Video Classification of Cricket Games

by Bhavana Madhuri Velakaturi

Submission date: 10-Dec-2021 11:03AM (UTC+0530)
Submission ID: 1726302277
File name: PW22SSS05-IEEE_Report.docx (551.42K)
Word count: 2562
Character count: 13456

Video Summarisation of Sports Games

Rohan Bennur
Computer Science and Engineering
PES University, R R Campus
Bangalore, India
rohan.bennur@gmail.com

J Jayanarayanan
Computer Science and Engineering
PES University, R R Campus
Bangalore, India
jay.jvpd@gmail.com

Bhavana Madhuri Velakaturi
Computer Science and Engineering
PES University, R R Campus
Bangalore, India
bhavanamv123@gmail.com

Eshwar P K
Computer Science and Engineering
PES University, R R Campus
Bangalore, India
eshwarpk2001@gmail.com

***Abstract*—There is no denying the crucial role highlight videos can play for many student-athletes in their recruiting process, or a busy person who just wants to view the gist of a particular match. Instead of spending hours on watching the full recording of a game, watching highlights would make that person's life much easier. A key ingredient of highlight generation is video classification. There exist multiple video classification algorithms, but a comparative study of these algorithms that work on classifying sports clips doesn't exist. In order to solve this problem, this work analyses multiple video classification methods such as CNN-LSTM, Random Forest, Decision Trees, Support Vector Machine and Naive Bayes, and compares each of them to determine which classification method provides the highest validation accuracy in a reasonable amount of time.**

***Keywords*— Video Classification, Acoustic Events, Feature Extraction, Classifier Module, Accuracy.**

I. Introduction

With the growth of the tech industry, new gadgets are being made everyday, with which an extensive amount of unorganized and unedited videos are generated. A favoured field corresponds to sports games, especially cricket. Sports games go on for a couple of hours to sometimes a couple of days. An ordinary consumer who wants to know the main events of the cricket match would not be interested in wasting hours of time to watch the entire recorded match. Therefore, technology to swiftly search and browse content that we want from wide-ranging videos is important. The need for automatic sports video summarisation is high, now more than ever.

Automatic video summarization is a major problem which requires extracting semantics from video clips. Extensive research has been done in the field of automatic video summarization using audio, video and text cues as features for generating sports highlights. A key ingredient of highlight generation is video classification. There exist multiple video classification algorithms, but a comparative study of these algorithms that work on classifying sports clips doesn't exist. This paper proposes to compare some of the most famous classification algorithms, and determine which model has the highest validation accuracy of correct classifications.

The target of our project is to compare some of the video classification methods such as CNN-LSTM, Random Forest, Decision Trees, Support Vector Machine and Naive Bayes, and determine which classification method provides the highest validation accuracy in a reasonable amount of time.

A. Project Scope

The goal of this project is to analyse and compare the multiple video classification algorithms to conclusively deduce which method provides the highest validation accuracy in a reasonable amount of time.

The objectives of this project include:

1. Understanding existing video classification algorithms that use acoustic features to perform the classification.
2. Procuring and preparing the video dataset.
3. Training and implementing the different classification models to compare the validation accuracies of each of them.
4. Designing a GUI for implementing the model with the highest validation accuracy.

B. Project Benefits

We gain a deeper understanding and appreciation for the significance of pre-existing video summarisation and video classification techniques. Some of the benefits of going with the acoustic event detection technique are as follows:

1. For a sport like cricket, in which the commentators' speech is rich with useful data, using an acoustic event detection method is more efficient and accurate compared to other techniques.
2. For a sport like cricket, where the crowd cheers only when the ball goes close to or beyond the boundary, or a wicket falls, the crowd's noise can also be used for the classification process.
3. When the above mentioned techniques are used together to perform key event classification, it can serve as a powerful tool for video classification with high accuracy and efficiency.

II. Literature Review

Video summaries generated through calculation of recall and precision rates and experimentally determined values of parameters were found not to be precise enough. Takahashi, et al (2004) concluded that personalisation in making video summaries was not considered to meet users' preferences [1].

A. Tejero-de-Pablos, et al (2018) proposed that summaries to be generated from UGSV datasets require for the videos to be recorded from a close distance which is not the case in most user-generated sports videos [2]. Shorter frames increase the computation and do not capture semantics of action completely whereas, longer frames lead to reduced accuracy. This infers the concern with frame length estimation.

Another paper included highlight generation based on detection of acoustic events. It was concluded by A. Baijal, et al (2015) that this approach was computationally less expensive and more accurate as compared to the above methods [3]. The surveyed paper concentrates on audio contents of Rugby matches to generate highlights.

