UC SANTA BARBARA
Early Research Scholars Program

Precision-Driven Sentence Filtering for Long Text Summarization

Alex Mei, Anisha Kabir, Rukmini Bapat, John Judge, Tony Sun, William Wang

Introduction

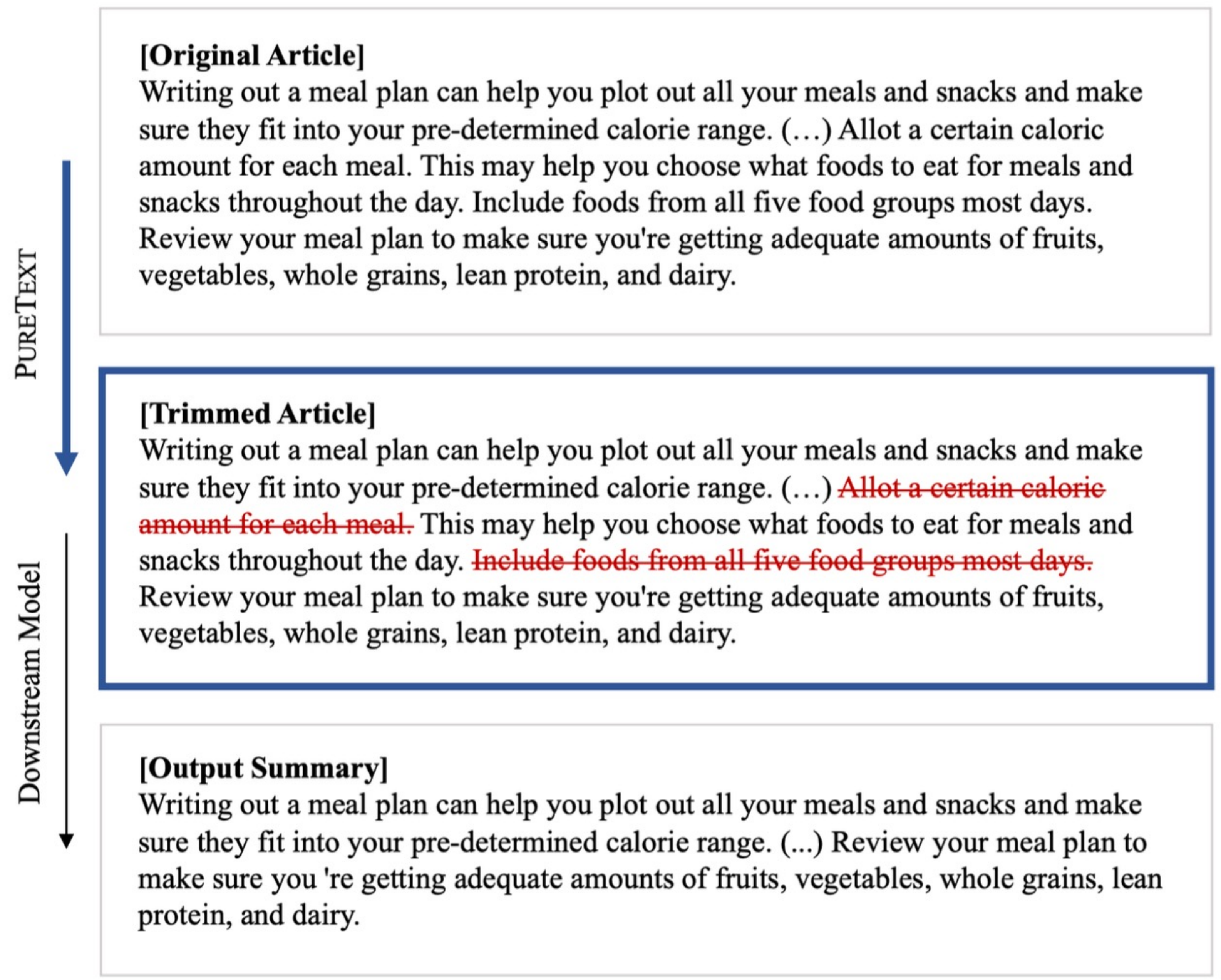
- Difficult to summarize long texts due to model input limitations
- Default truncation can result in incomprehensive summaries

Research Goals

- Improve performance and quality of long text summarization
- Explore dataset- and model-agnostic approaches to text summarization

Methodology

- PURETEXT is a lightweight layer for selecting high-quality sentences
- Fine-tuned a BERT-based model to classify sentences as either important or unimportant using a sentence's ROUGE score



[Original Article]
Writing out a meal plan can help you plot out all your meals and snacks and make sure they fit into your pre-determined calorie range. (…) Allot a certain caloric amount for each meal. This may help you choose what foods to eat for meals and snacks throughout the day. Include foods from all five food groups most days. Review your meal plan to make sure you're getting adequate amounts of fruits, vegetables, whole grains, lean protein, and dairy.

[Trimmed Article]
Writing out a meal plan can help you plot out all your meals and snacks and make sure they fit into your pre-determined calorie range. (…) ~~Allot a certain caloric amount for each meal.~~ This may help you choose what foods to eat for meals and snacks throughout the day. ~~Include foods from all five food groups most days.~~ Review your meal plan to make sure you're getting adequate amounts of fruits, vegetables, whole grains, lean protein, and dairy.

[Output Summary]
Writing out a meal plan can help you plot out all your meals and snacks and make sure they fit into your pre-determined calorie range. (...) Review your meal plan to make sure you 're getting adequate amounts of fruits, vegetables, whole grains, lean protein, and dairy.

