Group 14 Project Proposal

Github

https://github.com/RachaelC358/NLP_Group-14_Project

1. Project Title and Team Members

Title

Text Summarization of Sports Recaps for Game Analysis

Team Members

- Nathan Adams
- Jacob Benz
- Rachael Carpenter
- Zachary Smith

2. Goals and Objectives:

Motivation

Most major sports have a number of outlets which cover each team throughout their respective seasons. These sources of commentary and statistics are useful documents for understanding how well a team is performing in a season and may be the only resource that a fan might see for a particular game or event. For instance, fans of baseball will have 162 games a year to keep up with for a single team. While there are certainly diehard fans who will watch every minute of every game, most people will often rely on post-game recaps to understand what happened in a game they did not watch. Even in sports with fewer games in a season, like college football, there is no shortage of opinion articles or game breakdowns.

Noting that news for sporting events far exceeds the live coverage, a method for viewing recaps or articles in a condensed, bite-sized manner would be useful for someone trying to catch up on the latest game. By summarizing recaps, there is

potential for additional analysis to be done in a quick manner as well. For instance, sentiment analysis might be a function that is added on after the text summarization is complete. This is dependent on how much time is needed to complete the initial functionality, however.

Significance

Automatic text summarization can help save time in people's everyday lives by allowing them to make informed decisions with less effort. Manually creating summaries of large sets of data is so inordinately time consuming that it would be almost impossible to do without machine learning.

Sports, in general, have a massive following and as result there are almost too many resources to keep up with. Condensing the corpus of information for a particular team could help sports maintain interest as a season goes along. Many fans stay engaged in the early days of a new season but then later disengage as the games pile up. A programmatic way to deliver fast news could reduce fan falloff and give exposure to a larger variety of sources. This technique has been used for political or world news already to varying effect. Our idea is to essentially provide the analog of an academic article abstract in the world of sports. Another similar idea is that of a book summary on the back cover. Once a user reads the summary, it may prompt them to consume the article in its entirety.

As mentioned before, our idea may be extended beyond summarization. We may use our summaries and a simple sentiment analysis to provide quick information and context in a single package. This yields a result that not only explains what happened but might indicate the tone of the report. By collecting a number of sources for the same event, a fan could get some of the same information as they would have watching the game live.

Objectives

Text summarization has been done in multiple different ways, but we want to implement an accessible solution that caters towards sports coverage. Web scraping documents combined with utilizing neural networks and other NLP techniques will allow us to provide a brief summarization of any given article.

In our final iteration, our project could be hosted on a webpage, where the user can input a URL and our application will return a summarized version of the article. Since we

plan to use this specifically for sports, other information such as scores, teams, and important players/plays can be extracted, formatted, and displayed for the user as well.

While the primary objective is summarization, we want to attempt to gather as much information as possible so that our project can easily be extended upon each module's completion. If we can develop good web scraping tools, pull the score for each game, and then store that information, we can use that data to not only summarize text recaps, but to also create a database of statistics. This is a secondary objective but one which naturally follows our primary summarization goal.

Features

A primary feature of this project is its ability to gather a large amount of text information from varying sources on the internet. This is currently planned to be achieved via the earlier stated technique of web scraping, which, in addition to its above outlined importance, will play a major role in the data collection portion of the research and development.

One obvious feature of our design plan is the ML aspect, which includes a model/technique that will succinctly coordinate text processing and creation. As it currently stands, there are many paths forward to go about such a design, and details such as which exact model, method of normalization, and smoothing algorithm can be determined as we progress forward. Currently, the probable method we plan to utilize will involve a variation of a neural network.

Ideally the user interface component of the project should be as simple as possible, both in terms of creation and UI/UX design. This will involve an easy-to-read website with minimal extraneous features.

3. References

Huang, K. H., Li, C., & Chang, K. W. (n.d.). Generating Sports News from Live Commentary: A Chinese Dataset for Sports Game Summarization. *Tencent AI Lab*. https://aclanthology.org/2020.aacl-main.61.pdf

Priyadharshan, T., & Sumathipala, S. (2018, December). Text Summarization for Tamil Online Sports News Using NLP. *2018 3rd International Conference on Information Technology Research (ICITR)*. https://doi.org/10.1109/icitr.2018.8736154

Sblendorio, D. (2022, January 8). *How to do text summarization with deep learning and Python*. ActiveState. Retrieved October 7, 2022, from https://www.activestate.com/blog/how-to-do-text-summarization-with-python/

USMNT scoreless draw against Saudi Arabia adds more concerns before World Cup. (2022, September 28). Stars and Stripes FC. Retrieved October 7, 2022, from https://www.starsandstripesfc.com/usmnt-friendly/2022/9/28/23377078/usa-usmnt-scoreless-draw-friendly-saudi-arabiarecap-world-cup