Language-Model-based methods for Vietnamese Single-Document Extractive Summarization

1st Vo Le Cuong
Ha Noi University of Science and Technology
Viet Nam
cuong.vole@hust.edu.vn

2nd Vu Xuan Thinh
Ha Noi University of Science and Technology
Viet Nam

3rd Vu Duy Thien
Ha Noi University of Science and Technology
Viet Nam

4th Vo Sy Hung
Ha Noi University of Science and Technology
Viet Nam

5th Nguyen Xuan Viet
FPT Information System
Viet Nam

***Abstract*— Latterly, there has been a dramatic increase in the amount of text data which demands effective summarization. This paper proposes a method of using English text summarization frameworks and Vietnamese pre-trained language models for Vietnamese single-document extractive summarization. The experiments were conducted with three frameworks namely BertSumExt, MatchSum, and CoLoExt, and two pre-trained language models namely PhoBert and BartPho. Our models are evaluated on two well-known Vietnamese summarization benchmark datasets, namely Vietnews and Wikilingua, and achieved state-of-the-art results on Vietnews with a maximum ROUGE-1/2/L score are 57.15/26.23/39.76. The results on Wikilingua also show the effectiveness of our methods.**

Keywords*— *extractive text summarization, pre-trained language model, single-document

I. Introduction

In the last two decades, language models have been widely studied for language understanding and generation [1]. With the development of deep learning and self-supervised learning, many neural networks such as convolutional neural networks (CNNs), recurrent neural networks (RNNs) or graph-based neural networks (GNNs) [2], [3], [4] have been used for pre-trained language models. These models are able to learn the contextual representation of the word, which is the core objective of language model, but they still have some drawbacks. Convolutional models and recurrent models are not suitable for capturing the long-range interactions between words due to their focus on the local context of words [5] While with graph-based models, constructing a good graph structure is a challenging problem [6]. In 2017, Vaswani et al. [7] proposed Transformers architecture which opened a new stage for pre-trained language models. With non-recurrent sequence-to-sequence (seq2seq) architecture, self-attention mechanism, and other features Transformers can take advantage of parallel capabilities of high-performance computing devices [8], learn long-range dependencies of words in a sequence and make it possible to train a deeper network [7]. So far, many Transformers-based pre-trained language models have been proposed such as GPT [9], Bert [10], ROBERTa [11], Bart [12], and T5 [13] and achieved state-of-the-art results on almost all NLP tasks. Using Transformer-based pre-trained language models as the backbone in NLP tasks has become a standard procedure [8].

Despite the expeditious development of pre-trained language models in English, there are some papers that developed pre-trained language model for Vietnamese. In 2020, Nguyen and Nguyen [14] proposed PhoBert as the first public large-scale language model pre-trained for Vietnamese. PhoBert improved many state-of-the-art in Vietnamese NLP tasks such as POS tagging, Dependency parsing, Named-entity recognition, etc. Following PhoBert, Bui et al., 2020 [15] proposed viBERT and vELECTRA based on Bert and ELECTRA architectures. Recently, Nguyen et al. presented sequence-to-sequence models BARTpho based on BART [16] and Long et al. presented ViT5 based on T5 [17]. Both BARTpho and ViT5 outperformed previous works on abstractive text summarization task. The good performance of Vietnamese pre-trained language models motivates us to explore further the effectiveness of applying them for downstream NLP tasks. This study focuses on Extractive Text Summarization task.

Nowadays, the amount of textual material is growing extremely fast. Searching for information has become a time-consuming activity because of the large quantity of textual data which can include irrelevant content or noise [18]. Automatic text summarization (ATS) becomes one of the solutions to help people filter out unimportant data and save their time. ATS aims to condense the ideas from the input document into a shorter version while preserving the essential information from the original document. Text summarization has two main approaches: extractive and abstractive. With the abstractive approach, the summarizers analyze the main concepts in a document using NLP methods, then paraphrase the document in fewer words, so the generated summary can contain new words and phrases that do not appear in the source text [19]. With the extractive approach, the model chooses the most important words and sentences from a document and concatenates them to create the summary [18]. This paper focuses on extractive text summarization since it usually has higher accuracy in content and grammar because of direct extraction from the text input while computing is faster.

