Emotional Mario: A Games Analytics Challenge: MediaEval 2021

Mutaz Alshaer, Kseniia Harshina, Veit Isopp
Alpen-Adria Universität Klagenfurt, Austria
mutazal@edu.aau.at, k1harshina@edu.aau.at, veitis@edu.aau.at

ABSTRACT

Video games practice and experience, play a significant role to understand and analyze specific cases or scenarios of video games. Data and results that come from players' involvements during the gameplay, allow experiments and tasks to observe more about the game and methods. In the Mediaeval 2021 for Emotional Mario task, investigating the possible events through the biometric and facial emotion data for the popular old video game Super Mario Bros. Data of ten participants were used to show the results including players faces and gameplay, heart rate, interbeat intervals (IBI) and others were used to show the results.

1 INTRODUCTION

The main approach was to split the exercise into three approaches, Machine learning, finding outliers and analyzing emotional data. The idea was to combine all three approaches to get a reasonable result. This would be done by comparing the results of each approach and looking for matches.

The assumption is that if multiple results match, the likelihood of there being an event would increase. Finally, using the emotional dataset to determine which event might occur. Two different approaches were used, where the first approach was to compare all three results and look for matches only available on all three results.

The other approach was to check if at least two results match and if that is the case, take it as a match ignoring if the third result was also a match. The second approach might have more false positives but will also have more matches as the first approach will ignore anything that isn't matched by all 3 results.

2 APPROACH

2.1 Event Detection using Machine Learning

This approach focuses on trying to detect game events using Machine Learning (ML) algorithms. To achieve this the ground truth for the event data of the available participants was combined with the sensory participant data into a single data frame. Sensory data and event data are independent and dependent variables respectively. The first approach was to apply classification models to find the events. However, later it was decided to use regression models. To be able to use a regression model, the event data was transformed from event labels to probabilities, event frames corresponding to the 1.0 probability and the frames before and after corresponding to 0.9 for ten consecutive frames, then 0.8 and so on until 0.1. This way more event data was cultivated allowing us to use ML methods. Two regression models that were used were Random Forest and XGBoost.

2.2 Outliers of the Datasets

One of the approaches was to look for outliers of the datasets. To ensure that it doesn't give wrong outliers each dataset was looked at separately and the mean was taken from the dataset, then the standard deviation was used to check, whether there are a lot of outliers or not and then using this information narrow down the outliers. The assumption on this approach is that only outliers could be events, this is due to the assumption that the body of the person playing should react to stress, anxiety and happiness from the events that are being located. Then using the interquartile range the outliers were located. Finally, it was assessed that all outliers and the weaker outliers were included in the outliers. Here is to note that this approach could also only focus on the stronger outliers.

2.3 Facial Emotions and Gameplay

In this approach, we connected the facial emotions ("angry", "disgust", "fear", "happy", "sad", "surprise" and "neutral") of the 10 participants based on each frame during the gameplay. The aim is to recognize the potential key events such as the end of a level, power-up, extra life or Mario's death derived from the highest facial emotions. Since "neutral" would achieve the most identified emotion in frames, we decided to use the first and second highest emotion percentages and compare them with other approaches that match the same frame to determine the possible events to include in our analysis and results.

3 RESULTS AND ANALYSIS

3.1 Tables

The below tables represent the results, regarding frames and seconds of gameplay:

Table 1: Frame match +/-25 frames (match within 1 second)

Precision	Recall	F1
0.0175	0.0477	0.0256

MediaEval'21, December 13-15 2021, Online

Table 2: Frame match +/-125 frames (match within 5 seconds)

Precision	Recall	F1
0.0242	0.0812	0.0373

Table 3: Event match +/-25 frames (match within 1 second)

Precision	Recall	F1
0.0112	0.0057	0.0076

Table 4: Event match +/-125 frames (match within 5 seconds)

Precision	Recall	F1
0.0112	0.0849	0.0197

3.2 Figures

The Figure below is the example of the heart rate and specific event "new stage".

