TA-Sum: The Extractive Summarization Research Based on Topic Information

Fang Xie[1], Hao Li[1], Beiye Zhang[1], Jianan He[2], Xincong Zhong[3]*

[1] *School of Computer Science, Hubei University of Technology, China*
[2] *Shanghai Jiurong Information Technology Co., Ltd, China*
[3] *Southern Marine Science and Engineering Guangdong Laboratory, China*
thanks_xf@hotmail.com, lihbut@foxmail.com
zzby1366@163.com, he.ja@jiuirongapp.com, xczhong1997@gmail.com

Abstract

Text summarization is divided into extractive summarization and abstractive summarization. The extractive summarization technology aims to extract some main phrases and sentences from the original text to form a short summary for people to read quickly. However, extractive summarization is faced with problems such as poor sentence coherence and incomplete information, which makes it difficult to screen out important sentences from the source text. DNN (Deep Neural Network) is widely used for text summarization task. This paper proposes a TA-Sum model based on the neural topic model. Introducing the topic information can help people understand the relevant main content of source text quickly. We obtain the topic information using the neural topic model and implement the attention mechanism to fuse the topic information with the text representation, which improves the semantic integrity and completeness of the summary. The experimental results on the large-scale English data sets CNN/Daily mail are improved by 0.37%, 0.11%, and 0.17% respectively compared with BertSum, which demonstrates the effectiveness of our method.

Keywords: Extractive summarization, Topic discovery, Attention mechanism, Neural network

1 Introduction

In today's era of big data, the problem of information overload prevents people from quickly obtaining the information they want. The emergence of text summarization technology effectively helps us obtain the content that the text wants to express better and quickly. Text summarization techniques include extractive summarization and abstractive summarization. Extractive summarization refers to extracting key sentences that can represent main ideas from the source document to form a short summary, and the extracted sentences are all from the source document. Abstractive summarization is a text summary technique that generates new sentences by the model itself, which allows for the generation of new words or phrases with a certain probability. However, the grammatical errors are more obvious in the generated sentences. The extractive summarization is more feasible compared with abstractive summarization, so this paper mainly studies the related aspects of extractive summarization.

Text summarization technology initially used text keywords and other information to score sentences. Statistics-based text summary technique selects sentences containing more key words to form the summary. In recent years, sentences clustering, graph sorting, and other methods have been applied in the field of extractive summarization. TextRank constructs a graph model according to the co-occurrence relationship between words, takes the convergence property of Markov chain as the theoretical basis, and uses the voting mechanism to select the summary sentence. Some researchers have used topic models to obtain deeper textual information. The application of LDA (Latent Dirichlet Allocation) topic model mines the topics of text by clustering to guide the generation of summary [1-2]. However, the semantic features exploited by these methods are relatively superficial and lack the understanding of the deep meaning and implicit information about the text.

In recent years, deep learning is gradually playing a role in various fields [3]. Some automatic summarization techniques based on deep learning have also achieved certain results. However, due to the lack of deep understanding of text information (sentence semantics, positional relationship, topic information, etc.), many models cannot achieve a good effect on content extraction. To handle these problems, this paper proposes a new extractive model with topic information. This model judges the importance of sentences by combining topic information and text representation information, uses attention fusion mechanism to guide the model to focus on the topic information, and finally filters out sentences with high scores to form summaries. The main contributions of this work are given as follows:

1) A new model, TA-Sum, is proposed for extractive summarization, improving the accuracy of the summary.
2) A method is proposed to select the key sentences based on the topic model. Clustering the token embedding in the low dimensional space can filter some unnecessary information with topic discovery, which is convenient to select better sentences in the scoring extraction model.
3) Through the attention mechanism, the topic informa-

*Corresponding Author: Xincong Zhong; E-mail: xczhong1997@gmail.com
DOI: 10.53106/160792642023122407010

tion is integrated into the extraction scoring model to distinguish the sentences with high importance and ensure the accuracy and completeness of the summary.

