Aspect Based Sentiment Analysis for Game reviews

Raseshwari Pulle
Rochester Institute of Technology
rp3737@rit.edu

Hiten Gupta
Rochester Institute of Technology
hg1928@rit.edu

Prakhar Rawat
Rochester Institute of Technology
pxr7185@rit.edu

Sultan Mira
Rochester Institute of Technology
sfm2686@rit.edu

1 INTRODUCTION

In state of the art game development mechanism the developers have to consider the reviews of the users in order to make their product a success [9]. This relationship is hard to find as there is a large quantity of user reviews. The prescient relationship uses sentiment analysis [8] for different aspects mentioned in a user review to the success/failure of the video game. This relationship proves to be a quality metric for the team working on the game and empower them with a mindset of the users.

In this study, we have performed an Aspect Based Sentiment Analysis (ABSA) [6] based on data gathered from user reviews regarding two video game series, on Metacritic.com. Our purpose was to explore if the sentiment an aspect (commonly used words in the reviews) was used in, would reflect the overall rating from the reviewers. A positive result would imply that user reviews can be used to explain user attitudes (positive/negative sentiment) from a root-cause point of view (the aspects).

2 RELATED WORK

Most of the studies that has been performed, has been on professional reviews: Pinelle, Wong & Stach [5] used professional reviews as a source to find common video game issues, which they compiled into a set of design patterns, Zagal, Ladd & Johnson [9] found that game reviews often include design suggestions and serious discussions on game designer's intention and goals. User created reviews has been used as well, but not as frequently: Strååt & Verhagen [7] used user reviews to evaluate video game heuristics, Zagal & Tomuro[10] studied cultural differences and similarities in user created reviews from Japan and USA, and quite recently, Koehler, Arnold, Greenhalgh, Owens Boltz & Burdell's published their article "A Taxonomy Approach to Studying How Gamers Review Games" [4]. They used an existing theoretical model, a video game taxonomy, and compared user submitted reviews with the categories of the taxonomy. They found that users to a certain degree used the same concepts as the taxonomy, and that there was a difference in use of the concepts depending on the game rating.

3 MOTIVATION

There is no doubt that sentiment analysis provides insightful on a vast range of real world applications, because even if the overall sentiment of a text doesn't necessarily have to relate the root cause of an author's opinion. Aspect-Based Sentiment Analysis makes it relatively easier to identify and determine if the sentiment in the text points towards a specific aspect.

Even though game companies invest a substantial amount of time and money in user experience design, the efforts of user experience specialists are not enough to guarantee a successful game release.Even though the users enjoyed separate parts of the game, a narrative that did not meet the players' expectations spoiled their experience. This implies that no matter how well the work is done, if important aspects of the game are left without proper attention, the game experience may suffer. Collecting and understanding the users' opinion is a cornerstone in every user centered design, games being no exception, and discrepancies in expectations between developer and user can be harmful for the end product.

XXXX, XXXX
2016. 978-x-xxxx-xxxx-x/YY/MM…$15.00
DOI: 10.1145/nnnnnnn.nnnnnnn

4 BACKGROUND

4.1 Aspect Based Sentiment Analysis

An aspect based sentiment analysis (ABSA) [6] is performed when user sentiment of certain aspects of a multi-aspect entity is to be measured, in a dataset gathered from user comments, such as online forums or user created reviews. Video games have plenty of aspects that the user consider when playing, e.g. playability, graphics, storyline.

Aspects are words or phrases that exist either explicitly or implicitly in the dataset and provide information in terms of features or attributes. The sentiment analysis is then performed either through a scripted natural language processing algorithm, or through a manual read through. The result will show the sentiment for each aspect, for example in terms of positive, neutral, or negative sentiment.

4.2 Metacritic

Metacritic.com is a site that aggregates professional reviewer scores from various online media review sources. Television shows, movies, music and video games (various platforms) are examples of media that are presented.

4.3 Games in the Study

In order to perform Aspect Based Sentiment Analysis we have chosen reviews from three games *Devil May Cry* [] *Call of Duty III : Black Ops* [2] *Overwatch*[3]. These games are best games among their domain and each of them belong to different categories.

4.4 Aspects in the study

For the ease of simplicity and to keep it general we have considered 3 major aspects of a game :

- Gameplay - It encompasses the quality of the gameplay of the game, which is experienced by the user while playing the game.
- Combat -Since these were shooting game so the combat is the aspect which wraps up all the fighting experience the user has
- Action - It is the genre aspect of the game.

