Creating spoiler-free summaries of sitcom television screenplays

Caspar de Jong
caspar.de.jong
@student.uva.nl

Piter Nicolaas
piter.nicolaas
@student.auc.nl

Seb Olsen
seb.olsen2
@student.auc.nl

Abstract

This project aims to create spoiler-free summaries of TV show episodes using scripts from the sitcom Friends. By developing a model that identifies and omits the climax and resolution stages of the episodes, we seek to generate concise and engaging episode previews without revealing key plot points. This approach preserves the viewing experience, providing a glimpse into episodes without spoiling their outcomes.

1 Introduction

The recent success of large language models in different natural language processing tasks has led to excitement and adoption of this technology by many industries and individuals. One task where such models have been showing great performance and rise in accuracy is automated summarising. An interesting application of summarising is creating descriptions for movies and TV show episodes showed on streaming services and movie review websites. Summarising TV show episodes should take into account many factors of which one is to ensure the generated texts does not contain large spoilers.

While there is previous work done on automating summarising movies and TV show episodes, little research has been done on accounting for spoiler-free AI-generated summaries. Therefore, we will research how to adapt a language model to produce spoiler free summaries for TV show episodes.

Our approach is to develop a method with which we can recognize the climax stage in a TV show episode, based on the publicly available script. This information we give to a T5 summarisation model to create spoiler-free descriptions of TV-show episodes. Our research questions are therefore:

1. *How can we automate recognizing the climax stage of a story?*
2. *How can we teach a summarisation model to ignore the climax stage in its summary to avoid spoilers?*

2 Related Work

Spoilers and Climax TV show and movie stories are consecutive events where interactions happen between characters, their surroundings, and their ideas. The plot of a story is the logically connected sequence of these events. Research done on movie story narratives shows structural similarities across various stories. This suggests that most stories fall into a relatively small number of unique plot types (Liu et al., 2020). This simplification makes it easier to find the position of the stories' climax, as it is suggested that the climax often occurs at 3/4th of the plot.

In their (2024) paper researching plot extraction, DeBuse and Warnick define a notion of entity activity, which describes the concentration entities in a scatterplot they draw to describe the narrative of a story. After applying measures of smoothing, an entity activity line was drawn by summing the gaussian curves of each entities' activity. DeBuse and Warnick (2024) argue that the peak of this entity activity line towards the end of the story corresponds to the climax of the plot. They also mention that the other peaks seen in this activity line can be seen as sub-climaxes or crisis points in the story.

Automated summarising Automated summarisation is a core natural language processing (NLP) task for language models. Using the concept of transfer learning, Google developed a state-of-the-art summarisation model which they named T5 (Raffel et al., 2020). This model performs great on different NLP task and summarisation benchmarks. An extension of this model is T5-long

(Guo et al., 2021a). T5 can handle an input sequence of 512 tokens and T5-long can handle 4096 input tokens.

3 Preliminary exploration

To create spoiler free summarisations, our first idea was to fine-tune a language model on real-world episode descriptions from IMDb [1]. We wanted to use descriptions for the Friends episodes from IMDb. These descriptions are one to three sentences long. The idea behind this is that these episode descriptions are manually curated and would therefore be spoiler free. To retrieve this data, we built a web scraper to read off the descriptions for each episode from the IMDb website.

However, we stumbled across difficulties in making this fine-tuning work. First, we realized that this fine-tuning process does not generalize to other TV-scripts. It will learn the content of the Friends descriptions rather than how to make a spoiler-free summary. Secondly, we learned that fine-tuning would require to manually label the output by the model, which would cost too much time.

4 Data collection and description

For this project, we used the script of the TV-show Friends [2]. Friends is a popular American TV-show, which aired from 1994 to 2004 on NBC. The story tells the life of six young urban professionals: Rachel, Joey, Chandler, Phoebe, Ross, and Monica living in New York City. The TV-show consists of 10 seasons and 234 episodes in total. The scripts of every episode are available on Kaggle as .txt file [3]. Each .txt file is broken down in scenes, which are marked by square brackets as follows:

[Scene: Central Perk, Rachel is introducing Phoebe, who is playing her guitar for the crowd.]

The scenes consist mostly of dialogue lines by the characters, displayed as follows:

Chandler: Oh, great. This is just...

Other lines in the dataset are descriptions of the situations or stage directions, surrounded by brackets:

(Monica and Rachel look at Phoebe strangely.)

[1] https://www.imdb.com/title/tt0108778/episodes/?ref_=tt_eps_sm

[2] https://www.imdb.com/title/tt0108778/

[3] https://www.kaggle.com/datasets/blessondensil294/friends-tv-series-screenplay-script/data

5 Methods

5.1 Preprocessing data

To make the data usable for climax detection, we preprocessed it by breaking up episode scripts into scenes, and removing the names of the characters before their dialogue. Secondly, we removed the brackets of situation descriptions.

5.1.1 Climax detection

The research by DeBuse and Warnick (2024) mainly focuses on scene partitioning. Since our dataset is already partioned into scenes, we decided to create a simplified version of their entity activity line, by calculating relative entity activity for each scene, under the assumption that the climax of a sitcom episode takes place in a single scene. By counting occurrences of the TV shows' main characters relative to the total token count of each scene, a relative entity activity value was found. To make sure the summaries generated by our summarisation model do not contain the climax and subsequent resolution stages of the plot, we cut off the script of each episode before the scene with the highest relative entity activity, under the assumption that this is the scene containing the climax.