The TA-Sum model enjoys the following advantages: (1) it employs the pretrained model to bring linguistic knowledge to obtain more accurate and stable text representation on the CNN/Daily mail dataset; (2) TA-Sum supplements the insufficient text information in the process of generating summaries, strengthening the model's judgment on the importance of sentences and improving the accuracy of the summaries. The experimental results on the English data sets CNN/Daily mail show that the evaluation indicators ROUGE-1, ROUGE-2, and ROUGE-L have all been improved.

2 Related Work

Short information is more easily accepted by people, and text summarization technology makes it more convenient for people to read information. Extractive summarization is to extract sentences that can represent the main idea from the text according to the rules to form a summary. Therefore, judging the importance of sentences and obtaining more complete text representations have become a research hotspot.

Pretrained language models have made a great improvement in neural language processing, which is trained to learn and represent the text through large-scale data, so as to enhance the model's understanding of the input text and adapt to different downstream tasks through fine tuning. Early Skip-gram [4] and Glove [5] cannot capture the text context information completely. Skip-gram predicts context information based on the central word, but it lacks the capture of long-distance interdependencies in sentences. The ELMO model proposes a context representation method and can effectively deal with the problem of polysemy of one word [6]. The transformer model is more likely to capture the long-distance interdependencies in sentences and makes the text pay more attention to its own information, improving the ability to extract text features [7]. Jacob et al. proposed the BERT pretraining model, which adopts the encoding part of transformer, better captures the contextual information of long texts, and improves the semantic representation of texts [8]. ALBERT and RoBERTa are all finetuned based on the BERT model to better mine the text semantics. ALBERT uses bi-directional transformer to obtain text feature representations [9]. RoBERTa represents the text by using larger parameter settings and does not do any additional preprocessing or word segmentation on the inputs [10]. Yang et al. modified the input sequence and embedding of BERT to make it possible to extracting summaries [11]. Applying Bert to the text summary task can better capture the text semantics from a fine-grained aspect in the process of model construction. In order to obtain the more comprehensive text representation, we adopt the BERT to obtain the sentence embedding.

The absence of key information often leads to the inaccuracy and incompleteness of the final summary. Therefore, it is particularly necessary to mine and supplement the text information during the generation of the summary. The introduction of text information (such as keywords [12], sentence position [13], text structure [14], contextual information, and topic information) enriches the semantic representation of text and improves the effect of text summarization technology. Some research [15-16] based on neural topic model has brought changes to the field of text summarization. Gupta et al. used LSA and the TF-IDF algorithm to extract sentences [17]. NVDM encodes documents with variational posteriors in the latent topic space [18]. ETM marries LDA with word embedding to obtain the better topics [19]. Based on the above research, researchers have combined the topic model and extractive model to select sentences to form a text summary. Srikanth et al. used the BERT model to obtain sentence embedding vectors, performed dynamic clustering, and selected the sentence closest to the center of each cluster as the final summary [20]. Topic-GraphSum integrates the graph neural network with neural topic model to select sentences [21]. However, the above methods have some problems involving either incomplete text representation or lack of more guiding information. In this paper, the topic discovery is facilitated by clustering in the low-dimensional space. Then, the attention fusion mechanism is used to integrate the topic representation into the extractive model to extract important sentences. The extractive model can be guided by the topic information to focus on the more important information.

3 Proposed Method

Based on the above research, this paper proposes a TA-Sum model, which obtains the topics by using the neural topic model and uses an attention mechanism to fuse topic information with textual representation. The model structure is shown in Figure 1.