Recently, language model-based frameworks have been significantly developed for English text summarization tasks and achieved state-of-the-art performance on many datasets. With the extractive text summarization task, in 2019, Liu and Lapata [20] formulated it as a sentence-level sequence labeling task and proposed a general framework for text summarization. Their extractive model BertSumExt employed Bert [10] as its encoder. A special token [CLS] is inserted at the beginning of each sentence so that vectors of those [CLS] tokens from the encoder will be the representation of sentences. The representations of sentences then are fed to output layers which contain 2 inter-sentence Transformer layers [7] and a sigmoid classifier to capture document-level features for extracting summaries and scoring the sentences. In 2020, Zhong et al. [21] formulated the task as a semantic text matching problem and proposed a re-ranking framework called MATCHSUM which follows a two-stage paradigm. In the first stage, a "content selection module" chooses salient sentences to generate summary candidates. In the second stage, MATCHSUM used two BERTs with tied-weights as encoders to match input documents and summary candidates in a semantic space. Then, the output layer which is a cosine-similarity layer chooses the candidate which is closest to the input document. In 2022, An et al. [22] followed the idea of a summary-level framework and proposed a one-stage paradigm called COLO [22], their extractive model COLOEXT is a one-stage re-ranking model which combines both a summarizer to score sentences and a re-ranker to choose summary candidates. It deploys the encoder of BART as the encoder. For output layers, 3-layer MLP and a sigmoid classifier are used to implement the summarizer and re-ranker respectively. To train both the summarizer and re-ranker at the same time, candidates are sampled from changing model during training instead of sampling from fixed model.

Regarding Vietnamese, to the best of our knowledge, there are four studies on extractive text summarization. In 2016, Nguyen et al. [23] presented VSoLSCSum – a dataset for social context single document summarization. They validated their dataset on social context summarization methods and learning to rank (L2R) methods then concluded that formulating sentence selection as a L2R task benefits the summarization. In 2019, Nguyen et al. [24] were the first to publish a large benchmark dataset of single summarization called Vietnews. They evaluated several extractive and abstractive methods on the dataset. Their results showed that there were small margins among unsupervised and supervised learning methods: unsupervised method Sumbasic has the highest ROUGE-1 score (52.65) while supervised method SVR achieves the highest results in ROUGE-2 and ROUGE-L (23.67/35.02). In multi-document summarization task, Nguyen et al. [25] experimented with many extractive methods in three directions: unsupervised, supervised, and deep learning on two datasets, namely VN-MDS and ViMs. Their results showed that there was no method that obtained the best results in all cases, but L2R methods achieved very promising results in various settings. In 2021, Huy et al. [26] compared multilingual and monolingual BERT models in Vietnamese extractive multi-document summarization on VietnameseMDS dataset. The results indicated that BERT models are always better than other systems.

The above studies tend to use unsupervised methods, such as LSA, LexRank, TextRank, Luhn, KL, and Sumbasic, LexRank models. The commonly used supervised models are SVM, SVR, and RankBoost. Two deep learning models, CNN and LSTM, were tested in 3 out of 4 studies. There is one paper that used BERT-based models (PhoBERT and BERT4News) for multi-document text summarization problem and showed the potential of language model for Vietnamese extractive summarization task. It is clearly shown that no research has applied an effective way to use language model for single document summarization in Vietnamese, while in English, this approach has achieved many state-of-the-art results. This study proposes combining English text summarization frameworks with Vietnamese pre-trained language models and evaluates their effectiveness for Vietnamese extractive text summarization. Two Vietnamese pre-trained language models PhoBERT and BARTpho are respectively adopted as encoders of three well-known extractive text summarization models: BERTSUMEXT [20], MATCHSUM [21], and COLOEXT [22]. The experiments were conducted on two mainstream datasets to evaluate the effectiveness of our models: Wikilingua and Vietnews datasets. On the Vietnews dataset, our models outperformed previous methods.

