International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 11 | Nov 2022 www.irjet.net p-ISSN: 2395-0072

Automatic Report Generation of a Football Match

[1]Harsh Parikh, [2]Anirudh Navin, [3]Dhairya Panjwani, [4]Utkarsh Bhosekar, [5]Gunjan Sharma, [6] Lokesh Devnani

[12345]Student, Dept of Information Technology, Thadomal Shahani Engineering College, Mumbai, India
[6]Student, Dept of Information Technology, Vivekanand Education Society's Institute Of Technology, Mumbai, India

---

Abstract - *Football in particular needs more automation as more nations and clubs want to build their squads using data-driven strategies. Modern games are continually changing, from the creation of stats to the tactics used on the field. The system we propose takes an input of the raw video file of the game and can split the input file into two halves of the game for better performance. The system then uses algorithms to extract highlights from the competition. Outputs here are highlights or key moments that occur in the game. It does not include the time from the game where there are not many eventful actions happening in the game. We then get the audio of the complete game and convert the audio file to text using the speech-to-text model. This is the final output of the entire proposed system.*

Key Words: **Machine Learning, Football, Report Generation, Highlights, Twitter analysis, Voice modulation**

1. INTRODUCTION

Automation in the field of sports, especially football is a need of the hour as many clubs and countries are opting for a data-driven approach for their side. From generating statistics to tactics on-field the modern game is constantly evolving. As the profits from sports are increasing exponentially, teams are heavily investing more in gathering statistics on their players. Certain statistics, such as the distance a player ran during a match, the average position of a player on the field, offensive yards per carry, etc. can provide some hidden information on a player's performance. It may help discover some hidden talent the player might possess. These statistics are then studied by data analysts and managers to gain insight into a game. These are player-focused stats that help a team improve the quality of individuals but when preparing for a game a manager needs a complete summary of the multiple games such as the teams' performance in the last few games, opponents' performance in the last few games, a summary of a few games where the opponent lost, etc. [1][2]

This project will make this job easier as it can summarize a whole football match into a text file that distils all the important plays that happened in the match. Such information could be crucial to winning matches if well-analysed by the team's coach. Managers can look at these reports to understand the important moments in a game and while preparing for the game players can also find the weak link in the defence of the opponent and attack that link. Every top manager must go through the process of analysing the opponent. During that, they need to examine previous games of the opponent. We propose a model that condenses the action of a match to a report that is easily digested, it will reduce the time management has to spend analysing every game. Our project is not only applied to the analytics department of a football team but also the people covering football games, broadcasting companies, sports leagues, and TV channels. Delivering high-quality content as quickly as possible to the viewers is their top priority. Keeping that in mind, media companies are turning to technology to find a way to speed up this process. To speed up the process of sports content generation, media providers are looking into ways to have AI analyse the game footage and pick out the highlight-worthy moments automatically. This could help sports journalists as they will now have a rough summary of the game instead of making notes throughout the game.

Our proposed system takes the input of the raw video file of the game which usually lasts ninety minutes plus injury time, to improve performance we can divide the input file into two halves of the game. The system will then extract highlights from the competition by using machine learning and non-machine learning algorithms. The output here is a highlight or important moments that happened in the game this will remove the time when the game is slow and nothing important is happening. After that, we remove the video from the highlights so that we now only have the audio of the whole game. Thus, when the audio file is converted to a text file using speech-to-text models will generate a match summary of all the crucial bits and how it happened from a professional commentator. This is the final output of the whole proposed system.

2. FOOTBALL GAME STATES

Before getting into the workflow on how the pipeline works, we first need to understand a few basic concepts about football and how a game can be broken down into multiple phases. As football is evolving at a high speed and a lot of the time these phases merge into each other. But for our Machine learning model, we have used these phases and their outcome will decide the highlights video.[3]

Build-up

Build-up play means having possession of the ball and trying to build an attack through the opponent's organized defense. Build-up play not only means having a bulk of short passes in a row, among the teammates but also a long ball from the goalkeeper to the forward is a build-up. The reason for this phase is to slowly build an attack and arrive at the opposition box prepared. This is a slow and laborious process that does not make the highlight video.

Transition

Transition in the game is a moment where a team loses possession and the opponent team tries to take advantage of the opponents being unorganized with their positions. The team with the advantage makes quick and well-coordinated moves to change the phase of the game by taking the best out of this opportunity and scoring a goal. Transition attacks require a lot of accuracy with every move they make.

