Text Summarization with Pre-trained Language Models

Hoa Thi Nguyet
Hung Yen University of Technology and Education
hoathinguyet10082003@gmail.com

Abstract

This paper investigates the application of two pre-trained language models, BARTpho and ViT5, for text summarization tasks. It develops an effective framework that condenses long texts into concise summaries. The experiments, conducted across multiple datasets, demonstrate that these models can capture essential information and present it succinctly. The results indicate significant improvements in accuracy and consistency compared to traditional methods, as measured by ROUGE scores. The main conclusion is that BARTpho and ViT5, when fine-tuned, provide robust tools for generating high-quality text summaries, offering substantial benefits for information retrieval and content management systems.

Keywords: Text summarization

1 Introduction

Automatic text summarization is a method of condensing a large amount of information into a brief, concise summary that selects important information and omits redundant details(Gehrmann et al., 2018) . The summary result ensures that important information is retained, preserving the original text's meaning and spelling without distorting its content. Text summarization has always been a challenging task for Natural Language Processing (NLP). Research approaches in this field have evolved through three main techniques: simple sentences highlighting techniques (since the 1950s), classical machine learning (1990), and deep learning (2015) . The output of a text summarization system benefits many NLP applications, such as web search. For instance, Google's search engine often returns a short description of web pages corresponding to search queries, or online news providers offer highlights of web documents on their interfaces. This requires high-quality text summarization systems.

The text summarization problem typically employs two main method: Extractive Summarization and Abstractive Summarization (Sharma and Sharma, 2022). Besides these two techniques, deep learning is also a new method. Deep learning aims to find high-level abstract data models using a set of algorithms with multiple layers of processing and complex structures.

Recent successes of pre-trained language models offer a new approach to the text summarization problem. Although extractive and abstractive summarization have been applied to English with promising results, but the application of these techniques for Vietnamese text summarization remain limited. In this paper, we introduce a text summarization model using abstractive approaches based on BARTpho and ViT5. The main contributions of this paper are as follows:

- The paper utilizes a pre-trained language model for automatic text summarization, collects and utilizes a Vietnamese dataset, and trains two models BARTpho and ViT5.

- The paper compares the model's results of BARTpho and ViT5 models. The results show that the BARTpho and ViT5 summarization model yields promising outcome of pre-trained language models in processing the Vietnamese language

The remainder of this paper is organized as follows. Section 2 provides methods. Section 3 introduces the model. Next, Section 4 presents the setting and performance evaluation. Section 5 results and discussion, and finally section 6 conclusion.

The purpose of this study is not only to efficiently summarize text but also to collect and utilize datasets and to train two models, BARTpho and ViT5. By comparing the performance of these models with others, the research aims to provide deeper insights into the capabilities and challenges

of pre-trained language models in processing the Vietnamese language, by advancing the development of natural language processing technology for this language.

2   Related Work

In abstractive summarization(Sharma and Sharma, 2022), several studies have demonstrated innovative approaches to enhance text reduction efficacy. A notable example from 2018 involves the work by Gehrmann and colleagues(Gehrmann et al., 2018), developed a unique method incorporating a bottom-up content selector. This strategy was designed to identify key phrases from the source documents essential for inclusion in the summary. Following the selection process, a copying mechanism was selectively applied to these pre-chosen phrases during the decoding phase. The application of these techniques led to notable improvements in ROUGE scores across several well-established summarization datasets, showcasing the potential of their method in refining the summarization process.

In recent years, the use of pre-trained language models has significantly improved the performance of language generation tasks (Liu and Lapata, 2019) developed a Transformer-based encoder-decoder model to enable the use of pre-trained language models like BERT for abstractive summarization. Here, the authors proposed a novel document BERT-based encoder (BERTSum) and a general framework for extractive and abstractive summarization tasks. Building on BERTSum,(Dou et al., 2022) introduced GSum, which utilized various guidance signals to generate more accurate summaries. GSum achieved state-of-the-art performance on four popular English summarization datasets.