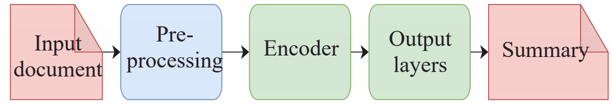

Figure 1. Overview architecture of extractive text summarization system

The contributions of this paper are:

- This is the first study to propose extractive methods that apply pre-trained language models for Vietnamese single-document summarization.
- Results on the popular Vietnews dataset show that our models outperform previous methods by a large margin.

II. METHOD

A. Overview architecture

In general, our extractive summarization system includes two modules: preprocessing module and summarizing module which contains one or two encoders and output layers as in Figure 1. The input document goes through a preprocessing module where it will be segmented into words and sentences before being fed into the summarizing module. At summarizing module, the preprocessed document is computed by the encoder to get the contextual embeddings of each sentence. Output layers calculate on them to score every sentence or summary candidate. The final summary will be based on those scores. Our models use two well-known Vietnamese pre-trained language models, specifically PhoBERT and BARTpho, to implement the encoder of the summarizing module. Following, we will give details of both modules.

B. Preprocessing module

The first module in our text summarization system is the preprocessing module. In Vietnamese writing system, a word can contain many syllables which are separated by spaces. For example, a 6-syllable text "Chúng tôi là những sinh viên" (We are students) is just a 4-word text "Chúng_tôi" (We), "là" (are), and "những", "sinh_viên" (students). Therefore, to preserve the meaning of the text, a Vietnamese document need to be segmented into words before it can be summarized. In this study, RDRSegmenter [30] from the VNCoreNLP toolkit [31] was used for word segmentation. The preprocessing module takes a raw document as input and returns a word-segmented document later will be fed into the next summarizing module.

C. Summarizing module

The summarizing module uses the output of the previous preprocessing module and returns a summary of the document. The summarizing module was experimented on three approaches: general approach, re-ranking one-stage approach and re-ranking two-stage approach. All of our models are listed in Table 1

1) General approach

With the general approach, this study uses the architecture of BERTSUMEXT which follows the architecture in *Figure*

1. PhoBERT, a Vietnamese pretrained language model and has the same architecture as BERT, is used to implement the encoder in the experiment. This model is called BERTSUMEXT- PhoBERT. The output layer's architecture of BERTSUMEXT is reused. These output layers generate sentence-level scores for each sentence, and sentences with the highest score are selected to form the final summary. This strategy could make the selected sentences share the same meaning. To overcome this drawback, the re-ranking one-stage architecture of COLOEXT was considered.

2) Re-ranking one-stage approach

A re-ranking one-stage model contains both a summarizer to score sentences and a re-ranker to choose summary candidates. In this approach, COLOEXT's architecture was used for our models. PhoBERT and encoder of BARTpho were applied to implement the Vietnamese encoder. Our models are named COLOEXT- PhoBERT and COLOEXT- BARTpho. For output layers, our models use the same architecture as COLOEXT's output layer which contains summarizer's output layers and re-ranker's output layers. Both summarizer model and re-ranker model share the same encoder, that means our pre-trained language models are fine-tuned for two tasks. In spite of benefits in training and inference efficiency, one-stage paradigm could decrease performance on accuracy. Therefore, the third approach: re-ranking two-stage paradigm was experimented.

3) Re-ranking two-stage approach

Re-ranking two-stage paradigm distinguishes between summarizer and re-ranker. At the first stage, the summarizer or "content selection module" will choose salient sentences independently to generate summary candidates. In the training phase, a salient score of each sentence is used to select and generate summary candidates for the second stage. Given the $sent_i$ which is the *i*-th sentence in the document, its salient score is:

$$score_i = R1(sent_i, G) + R2(sent_i, G) \qquad (1)$$

where *R1* and *R2* compute ROUGE F1 unigram and bigram respectively of *i*-th sentence with gold summary *G*. In testing phase, the summarizer of our COLOEXT- BARTpho model is used to generate score for each sentence to compare re-ranker's performance of our one-stage models and our two-stage models. At the second stage, we use MATCHSUM's architecture, both PhoBERT and BARTpho are applied as the encoder. These models are named MATCHSUM-PhoBERT and MATCHSUM-BARTpho. The same output layer architecture of MATCHSUM is used for the re-ranker models.