Automatic highlight generation approach which uses video shot and replay detection. Extracted score bar to highlight key events was concluded by M. E. Anjum, et al (2013) as an accurate method and was kept as an alternative approach [4]. Accuracy of OCR software can influence time taken to classify key events.

Event based and excitement based features were also taken into consideration to identify key events as put forward by P. Shukla, et al (2018) [5].

III. Data

Although cricket is a famous sport in the world and there are a lot of cricketing videos on the internet , there isn't much research done on cricket as a sport . Due to lack of research work in the field of cricket there is no labelled and annotated refined data that can be used for supervised learning. Hence we've been continuously constructing labelled and annotated data.

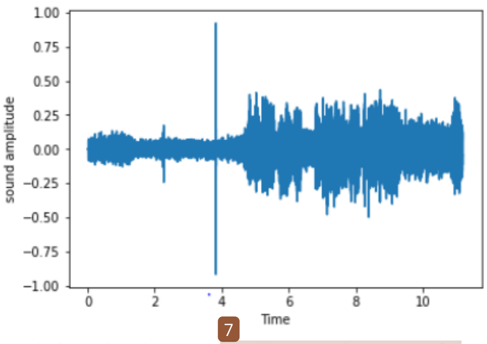
The dataset under construction has been divided under three major labels i.e four, six, wicket. Each of the labels consists of short video clips of around 10 seconds.The video clips collected are sourced from various tournaments like IPL , bilateral series between cricket playing nations , trilateral series and even the world cups. The videos are also sourced from all the formats of the game i.e T20, ODIs and test matches.

- T20 :
  T20 or twenty twenty cricket match is a short format cricket match limited to 20 overs of gameplay per each side lasting for about one and half hours for each innings.
- ODI :
  ODI short for One Day International is a form of limited overs cricket match between cricket playing nations where each side bats for 50 overs , and the game lasts for about 4 hours per innings and 8 for the whole match
- Test matches :
  Test match is the longest format of a cricket match where each side is supposed to play two batting innings per match and could last for 5 days.

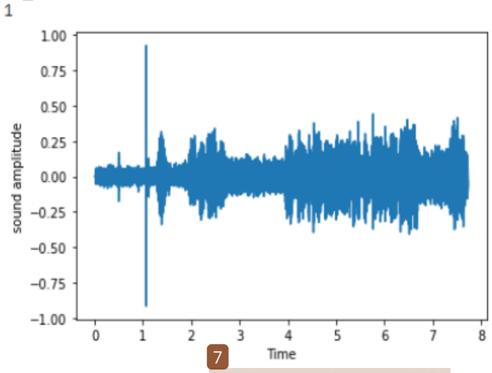
IV. Visualisations

EDA and visualisations :

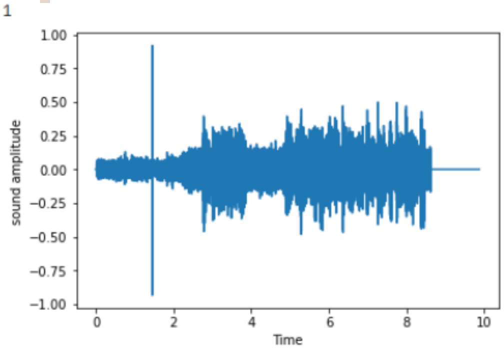
The below plot shows an amplitude vs time graph for when a wicket has fallen.



The below plot shows an amplitude vs time graph for when it is a six.



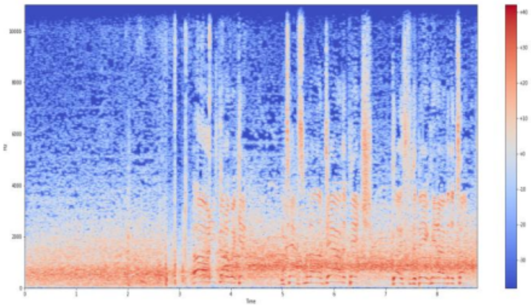
The below plot shows an amplitude vs time graph for when it is a four.



After careful inspection and comparing the graph with its audio data, it can be interpreted that the clear spike in the amplitude vs time graph occurs when the batsman has hit the shot with the bat. The spike is the meeting of the ball with the data. Since, this is when action begins, it can be taken as a cue for the start of the action. The commentator's speech before the batsman hits a shot is useless and therefore only the data after the spike is used in the process of data classifying.

A spectrogram is a visual representation of the signal strength or "**loudness**" and is usually depicted as a heatmap. This shows not only if there is more or less energy at, for example, 2 Hz vs 10 Hz, but also how energy levels vary over time.

The below figure shows a spectrogram for when a wicket, six or four has taken place.