Meanwhile, there are few studies on Vietnamese text summarization, with most focusing on extractive methods compared various extractive approaches, including unsupervised ranking (e.g., LexRank, LSA), supervised learning with TF-IDF and classifiers (e.g., SVM, AdaBoosst), and deep learning methods. Similarly, (Nguyen et al., 2019) evaluated extractive methods on their publicly released dataset as a benchmark for future research. Recently, the BARTpho model (Tran et al., 2021)has marked a significant advancement in Vietnamese text summarization. This model, a specialized adaptation of the broader BART frame-

work (Lewis et al., 2019), serves as a substantial example of how large pre-trained sequence-to-sequence models can be effectively utilized in language-specific applications. The BARTpho model's performance in Vietnamese abstractive summarization tasks highlights its capacity to handle complex linguistic nuances and generate coherent summaries that are both concise and contextually accurate. Despite the successes of the BARTpho model, ongoing research continues to explore alternative architectures that might offer improvements. Notably, studies (Phan et al., 2021) illustrate the potential advantages of the T5 architecture, suggesting that it might surpass BARTpho in certain scenarios. These findings are part of a broader conversation in the NLP community regarding optimizating language models for specific languages, which is crucial given the unique challenges posed by different linguistic structures.

3   Method

3.1   *BARTpho*

BARTpho (Tran et al., 2021) is a version of the BART (Bidirectional and Auto-Regressive Transformers) and is the first public large-scale monolingual sequence-to-sequence model pre-trained for Vietnamese. BARTpho uses the "large" architecture and the pre-training scheme of the sequence-to-sequence denoising autoencoder BART, thus it is especially suitable for generative NLP tasks. BARTpho employs the encoder-decoder Transformer architecture for its pre-training phase.

The pre-training of BARTpho involves a two-stage process similar to that of BART. The first stage consists in corrupting the input text using a noising function, and the second stage focuses on learning to reconstruct the original text. This encoder-decoder learning process allows BARTpho to operate in a left-to-right manner, which helps in effectively reconstructing the original text from encoded representations. BARTpho integrates the strengths of both BERT, with its bidirectional encoder, and GPT, recognized for its left-to-right decoder, positioning it as a generalized model that enhances text processing capabilities. The model's comprehensive pre-training and fine-tuning strategies enable it to decode complex linguistic patterns, making it highly effective for various NLP tasks including text summarization, question answering, and language generation in Vietnamese.

3.2 *ViT5*

ViT5 (Phan et al., 2022)represents a specialized adaptation of the renowned T5 model, meticulously engineered to cater specifically to the Vietnamese language. This adaptation goes beyond simple translation or language support, it encompasses a deep understanding of the linguistic structures and cultural nuances embedded within the Vietnamese language. The model employs a self-supervised pre-training methodology that mirrors the proven T5 approach, yet it distinctively utilizes a wide-ranging and high-quality corpus of Vietnamese texts. This strategic selection ensures that ViT5 is not only robust but also highly attuned to the specific characteristics of Vietnamese textual data. The pre-training of ViT5 involves a carefully curated dataset comprising literary works, official documents, web articles, and daily communication exchanges that reflect the rich diversity of the Vietnamese language. This dataset facilitates a comprehensive exposure to various linguistic elements, including idiomatic expressions, technical jargon, and colloquial language, all of which are critical for a deep learning model to effectively process and understand a language in its entirety.

4 Setting And Performance Evaluation

4.1 *Data preparation*

This study focus on collecting data from five prominent Vietnamese news websites: Dan Tri, Tien Phong, Tuoi Tre, Nhan Dan, and An Ninh Thu DoDo. The goal is to gather articles and analyze them based on the following columns:

Title: The headline of the article.

Abstract: A brief summary or abstract of the article.

Content: The full content of the article.

We utilize Python libraries such as Pandas, Numpy, BeautifulSoup, and Selenium to automate the data collection process. BeautifulSoup helps in parsing HTML content, while Selenium is used to interact with dynamic elements of the web pages. After gathering the data, we structure it into a JSON file for further analysis.