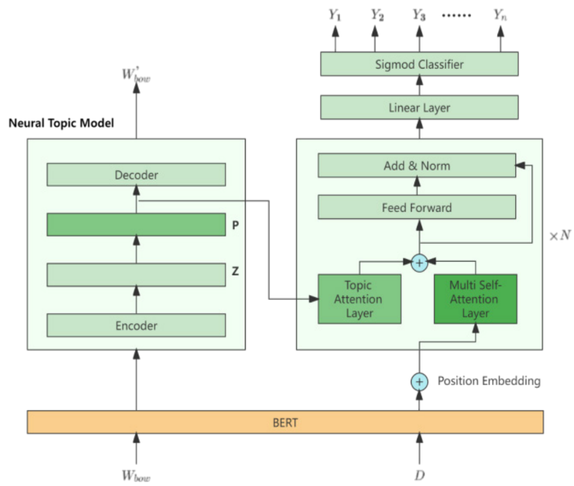

Figure 1. Overall framework of TA-Sum model
(Z represents the low-dimensional spherical space with K topic clusters, and P represents the topic-word distribution.)

This section introduces the process of data preprocessing and the discovery of topic information by the neural topic model. Then, the fusion process of attention mechanism is explained, followed by a detailed description of the extraction process by generative model.

A. Data Preprocessing

Before sending the data into the pre-training model, we first remove punctuation and stop words (do, does, an, etc.) to save storage space and improve the computational efficiency of Deep Neural Network.

After text preprocessing, BERT is fine-tuned to obtain text representation. As illustrated in Figure 2, we follow Yang et al. [11] to obtain the text embedding. [CLS] and [SEP] tags are added at the beginning and end of the sentence to distinguish sentences.

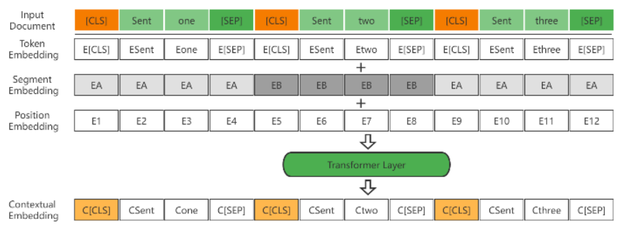

Figure 2. Overview architecture of fine-tuning Bert (Using different colors to sign the different sentences. The final C [CLS] represents the semantic information that contains its context.)

BERT represents the text through different input embedding, which mainly includes token embedding, segment embedding, and position embedding. Token embedding represents the result of vectorization on each word, while segment embedding distinguishes the different sentences. Position embedding is used to distinguish word vectors in different positions, and the calculation process is given in Equations (1) and (2).

$$PE_{(pos,2t)} = \sin(pos / 10000^{2t/d}). \tag{1}$$

$$PE_{(pos,2t+1)} = \cos(pos / 10000^{2t/d}). \tag{2}$$

Pos denotes the position index, t denotes the dimension index, and d represents the vector's dimension.

B. Neural Topic Model

The topic model aims to make better use of the semantic relationship between words and discover latent topics from word co-occurrence. In the text summary task, the text summarization model can obtain more accurate summary under the guidance of topic information. In this paper, a neural topic model is applied to the downstream tasks based on Meng et al. [22]. We obtain the topic information of CNN/Daily mail through the neural topic model and combine the extraction model to extract summary sentences. Firstly, the input bag of words is D = [W_1, W_2, W_3, …, W_n], where D is the document, W represents the words per article, and n is the total number of words in the article. We use the language model BERT to obtain the word representation of each document $h^{(w)}$ and then the original vector representation is mapped to the latent space Z (a pre-defined low-dimensional spherical space with K topic clusters, which is suited for the reduced-dimension embedding for clustering) through the mapping function. We can obtain the latent embedding Z^w in Equation (3) in the latent space. As shown in Equation (4), we then calculate the initial distribution of topic-words in the latent space by Von Mises-Fisher (VMF) [23].

$$Z^w = f(h^{(w)}). \tag{3}$$

$$p(z_i^{(w)} | t_k) = vMF_{r'}(t_k, K). \tag{4}$$

The mapping function fuses the encoding layer of the Auto-Encoder. t_k is a topic center vector, sampled from a uniform distribution over the K topics. K is a concentration parameter, and its value is set as 10. r' is the latent space dimension, and its value is set as 100. The number of topics K=200.