Lastly, the cut off scripts were tokenized and subsequently lemmatized, before being fed into our summarisation model.

5.2 Summarization

Because of the average script length in our dataset, the LongT5 model was chosen for the summarisation task. First introduced by (Guo et al., 2021b), this model makes use of a new attention mechanism to reduce the steep increase in cost when working with longer inputs. The LongT5 model is based on the T5 transformer model, but is able to produce state of the art results and outperform the original T5 model on multiple summarization tasks.

6 Results and Findings

The results of the climax detection method are shown in Figures 1 and 2. They show that there are three small spikes in climax detection around 40, 60, and 85 percent. There is also a large climax spike at the end of the episode.

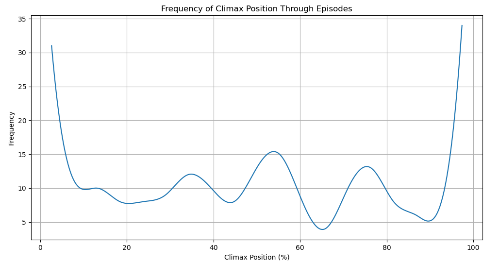

Figure 1: Climax position based on absolute entity activity summed for all episodes

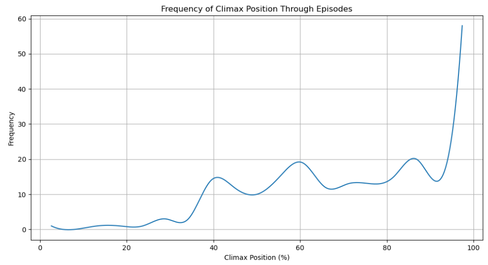

Figure 2: Climax position based on relative entity activity, summed for all episodes

With our results from the climax detection method, we decided to ignore the first 40 percent of the climaxes detected, assuming that these contain the scenes where the problem of the episode is formed, and not the climax or resolution stage of the plot. This was done to ensure the input content for the model would contain at least some useful story information. Below are the results for summarising episode 1 and 2 with the cut-off script.

Summary of episode 1: Yeah, yeah., yeah , yeah, yeah, yeah, yeah..!! I'm not going to do

Summary of episode 2: ...., Carol, Carol, Carol, Carol, Carol, Carol, Carol, Carol, Carol, Carol, Carol, Carol, Carol, Carol, Carol, Carol, Carol, Carol, Carol,?? Carol, Carol,? Carol? Carol, Carol, Carol

As these summarisation results are not accurate, we performed additional extensive summarisation as a preprocessing step. This method selected the most relevant sentences of the script, which were then given to the T5-long model. This yielded the following results:

Summary of episode 20: Can't believe two cows made the ultimate sacrifice so you guys could watch TV with your feet up. Yeah, since the divorce, when anybody asks me how I am, it's always with a sympathetic head tilt.

The summarisation model performed better in creating correct sentences but it did not accurately create a summary of the script.

7 Discussion and Conclusion

In this research, we aimed to create spoiler free summaries of TV-show episodes by using their script. We successfully performed a climax detection method on the script of the TV-show Friends. With this climax detection information, we cut off the scripts and forwarded them to a T5-long summarisation model. Unfortunately, the results from this summarisation model were not very accurate. At first, the model's output was very repetitive and did not make any sense. After additional extensive summarisation, the T5-long model created more correct sentences. Although the output was not an accurate description of the episode, as it was a summarisation of dialogue. This shows a clear limitation of our research, as our data consisted mostly of dialogue, which does not fully describe what happens in these television episodes. Unfortunately, we did not have the time and the resources to use a dialogue summarisation model which might have yielded more promising results. Another limitation is that the scripts consisted of both dialogues and descriptions or stage directions, for which there are no summarisation models that work well with both. Therefore, future work should focus on summarisation models that can handle script input. A method for future research could be a way to incorporate stage directions as a separate input into a summarisation model, to try and draw a better context as to where and under which circumstances the scene takes place to allow for better summarising of what happens in the episode. This would however make the model less universal, since not all datasets of screenplays or scripts contain stage directions as well.

8 Contributions

Piter researched climax detection, wrote code to handle climax detection and wrote the section on climax detection, relevant literature context and methods in the report. Seb wrote the code to train the model and handle preprocessing and generate results. Caspar wrote the introduction, data collection, conclusion and generally oversaw the editing of the report.

References

DeBuse, M. A., & Warnick, S. (2024). Plot extraction and the visualization of narrative flow. *Natural Language Engineering*, *30*(3), 480–524. doi: 10.1017/S1351324923000232

Guo, M., Ainslie, J., Uthus, D., Ontanon, S., Ni, J., Sung, Y.-H., & Yang, Y. (2021a). Longt5: Efficient text-to-text transformer for long sequences. *arXiv preprint arXiv:2112.07916*.

Guo, M., Ainslie, J., Uthus, D., Ontanon, S., Ni, J., Sung, Y.-H., & Yang, Y. (2021b). Longt5: Efficient text-to-text transformer for long sequences. *arXiv preprint arXiv:2112.07916*.

Liu, C., Shmilovici, A., & Last, M. (2020, 02). Towards story-based classification of movie scenes. *PLOS ONE*, *15*, e0228579. doi: 10.1371/journal.pone.0228579

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., … Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research*, *21*(140), 1–67.