This paper used the article content as input and aim to generate the abstract or title as output, thereby enabling automatic summarization of the articles.

News Source	Number of Samples
Dan Tri	9,846
Nhan Dan	33,731
Tuoi Tre	26,151
An Ninh Thu Do	33,556
Tien Phong	33,293

Table 1: Number of samples collected from Vietnamese news websites

4.2 *Implementation Detail*

After collecting the datasets from the five different Vietnamese news sources, this paper proceeded with splitting each dataset into training, testing, and validation sets

- 80% training data
- 10% validation data
- 10% test data

Once the individual splits were completed, this paper merged the corresponding training, validation, and testing sets from all five sources to form consolidated datasets:

Table 2: Train test validation split of Dataset

	Dataset
Train	109258
Validation	13658
Test	13661

4.3 *Settings and Evaluation Metrics*

Settings: This paper trained our models using the training data, fine-tuned and selected the most effective models with the validation set, and finally, this paper assessed the model performance using the test data.

This paper experiment with many different hyperparameters to fine-tune the model:

Table 3: Model's hyperparameters setting

Settings	Models	Hyperparameters settings
Setting 1	BARTpho	batch_size=2, lr=1e-4, epoch=1
	ViT5	batch_size=2, lr=1e-4, epoch=1
Setting 2	BARTpho	batch_size=2, lr=1e-4, epoch=5
	ViT5	batch_size=2, lr=1e-4, epoch=5
Setting 3	BARTpho	batch_size=2, lr=1e-4, epoch=10
	ViT5	batch_size=2, lr=1e-4, epoch=10

Table 4: Comparison of ROUGE scores for BARTpho and ViT5 across different settings

SETTING	OUTPUT	MODELS	ROUGE-1	ROUGE-2	ROUGE-L
1	Title	BARTpho	51.55	26.73	40.14
		ViT5	53.6	29.92	42.72
	Abstract	BARTpho	57.37	28.26	38.16
		ViT5	57.43	29.58	39.05
2	Title	BARTpho	53.37	28.83	41.82
		ViT5	55.68	32.28	44.61
	Abstract	BARTpho	59.02	30.25	39.55
		ViT5	59.45	30.2	39.6
3	Title	BARTpho	54.08	29.57	42.37
		ViT5	56.32	32.23	45.54
	Abstract	BARTpho	<u>59.42</u>	<u>33.37</u>	<u>44.62</u>
		ViT5	**59.67**	**33.54**	**45.45**

Evaluation Metrics: In the context of evaluating text summarization techniques, ROUGE-N metrics serve as a fundamental tool. By utilizing N=1, 2, and L, your study encompasses a comprehensive analysis ranging from single words to bi-grams, and also incorporates the longest common subsequence, which is covered by ROUGE-L. These metrics provide a view of how well the generated summaries replicate the range and structure of the reference texts, which are considered the gold standard in content accuracy and relevance. ROUGE-1 and ROUGE-2 focus specifically on the precision of word and bi-gram overlap, respectively. This is critical because it shows not only the ability of the summarization models to capture key terms and phrases but also their effectiveness in preserving the syntactic structures that contribute to the cohesiveness of the summary.

$$\text{ROUGE-N} = \frac{\sum_{s \in S_{\text{ref}}} \sum_{\text{gram}_n \in s} \text{Count}_{\text{match}}(\text{gram}_n)}{\sum_{s \in S_{\text{ref}}} \sum_{\text{gram}_n \in s} \text{Count}(\text{gram}_n)} \quad (1)$$

"n" represents the length of the n-grams used, with Count match(gram n) indicating the maximum count of these n-grams found both in the candidate summaries and the reference texts. Conversely, Count(gram n) quantifies the total n-grams present in the reference materials. For our evaluations, we utilized the ROUGE 1.5.5 toolkit, implementing it through the Python pyrouge package. Given the variations in the lengths of the summaries and the reference texts, we opted to use the F-score to maintain a balance between precision and recall metrics.