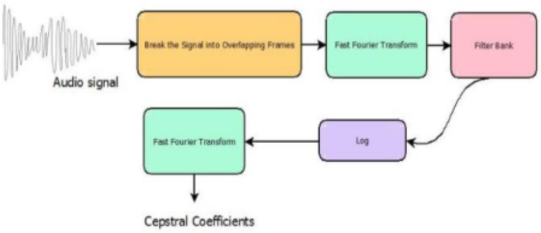
After careful inspection it can be seen there is an increase in energy or loudness of the commentators after 3-4 seconds of the video clip marking the excited nature of the commentator's speech.

V. Pre-processing and Initial Implementation

Pitch is one of the characteristics of a speech signal and is measured as the frequency of the signal.The *mel scale* is a scale that relates the perceived frequency of a tone to the actual measured frequency. It scales the frequency in order to match more closely to what the human ear can hear.

$$\mathrm{Mel}(f) = 2595 \log\left(1 + \frac{f}{700}\right)$$

As a part of pre-processing of data, the initial implementation includes the extraction of audio from a sport's video that has to be classified. From the extracted audio file, the MFCC features are extracted using the python library, librosa. The extracted MFCC features are then normalised and fed into the designed deep neural network models. The MFCC feature extraction technique basically includes windowing the signal, applying discrete fourier transform, taking log of the magnitude and then warping the frequencies on a *Mel scale*. It is followed by applying an inverse discrete cosine transform.



Initially the Moviepy module of python is installed which is majorly used for editing videos using python. A multimedia framework called FFmpeg is used for handling audio, video and other streams. The librosa library was then imported and the audio data was extracted from the videos and a new mp3 format was saved to a google drive. Using librosa and the newly stored mp3 audio file, MFCC features are extracted. Since the length of the videos in the dataset vary, the shape of the MFCC features for each audio file also varied. Hence all the MFCC features are resized to a constant shape of (20, 50). The MFCC features that were extracted are then stored in a speech_data array. These extracted features are to be normalised and then used as an input to CNN+RNN model.

VI. About The Model

Multiple video summarization methods were implemented and were compared to determine which summarisation method provides highest accuracy in a reasonable amount of time. They are :-

- Random Forest
- Decision Trees
- Support Vector Machine
- Naive Bayes
- CNN LSTM method

The Random Forest Classifier module, the DecisionTreeClassifier module and svm module, are all a part of the sklearn library. The GaussianNB function is imported from sklearn.naive_bayes library. The classifiers take the number of trees in the forest as an input. They are fitted with 2 arrays: an array holding the training samples, and an array that holds the class labels for the corresponding training samples. The model is then trained, after which the testing sample is given as input to the predict function to predict the labels for the test dataset. Finally, sklearn's metrics module is used to calculate the model accuracy of the classifier. For the SVM implementation kernel='linear' is chosen since we're making an SVM for data that is linearly separable. It can however be altered for non-linear data.

CNN-LSTM model consists of a single Conv1D layer with 4 LSTM layers. The LSTM layers help solve the problem of vanishing gradients which is normally found in a simple RNN and also help maintain the semantics over long ranges of data. The CNN layers are not only used to recognise patterns present in images but they are also used for spatial data analysis, computer vision and natural language processing. The model also consists of 4 dense layers with relu activation, whereas the output layer has a softmax activation for the classification task. The output layer has 3 neurons where each neuron represents one of the classes 4 runs, 6 runs or a wicket. Outputs are normalised, converting them from weighted sum values into probabilities that sum to one. The probabilities of each value are proportional to the relative scale of each value in the vector. The output values are between the range [0,1] which is helpful in order to avoid binary classification and accommodate as many classes or dimensions in the neural network model. Sparse categorical cross entropy loss function was used to calculate the loss of the model. The loss function calculates loss by computing the sum, where y_i hat is the i-th scalar value in the output and y_i is the corresponding target value.

$$\text{Loss} = -\sum_{i=1}^{\substack{\text{output}\\\text{size}}} y_i \cdot \log \hat{y}_i$$

Adam optimization was used so as to control the gradient rate in such a way that there is minimum oscillation when it reaches the global minimum while taking big enough steps to pass the local minimum hurdles. The model has therefore fared well with Adam as the optimizer and sparse categorical cross-entropy as the loss function. Model was trained for 100 epochs and achieved a training and testing accuracy of 93.89% and 74.25% respectively.