In order to obtain better topic-words distribution, the EM (Expectation-Maximization) algorithm is used to sharpen the topic-words distribution, which makes the distribution more uniform and accurate and highlights the relevant topic information. In the E-step, we obtain the initial distribution of topic-words. In the M-step, we update the distribution parameters continuously according to the clustering results. We firstly calculate the original posterior topic distribution p by the Bayesian rule, and then calculate the new topic distribution q for updating the model. This process is to obtain a more balanced topic-words distribution. The specific process is as follows:

$$p(t_k \mid z_i^{(w)}) = \frac{p(z_i^{(w)} | t_k) p(t_k)}{\sum_{k'=1}^{K} p(z_i^{(w)} | t_{k'}) p(t_{k'})}. \tag{5}$$

$$S_k = \sum_{k'=1}^{K} p(t_k | z_i^{(w)}). \tag{6}$$

$$q(t_k \mid z_i^{(w)}) = \frac{p(t_k | z_i^{(w)})^2 / S_k}{\sum_{k'=1}^{K} p(t_{k'} | z_i^{(w)})^2 / S_{k'}}. \tag{7}$$

N represents the total number of words, and the clustering loss is calculated as follows:

$$Loss_{clus} = -\sum_{i=1}^{N} \sum_{k=1}^{K} q \log p. \tag{8}$$

The total loss is:

$$L_{total} = \lambda Loss_{clus} + Loss_{pre} + Loss_{rec}. \tag{9}$$

$Loss_{pre}$ and $Loss_{rec}$ represent the autoencoder model loss and document reconstruction loss, respectively. The loss of the autoencoder model is obtained by calculating the difference between the output of the autoencoder and the original input. The document reconstruction loss is the difference between the reconstructed document representation and the original document representation, which aims to learn more meaningful topics. λ is the loss ratio factor, and its value is set as 0.1.

C. Attention Mechanism

The attention mechanism focuses on the information that is more critical to the task at hand, reduces attention to other information, and even filters out other irrelevant information. The attention mechanism generates the final sequence by assigning different weights to different parts of the sentence. By integrating the external information (such as topic information) into the attention mechanism, we can achieve the attention of topic information in the process of sentence extraction and improve the accuracy of the final summary.

The topic attention score is obtained by the cosine similarity between the input text and the topic vector T. Then, it is fused with the text multi attention mechanism by linear superposition, so that the model can take both topic relevance and context relevance into account when decoding. The calculation process is shown in Figure 3. The detailed calculation can be seen at the Chapter Generative Model.

T represents the topic-word distribution, and H represents the sentence embedding. Cross attention captures the semantic relationship between the topic words and sentences, and increases the attention paid to the topic words. Self-attention can better mine the complete text semantics by paying attention to its own information and contextual information.

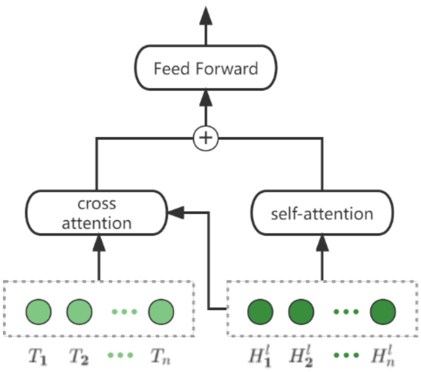

Figure 3. Attention fusion process based on topic information and text information

D. Generative Model

Yang et al. inserted a [CLS] tag before each sentence and a [SEP] tag after each sentence [11]. [CLS] token embedding will be used as sentence representation. We use segment embedding to indicate different sentences in a document. X is the word embedding with the BERT layer, and PE is the sentence position embedding, H^0 is the result of adding the two:

$$H^0 = X + PE. \tag{10}$$

$$H^l = Trans(H^0). \tag{11}$$

Trans is the computing block unit of the transformer model, and H^l is the hidden layer output after l layer computation.