Background

- Weakly-Supervised: a supervised task using non-human annotated labels.
- ROUGE: a recall-based summary evaluation metric that reports similarity between candidate and ideal summaries

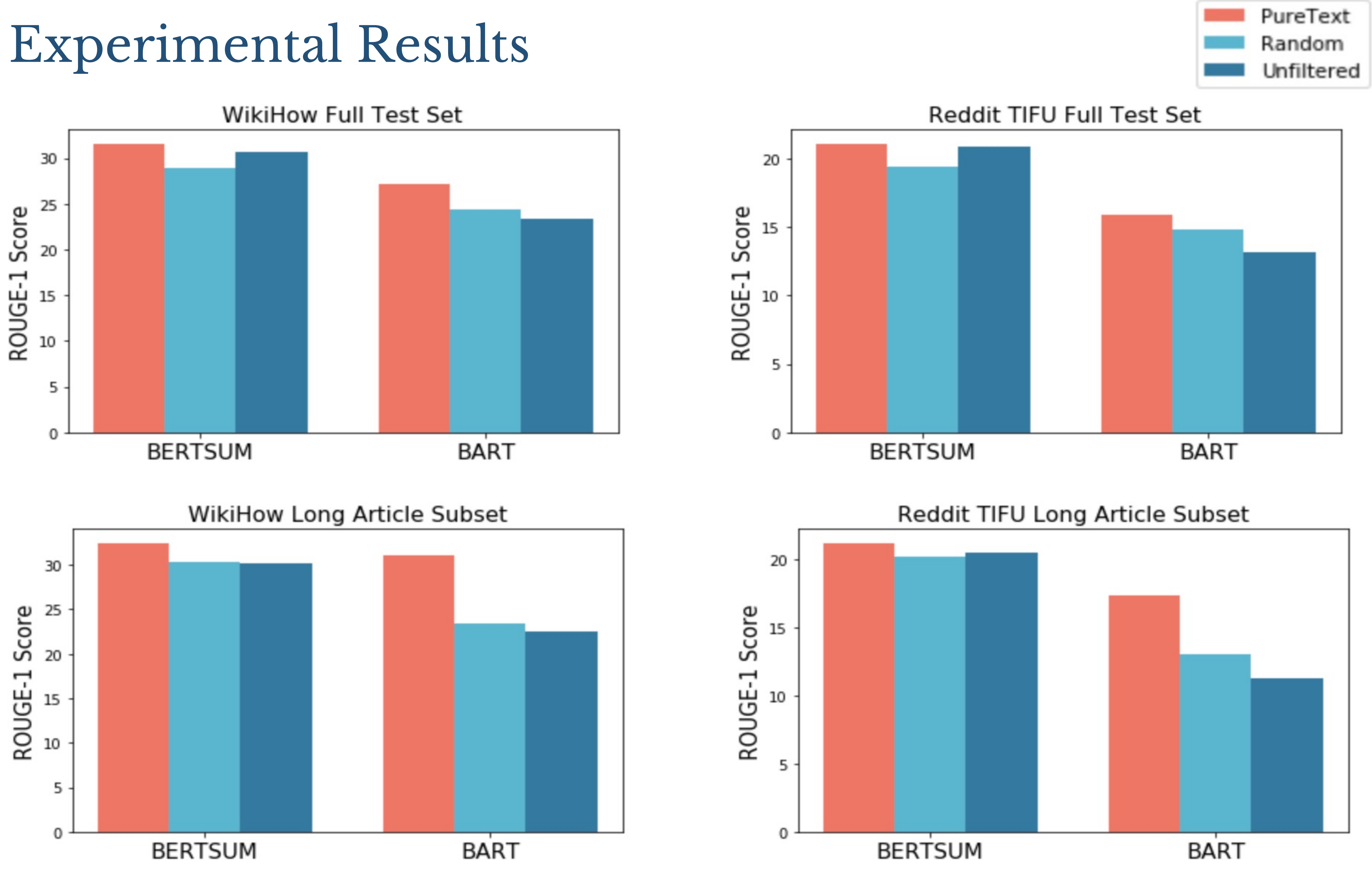
Setup

- Used datasets WikiHow and Reddit TIFU, with downstream models BERTSUM and BART
- Experimented on the full test and long article subset for each dataset
- Tested against baselines without PURETEXT: unfiltered, head, tail, head+tail, random dropping

Related Work

- Extract-and-Generate paradigms seek to identify important text components before summarizing
- Extractive-then-Abstractive methods produce summaries in a multi-step process

Experimental Results



Analysis

- Up to 0.83- and 3.84-point full dataset improvement on BERTSUM and BART respectively
- 3x improvement on long article subset over full datasets
- Statistically significant evidence ($p < .05$) PURETEXT improves long article summarization
- Particularly effective on long articles since arbitrary truncation removes important sentences
- Most applicable to datasets like WikiHow and Reddit, where key sentences are evenly distributed

Conclusion

- We utilize a BERT-based model trained with weakly-supervised learning to distinguish high-quality sentences as part of a layered-architecture approach, which are then passed to a downstream summarization model
- PURETEXT can greatly improve upon downstream model baselines for multiple datasets and models and excels at long article summarization
- We encourage future work to continue exploring the dataset- and model-agnostic nature of such a sentence filtering approach

Acknowledgements

We would also like to thank Diba Mirza, William Eiers, Aarti Jivrajani, Yoni Halpern and the Early Research Scholars Program for their continued mentorship