VII. Results

After performing the experiments and implementing the models, the accuracy of each model is as follows:

1. Using the CNN - LSTM method, we obtained a validation accuracy of **74.25%**
2. Using the Decision Trees classifier method, we obtained a validation accuracy of **62.01%**
3. Using the Random Forest classifier method, we obtained a validation accuracy of **60.23%**
4. Using the Support Vector Machine method, we obtained a validation accuracy of **58.9%**
5. Using the Naive Bayes method, we obtained a validation accuracy of **51.09%**

VIII. Conclusion

As the CNN - LSTM algorithm produced the output with the highest accuracy, we proceeded with using this algorithm to perform the video classification in our product. We developed an easy-to-use and elegant GUI for implementing video classification using CNN-LSTM so that users can easily upload clips and check if the video clip is a four, a six, or a wicket.

IX. Acknowledgement

We would like to thank our instructor Dr. Shylaja S S for her guidance and support throughout this project and PES university for giving us the opportunity.

I am grateful to the project coordinator, Prof. Silviya Nancy J, for organizing, managing, and helping with the entire process.

We are deeply grateful to Dr. M. R. Doreswamy, Chancellor, PES University, Prof. Jawahar Doreswamy, Pro-Chancellor – PES University, Dr. Suryaprasad J, Vice-Chancellor, PES University for providing to us various opportunities and enlightenment every step of the way. Finally, this project could not have been completed without the continual support and encouragement we have received from our family and friends.

X. References

[1] Takahashi, Yoshimasa & Nitta, Naoko & Babaguchi, Noboru. (2004). Automatic Video Summarization of Sports Videos Using Metadata. 3332. 272-280. 10.1007/978-3-540-30542-2_34.

[2] A. Tejero-de-Pablos, Y. Nakashima, T. Sato, N. Yokoya, M. Linna and E. Rahtu, "Summarization of User-Generated Sports Video by Using Deep Action Recognition Features," in IEEE Transactions on Multimedia, vol. 20, no. 8, pp. 2000-2011, Aug. 2018, DOI: 10.1109/TMM.2018.2794265.

[3] A. Baijal, Jaeyoun Cho, Woojung Lee and Byeong-Seob Ko, "Sports highlights generation based on acoustic events detection: A rugby case study," 2015 IEEE International Conference on Consumer Electronics (ICCE), 2015, pp. 20-23, DOI: 10.1109/ICCE.2015.7066303.

[4] M. E. Anjum, S. F. Ali, M. T. Hassan and M. Adnan, "Video summarization: Sports highlights generation," INMIC, 2013, pp. 142-147, DOI: 10.1109/INMIC.2013.6731340.

[5] P. Shukla, et al., "Automatic Cricket Highlight Generation Using Event-Driven and Excitement-Based Features," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2018, pp. 1881-18818, DOI: 10.1109/CVPRW.2018.00233.

Video Classification of Cricket Games

ORIGINALITY REPORT

14%	9%	5%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Mithali Shashidhar, Aishwarya Ramanath, Suresh Jamadagni. "Recommendation System for Farmers using Demand Forecasting", 2021 Asian Conference on Innovation in Technology (ASIANCON), 2021 Publication	2%
2	medium.com Internet Source	2%
3	www.kdnuggets.com Internet Source	1%
4	www.geeksforgeeks.org Internet Source	1%
5	Submitted to PES University Student Paper	1%
6	www.ncsasports.org Internet Source	1%
7	Emmanuelle R. Biglete, Mark Christian E. Manuel, Jonas D. Gampon, Daniel R. Ngo et al. "Taguchi-Based Experimental Investigation of Vortex-Induced Vibration (VIV) Energy	1%

	Converter under a Controlled Environment", 2020 IEEE 12th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM), 2020 Publication	
8	Submitted to University of Wales Institute, Cardiff Student Paper	1%
9	Yonas Tefera, Maarten Meire, Stijn Luca, Peter Karsmakers. "Chapter 11 Unsupervised Machine Learning Methods to Estimate a Health Indicator for Condition Monitoring Using Acoustic and Vibration Signals: A Comparison Based on a Toy Data Set from a Coffee Vending Machine", Springer Science and Business Media LLC, 2020 Publication	1%
10	towardsdatascience.com Internet Source	1%
11	machinelearningmastery.com Internet Source	<1%
12	repository.tudelft.nl Internet Source	<1%
13	Jiande Pi, Yunliang Qi, Meng Lou, Xiaorong Li, Yiming Wang, Chunbo Xu, Yide Ma. "FS-UNet:	<1%

	Mass segmentation in mammograms using an encoder-decoder architecture with feature strengthening", Computers in Biology and Medicine, 2021 Publication	
14	Submitted to Rutgers University, New Brunswick Student Paper	<1 %
15	cds.iisc.ac.in Internet Source	<1 %
16	arxiv.org Internet Source	<1 %

Exclude quotes	On	Exclude matches	< 5 words
Exclude bibliography	On		