

Abstractive Text Summarization Using Transformer-Based Models: A Comparative Study

Sakshi Pujari, Isha Ruplag, Mahesh Vanjire, Riya Patankar Sanjay Bhokare
Group of Institute Miraj

Abstract- Automatic text summarization aims to condense lengthy documents into concise summaries while preserving key information. With the rise of transformer-based architectures, abstractive summarization has made significant advances. This paper presents a comparative study of popular models such as BART, T5, and PEGASUS on standard datasets like CNN/DailyMail and XSum. We evaluate their performance using ROUGE and BERTScore, and analyze their strengths and limitations.

1 INTRODUCTION

Text summarization is a vital task in natural language processing (NLP) that reduces a text to its essential information. Traditional extractive methods select sentences verbatim, while abstractive summarization generates novel phrases. Recent transformer-based models, trained on large corpora, have shown superior performance in abstractive tasks.

2 RELATED WORK

Early work focused on extractive approaches using algorithms like TextRank and LSA. The advent of encoder-decoder architectures, particularly with transformers, led to models such as BART [1], PEGASUS [2], and T5 [3], which achieved state-of-the-art results in abstractive summarization.

3 METHODOLOGY

We compare three transformer models:

- BART: A denoising autoencoder for pretraining sequence-to-sequence models.
- T5: Reformulates NLP tasks into a unified text-to-text format.

- PEGASUS: Pretrained specifically for summarization using gap-sentence generation.

4 DATASETS

CNN/DailyMail and XSum are used for evaluation. Both are widely used for summarization benchmarks.

5 EVALUATION METRICS

We use ROUGE-1, ROUGE-2, and ROUGE-L for lexical overlap, and BERTScore for semantic similarity.

6 EXPERIMENTS

Each model was fine-tuned using Hugging Face's Transformers library. Training was conducted on NVIDIA GPUs for 3 epochs with batch size 8. Summaries were generated using beam search (beam size = 4).

7 RESULTS

Model	ROUGE-1	ROUGE-2	BERTScore
BART	44.16	21.28	0.85
T5	42.53	19.84	0.83
PEGASUS	45.10	22.06	0.86

Table 1: Performance comparison on CNN/DailyMail

8 CONCLUSION

PEGASUS slightly outperforms other models in ROUGE and BERTScore. However, each model has unique strengths in handling different input types. Future work will explore multilingual summarization and factual consistency improvements.

ACKNOWLEDGEMENTS

We thank Hugging Face and Google for providing pretrained models and datasets.

REFERENCE

- [1] Lewis, M., et al. (2019). BART: Denoising sequence-to-sequence pre-training for natural language generation.
- [2] Zhang, J., et al. (2020). PEGASUS: Pre-training with extracted gap- sentences for abstractive summarization.
- [3] Raffel, C., et al. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer.