Commentary generation for video game live-streaming

Cai Yulin†　Thawonmas Ruck†　Tomohiro Harada†

1. Abstract

With the development of live-streaming industry, many people tend to watch video game and sports in the live-streaming platform such as Twitch.com and Youtube Live[7]. The benefits of watching live-streaming of a video game are the fun made by commentaries and interaction between viewers. However, it is impossible for every live-streaming channel to have one commentator. Thus, this paper presents a novel solution for automatic commentary generation for the live-streaming video game.

2. Introduction

Usually, commentators bring us much enjoyment in sports competitions or events we watched on TV and listened to radio. Those people are the expert of mood making and explaining what happened during the game. Because of their exciting and professional commentaries, with and without the commentator, the experience is different. There are two types of commentator, play-by-play commentator, and color commentator. The play-by-play commentator usually comments about what is happening in the game while the color commentator comments about the background information of the players as well as some stories related to the game or sometimes gives light humor jokes. Together they bring enjoyment and entertainment to the audience[3]. Same things also happened in Esports live streaming. With the development of live streaming technology, the market is proliferating and reaching the millions level. Among the tons of live streaming contents, video games occupy the majority shares. Similar to the traditional sports commentator, the game commentator also brings our enjoyment as much as the traditional one[5]. However, it is a challenge for the streamer to make commentaries while playing the game that required high concentration to play, such as fighting game and real-time strategy game. Without the commentaries, it decreases the level of entertainment and rises the necessity to come out with a solution to solve the issue[2]. Another motivation for this research is because of the rise of procedural play generation(PPG)[8]. The PPG is about auto-generation of content for spectators who like watching live streaming. Currently, the generated gameplays lack commentaries so that there is a need to help spectators to understand the situations or make the gameplay more enjoyable by generating commentary in real-time. Inspired by the effects of the commentator in the traditional sports area, we explore the need of automatically commentary generation system for the live-streaming.

†立命館大学大学院情報理工学研究科,Graduate School of Information Science and Engineering, Ritsumeikan University

3. Related work

Fei Yan et al. proposed the commentaries generation for tennis video[9]. They built a dataset containing the tennis match videos from YouTube.com and commentaries from other websites. Based on the commentaries, they sliced the match videos to few seconds varied video clips. They used the support vector machine(SVM), Kernelised correlation filters to extract different visual features from the video. These features include the types(forehand or backhand) of the first and last shots in a point, the locations of the first and last bounces, and whether there is a net event. As for the commentary, they applied bags-of-words techniques to represent the word and weighted by term frequency-inverse document frequency(TF-IDF). So every commentary they collected can be represented by weighted vectors. They trained a recurrent neural network(RNN) to generate commentaries. The RNN is a type of artificial neural network, which performs super in text generation and time-series data. In this case, the input of the network is the visual features, and output is the probability of word tokens. The comment sentence is constructed by the highest probability of tokens. However, the processing time is a problem for the system. Feeding the model with video clips is not applicable in live-streaming, we need to improve their work to fit the requirement for the fast generation.

To help visually impaired people to enjoy TV broadcasting, Kiyoshi et al. proposed a solution for generating audio descriptions for sports live-streaming[6]. Their system had already tested in commentary generation for Judo match in Rio Olympic and made it possible to attach commentary automatically to international TV signals. There are four steps to generate the commentary. Firstly, they maintained the latest data containing sports facts such as current records, scores, and play-by-play event descriptions. Secondly, by detecting the change of new coming data, the system determined the events to make commentary. Lastly, using the pre-defined templates, the system composed the commentary according to the events and conveyed them to voice by text to voice techniques. Because the templates are made by the professional commentator, the quality of generated commentary is acceptable, and the processing speed is fast. However, in the video game live-streaming, such information is hard to obtain. To find the highlight moment of the video game, Ishii et al. proposed a method to generate highlights for video fighting game[4]. In their algorithm, the evaluation function in MCTS(Monte-Carlo Tree Search) is the combination of highlight cues. The highlight cues in the fighting game are the combination of score difference, action, and distance. They invited 66 participants to the experiment about comparing the effects of human selected and computer-generated highlights. The results indicated that the majority prefer the computer-generated one. From his work, we can see that the value of action, score difference, and the distance does play an essential role in the determination of Fighting game's highlights.

4. Methodology

This section addresses the idea proposed in this proposal. We choose FightingICE as the platform for commentary generation[10]. The reason is that FightingICE, a fast-paced video game, is used in an international game AI competition for many years. Also, it requires high concentration while playing, which results in the player only focus on the gameplay instead of commentating. The purpose of this research is to generate game commentary for the fast-paced video game. To achieve it, we propose the following steps. Firstly, we download the professional commentaries of fighting game competitions from Youtube.com. To generate the commentary template, we use SVM(support vector machine) to classify the commentary into play-by-play commentary and color commentary. Secondly, we combined the rule-based system and highlight detection techniques to detect the events in the game. Lastly, by matching the conditions pre-defined, the templates are used to generating the commentary text and convey to the vocal by text to speech algorithm.

