Deep Reinforced Self-Attention Masks for Abstractive Summarization (DR.SAS)

Ankit Chadha and Mohamed Masoud

Introduction

Our DR.SAS model tackles the abstractive summarization problem based on the CNN/DM dataset [1]. We have tested the limits of learning fine-grained attention in Transfomers [3] to improve the summarization quality. UniLM applies attention to the entire token space in a global fashion. We propose DR.SAS which applies the Actor Critic (AC) algorithm [2] to learn a dynamic self-attention distribution over the tokens to reduce redundancy and generate concise and coherent summaries. Our codebase will be publicly available on our github.

After performing hyperparameter tuning, we achieved better ROUGE results compared to the baseline. Our model tends to be more extractive yet coherent in details because of optimization over ROUGE rewards.

Model

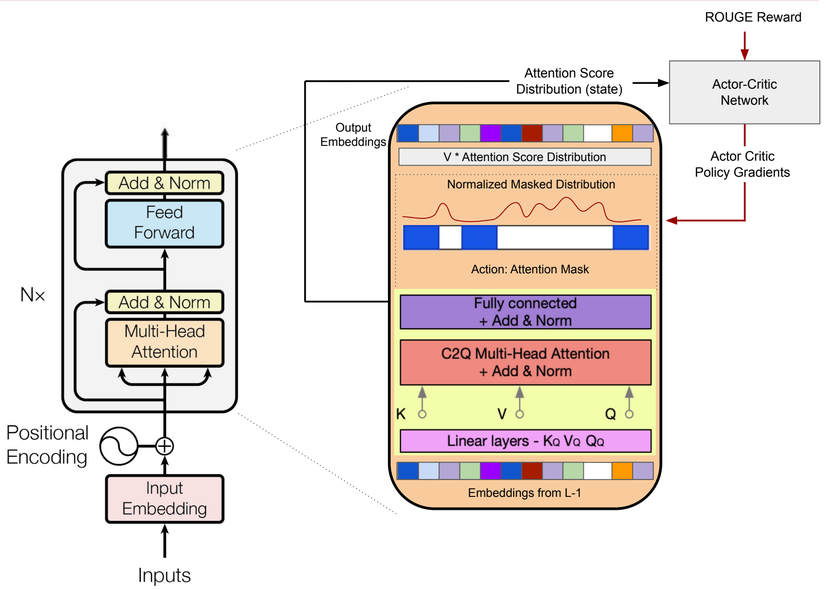

Fig 1: Encoder with AC Policy Learning for Self-Attention Masks

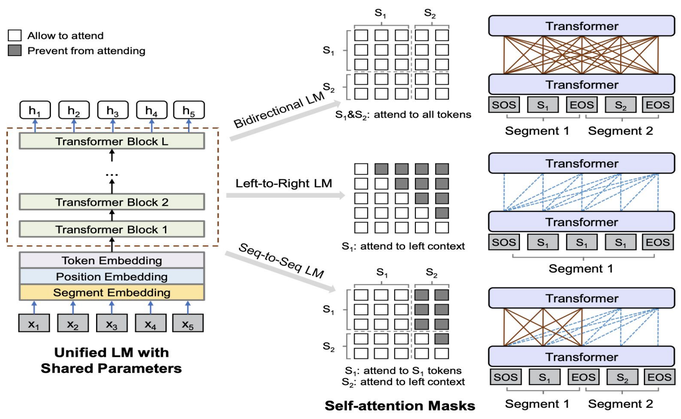

Fig 2: UniLM Architecture using task-specific Self-Attention Masks

Method

- We apply Reinforcement Learning to generate Self-Attention masks based on the attention details for coherent summarization
- We use the Actor Critic (AC) Algorithm which provides less noisy gradients as compared to Vanilla Policy Gradients
- Action: Mask particular token depending on state of attention distribution
- The model uses the ROUGE metric for evaluation and more importantly as a reward indicator for better summaries
- ROUGE stands for Recall-Oriented Understudy for Gisting Evaluation. It is a set of metrics for evaluating automatic summarization of texts as well as machine translation. It works by n-gram comparison between a generated and and reference summary

Results

Table 1: ROUGE Metrics for Baseline, Fine-tuned UniLM and Our Method

Metrics (F1 score)	Baseline	Fine-tuned UniLM	AC + Fine-tuned UniLM
ROUGE-1	21.0	40.79	**41.89**
ROUGE-2	6.5	19.01	**19.22**
ROUGE-L	19.2	38.10	**39.28**

Table 1: Results for TextRank baseline and Fine-Tuned UniLM compared to actor-critic policy gradient additional self-attention mask model. The model has marginally better results. The marginal improvement can be attributed to the bias of the reinforced actor-critic policy gradient towards better ROUGE (AC - Reward)

Conclusion

We present a novel NLP and RL architectural scheme to help learning dynamic self-attention masks which has the potential to improve over UniLM. We also learned and experimented with various choices of state space representation to effectively learn an objective that would generalize for the task of abstractive summarization. These learnings can be extended to other natural language tasks. As future work, we would like to experiment with layer specific self-attention to better understand the granularity of features most effective for summarization.

