Brandão, "Evaluating acceptance of video games using convolutional neural networks for sentiment analysis of user reviews," in *Proceedings of the 30th ACM Conference on Hypertext and Social Media*, 2019, pp. 273–274.
[5] A. d. C. Vieira and W. C. Brandão, "GA-Eval: a neural network based approach to evaluate video games acceptance," in *Proceedings of the 18th Brazilian Symposium on Computer Games and Digital Entertainment*, ser. SBGAMES'19, Oct. 2019, pp. 595–598.
[6] F. M. D. P. R. Chaves, "Video game success and consumer acceptance: a sentiment classification of gamer reviews," Ph.D. dissertation, 2019.
[7] NewZoo, "Global games market report," 2019.
[8] Y. HaCohen-Kerner, D. Miller, and Y. Yigal, "The influence of preprocessing on text classification using a bag-of-words representation," *PLOS ONE*, vol. 15, no. 5, pp. 1–22, 05 2020. [Online]. Available: https://doi.org/10.1371/journal.pone.0232525
[9] K. Kowsari, K. J. Meimandi, M. Heidarysafa, S. Mendu, L. E. Barnes, and D. E. Brown, "Text classification algorithms: A survey," *CoRR*, vol. abs/1904.08067, 2019. [Online]. Available: http://arxiv.org/abs/1904.08067

TABLE I: Proposed Dataset Statistics

	Reviews	Positive reviews	Negative reviews
total number of reviews	2000	1000	1000
vocabulary size	16705	11325	10017
mean words per review	79.14	85.235	73.05
mean sentences per review	3.77	4.06	3.49
mean words per sentence	20.88	20.04	21.72

TABLE II: Experimental results.

	Accuracy	Precision	Recall	F-measure	Execution Time
SVM	0.7488 ± 0.0343	0.7152 ± 0.0474	**0.8283 ± 0.0387**	0.7666 ± 0.0343	2.2864 ± 0.4422
RF	**0.7952 ± 0.0292**	0.7807 ± 0.0453	**0.8211 ± 0.0400**	**0.7994 ± 0.0330**	1.9811 ± 0.0879
LR	0.8132 ± 0.0291	**0.8097 ± 0.0437**	0.8185 ± 0.0402	0.8132 ± 0.0329	0.9034 ± 0.3565

Fig. 2: Most frequently unigrams and bigrams in positive reviews

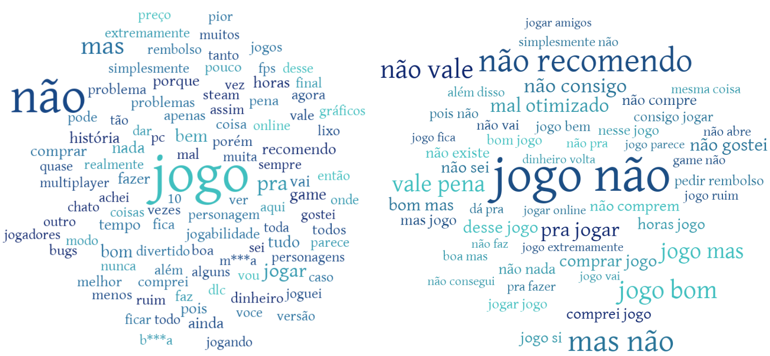

Fig. 3: Most frequently unigrams and bigrams in negative reviews

[10] N. Pervan and H. Keles, "Sentiment analysis using a random forest classifier on turkish web comments," vol. 59, 12 2017.

[11] H. Parmar, S. Bhanderi, and G. Shah, "Sentiment mining of movie reviews using random forest with tuned hyperparameters," 07 2014.

[12] R. Bhatnagar, *Machine Learning and Big Data Processing: A Technological Perspective and Review*, 01 2018, pp. 468–478.

[13] A. Criminisi, E. Konukoglu, and J. Shotton, "Decision forests for classification, regression, density estimation, manifold learning and semi-supervised learning," 2011.

[14] M. Ahmad, S. Aftab, and I. Ali, "Sentiment analysis of tweets using svm," 2017.

[15] M. Fikri and R. Sarno, "A comparative study of sentiment analysis using svm and sentiwordnet," 2019.

[16] P. Dubey, A. Mishra, and B. K. Saha, "Sentiment analysis using svm and deep neural network," in *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, 2019, pp. 952–957.

[17] V. Vapnik, "Principles of risk minimization for learning theory," in *Proceedings of the 4th International Conference on Neural Information Processing Systems*, ser. NIPS'91. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1991, pp. 831–838. [Online]. Available: http://dl.acm.org/citation.cfm?id=2986916.2987018

[18] S. Bergsma, D. Jung, R. Lau, Y. Wang, and S. Wang, "Machine learning approaches to sentiment classification cmput 551 : Course project winter , 2005," 2005.

[19] A. Tyagi and N. Sharma, "Sentiment analysis using logistic regression and effective word score heuristic," *International Journal of Engineering and Technology(UAE)*, vol. 7, pp. 20–23, 04 2018.

[20] W. P. Ramadhan, S. T. M. T. A. Novianty, and S. T. M. T. C. Setianingsih, "Sentiment analysis using multinomial logistic regression," in *2017 International Conference on Control, Electronics, Renewable Energy and Communications (ICCREC)*, Sep. 2017, pp. 46–49.