Spoiler Alert: Using Deep Learning to Detect Spoilers in Text and Generate Similar Spoiler-Free Text

Abolfazl Eshagh, Mobina Salimipanah, Parnian Razavi, Ali Nikkhah, Sarina Zahedi, Ramtin Khoshnevis

Abstract

In this work, we explore the use of deep learning techniques to detect spoilers in text and generate spoiler-free versions of the content. We focus on two tasks: first one being generating spoiler-free plot summaries from movie synopses and generating spoiler-free review summaries from user reviews, and the other one being classifying text with spoilers.. By fine-tuning a BART model, we demonstrate the potential of conditional generation methods to effectively create spoiler-free text. Our results show that while smaller datasets lead to overfitting, larger datasets allow for more robust fine-tuning, resulting in improved performance.

1. Introduction

Spoilers in text-based content such as movie reviews and synopses can significantly impact the reader's experience. The ability to automatically detect and remove spoilers from text is a valuable tool in preserving the enjoyment of narratives. This paper presents a deep learning approach to both detect and generate spoiler-free text using state-of-the-art models. By leveraging large-scale datasets and powerful transformers like BART, we aim to provide a comprehensive solution to this problem.

2. Related Work

Research in spoiler detection has primarily focused on classification tasks, where machine learning models are trained to identify spoiler content. However, less attention has been given to the generation of spoiler-free text. Our work builds on previous studies by not only detecting spoilers but also generating alternative text that retains the original meaning while removing spoilers.

3. Datasets

For this work, we utilized the IMDB Spoiler Dataset, which is divided into two parts::

1. **IMDB Reviews Dataset**: This dataset contains over 450,000 user reviews from the IMDB website, including both spoiler and non-spoiler summaries. We used this dataset for the review summary generation task.
2. **Movie Plot Synopses Dataset**: A subset of the IMDB dataset focusing on movie plot summaries. This dataset was used for the plot summary generation task. Rows without values in the `plot_synopsis` column were removed during preprocessing.

4. Method

Dataset preprocessing

The first step involved cleaning and preparing the datasets for model training. For both datasets, we removed irrelevant columns and filtered out any entries with missing values in key columns such as plot_synopsis and review_summary.

1. **Spoiler Classification:**

:

https://github.com/NLP-Final-Projects/Spoiler-Alert

2. **Plot Summary Generation:**
    - We started by removing rows with empty values in the *plot_synopsis* column.
    - We then fine-tuned the BART model (*BartForConditionalGeneration*) on this dataset. Due to the small size of the available data, the model quickly overfitted, achieving the best performance within 10 epochs.
3. **Review Summary Generation**:
    - Similar preprocessing steps were taken, focusing on the *review_text* and *review_summary* columns.
    - Given the larger size of this dataset (over 450,000 samples), we randomly downsampled the data to fit within our computational resources.
    - The BART model was then fine-tuned on this downsampled dataset. Unlike the plot summary task, the larger dataset allowed the model to continue improving across multiple epochs, avoiding early overfitting.

4.1. Classification

…………………………………………………………
…………………………………………….

4.2. Generation

In the generation phase, we have undertaken multiple tasks, focusing on generating spoiler-free plot summaries and classifying plot summaries using two different models: BART and LED.

Spoiler-Free Plot Summary Generation
First, we performed initial preprocessing on our dataset. It is important to note that some entries lacked values in the *plot_synopsis* column, necessitating the removal of those samples. We then defined our model using the *BartForConditionalGeneration* model. This model was fine-tuned over 10 epochs. Unfortunately, due to the limited size of our available dataset, the model quickly overfitted, making further fine-tuning beyond 10 epochs impractical. The best state of the model, captured during this training session, was saved and utilized for evaluation. The results for this task are evaluated using the metrics outlined in Section 5 and are presented in Table 2.

Spoiler-Free Review Summary Generation
Similar to the plot summary generation process, we began by preprocessing this part of the dataset. The only two features needed were the *review_text* and *review_summary* columns, so all other columns were discarded. It is worth mentioning that this part of the dataset is significantly larger, containing more than 450,000 samples of review text and corresponding summaries. Due to computational resource constraints, we had to downsample this dataset randomly. Next, we loaded the *BartForConditionalGeneration* model and fine-tuned it on the prepared data. Unlike the plot summary task, the larger dataset size allowed the BART model to continue achieving lower losses with each epoch. The results for this task are evaluated using the metrics outlined in Section 5 and are shown in Table 3.

5. Metrics

To evaluate the performance of our models, we used standard metrics such as ROUGE and BLEU scores, which measure the overlap between the generated text and reference summaries. Additionally, we tracked the model's loss over epochs to monitor overfitting and generalization.

6. Results

The results indicate that while the plot summary generation model struggled with overfitting due to the small dataset size, the review summary generation model benefited from the larger dataset, achieving lower loss and better performance metrics. Detailed results are presented in Tables 2 and 3.

7. Future Work

…………………………………………………………
…………………………………………………………
…………………………………………………………
…………………………………………………………
………

Metric	Value
BLEU Score	0.0163
Brevity Penalty	0.4838
Length Ratio	0.5794
Translation Length	9799
Reference Length	16913
Precisions 1-gram	0.3294
Precisions 2-gram	0.0559
Precisions 3-gram	0.0135
Precisions 4-gram	0.0051
ROUGE-rouge1 (Precision)	0.3795
ROUGE-rouge1 (Recall)	0.2223
ROUGE-rouge1 (F-Measure)	0.2744
ROUGE-rouge2 (Precision)	0.0785
ROUGE-rouge2 (Recall)	0.0444
ROUGE-rouge2 (F-Measure)	0.0555
ROUGE-rougeL (Precision)	0.2330
ROUGE-rougeL (Recall)	0.1368
ROUGE-rougeL (F-Measure)	0.1685
ROUGE-rougeLsum (Precision)	0.2333
ROUGE-rougeLsum (Recall)	0.1370
ROUGE-rougeLsum (F-Measure)	0.1686
BERTScore Precision	0.6984
BERTScore Recall	0.6610
BERTScore F1	0.6790

Table 2. The results for *plot summary* generation using *BartForGeneration* model on multiple metrics.

8. Conclusion

Our study demonstrates the potential of using deep learning models like BART for both detecting and generating spoiler-free text. While challenges remain, particularly with smaller datasets, the results are promising and suggest that with further refinement, these models could be effectively deployed in real-world applications.

9. References

[1] R. Misra, "IMDB Spoiler Dataset," Kaggle, 2020. [Online]. Available: https://www.kaggle.com/datasets/rmisra/imdb-spoiler-dataset. [Accessed: Aug. 10, 2024].

[2] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, and L. Zettlemoyer, "BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension," in *Proc. 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2020, pp. 7871-7880. [Online]. Available: https://arxiv.org/abs/1910.13461

Metric	Value
BLEU Score	0.0219
Brevity Penalty	0.8811
Length Ratio	0.8876
Translation Length	11199
Reference Length	12617
Precisions 1-gram	0.0797
Precisions 2-gram	0.0293
Precisions 3-gram	0.0159
Precisions 4-gram	0.0103
ROUGE-rouge1 (Precision)	0.1309
ROUGE-rouge1 (Recall)	0.1167
ROUGE-rouge1 (F-Measure)	0.1110
ROUGE-rouge2 (Precision)	0.0495
ROUGE-rouge2 (Recall)	0.0433
ROUGE-rouge2 (F-Measure)	0.0415
ROUGE-rougeL (Precision)	0.1260
ROUGE-rougeL (Recall)	0.1124
ROUGE-rougeL (F-Measure)	0.1067
ROUGE-rougeLsum (Precision)	0.1257
ROUGE-rougeLsum (Recall)	0.1125
ROUGE-rougeLsum (F-Measure)	0.1066
BERTScore Precision	0.6686
BERTScore Recall	0.6591
BERTScore F1	0.6630

Table 3. The results for *review summary* generation using *BartForGeneration* on multiple metrics. (after 1 epoch)

[3] "Bart — transformers 4.25.1 documentation," Hugging Face. [Online]. Available: https://huggingface.co/docs/transformers/en/model_doc/bart . [Accessed: Aug. 10, 2024].