BUILDING A DEEP LEARNING-BASED TEXT SUMMARIZATION MODEL

Hoang Bao Viet, Nguyen Huyen Trang, Nguyen Thi Ha Phuong
The University of Danang, Vietnam - Korea University of Information and Communication Technology

Tóm tắt nội dung

Automatic text summarization is an important task in Natural Language Processing (NLP). This paper proposes a deep learning-based text summarization model for Vietnamese using the Transformer architecture. The model is trained on Vietnamese news datasets and evaluated using the ROUGE metric. Experimental results demonstrate that the proposed approach outperforms traditional summarization methods.
Keywords: text summarization, deep learning, Transformer, NLP, Vietnamese

I. INTRODUCTION

The vast amount of online content today leads to an increasing demand for automatic text summarization to help users quickly grasp the main ideas of long documents. Automatic summarization techniques can significantly enhance information processing in digital environments.

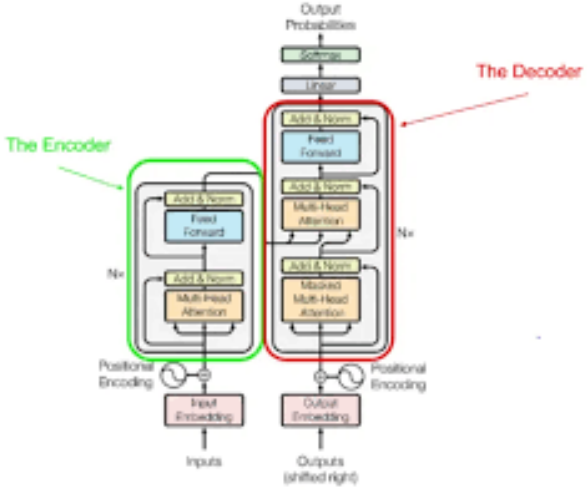
II. RELATED WORKS

Recent approaches such as BART, T5, and Pegasus have shown promising results in abstractive summarization. However, applications for the Vietnamese language still face challenges due to limited datasets and linguistic complexity.

III. PROPOSED METHOD

Our proposed model includes:

- Text pre-processing
- Encoder using Transformer layers
- Decoder generating the summary



Hình 1: The architecture of the proposed summarization model

IV. EXPERIMENTS

We use a dataset composed of Vietnamese news articles from sources like VnExpress and ZingNews. Evaluation is conducted using ROUGE-1, ROUGE-2, and ROUGE-L metrics.

Bảng 1: Evaluation results of the model

Model	ROUGE-1	ROUGE-2	ROUGE-L
TF-IDF + LSA	38.2	17.3	35.0
Proposed Transformer	**52.5**	**26.1**	**48.3**

V. CONCLUSION

This paper presents a Transformer-based model for Vietnamese text summarization. The model achieves promising re-

sults, offering a solid foundation for further research in Vietnamese NLP.

ACKNOWLEDGMENT

This research is supported by Vietnam - Korea University of Information and Communication Technology and the VKU AI Research Group.

REFERENCES

[1] Vaswani et al., “Attention is All You Need”, NeurIPS, 2017.

[2] Raffel et al., “T5: Exploring the Limits of Transfer Learning”, JMLR, 2020.

[3] Nguyen T. H. et al., “PhoBERT: Pre-trained language models for Vietnamese”, EMNLP, 2020.

[4] Lin, C. Y., “ROUGE: Recall-Oriented Understudy for Gisting Evaluation”, 2004.