Subsequently, encoder of our models in Table 1 are further fine-tuned while output layers are trained from scratch on Vietnamese text summarization benchmark datasets, specifically Vietnews and Wikilingua, using the same training pipeline of the original models which our models are based on.

III. EXPERIMENT

A. Datasets

This study evaluated all models on two benchmark datasets, namely Vietnews [24] and Wikilingua [29]. Table 2 shows the statistics of these datasets. Vietnews [24] is a single-document abstractive summarization benchmark. It contains articles from three sources: tuoitre.vn, vnexpress.net, and nguoiduatin.vn. The dataset's authors filter out articles related to questionnaires, admissions, analytical comments, and weather forecasts which are not so important to document summarization. The default splits of Vietnews's authors for training, validation, and testing (105,418/22,642/22,644) was used. Wikilingua [29] is a large-scale, multilingual dataset for abstractive summarization systems. The dataset consists of article and summary pairs in 18 languages from WikiHow, including Vietnamese. The Vietnamese guides on WikiHow are translated from the corresponding English versions by human writers and further reviewed by WikiHow's international translation team to ensure quality. The preprocessed dataset and splits of [17] for training, validation and testing (13707/3926/1957) was used in the experiments. In Vietnews, the text is already segmented and can be separated by white spaces. With Wikilingua, RDRSegmenter [30] from the VNCoreNLP toolkit [31] was used for word segmentation. The input documents were truncated based on the maximum position embedding of each encoder.

Table 1. BASE ARCHITECTURES AND ENCODERS OF OUR PROPOSED METHODS

Name of our proposed model	Base architecture	Vietnamese language model
BERTSUMEXT- PhoBERT	BERTSUMEXT [20]	PhoBERT [14]
MATCHSUM-PhoBERT	MATCHSUM [21]	PhoBERT [14]
MATCHSUM-BARTpho	MATCHSUM [21]	BARTpho [16]
COLOEXT- PhoBERT	COLOEXT [22]	PhoBERT [14]
COLOEXT- BARTpho	COLOEXT [22]	BARTpho [16]

B. Implementation details

For BERTSUMEXT- PhoBERT, PhoBERT-base version was used to implement the encoder. The summary is created by selecting the top-2 sentences for Vietnews and top-3 sentences for Wikilingua. For models based on COLOEXT and MATCHSUM framework, both PhoBERT-base and encoder of BARTpho-word-base version of BARTpho were used to implement their encoders. In the inference process, the number of top sentences *k* to create summary candidate and number of selected sentences for a candidate *N* are set respectively to 5, {3,4,5} for Wikilingua and 5, {2,3} for Vietnews.

To the best of our knowledge, there are not any works that experiment extractive text summarization on Wikilingua and only one experiment on Vietnews which is from the authors of Vietnews. So, this study intends to compare the results of our models with the results from the authors of Vietnews. To ensure a fair comparison and evaluate the effect of oracle summary on our models, two kinds of oracle summary were generated for each document in Vietnews to train our models. The first one is called ORACLE-Sentence which is obtained by salient score of sentences used in [24]. The second one is called ORACLE-Candidate which is obtained by a greedy algorithm [27][20]. The algorithm tries to choose a candidate which maximize sum of ROUGE 1 and 2 againts the gold summary. ORACLE-Sentence is used to train PHOBERTSUMEXT, COLOEXT-BARTpho and COLOEXT-PhoBERT on Vietnews datasets, ORACLE-Candidate is used to train all our models on both Vietnews and Wikilingua datasets. The experiments use the same optimizer and learning

Table 2. DATASET OVERVIEW

Datasets	Train	# Docs Valid	Test	Avg. doc length[a] Words	Sentences	Avg. summary length[a] Words	Sentences
Vietnews	105,418	22,642	22,644	418.74	17.72	28.59	1.23
Wikiligua	13,707	3,926	1,957	416.14	24.58	34.45	5.21

[a] Avg. doc length and Avg. summary length indicate the average length of document and summary in test set.

rate schedule of the original models that our models follow. Our models are trained on single 11GB GEFORCE GTX 1080 Ti GP.