After obtaining the initial input of the calculation block unit, in order to strengthen the perception of the topic information by the extraction module, the topic representation T and the intermediate text representation H are then calculated interactively with the attention mechanism. The calculation process of a single Trans is as follows:

$$H_1^l = LN(MHA(H^l, H^l, H^l) + H^l). \tag{12}$$

$$H_2^l = LN(MHA(H_1^l, T, T) + H_1^l). \tag{13}$$

Then, we accumulate the attention results, and the final output representation is obtained through the feed-forward neural network layer and the normalization layer. The calculation is as follows:

$$H_3^l = H_1^l + H_2^l. \tag{14}$$

$$H^{l+1} = LN(FFN(H_3^l) + H_3^l). \tag{15}$$

MHA is the multi-head attention layer, LN is the layer normalization operation, and FFN is the feed forward neural network.

We use a sigmoid classifier to determine the final summary sentences:

$$\hat{Y}_i = \sigma(W_o H^L + b_o). \tag{16}$$

L is the transformer layer. The layer setting of the transformer model is mainly adjusted according to the size of the data set. The experiment performs best with 2 layers of Transformer. σ is the sigmoid function, W_o is the weight matrix, and b_o is the bias.

4 Experiments

A. Data Sets

This paper uses the public large-scale English data CNN/Daily mail as the experimental data set. It contains 312,085 news articles and corresponding artificial summaries. The artificial summaries can accurately summarize the main ideas of text, which is convenient for comparing the results with the extracted summaries. We adopt the method of Hermann et al. [24] to split the data set into 287227/13368/11490 for the training set, validation set, and test set respectively and preprocess the data set following the method used by See et al. [25]. The data set allocation is listed in Table 1.

Table 1. Allocation of datasets

Data sets	Train	Validation	Test	Avg. doc	Avg. sum
CNN	90266	1220	1093	761	46
Daily mail	196961	12148	10397	653	55

B. Experimental Details

For the neural topic model, some parameter settings follow the experimental settings of Meng et al. [22]. The training steps are 20, and the batch size is 32. The Auto-Encoder neural network hidden dimensions are 500-500-1000. We choose nouns, verbs, and adjectives as the topic words. For extractive model, this model is trained by 2 GPUs (RTX 2080Ti). The size of vocabulary is 30522, the length of each text we intercept is 512, and the hidden layer of BERT is 768. The feedforward neural network's dimension is 2048. The Adam method is used to update the learning rate, and the momentum β_1 = 0.9, β_2 = 0.999. In order to solve the instability of neural network in the initial training stage, we follow Vaswani [7] by warming-up on the first 10000 steps:

$$Ir = 2e^{-3} * \min(\text{step}^{-0.5}, \text{step} * \text{warmup}^{-1.5}). \tag{17}$$

The training steps are 50000 steps. In the evaluation stage, we set the learning rate as 2e-3 to ensure the optimality of the model. Finally, the model will select top-3 sentences for summary.

C. Evaluation Indicators

This experiment uses ROUGE (Recall-Oriented Understudy for Gisting Evaluation) [26] to evaluate the proposed model's effectiveness. ROUGE is a method for evaluating summaries based on the co-occurrence information of n-grams between reference summaries and candidate summaries. We will use ROUGE-1 (Unigram), ROUGE-2 (Bi-gram) and ROUGE-L (Longest Common Sequence) to evaluate the quality of summaries.

$$Rouge-N = \frac{\sum_{S \in RefSum} \sum_{n-gram \in S} Count_{match}(gram_N)}{\sum_{S \in RefSum} \sum_{n-gram \in S} Count(gram_N)}. \tag{18}$$

Recall and Precision are other evaluation metrics of the quality of summary.

$$Recall_{LCS} = \frac{LCS(RefSum, CandiSum)}{Length(RefSum)}. \tag{19}$$

$$Precision_{LCS} = \frac{LCS(RefSum, CandiSum)}{Length(CandiSum)}. \tag{20}$$

$$F_{LCS} = (1+\theta^2)\frac{Recall_{LCS} * Precision_{LCS}}{Recall_{LCS} + Precision_{LCS}}. \tag{21}$$

For θ, we usually take θ equals 1, that is, the F1 score.