Attack

Attack refers to the movement of the team having possession of the ball, making their moves flow through the opponent's midfield and defence, and attempting to score a goal. The team with the possession, playing passes, and making moves can be said as building an attack. Any successful attack that leads to a goal or a near-goal opportunity is a part of the highlight video.

Counter-attack

Counter-attack is one of the quickest ways to score goals in football. It starts with the transition in the game where a team loses control over the ball and the opponent tries to counter by making important passes leading the strikers to make the run. With a minimal number of passes, having an attempt on the goal counter-attacking style of football is said to be the most perilous as it can give maximum damage within a few seconds.

Press

Pressing is a term where the defending team intends to put pressure on the opponent team, in possession of the ball. The main idea here is to give the team in possession as less time as possible to make their moves for attacking. Pressing is not just running toward the player having the ball. It is done in a planned and structured manner.

Dead-ball

Dead-ball events in a football match are the moments when the play is stopped temporarily and the ball is not in motion. This happens when a foul is committed for example a handball, a failed tackle, or an offside. In these situations, the opponent team is awarded a free kick in the position where the foul is committed on the pitch. In case of a foul inside the box near the goalkeeper, a penalty kick is awarded.

Defense

Defense is an action where the players at the back (the center-backs and the full-backs) try to stop the opposing team's players from scoring the goal. In this, the team defending sits back and endures waves of attack from the opponent team. Anything fruitful in this will go in our highlights video.

3. Proposed workflow

We propose a system where we take the whole game and break it into chunks of 15-30 second videos the time is an arbitrary number and can be changed. The system will feed it to the next step where we try and identify the phase of the game and decide whether that makes into the highlights video. The system is not reliant on one model and another Non-Machine Learning model should be used to improve the accuracy. After that the video is dropped from the highlights. And audio is extracted and in the final stage using an ANN model we can generate a report from the audio file. This can be observed in Fig 1.

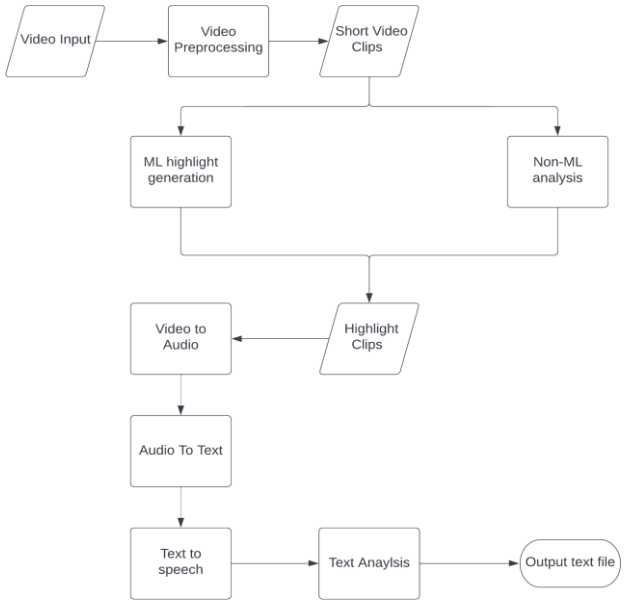

Figure 1. Proposed workflow diagram

3.1. Video Preprocessing

The video is divided into multiple smaller chunks and processed in a parallel manner. This is one of the primary methods for reducing time. It is possible to build it using video time pointers. Splitting is virtual in this manner and does not result in actual sub-file formation. However, the result contains multiple video files that must be integrated

into one. These limits at least make the method unpleasant when it comes to the duties at hand, especially when employing neighbouring frames for computation. To deal with this, we had to divide the video with an overlap such that no information was lost at junction points. The process of converting a coloured video to a black-and-white video is required because it reduces the time needed to generate the highlights from the raw video file which can be 90 to 120 minutes. Our goal here is to generate a report of the whole game rather than a highlight of the whole match so we can reduce the quality of the video.