C. *Results*

ROUGE-1, ROUGE-2, and ROUGE-L [33] are used to evaluate the summarization quality of models. ROUGE-1 and 2 measure unigram and bigram overlap. ROUGE-L measures the longest common subsequence against gold summary.

Results on Vietnews are shown in Table 3. The first section includes LEAD-2[1], ORACLE-Sentence, two models which have the highest ROUGE score in [24] and our models training on ORACLE-Sentence. Our models outperform both SVR and SumBasic methods by a large margin. The second section includes ORACLE-Candidate and our models which are training on ORACLE-Candidate. It is not a surprise that models based on MATCHSUM and COLOEXT's architecture have higher scores than BERTSUMEXT-PHOBERT. As in II.C.1, BERTSUMEXT- PHOBERT does not care about the summary-level score of candidates, this drawback affects its performance compared with the re-ranking system. The results prove the effective of re-ranking strategy on Vietnamese summarization task but still fall behind oracle summary. MATCHSUM-based model's ROUGE-1 score is two point higher than COLO-based model, that means training re-ranking model independently still benefits accuracy. Additionally, ORACLE-Sentence's ROUGE score is five to six points lower than ORACLE-Candidate, the reason could be the algorithm of ORACLE-Candidate which generates summary-level score is better than the algorithm of ORACLE-Sentence which generate sentence-level score. It is observed that with better oracle summary, our models get better results.

Results on Wikilingua are shown in Table 4. The first subsection includes LEAD-3 and ORACLE-Candidate and the second subsection includes the results of our models. Because of limits in computing resources, MATCHSum-BARTpho was only trained on Wikilingua dataset. Models which are based on MATCHSum and COLOEXT's architecture still have the highest results. ROUGE-1 and ROUGE-L of MATCHSum-BARTpho are about 2 scores higher than COLOEXT-based models. Additionally, on both dataset, models using encoder of BARTpho are slightly better than ones using PhoBERT while BARTpho and PhoBERT are pre-trained on the same corpus. The reason may be the position embeddings in the BARTpho model have maximum length of 1024, this helps models using encoder of BARTpho have access to most of full input documents and therefore be able to learn the whole input document representations and sellect

Table 3. ROUGE F1 RESULTS ON THE VIETNEWS TEST SET

Model	**R-1**	**R-2**	**R-3**
LEAD-2 [24]	5.86	4.77	5.80
ORACLE-Sentence	57,28	34,69	43,11
Sumbasic [24]	<u>52.65</u>	19.13	26.32
SVR [24]	50.41	<u>23.67</u>	<u>35.02</u>
BERTSUMEXT- PHOBERT(*len*[2] = 256)	52.79	25.61	36.08
COLOEXT- PhoBERT (*len* = 256)	54.42	26.75	38.57
COLOEXT- BARTpho (*len* = 1024)	**54.84**	**26.79**	**38.69**
ORACLE-Candidate	63,24	39,83	48,18
BERTSUMEXT- PHOBERT(*len* = 256)	54.29	26.15	37.03
COLOEXT- PhoBERT (*len* = 256)	54.66	**26.90**	38.76
COLOEXT- BARTpho (*len* = 1024)	55.05	26.82	38.89
MATCHSUM-PhoBERT (*len* = 256)	57.02	26.50	38.24
MATCHSum-BARTpho(*len* = 1024)	**57.15**	26.23	**39.76**

Table 4. ROUGE F1 RESULTS ON THE WIKILINGUA TEST SET

Model	**R-1**	**R-2**	**R-L**
LEAD-3	51.68	20.32	46.29
ORACLE-Candidate	60,92	33,34	55,46
BERTSUMEXT- PHOBERT(*len* = 256)	54.14	22.51	47.82
COLOEXT- PhoBERT (*len* = 256)	54.61	22.88	48.39
COLOEXT- BARTpho (*len* = 1024)	54.90	23.39	48.83
MATCHSum-BARTpho(*len* = 1024)	**56,53**	**23,93**	**50,53**