5 APPROACH

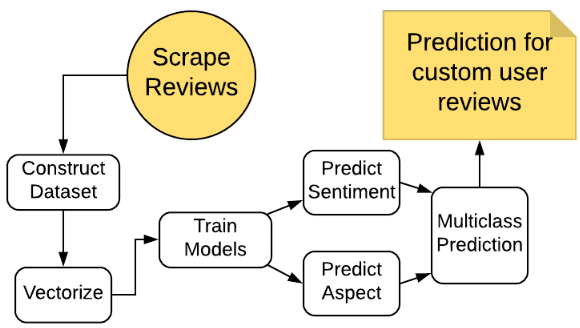

Figure 1: Task Flow in Approach

5.1 Approach Overview

Our approach consists of seven steps as can been seen in Fig 3, the following sections discuss in detail all the steps we performed.

5.1.1 ***Scrape Reviews.*** The extraction of reviews for the games from Metacritic we implemented a scrapper in python to crawl the reviews posted for one game at a time (the website scrapping script can be found at our github repository of this project) [1] The meta-data of the reviews we scraped comprised of timestamps of the review, user name, reviewer game rating, and review text.

5.1.2 ***Construct Dataset.*** We combined reviews of three games scrapped in previous step and created one consolidated file. We cleaned the reviews in consolidated file and created a dataset of reviews manually. To construct this dataset, two additional attributes of polarity and category were added to each review.

Polarity can be two values either positive or negative and category can be either gameplay, combat or action (the values of category are aspect labels which we are considering for our project) Each review was analyzed for either having polarity. The final data set used for this project is such that each row consists of details - game name, polarity, category, rating and review text.

An example of a review stored in a row of our dataset - *['Overwatch', 'positive', 'gameplay', 9, 'This is a great game! Just love it!!']*

5.1.3 ***Vectorize.*** We used Pandas API available in python for creating dataframes from out dataset and CountVectorizer from sklearn.feature_extraction.text API to convert review text into tokens/features. To remove irrelevant words from the reviews we used standard stopword removal . The max_df=0.80 and min_df=4 were set in CountVectorizer to ignore terms that appear in more than 50 percent of the reviews and the ones that appear in more than 4 reviews respectively. This was done to getter better tokens and reduce noise.

5.1.4 ***Train Models.*** In this step we are using four different classification algorithms - Naive Bayesian, Logistic Regression, Linear SVM and K-Nearest Neighbors and then depending on the algorithm having best accuracy we trained our model. Two algorithms which performed well on our dataset were Naive Bayesian and Logistic Regression, we used them to train 2 models one for predicting review sentiment and one for predicting the review aspect. We provide as input our train and test datasets formed by the 70:30 split. We are using the sklearn API for the algorithms.

Table 1: Sentiment Prediction Tokens

Token	Positive	Negative
Addictive	6.0	0.0
Worst	1.0	14.0
Balanced	11.0	4.0
Aweful	0.0	6.0

5.1.5 ***Predict Sentiment.*** In order to verify correct classification of features/tokens we applied the best performing algorithm for sentiment prediction Naive Bayesian) to print the predictions for each token/feature.The total of 399 tokens/features were achieved, of which 276 were classified as positive and 123 were classified as negative. Table 1 depicts this classification in simple nature. Tokens like 'addictive', 'balanced' are correctly classified as positive and tokens like 'worst', 'awful' are correctly classified as negative. We perform sentiment prediction on the review by training our model using Naive Bayesian classifier.

5.1.6 ***Predict Aspect.*** To verify that the features/tokens are correctly classified into atleast one of the aspect category we used Naive Bayesian classifier to print the predictions made for each token/feature. We got a total of 13 features which were classified to belong to either 'combat', 'gameplay' or 'action' aspects. An example of the classification of tokens can be seen in Table 2. Tokens like 'team' are predicted to have aspect 'gameplay', 'dmc' is predicted to have aspect 'action' and so on.

Table 2: Aspect Predicition Tokens

Token	Combat	Gameplay	Action
Team	0.0	7.0	0.0
dmc	0.0	0.0	6.0
play	2.0	3.0	2.0
story	0.0	2.0	2.0

5.1.7 ***Multiclass Prediction.*** In this step we combine the predictions made by the two different models in previous steps - Predict Sentiment and Aspect and make predictions for a new custom review input by the user. Examples of these can be seen in Fig The output of this step is the prediction performed on custom user reviews which are not present in the data set.

6 EVALUATION

For sentiment prediction we compared the prediction results of four classifiers to decide which one performs best for our dataset. Overall

all the classifiers performed well on our dataset for predicting the sentiment of the review as - positive or negative. The precision, recall, f1-score and accuracy values for each classifier can be seen in Table 3. Naive Bayesian has accuracy 65% and better F1-score as compared to Logistic Regression hence we selected Naive Bayesian to train our model and make predictions on custom review input by the end-user.

