

THE
UnKnowns

Arjun Sivaraman

Pratham Agrawal

Prasoon Soni

Shivam Dheer

Problem Statement

- ➔ Publishing highlights after a sports game is a mandatory practice followed in the broadcast industry. A highlights video provides a quick snapshot of the event highlighting the interesting events in the game and thus provides the viewers an opportunity to get a quick summary of the game.
- ➔ Use AI enabled methods to automatically generate highlights data feed from an input video file. The expected output data feed may contain start/end time stamps of interesting clips from the given video feed. For example, for a cricket game the clips may contain fall of wickets, exceptional fielding, batsman hitting boundaries etc.



Sportlight

A solution for highlights extraction for sports games by **The Unknowns**.

- The application uses AI enabled methods to automatically generate highlights data feed from an input video file.
- The expected output data feed contains start/end time stamps of interesting clips from the given video feed. For example, for a cricket game the clips may contain fall of wickets, exceptional fielding, batsman hitting boundaries etc.
- Finally, the highlights of that particular game would be displayed to the user sequentially.

How we built it ?

1. We would be building a web application, where the user can directly upload the video to the web server built using Node.js.
2. The video will be processed frame by frame. The processing will be convert the audio to text form by the use of symbl.ai using an API call.
3. The obtained texts along with the time frames will be processed by a python script consisting of technologies like NLTK, wordnet and NLP to carefully understand the text.
4. Firstly, the script would pre-process the data by removing the unwanted literals such as punctuation, special characters, connectives and other unnecessary words.
5. The script will then convert the phrase into its vector form. A phrase can be represented in n-dimensional vector space. A wordnet of some common cricket highlight words would be created and stored in for later use.
6. All of this is carried keeping in mind the emotions of the commentators in the audio and the words they speak. For example, “bowled”, “that’s a six”, and others (for cricket).
7. The vector would be parsed in a TF-ID vectorizer. Different similarity checks would be carried out in order to check the Similarity between the received text and the highlight words.
8. Once we get the check results, we can use it to decide if that particular frame can be used or not. If yes, we would be giving the start and end timestamp of the highlight (considering a 5-sec time interval).

Challenges we ran into...



- One major challenge we faced is to fix the values for different similarity checks in order to see if the given phrase is of a highlight or not.
- Different TF-ID vectors should be created for different sports in order to encompass all kinds of sports highlights.
- Segregating different phrases belonging to the same highlight so as to avoid repetitive highlights.

Current Implementation

- ★ Initial setup of the server with the project structure and git repository.
- ★ Constructed the wordnet cloud and TF-ID vector.
- ★ Established a connection between the symbl.ai API and the web-server to transfer and analyse the uploaded video.
- ★ Constructed pipelines for the smooth functioning of the python script.
- ★ Integration of the web application, and the python pipeline.

what's Upcoming

1

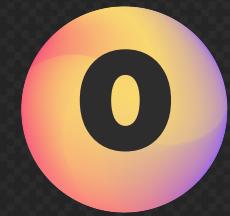
Optimising the time taken for the video to be processed by the API (symbl.ai) and the NLTK model.

2

Displaying the highlights from the obtained timestamps to the user through a user friendly interface.

Business Model

Subscription Based Model



Level 0

Access to all the users for free



Level 1

Paid Subscription



Level 2

Paid Subscription

Level 0 Subscription

Access to all the members for free.

Benefits :

- Save on your time by getting the match highlights.
- Upload video and get the generated match highlights.
- Get access to 1000 minutes of free video upload.

You will be given access to the above immediately after sign up to our portal.

Level 1 Subscription

Paid Subscription.

\$3.99 for 1 month OR \$20.99 for 6 months OR \$40.99 for 12 months

Level 0 Benefits + the following...

- Get unlimited minutes of video uploads to get the generated highlights.
- Get access to the key moments of both innings.
- You will get an option to choose the length of the highlights generated.
(Short / Medium / Detailed).
- Access to downloadable exclusive match videos. Save videos and watch them from anywhere offline.
- Access to downloadable exclusive match report.

Level 2 Subscription

Paid Subscription.

This level would be included in the upcoming implementation of the project which would be based on the feedback received from users and with the trends upcoming.

Proposed Revenue Model

\$4.99 for 1 month OR \$26.99 for 6 months OR \$51.99 for 12 months

Level 0 Benefits + Level 1 Benefits + the following...

- Video uploaded can be in any language.
- Video Quality Options to be available.
- Audio Quality Options to be available.

**other features to be added based on the feedback received and with the upcoming new transitions in sports.



Thank You
THE UNKNOWNS...