D. Results and Comparison

In order to illustrate the superiority of the model in this dataset, we select 5 benchmark models to compare the indicators.

- NeuSUM [27] integrates selection strategies into the scoring model for sentence selection. It selects one sentence at a time and scores the sentence based on the output summary and the current state.
- HER [28] proposes a two-stage method that adopts the convolutional neural network to encode the gist of document and a decision-making policy with adapted termination mechanism to form summaries.
- HIBERT [29] applies Hierarchical Bidirectional Encoder Representations from Transformers for document encoding, improving the ability of text representation.
- PNBERT [30] leverages the transferable knowledge and learning schemes to promote the performance of the model.
- BERTSUM [11] modifies the input sequence of Bert and uses transformers to mine the relationship between sentences and extract sentences.

Table 2 shows the experimental results of our model on CNN/Daily mail. Recall indicates the degree of overlapping words in the reference sentence. Precision indicates the proportion of overlapping words in the generated sentence. F_{LCS} is the final point about the comparison of reference summary and generated summary.

Table 2. Precision, recall, and F_{LCS} scores of our model on CNN/Daily mail (%)

Rouge Points	Recall	Precision	F_{LCS}
Rouge-1	58.84	36.67	43.62
Rouge-2	24.94	17.18	20.35
Rouge-L	49.10	33.46	39.80

From Table 3, comparing with BERTSUM, we can see that the scores of ROUGE-1, ROUGE-2, and ROUGE-L are improved by 0.37%, 0.11%, and 0.17%, respectively. This experiment proves that the topic information can help the summary model generate high-quality summaries.

Table 3. Comparison of the ROUGE scores of other models on CNN/Daily Mail (%)

Method	ROUGE-1	ROUGE-2	ROUGE-L
NeuSUM	41.59	19.01	37.98
HER	42.30	18.90	37.90
HIBERT	42.37	19.95	38.83
PNBERT	42.69	19.60	38.85
BERTSUM+Transformer	43.25	20.24	39.63
TA-Sum	43.62	20.35	39.80

Table 4 shows the comparison of the reference summary and the final candidate summary. We can see that the ideas expressed by the sentences are highly consistent.

Table 4. Comparison of reference summary and candidate summary by our model on CNN/Daily Mail

Example
Reference Summary
Thomas piermayr has been training with blackpool this week. Austrian defender is a free agent after leaving mls side colorado rapids. Blackpool are bottom of the championship and look set to be relegated.
Candidate Summary
Piermayr is a free agent and had been playing for colorado rapids. Blackpool are in talks to sign austria defender thomas piermayr. The 25-year-old has been training with the championship club this week and they are keen to get him on board for what is expected to be confirmed as a campaign in league one next season.

5 Conclusion

In this paper, a summary extraction model based on topic information is proposed, which introduces the text topic information and combines the topic representation with the text representation by using the attention mechanism to capture the implicit information of text. The neural topic model applies the low-dimensional latent space to mine the implicit topic information of text. This not only effectively alleviates the grabbing of some unnecessary information in the high-dimensional space, but also helps obtain more balanced cluster, so as to infer more accurate topic information. The TA-Sum model can express the more complete semantics of the input text and use the attention mechanism to focus on the topic information of the text, thus improving the accuracy of the summary extraction process. Results on the CNN/Daily mail data sets show that the model improved the ROUGE score of generated summaries. The actual summaries also show the advantages of the model.