3.2. Highlight extraction

3.2.1. Machine Learning Algorithms

Video shot identification is a critical component of sports video summary and highlight generation. A football match may be split down into a series of video clips. As a result, events in a football match can be thought of as video clips played in a sequence when combined can generate the whole match. Because highlights are typically a compilation of significant occurrences, video clips can be utilized to generate them. Processing video shots instead of the whole game will make the model more time efficient. It should be remembered that a football match might go on for quite some time, generally up to ninety minutes, and some extra injury time. These clips can be of arbitrary size and can be made in the video preprocessing stage. These clips can then be used here and the model can classify them into 5 possible events. In a football match, a replay is video footage of one or more moments from a match aired more than once. Because they are regularly utilized in-game highlights, replays give crucial indicators for critical occurrences. Furthermore, replays are especially significant in football since video assistant referees utilize them to affirm or evaluate a judgment. The model can take into notice that if the same event is replayed then there is a chance that the event is a part of the replay and is an important bit of the game and should be a part of the highlights.

In the highlights, we generally want successful attacks, Dead ball events, fouls that lead to a yellow or red card, and some unsuccessful attacks. This can come from attacks, counter-attacks, and dead-ball events. Dead ball events can be from any part of the field, our highlight cut needs to be from a free kick in the opposition half near the goal and penalty kicks. It can be a naive idea to just put those in these events in our highlight cut but there may come a situation where an exciting event happens in the buildup phase which should not be a part of the highlight. The most common thing that can happen in this phase of the game is when team A is in the buildup and suddenly due to a poor pass team B gets possession high up the pitch which results in an easy goal-scoring opportunity. Another major clue that can help us are celebrations that happen after scoring a goal. Celebrations are essential indications for identifying significant moments in a football match. We could develop a model to recognize a player's celebration. The positive class is made up of photographs in which a player has just scored a goal or earned a penalty, and the negative class is made up of images that would ordinarily occur throughout a football match i.e., an exceptional defensive play or a VAR (Video Assistant Referee) turning a decision.

A CNN and RNN architecture can be adapted to classify keyframes and remember previous clips to generate some continuity. For example, if the phase is divided into multiple clips, internal memory in RNN can keep track of previous clips. In this way, we can generate a Machine learning model that can classify each clip and categorize them into multiple phases then the RNN model will look at the previous classified clip and check if the current state of the game is interesting enough to be added to the highlights video. As CNN works well with image and video classification, we have chosen this and with a need to store the previous plays of the game which requires some memory we have chosen RNN. The architecture that we propose in this is CNN architecture like VGG-19 and for RNN we suggest GRU architecture as it shows better accuracy than LSTM because it is easy to modify and doesn't need memory units, therefore it is faster to train than an LSTM model and gives similar performance. [4]

3.2.2. Non-Machine Learning algorithms

The highlight generator should not rely solely on one parameter such as the model prediction of the clip. It should have multiple other methods such as voice modulation in the audio file as any important event is followed by a change in amplitude in the voice of the commentator or Twitter analysis of the match. Relying solely on one prediction can lead to some incorrect predictions which impact the insights that can be extracted. Even a small python code can be written that tracks the score line and cuts 1 minute before the goal is scored so that we catch the build-up and the whole attack. A combination of a machine learning model along with any of the non-machine learning models and taking a weighted average of them will reduce the risk of misclassification. Non-Machine Learning algorithms include the following:

Twitter Analysis

Techniques for automatically summarizing any sport using Twitter analysis have grown in popularity in recent years. When summarizing a sports video, we should consider the viewers' perspective to ensure that they comprehend not only the game flow but also the atmosphere of the game, such as the enthusiasm at each incident of the game. A video that has been summarized is referred to as a "highlight video" in this context. Events such as a goal, a play that aids a goal, and a goal-scoring opportunity thrill viewers since they are directly tied to a game's success or loss.[5]

As soon as any major event in the game happens fans go to their Twitter account and tweet about it on a particular hashtag. If we plot a time series graph concerning the frequency of tweets received, we can have a rough estimate of the key moments of the game. These are the moments that fans of the sport have deemed important and should be a part of the highlight video.

Audio analysis/ Voice modulation

Audio intensity is a valuable indicator for identifying significant moments in a football game. The enthusiasm of the audience, player appeals, and furious commentary are all examples of major audio cues that are frequently connected with key developments in a match. Audio characteristics have been utilized to identify goals and other situations such as fouls etc. [6] This method is not only concerned with just recognizing what is occurring on the field, but also with how the audience reacts to it. By analyzing the sound, the model detects anything out of the ordinary such as an audience cheering or a referee disagreeing with a player, and then selects the appropriate clips for inclusion in the compilation. However, this strategy may not be appropriate for games with less noisy audiences, such as chess, where there are no applauding fans. Speaking of crowds, they are not always a reliable method of detecting important events. In football at random intervals of time fans chant for a player or sing the club's anthem to motivate the home team.