Some challenges include:

- Encoding embeddings or the meaning of tokens in state information would be a more precise way to take the methodology forward
- A precise state space could lead to a very high number of parameters
- UniLM + AC is compute heavy and requires a lot of GPU resources to converge and generalize
- The reinforced model utilizes ROUGE metric as a reward. ROUGE is biased towards more overlap between the reference summary and the model generated summary. That may add bias towards extracting more information from the original article (not necessarily essential points)

We would like to thank the CS221 Team for all the support and guidance throughout the course and Cloud access.

References

[1] Dong et. al. - NeurIPS 2019, Unified Language Model Pre-training for Natural Language Understanding and Generation
[2] Sutton, McAllester, Singh, Mansour (1999). Policy gradient methods for reinforcement learning with function approximation: actor-critic algorithms with value function approximation
[3] Vaswani et. al. - NIPS17, Attention is all you need

Discussion

Article: (CNN) Where do you go from here ? [SEP_0] The fourth season of " Game of Thrones " saw massive battles , major deaths (Tywin !) and White Walkers , but what can fans expect Sunday as we head into a fifth season of one of the most popular shows in HBO history ? [SEP_1] It ' s the most high - profile premiere yet , airing simultaneously in 170 countries for the first time . (HBO is a Time Warner company , like CNN .) . [SEP_2] We sought out " Thrones " aficionado Doug Gross , a writer for Nerdwallet and a former CNN employee , who had a few thoughts on the matter (beware , TV fans , he has read the books) . [SEP_3] " We ' re going to start seeing some of the show ' s major story arcs coming together , " Gross said (as confirmed by the executive producers) .

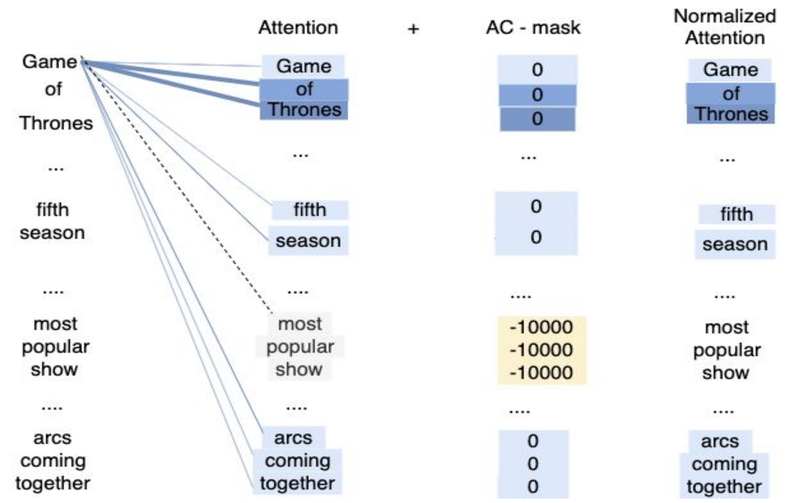

Fig 3: Visualization of Attention Score Distribution with AC Learning

Summary (UniLM): The fifth season of " Game of Thrones " will air Sunday . [X_SEP] It ' s the most high - profile premiere yet , airing simultaneously in 170 countries for the first time .

Summary (UniLM + AC): Game of Thrones " returns for fifth season on Sunday [X_SEP] Doug Gross , a writer for Nerdwallet and a former CNN employee , says he will see some of the show ' s major story arcs coming together.

Observations:

- Both UniLM and UniLM + AC produce abstractive summarization of the article by extracting the main points of the article, rewording and connecting the point in coherent way.
- The UniLM + AC method's inclination to increase the ROUGE (reward) tends to be add extractive features in the form of details to the output summary. The n-gram overlap helps increases ROUGE
- The bias towards a higher ROUGE score may increase the likelihood that the proposed method extracts more potentially less essential points depending on the application.

Video Description - https://drive.google.com/file/d/1bfvUoETNP2i0LKmhHjzXtVKTjmLqh0Yl/view?usp=sharing