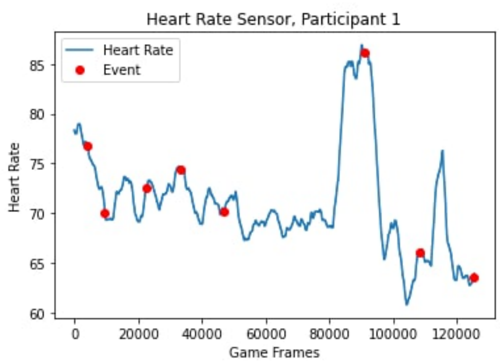

Figure 1: Heart Rate Sensor, Participant 1.

The figure depicts the heart rate of participant one throughout their gaming sessions. The red dots indicate when the "new stage" event occurs. Throughout this particular session participant reaches a new stage a total of 8 times. Some of the heart rate spikes indicate a possible correlation between the player's heart rate sensory data and reaching a new stage of the game.

4 CONCLUSIONS

The above-described methods were used to create multiple attempts to determine specific event locations in the participant videos and at the same time try to recognize the specific event as well. As a total of 5 approaches could be submitted, the following setup was used. As described in the Introduction the two separate methods either compare all three-event results approach or only compare two of the event results and find matches followed by comparing then two others and so on. In addition to these two methods, it was possible to increase the accuracy of the ML approach meaning the percentage and likelihood of it being an event according to the ML results. It was also possible to increase the threshold of the outlier approach. In the end, only the accuracy of the ML approach was used to check for better accuracy. Using the 2 methods and the 3 different approaches in addition to the changing in value for the ML results, into either more than 50% accuracy, more than 70% accuracy or more than 90% accuracy, a total of 6 possible results were found. The results from method two with a 90% ML accuracy returned the best results.

Looking at each of the above-mentioned approaches the error rate is high due to the many possible areas, were changing the values might affect the total outcome. Looking at the outlier approach it is very clear that by using the method of comparing only two approaches at a time, it is more likely to have a match with outliers. This might create more matches than should be possible, and changing the values on the outlier approach might have increased the accuracy. As depending on whatever weak outliers or strong outliers should be considered outliers. In addition to this depending on how high or low the threshold for the outlier approach was set the results might have also variated.

Another area for errors was the ML approach as it hasn't provided the expected accuracy required for the goal of the project, however perhaps with further data preparation techniques and/or trying alternative ML regression models the accuracy could be increased. Another route could be trying to apply deep learning to the problem. A possible reason for low accuracy with this approach could be that the number of events is too low to merit the use of ML, which usually requires large amounts of data. However, it is possible that with further research the approach could have the potential to provide more accurate solutions for similar problems.

On the other hand, in the facial emotion and gameplay approach, some challenges to recognize a specific event due to unusual or unexpected emotions by players' faces were encountered. For instance, a participant reacts to Mario's death with a happy emotion instead of sadness or anger. That leads to the emotional analysis of the players showing inaccurate results in some parts.

In conclusion, it is clear that more time would need to be used to tweak the threshold to increase accuracy on measurements. In addition, it needs to be noted that a total of 10 participants might also be to a small amount to create accurate approaches as it is unclear if any of the participants have completely different reactions to the other participants. This would highly reduce the accuracy for once in regard to the correct threshold set for the outliers, but also in addition to the ML approach.

ACKNOWLEDGMENTS

We would like to thank Dr. Mathias Lux for his support and help.

REFERENCES

[1] Aguinis, Herman, Ryan K. Gottfredson, and Harry Joo. “Organizational Research Methods Best-Practice Reprints and ...” Best-Practice Recommendations for Defining, Identifying, and Handling Outliers. Organizational Research Methods. Accessed November 27, 2021. http://www.hermanaguinis.com/ORMoutliers.pdf.

[2] Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785-794. 2016.

[3] Dekking, F.M, C. Kraaikamp, Lopuhaä H.P, and L.E Meester. A Modern Introduction to Probability and Statistics: Understanding Why and How. Springer-Verlag London, 2005.

[4] Ho, Tin Kam. "Random decision forests." In Proceedings of 3rd international conference on document analysis and recognition, vol. 1, pp. 278-282. IEEE, 1995.