Algorithm	Sentiment	Precision	Recall	F-1 Score	Accuracy
Naive - Bayesian Algorithm	Positive	.66	.90	.76	65.44%
	Negative	.60	.25	.36	
Logistic Regression	Positive	.67	.85	.75	64.33%
	Negative	.55	.31	.40	
Linear SVM	Positive	.67	.79	.73	62.86%
	Negative	.51	.36	.42	
K - Nearest Neighbour	Positive	.59	.62	.61	52.205%
	Negative	.41	.38	.39	

Table 3: Statistical Results for Sentiment Prediction

Table ?? depicts the results of confusion matrix generated for Naive Bayesian which clearly shows high value for true negative - 152 and lower number for False positive - 17. These results verify our decision to use Naive Bayesian for predicting review sentiment.

	Combat	**Gameplay**	**Ation**
Combat	5	1	1
Gameplay	2	1	1
Action	2	0	4

Table 4: Confusion Matrix of Naive Bayesian

Similarly for aspect prediction we compared the prediction results of four classifiers to decide which one performs best for our dataset. Logistic Regression performed well as compared to other three classifiers as can be seen from the results displayed in Table 2. Hence, we trained our model using Logistic Regression for predicting aspect of custom review input by the end-user.

	Combat	**Gameplay**	**Ation**
Combat	5	1	1
Gameplay	2	1	1
Action	2	0	4

Table 5: Confusion Matrix for Logistic Regression

Table ?? represents the results of confusion matrix generated for Logistic Regression which show higher prediction for 'combat' related tokens being classified as 'combat'. We see 'gameplay' tokens having higher value as 'combat', this can be mainly because of two reasons - we built the data set manually and there were not enough sample reviews in our data set to train the classifier more efficiently. However, for 'action' we see correct predictions. The confusion matrices for other classifiers are available in our github repository [1] for review.

7 THREATS TO VALIDITY

Construct Validity The dataset is constructed file manually by consolidating reviews of three different games and manually added the attributes of polarity and aspect category to each review.

Table 6: Aspect Based Sentimental Analysis Results

Algorithm	Aspect	Precision	Recall	F-1 Score	Accuracy
Naive - Bayesian Algorithm	Combat	.50	.29	.36	35.29%
	Gameplay	.27	.75	.40	
	Action	.50	.17	.25	
Logistic Regression	Combat	.56	.71	.63	58.82%
	Gameplay	.50	.25	.33	
	Action	.67	.67	.67	
Linear SVM	Combat	.42	.71	.53	41.17%
	Gameplay	.33	.25	.29	
	Action	.50	.17	.25	
K - Nearest Neighbour	Combat	.14	.25	.18	35.29%
	Gameplay	0	0	0	
	Action	.83	.42.	.56	

External Validity We have collected reviews from only three games and the results may vary for larger datasets. But our results have good prediction accuracy and show potential. We have considered four classifiers and compared them with each other to select the most efficient one, there can be other classifiers too, which may give better results.

8 CONCLUSION

In our project we have implemented aspect based sentiment analysis on game reviews scrapped from metacritic website. To verify our approach we have used four different classifiers and used the best performing one to train our two models for predicting the sentiment and aspect of the review. We were able to get 65% accuracy while predicting sentiments and 59% accuracy while predicting aspects. Though we have manually created the dataset our results show potential for further extending this work by using reviews of more games and evaluating different classifiers.

REFERENCES

[1] Github repo. https://github.com/Hiten1928/SentimentalAnalysis.

[2] Activision. Call of duty: Black ops, 2010.

[3] B. Entertainment. Overwatch, 2016.

[4] M. J. Koehler, B. Arnold, S. P. Greenhalgh, and L. O. Boltz. A taxonomy approach to studying how gamers review games. *Simulation & Gaming*, 48(3):363–380, 2017.

[5] D. Pinelle, N. Wong, and T. Stach. Heuristic evaluation for games: usability principles for video game design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1453–1462. ACM, 2008.

[6] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, A.-S. Mohammad, M. Al-Ayyoub, Y. Zhao, B. Qin, O. De Clercq, et al. Semeval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*, pages 19–30, 2016.

[7] B. Strååt and H. Verhagen. Vox populi–a case study of user comments on contemporary video games in relation to video game heuristics. In *Game on, 9-11 September 2014, University of Lincoln, Lincoln, United Kingdom*, pages 5–9. University of Lincoln, United Kingdom, 2014.

[8] B. Strååt and H. Verhagen. Using user created game reviews for sentiment analysis: A method for researching user attitudes. *on CEUR in Proceedings of GHItaly*, 2017.

[9] J. P. Zagal, A. Ladd, and T. Johnson. Characterizing and understanding game reviews. In *Proceedings of the 4th international Conference on Foundations of Digital Games*, pages 215–222. ACM, 2009.

[10] J. P. Zagal and N. Tomuro. Cultural differences in game appreciation: A study of player game reviews. In *FDG*, pages 86–93, 2013.