3.3. Video to audio of the highlights video

In the pipeline, we processed the raw video file to generate a highlights video of the whole match by using a combination of machine learning algorithms or non-machine learning models. In this step we are trying to extract the audio from the highlights this will help in the upcoming step where we use Automatic Speech Recognition (ASR)[8] to generate the expected report. There are several libraries and techniques available in Python for the conversion of Video to Audio. One such library is MoviePy. MoviePy is a Python library for video editing, including slicing, concatenation, title insertion, video compositing, video processing, and effect generation.[9] Some instances of the application may be found in the gallery. The python script required to fulfil this is straightforward. It is as simple as accessing the audio member of the VideoFileClip object that we created to extract the audio. The extracted audio must then be written to a new file, the name of which must be supplied.

3.4. Speech to text

At its most fundamental, speech is nothing more than a sound wave. The acoustic properties of these sound waves or audio signals include amplitude, crest, peak, wavelength, trough, frequency, cycle, and trough. Most today's speech recognition systems use what is known as a Hidden Markov Model (HMM). This method is based on the idea that, when analyzed over a short enough duration (say, ten milliseconds), a speech signal might well be properly represented as a stationary process—that is, a process whose statistical features do not vary over time. Speech recognition or converting the audio file received from the previous step into a report can be done by creating an Artificial Neural Net or by simply using a python library.

Speech Recognition translates spoken words to text. Most people use Deep learning and Neural networks for speech recognition. We can use some of the packages from python for speech recognition such as, 'apiai', 'assemblyai', 'Google-cloud-speech', 'SpeechRecognition', 'Pocketsphinx', 'wit', etc. Google speech recognition API is an easy method to convert speech into text, but it requires an internet connection to operate. We can use the 'SpeechRecognition' module in our use case. The flexibility and ease of use of the SpeechRecognition package make it an excellent choice for any Python project. SpeechRecognition has a Recognizer class, it is the place where everything happens. It has instances that are used to recognize the speech. Each model has seven methods that can read various audio sources using the different APIs, recognize_bing(), recognize_google, recognize_google_cloud(), recognize_houndify(), recognize_ibm(), recognize_sphinx(), recognize_wit().

4. CONCLUSION

This project will be revolutionary in the field of sports analysis as it not only gives the manager and the analyst video highlights of the whole game distilled into a highlight video that lasts only a few minutes but also a report of the whole game in the form of text. There can be another step added to the pipeline where the generated report goes in through text analysis to see how the team won or any significant pattern that emerges from it. Another place where we can improve is modelling selection and training. this workflow can be adapted to any other sport like hockey or rugby where the score is kept and the phases of play are clear and distinct. Finally, we have proposed a combination of CNN and RNN can be useful but looking into other models and architecture that improve the classification accuracy and maintain continuity can improve the final product considerably.

REFERENCES

[1] Y. Lee, H. Jung, C. Yang, and J. Lee, "Highlight-Video Generation System for Baseball Games," 2020.

[2] P. Shukla et al., "Automatic cricket highlight generation using event-driven and excitement-based features," 2018.

[3] T. Worville, "Phases of play – an introduction," Stats Perform, Apr. 30, 2019. https://www.statsperform.com/resource/phases-of-play-an-introduction/ (accessed Nov. 24, 2022).

[4] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," 2014.

[5] Researchgate.net. https://www.researchgate.net/publication/287050282_Event_Detection_based_on_Twitter_Enthusiasm_Degree_for_Generating_a_Sports_Highlight_Video (accessed Nov. 24, 2022).

[6] M. R. Islam, M. Paul, M. Antolovich, and A. Kabir, "Sports highlights generation using decomposed audio information," 2019.

[7] B. Guven, "Extracting Audio from Video using Python," Towards Data Science, Nov. 29, 2020. https://towardsdatascience.com/extracting-audio-from-video-using-python-58856a940fd (accessed Nov. 24, 2022).

[8] K. Doshi, "Audio deep learning made simple: Automatic Speech recognition (ASR), how it works," Towards Data Science, Mar. 25, 2021. https://towardsdatascience.com/audio-deep-learning-made-simple-automatic-speech-recognition-asr-how-it-works-716cfce4c706 (accessed Nov. 24, 2022).

[9] "Moviepy," PyPI. https://pypi.org/project/moviepy (accessed Nov. 24, 2022).