Based on the previous works, we also considered score difference, action, and distance as the variable for highlight detection. The formula defined as follows:

$$Highlight = \alpha \text{Score} + \beta Action + \theta Distance$$

According to Ishii' work, assigning 0.3 to the parameter value of α, β and θ produced an excellent output for highlight generation. Thus, the value also used in our formula.

The difference in our work is the formula we used to calculate the score difference. We assume that when the game comes to an end, the excitement level will increase, based on the experience of the sports competition[1]. Therefore, the formula for calculating the score difference is showed below:

$$\frac{1}{|Damage| + 1} \times Remain\ time$$

As for the calculation for distance and action, we follow Ishii's formula.

$$Distance = 1 - \left|\frac{center - Xpros}{center}\right|$$

Center is the *x* coordinate at the center of the screen, ans *Xpros* is the player's *x* coordinate. *Distance* will have a higher value when the player is positioned closer to the center of the screen.

$$Action = \begin{cases} \frac{1}{2^{Rank-1}}, & belongs\ to\ RankAct \\ 0, & otherwise \end{cases}$$

RankAct is a list of actions, and *Rank* is the value associated to each action in the list. The higher the action rank in the list, the higher the value will get.

5. User evaluation

To evaluate our proposal system, we will conduct online questionnaires by asking the participants to watch generated highlights and give the score from 0 to 5. 0 means no interesting, and 5 means the most exciting clips. Each video clip is between thirty and ninety seconds in length. After each clip, participants will answer questions related to their enjoyment of the clip. We will prepare three types of the video clip, which are "the original commentary", "no commentary," and "the commentary generated by our system."

6. Conclusion

Anto commentaries generations can bring great enjoyment to the audience for esports live-streaming. Currently, real-time commentaries generation for esports is under-researched. In this proposal, we reviewed similar systems in commentaries generation and proposed an auto commentaries generation system for fighting game live-streaming. To evaluate our proposed system, we will ask participants to answer an online questionnaire.

7. Reference

1. Vinay Bettadapura, Caroline Pantofaru, and Irfan Essa. 2016. Leveraging contextual cues for generating basketball highlights. In Proceedings of the 24th ACM international conference on Multimedia, 908–917.
2. Sven Charleer, Kathrin Gerling, Francisco Gutiérrez, Hans Cauwenbergh, Bram Luycx, and Katrien Verbert. 2018. Real-Time Dashboards to Support eSports Spectating. In Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play, 59–71. https://doi.org/10.1145/3242671.3242680
3. Jeff Huang Cheung, Gifford. 2011. Starcraft from the Stands Understanding the Game Spectator. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 763–772.
4. Ryota Ishii, Suguru Ito, Tomohiro Harada, and Ruck Thawonmas. 2019. A Fighting Game AI Using Highlight Cues for Generation of Entertaining Gameplay. In 2019 IEEE Conference on Games (COG), 1–8.
5. Seth E Jenny, Margaret C Keiper, Blake J Taylor, Dylan P Williams, Joey Gawrysiak, R Douglas Manning, and Patrick M Tutka. 2018. eSports Venues: A New Sport Business Opportunity. Journal of Applied Sport Management 10, 1.
6. Kiyoshi Kurihara, Atsushi Imai, Nobumasa Seiyama, Toshihiro Shimizu, Shoei Sato, Ichiro Yamada, Tadashi Kumano, Reiko Tako, Taro Miyazaki, and Manon Ichiki. 2019. Automatic Generation of Audio Descriptions for Sports Programs. SMPTE Motion Imaging Journal 128, 1: 41–47.
7. Karine Pires and Gwendal Simon. 2015. YouTube live and Twitch: a tour of user-generated live streaming systems. In Proceedings of the 6th ACM multimedia systems conference, 225–230.
8. R Thawonmas, T Harada - on What's next for A I in games, undefined San, and undefined 2017. 2017. AI for Game Spectators: Rise of PPG. Ice.Ci.Ritsumei.Ac.Jp: 1032–1033. Retrieved from http://www.ice.ci.ritsumei.ac.jp/~ruck/PAP/wnaig2017.pdf
9. Fei Yan, Krystian Mikolajczyk, and Josef Kittler. 2016. Generating commentaries for tennis videos. In Pattern Recognition (ICPR), 2016 23rd International Conference on, 2658–2663.
10. Shubu Yoshida, Makoto Ishihara, Taichi Miyazaki, Yuto Nakagawa, Tomohiro Harada, and Ruck Thawonmas. 2016. Application of Monte-Carlo tree search in a fighting game AI. In 2016 IEEE 5th Global Conference on Consumer Electronics, 1–2.