Although this model has made some improvement in extractive summarization, there are some limitations. TA-Sum faces challenges in long text extractive summarization, and long sentences cannot be represented completely. In the future, there will be more work to improve our model's generalization ability, such as by representing the long sentences more completely and applying the model to Chinese text summarization.

Acknowledgement

This work is supported by the Key Project of Hubei Education Department under Grants No. D20201402 and D20191406, the Teaching Research Project of Hubei Education Department under Grants No. 2021295 and 2022286; the Science Start-up Foundation for High-level Talents of HBUT under Grants No. 430100391 and 337396, and the Research and Demonstration of Digital Technology for Urban Renewal and Future Community Development sponsored by China Railway Construction Corporation Limited.

References

[1] V. T. Lokare, A. W. Kiwelekar, L. D. Netak, Analysis of the Influence of Physiological and Cognition Measures in Higher Education using Machine Learning, *International Journal of Performability Engineering*, Vol. 18, No. 2, pp. 117-127, February, 2022.

[2] D. M. Blei, A. Y. Ng, M. I. Jordan, Latent dirichlet allocation, *The Journal of Machine Learning Research*, Vol. 3, pp. 993-1022, March, 2003.

[3] S. Alagarsamy, V. James, RNN LSTM-based Deep Hybrid Learning Model for Text Classification using Machine Learning Variant XGBoost, *International Journal of Performability Engineering*, Vol. 18, No. 8, pp. 545-551, August, 2022.

[4] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, N*IPS'13: Proceedings of the 26th International Conference on Neural Information Processing Systems*, Lake Tahoe Nevada, 2013, pp. 3111-3119.

[5] J. Pennington, R. Socher, C. Manning, Glove: global vectors for word representation, *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, 2014, pp. 1532-1543.

[6] M. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, L. Zettlemoyer, Deep contextualized word representations, *Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, New Orleans, Louisiana, 2018, pp. 2227-2237.

[7] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, California, USA, 2017, pp. 6000-6010.

[8] J. Devlin, M. W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional Transformers for language understanding, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Vol. 1, Minneapolis, MN, USA, 2019, pp. 4171-4186.

[9] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, R. Soricut, *ALBERT: a lite BERT for self-supervised learning of language representations*, ArXiv: 1909.11942, September, 2019. https://arxiv.org/abs/1909.11942

[10] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, *RoBERTa: a robustly optimized BERT pretraining approach*, arXiv:1907.11692, July, 2019. https://arxiv.org/abs/1907.11692

[11] Y. Liu, *Fine-tune BERT for extractive summarization*, arXiv:1903.10318, March, 2019. https://arxiv.org/abs/1903.10318

[12] F. Xie, J. Wang, R. Xiong, N. Zhang, Y. Ma, K. He, An Integrated Service Recommendation Approach for Service-based System Development, *Expert Systems with Applications*, Vol. 123, pp. 178-194, June, 2019.

[13] N. Hayatin, G. I. Marthasari, S. Anggraini, Improvement of cluster importance algorithm with sentence position for News summarization, *2018 5th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, Malang, Indonesia, 2018, pp. 483-488.

[14] R. Reztaputra, M. L. Khodra, Sentence structure-based summarization for Indonesian news articles, *2017 International Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA)*, Denpasar, Indonesia, 2017, pp. 1-6.

[15] M. A. I. Talukder, S. Abujar, A. K. M. Masum, S. Akter, S. A. Hossain, Comparative study on abstractive text summarization, *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kharagpur, India, 2020, pp. 1-4.

[16] J. L. Chen, N. O. Hembara, M. M. Hvozdyuk, Nonstationary Temperature Problem for a Cylindrical Shell with Multilayer Thin Coatings, *Materials Science*, Vol. 54, No. 3, pp. 339-348, November, 2018.

[17] H. Gupta, M. Patel, Method of text summarization using Lsa And sentence based topic modelling With Bert, *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, Coimbatore, India, 2021, pp. 511-517.