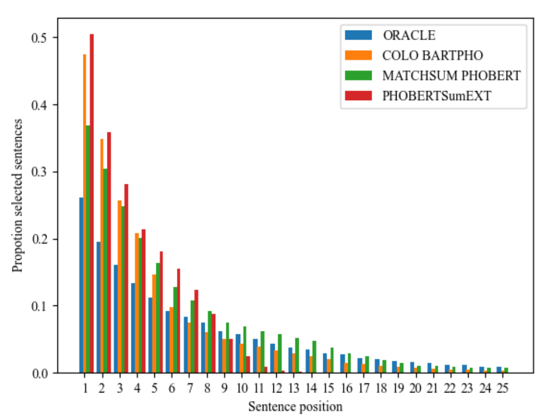

Figure 2. Proportion of selected sentences according to their position in the original document in test set of Vietnews

[1] *is common baseline in extractive text summarization task, it means selecting the first two sentences of a document to form a summary*

[2] *Indicates max input length of the document*

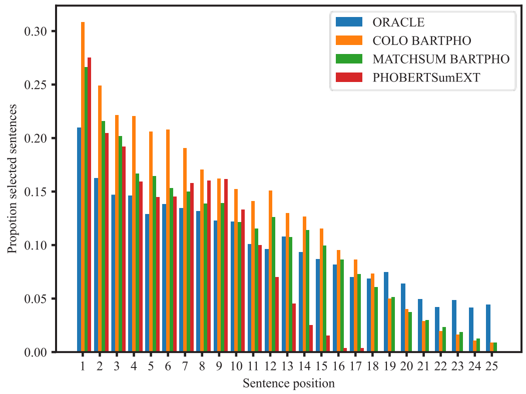

Figure 3. Proportion of selected sentences according to their position in the original document in test set of Wikilingua

better candidates.

D. Discussion

The position of selected sentences for summary is important, it can show how well our models learn the whole input document representations. Figure 2 and Figure 3 the proportion of selected sentences according to their position in the original document on test set of Vietnews and Wikilingua respectively. As it can be seen in the figures that the sentences in ORACLE summary are widely distributed across documents and the probability of being selected decreases quite smoothly when the position of sentence increases. Although our models concentrate on the first sentences of the input document, the probability of being selected still decreases smoothly when the position of sentence increases. CoLoExt- BARTpho and MatchSum-BARTpho use the same summarizer to choose summary candidates, but MatchSum-BARTpho tends to select more widely than CoLoExt- BARTpho and its behavior is the most like ORACLE-Candidate.

IV. Conclusion

This paper proposes several text summarization methods for Vietnamese documents by using pre-trained PhoBERT and BARTpho as encoders in three state-of-the-art extractive text summarization frameworks. This is the first study that fine-tunes pre-trained language model for Vietnamese extractive single document text summarization. Experimental results of two datasets show that our models achieve state-of-the-art for the Vietnamese extractive single document text summarization task on the Vietnews dataset and good result on Wikilingua dataset. In the future, we would like to explore the capability of language model for abstractive text summarization which we believe will be the main focus of interest in ATS field.

References

[1] Zhao, W.X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., Liu, P., Nie, J., & Wen, J. (2023). A Survey of Large Language Models. ArXiv, abs/2303.18223.

[2] Bin He, Di Zhou, Jinghui Xiao, Xin Jiang, Qun Liu, Nicholas Jing Yuan, and Tong Xu. 2020. Integrating Graph Contextualized Knowledge into Pre-trained Language Models. In Findings of the Association for Computational Linguistics: EMNLP 2020, Online Event, 16-20 November 2020 (Findings of ACL, Vol. EMNLP 2020). Association for Computational Linguistics, 2281–2290

[3] Tianxiang Sun, Yunfan Shao, Xipeng Qiu, Qipeng Guo, Yaru Hu, Xuanjing Huang, and Zheng Zhang. 2020a. CoLAKE: Contextualized Language and Knowledge Embedding. In Proceedings of the 28th International Conference on Computational Linguistics, COLING 2020, Barcelona, Spain (Online), December 8-13, 2020. International Committee on Computational Linguistics, 3660–3670.

