SPORTS ANALYSIS

Project Proposal

Course: CSCE 5290 Natural Language Processing, Spring 2022

Instructor: *Dr. Sayed Khushal Shah (sayed.shah@unt.edu)*
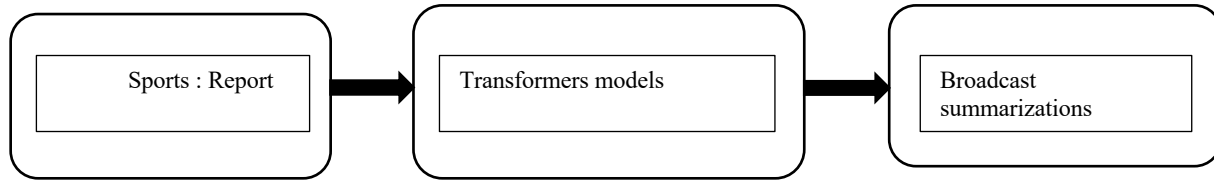
Member:

Sreeja Bellamkonda *(sreejabellamkonda@my.unt.edu)*

Github Link:

https://github.com/SreejaBellamkonda/NLP--Sport-Analysis

Idea Description

Sports : Report	→	Transformers models	→	Broadcast summarizations

Every day, tens of thousands of athletic activities take place. Despite their pattern form, most of the sports news (results of sports tournaments) is written by hand. We want to see if it's feasible to produce news based on the broadcast - a series of remarks that characterize the game in real-time - in this project. This problem is classified as a summarization problem, and it is solved by using the Natural Language Processing Models.

Approaches that are both extractive and abstract We'll start with extractive models and then go on to using the Bidirectional Encoder Representations from Transformers (BERT) as an encoder and text augmentation using a thesaurus. Other types of encoders, on the other hand, do not exhibit considerable gains.

The basic idea of our project is to build a model that can generate summarizations of sports commentary.

Motivation & Significance:

Sports material has gone viral in recent years. Some events are popular, with billions of individuals tuning in to witness them. Thousands of events happen daily throughout the globe that piques the curiosity of a lot of people. Online sports materials appeal to a large range of individuals. whether or not an individual has just seen a game, he's inquisitive about reading the news since it contains additional information. As a result, human resources are in high demand to put in writing this news or multiple for every athletic event. Cutting expenses while improving the standard of stories has been a priority for media firms. Some well-known newspapers, like the Associated Press, Forbes, The New York Times, and the la Times, generate news automatically in easy areas and can also conduct research to extend the standard of such news. the subsequent are a number of the benefits of manufacturing sports news:

- The algorithm will write news faster and make the minimum number of errors.
- The algorithm is ready to generate identical news from a distinct point of view, in several languages and personalize them to a private reader's preferences.

As we see, the above needs provide the first motivation for our work. We decided to find out the way to generate news, entirely automatically, without highlighting the fit about the match, based only on textual comments of the match. Text comments are expressions, and sentences that describe a game at a specific point in time.

Objectives & Features:

The objective of our project is to develop a very trained model that will summarize the knowledge and comparatively perform on par with the manual summarizations.

Technical Goals:

- Reviewing previous work on text generation using extractive and abstractive approaches.

- Apply unsupervised techniques to calculate extractive summarization and oracle extractive model. Later apply two neural seq2seq approaches.

- Compare the results of the study and explain why different approaches give such results and discuss the constraints within this task and also the further actions.

References

1. Theanirudhvyas. "Theanirudhvyas/Sports-Commentary-Summarization: Creating News Articles from Cricket and Football Commentaries." *GitHub*, https://github.com/theanirudhvyas/Sports-Commentary-Summarization.

2. Engdahl, Sylvia. "Blogs." *Amazon*, Greenhaven Press/Gale, 2008, https://aws.amazon.com/blogs/machine-learning/enhance-sports-narratives-with-natural-language-generation-using-amazon-sagemaker/.

3. "Sports Analytics: Analyzing Cricket Commentary." *Analytics Vidhya*, 25 June 2020, https://www.analyticsvidhya.com/blog/2020/02/sports-analytics-generating-actionable-insights-using-cricket-commentary/.

.