[18] Y. Miao, L. Yu, P. Blunsom, Neural variational inference for text processing, *Proceedings of The 33rd International Conference on Machine Learning*, Vol. 48, New York, NY, USA, 2016, pp. 1727-1736.

[19] A. B. Dieng, F. J. R. Ruiz, D. M. Blei, Topic modeling in embedding spaces, *Transactions of the Association for Computational Linguistics*, Vol. 8, pp. 439-453, January, 2020.

[20] A. Srikanth, A. S. Umasankar, S. Thanu, S. J. Nirmala, Extractive text summarization using dynamic clustering and co-Reference on BERT, *2020 5th International Conference on Computing, Communication and Security (ICCCS)*, Patna, India, 2020, pp. 1-5.

[21] J. L. Chen, J. Su, O. Kochan, M. Levkiv, Metrological Software Test for Simulating the Method of Determining the Thermocouple Error in Situ During Operation, *Measurement Science Review*, Vol. 18, No. 2, pp. 52-58, April, 2018.

[22] Y. Meng, Y. Zhang, J. Huang, Y. Zhang, J. Han, Topic discovery via latent space clustering of pretrained language model representations, *Proceedings of the ACM Web Conference 2022*, Virtual Event, Lyon France, 2022, pp. 3143-3152.

[23] S. Gopal, Y. Yang, Von Mises-Fisher clustering models, *PMLR: Proceedings of Machine Learning Research*, Vol. 32, No. 1, pp. 154-162, 2014.

[24] K. M. Hermann, T. Kocisky, E. Grefenstette, L. Espeholt, W. Kay, M. Suleyman, P. Blunsom, Teaching Machines to Read and Comprehend, *NIPS'15: Proceedings of the 28th International Conference on Neural Information Processing Systems*, Vol. 1, Montreal, Canada, 2015, pp. 1693-1701.

[25] A. See, P. J. Liu, C. D. Manning, Get to the point: summarization with pointer-generator networks, *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, Vancouver, Canada, 2017, pp. 1073-1083.

[26] C.-Y. Lin, ROUGE: a package for automatic evaluation of summaries, *Text Summarization Branches Out*, Barcelona, Spain, 2004, pp. 74-81.

[27] Q. Zhou, N. Yang, F. Wei, S. Huang, M. Zhou, T. Zhao, Neural document summarization by jointly learning to score and select sentences, *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, Vol. 1: Long Papers, Melbourne, Australia, 2018, pp. 654-663.

[28] L. Luo, X. Ao, Y. Song, F. Pan, M. Yang, Q. He, Reading like HER: human reading inspired extractive summarization, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, China, 2019, pp. 3031-3041.

[29] X. Zhang, F. Wei, M. Zhou, HIBERT: document level pre-training of hierarchical bidirectional transformers for document summarization, Florence, Italy, 2019, pp. 5059-5069.

[30] M. Zhong, P. Liu, D. Wang, X. Qiu, X. Huang, Searching for effective neural extractive summarization: what works and what's next, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, 2019, pp. 1049-1058.

Biographies

Fang Xie received the PhD degree in computer science and technology from Wuhan University. She is a lecture in Computer Science School, Hubei University of Technology. Her research focuses on big data processing, and service computing.

Hao Li was born in Henan, China. majoring in computer technology. His research focuses on text summarization.

Beiye Zhang was born in Hubei, China. majoring in computer technology. Her research interest is software engineering.

Jianan He is the general manager and Chief Engineer of Shanghai Jiurong Information Technology Co., Ltd. His research interests include intelligent control, unmanned equipment, and software engineering.

Xincong Zhong received the MS degree in electronics and electrical engineering from the University of Sheffield. He is currently a research assistant at China Southern Marine Science and Engineering Guangdong Laboratory (Zhanjiang). His research interests include computer vision, deep learning and underwater wireless communication.