[4] Wenhu Chen, Yu Su, Xifeng Yan, and William Yang Wang. 2020b. KGPT: Knowledge-Grounded Pre-Training for Data-to-Text Generation. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020. Association for Computational Linguistics, 8635–8648

[5] Qiu, X., Sun, T., Xu, Y. et al. Pre-trained models for natural language processing: A survey. Sci. China Technol. Sci. 63, 1872–1897 (2020). https://doi.org/10.1007/s11431-020-1647-3

[6] Lingfei Wu; Yu Chen; Kai Shen; Xiaojie Guo; Hanning Gao; Shucheng Li; Jian Pei; Bo Long, Graph Neural Networks for Natural Language Processing: A Survey , now, 2023.

[7] Vaswani, Ashish & Shazeer, Noam & Parmar, Niki & Uszkoreit, Jakob & Jones, Llion & Gomez, Aidan & Kaiser, Lukasz & Polosukhin, Illia. (2017). Attention Is All You Need.

[8] Xu, Han & Zhengyan, Zhang & Ding, Ning & Gu, Yuxian & Xiao, Liu & Yuqi, Huo & Jiezhong, Qiu & Liang, Zhang & Wentao, Han & Minlie, Huang & Qin, Jin & Yanyan, Lan & Yang, Liu & Zhiyuan, Liu & Lu, Zhiwu & Qiu, Xipeng & Ruihua, Song & Jie, Tang & Ji-Rong, Wen & Jun, Zhu. (2021). Pre-Trained Models: Past, Present and Future.

[9] Radford, A., Narasimhan, K., Salimans, T. & Sutskever, I. (2018). Improving language understanding by generative pre-training. , .

[10] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

[11] Liu Zhuang, Lin Wayne, Shi Ya, and Zhao Jun. 2021. A Robustly Optimized BERT Pre-training Approach with Post-training. In Proceedings of the 20th Chinese National Conference on Computational Linguistics, pages 1218–1227, Huhhot, China. Chinese Information Processing Society of China.

[12] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 7871–7880, Online. Association for Computational Linguistics.

[13] Raffel, Colin & Shazeer, Noam & Roberts, Adam & Lee, Katherine & Narang, Sharan & Matena, Michael & Zhou, Yanqi & Li, Wei & Liu, Peter. (2019). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer.

[14] Dat Quoc Nguyen and Anh Tuan Nguyen. 2020. PhoBERT: Pre-trained language models for Vietnamese. In Findings of the Association for Computational Linguistics: EMNLP 2020, pages 1037–1042, Online. Association for Computational Linguistics.

[15] The Viet Bui, Thi Oanh Tran, and Phuong Le-Hong. 2020. Improving Sequence Tagging for Vietnamese Text using Transformer-based Neural Models. In Proceedings of the 34th Pacific Asia Conference on Language, Information and Computation, pages 13–20, Hanoi, Vietnam. Association for Computational Linguistics.

[16] Tran, N.L., Le, D.M., & Nguyen, D.Q. (2021). BARTpho: Pre-trained Sequence-to-Sequence Models for Vietnamese. ArXiv, abs/2109.09701.

[17] Long Phan, Hieu Tran, Hieu Nguyen, and Trieu H. Trinh. 2022. ViT5: Pretrained Text-to-Text Transformer for Vietnamese Language Generation. In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Student Research Workshop, pages 136–142, Hybrid: Seattle, Washington + Online. Association for Computational Linguistics.

[18] M. F. Mridha, A. A. Lima, K. Nur, S. C. Das, M. Hasan and M. M. Kabir, "A Survey of Automatic Text Summarization: Progress, Process

and Challenges," in IEEE Access, vol. 9, pp. 156043-156070, 2021, doi: 10.1109/ACCESS.2021.3129786.

[19] El-Kassas, Wafaa & Salama, Cherif & Rafea, Ahmed & Mohamed, Hoda. (2020). Automatic Text Summarization: A Comprehensive Survey. Expert Systems with Applications. 165. 113679. 10.1016/j.eswa.2020.113679.