5 Results And Discussion

The table 4 compares BARTpho and ViT5 models across three settings for title and abstract generation, using ROUGE-1, ROUGE-2, and ROUGE-L metrics.

Setting 1: ViT5 outperforms BARTpho in title generation, but their performance on abstracts is similar.

Setting 2: ViT5 performs better in both title and abstract generation across all ROUGE metrics.

Setting 3: BARTpho slightly leads in title generation, but ViT5 dominates in abstract generation.

6 Demo System

Figure 1: Text summarization

7   Conclusions

The results from our experiments highlight significant differences between these models in terms of performance and specific situational applications. The ViT5 model, with its Transformer-based architecture, demonstrated superior capability in most tested settings, particularly in summarizing content from the dataset and generating accurate titles from the dataset. ViT5's strength lies in its ability to understand and process contextual information efficiently, allowing it to produce high-quality summaries with high ROUGE-1, ROUGE-2, and ROUGE-L scores.

Limitations

The main goal of the paper is to present how these models can be used to generate effective text summaries. It will likely include the algorithms, techniques, and methods used in the summarization process.

Context Sensitivity: Both BARTpho and ViT5 models heavily rely on the context provided in the input documents. They may encounter difficulties with texts where key information is subtly implied or spread across different sections, resulting in summaries that might overlook critical nuances or misinterpret the main themes.

Domain-Specific Performance: These models are primarily trained on general datasets such as Vi-wikisum and Vietnews, which may not fully capture the linguistic and stylistic nuances of specialized or technical domains. Consequently, their performance may degrade when applied to niche fields like legal documents, medical records, or highly technical academic articles without further domain-specific training.

Language Limitations: Although BARTpho is specifically designed for Vietnamese, its training on a broad corpus might not fully capture regional dialects or the entire spectrum of colloquial expressions. This can lead to less accurate summaries when dealing with texts that heavily use regional vernacular or slang.

Update and Maintenance: The dynamic nature of language and ongoing developments in the field necessitate continual updates and retraining of models to maintain their efficacy. Both BARTpho and ViT5 require ongoing maintenance to incorporate new text data and evolving language trends, which can be resource-intensive.

References

Zi-Yi Dou, Yichong Xu, Zhe Gan, Jianfeng Wang, Shuohang Wang, Lijuan Wang, Chenguang Zhu, Pengchuan Zhang, Lu Yuan, Nanyun Peng, Zicheng Liu, and Michael Zeng. 2022. An empirical study of training end-to-end vision-and-language transformers.

Sebastian Gehrmann, Yuntian Deng, and Alexander Rush. 2018. Bottom-up abstractive summarization. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4098–4109, Brussels, Belgium. Association for Computational Linguistics.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension.

Yang Liu and Mirella Lapata. 2019. Hierarchical transformers for multi-document summarization. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5070–5081, Florence, Italy. Association for Computational Linguistics.

Van-Hau Nguyen, Thanh-Chinh Nguyen, Minh-Tien Nguyen, and Nguyen Xuan Hoai. 2019. Vnds: A vietnamese dataset for summarization. In *2019 6th NAFOSTED Conference on Information and Computer Science (NICS)*, pages 375–380. IEEE.

Long Phan, Hieu Tran, Daniel Le, Hieu Nguyen, James Anibal, Alec Peltekian, and Yanfang Ye. 2021. Cotext: Multi-task learning with code-text transformer.

Long Phan, Hieu Tran, Hieu Nguyen, and Trieu H. Trinh. 2022. Vit5: Pretrained text-to-text transformer for vietnamese language generation.

Grishma Sharma and Deepak Sharma. 2022. Automatic text summarization methods: A comprehensive review. *SN Computer Science*, 4(1):33.

Nguyen Luong Tran, Duong Minh Le, and Dat Quoc Nguyen. 2021. BARTpho: Pre-trained Sequence-to-Sequence Models for Vietnamese. *arXiv preprint*, arXiv:2109.09701.