[20] Yang Liu and Mirella Lapata. 2019. Text Summarization with Pretrained Encoders. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 3730–3740, Hong Kong, China. Association for Computational Linguistics.

[21] Ming Zhong, Pengfei Liu, Yiran Chen, Danqing Wang, Xipeng Qiu, and Xuanjing Huang. 2020. Extractive Summarization as Text Matching. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 6197–6208, Online. Association for Computational Linguistics.

[22] Chenxin An, Ming Zhong, Zhiyong Wu, Qin Zhu, Xuanjing Huang, and Xipeng Qiu. 2022. CoLo: A Contrastive Learning Based Re-ranking Framework for One-Stage Summarization. In Proceedings of the 29th International Conference on Computational Linguistics, pages 5783–5793, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

[23] Minh-Tien Nguyen, Dac Viet Lai, Phong-Khac Do, Duc-Vu Tran, and Minh-Le Nguyen. 2016. VSoLSCSum: Building a Vietnamese Sentence-Comment Dataset for Social Context Summarization. In Proceedings of the 12th Workshop on Asian Language Resources (ALR12), pages 38–48, Osaka, Japan. The COLING 2016 Organizing Committee.

[24] V. -H. Nguyen, T. -C. Nguyen, M. -T. Nguyen and N. X. Hoai, "VNDS: A Vietnamese Dataset for Summarization," 2019 6th NAFOSTED Conference on Information and Computer Science (NICS), Hanoi, Vietnam, 2019, pp. 375-380, doi: 10.1109/NICS48868.2019.9023886.

[25] M. -T. Nguyen, H. -D. Nguyen, T. -H. -N. Nguyen and V. -H. Nguyen, "Towards State-of-the-art Baselines for Vietnamese Multi-document Summarization," 2018 10th International Conference on Knowledge sand Systems Engineering (KSE), Ho Chi Minh City, Vietnam, 2018, pp. 85-90, doi: 10.1109/KSE.2018.8573420.

[26] Huy Quoc To, Kiet Van Nguyen, Ngan Luu-Thuy Nguyen, and Anh Gia-Tuan Nguyen. 2021. Monolingual versus multilingual BERTology for Vietnamese extractive multi-document summarization. In Proceedings of the 35th Pacific Asia Conference on Language, Information and Computation, pages 692–699, Shanghai, China. Association for Computational Lingustics.

[27] Nallapati, Ramesh & Zhai, Feifei & Zhou, Bowen. (2016). SummaRuNNer: A Recurrent Neural Network Based Sequence Model for Extractive Summarization of Documents. Proceedings of the AAAI Conference on Artificial Intelligence. 31. 10.1609/aaai.v31i1.10958.

[28] Jiacheng Xu, Zhe Gan, Yu Cheng, and Jingjing Liu. 2020. Discourse-Aware Neural Extractive Text Summarization. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 5021–5031, Online. Association for Computational Linguistics.

[29] Ladhak, Faisal & Durmus, Esin & Cardie, Claire & McKeown, Kathleen. (2020). WikiLingua: A New Benchmark Dataset for Cross-Lingual Abstractive Summarization. 4034-4048. 10.18653/v1/2020.findings-emnlp.360.

[30] Dat Quoc Nguyen, Dai Quoc Nguyen, Thanh Vu, Mark Dras, and Mark Johnson. 2018. A Fast and Accurate Vietnamese Word Segmenter. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan. European Language Resources Association (ELRA).

[31] Thanh Vu, Dat Quoc Nguyen, Dai Quoc Nguyen, Mark Dras, and Mark Johnson. 2018. VnCoreNLP: A Vietnamese Natural Language Processing Toolkit. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations, pages 56–60, New Orleans, Louisiana. Association for Computational Linguistics.

[32] Paulus, Romain & Xiong, Caiming & Socher, Richard. (2017). A Deep Reinforced Model for Abstractive Summarization.

[33] Chin-Yew Lin. 2004. ROUGE: A Package for Automatic Evaluation of Summaries. In Text Summarization Branches Out, pages 74–81, Barcelona, Spain. Association for Computational Linguistic.