Automatic Chinese Text Summarization for Emergency Domain

Mingshuai Liu¹, Zhongju Wang¹ and Long Wang¹*

¹School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China
²Corresponding author’s e-mail: lwang@ustb.edu.cn

Abstract. With the rapid development of the global economy, natural disasters and emergencies frequently occur. A large amount of disaster accident data and emergency case handling measures on the Internet can be used to provide technical reference and auxiliary decision-making when the various social emergency incident occurs. This study establishes an accurate Chinese text automatic short summarization model to automatically obtain summary information from accident cases. In the proposed model, Generative Pre-Training 2.0 (GPT2), which is excellent in generating tasks, is employed as the basic network structure. Adabound algorithm is used to optimize the model so that the model is not disturbed by extreme learning rates, and it converges to the global minimum at the end of training. It solves the problem that the Adam optimization algorithm causes the model to converge to the local minimum due to the extreme learning rate. Meanwhile, Jaya algorithm is utilized to optimize the hyperparameters of the Adabound for a good performance. Experimental results demonstrated that the proposed method has a significant improvement in terms of Recall-Oriented Understudy for Gisting Evaluation (ROUGE).

1. Introduction
Natural disaster and emergency accident greatly affect the development of the global economy, and they are highly random and unpredictable [1]. Therefore, disaster prevention and emergency rescue have become an urgent task in the process of economic development. With the rapid development of information technology, a large amount of disaster accident data and emergency case handling measures on the Internet can be used to provide technical reference and auxiliary decision-making when the various social emergency incident occurs. Using Chinese automatic text summarization technology to extract summaries can not only save time but also improve the efficiency of the use of key information in the text plan. This has important significance and social value for emergency management research, disaster prevention, and emergency rescue [2].

Automatic text summarization technology can be divided into extractive summaries and generative summaries according to the method of abstract generation [3]. Extractive abstract extracts keywords based on the importance of words to form abstract. However, it only considers the word frequency of the word instead of the semantic information of the sentence, which results in poor coherence of the sentence. The generated summary concludes important information of the sentence through paraphrase and synonymous substitution. Compared with the extractive summary, the generative summary has better representation capabilities, and it can understand the contextual semantics of sentences. In the task of automatic text summarization, since the input and output are both text sequences, the semantic
order and sequence coherence of the generated sequences require the model to pay more attention to the relationship between sentences.

In recent years, with the introduction of large-scale deep natural language processing models, deep neural network models can gradually meet the requirements of generative text summarization. Rush et al. [4] applied neural networks to text summary generation and introduced an attention mechanism into the model. The experiments showed that the model reached an accuracy of 31% for a single word on the DUC-2004 and Gigaword datasets. Lopyrev [5] used an encoder-decoder structure model to generate news headlines. The authors proposed a simplified version of the attention mechanism that divides the last layer of the encoder's hidden layer into two parts, and one part is used for calculation the attention weight, while the other part is used to generate the context vector. Chopra et al. [6] proposed a convolutional attention mechanism to ensure that the appropriate input can be paid attention to at each step of generating words. The model only depends on the learned features, and it is easy to train on large-scale data. Nallapati et al. [7] introduced the large vocabulary trick (LVT) technology to the text summarization problem. In this method, the decoder word vocabulary consists of a certain number of high-frequency words, which solves the problem of excessively large decoder vocabulary.

Due to the lack of large-scale and high-quality Chinese text data sets, Chinese automatic text summarization is a very difficult problem. Hu et al. [8] constructed a large-scale Chinese short text summary data set based on Sina Weibo, which provides a benchmark for future research on generative Chinese abstracts. The dataset contains 2 million real Chinese short text data and a manual reference summary for each text. The author tested the recurrent neural network on this data set to generate abstracts and got good results. Tan et al. [9] proposed a graph-based attention neural network model based on the correctness, significance and fluency of the information in the abstract. In the classic encoder-decoder model, the authors introduced a graph attention mechanism to improve the model’s ability to adapt to the saliency of sentences, which improves the novelty, information accuracy and fluency of the abstract. To improve the performance of the generative summary, Zhou et al. [10] proposed a selective coding model. The model includes a sentence encoder, a selection gate network and a decoder with an attention mechanism. Among them, the sentence encoder and decoder adopt a recurrent neural network. The selection gate network constructs an additional layer of information representation by controlling the flow of information from the encoder to the decoder. This layer representation constructs a tailored semantic representation for the sentence summary. The model was tested on three generative sentence summary datasets. The experimental results show that the performance of the selective coding model is significantly improved compared to the optimal baseline model. Celikyilmaz et al. [11] segmented the long text into many short texts. This method treats each paragraph as a shorter text and assigns an encoder to each text. At the same time, the parameters of each other are updated through the information transfer between multiple encoders to achieve more cooperation of the two encoders, which makes the model better understand long text information.

Most of the research on generative abstracts is about sequence-to-sequence encoder-decoder structure, by adding various attention mechanisms, pointer-generating mechanisms, and covering mechanisms, or using convolutional neural networks to replace cyclic neural networks to solve various problems in the abstract generation process. However, the feature extraction capabilities of convolutional neural networks or recurrent neural networks are far inferior to the Transformer model, so scholars focus more on large-scale natural language models based on the Transformer model, such as the BERT model [12] and the GPT2 model [13]. The BERT model uses the context of the word to predict the word, while the GPT2 model predicts the following through the above. Therefore, the BERT model is suitable for natural language understanding tasks, and the GPT2 model is more suitable for natural language generation tasks. Therefore, this paper proposes an automatic summary model of Chinese text based on the GPT2 model and optimizes the training process of the model step by step to get a better performance model.
2. Materials and Methods
This paper mainly studies the acquisition of summary information from a large number of disaster accident data texts on the Internet and the disposal measures of various emergency cases, establishes accurate Chinese text automatic summary generation, and studies methods to improve the performance of the system model to respond to and deal with various accidents. Provide technical reference and auxiliary decision-making for social emergencies, and improve the handling efficiency after emergencies. The specific research content is as follows:

(1) Collect online disaster accidents and emergencies data to form a summary data set of Chinese text dedicated to the field of emergency, providing a basis for the extraction of key information in the field of emergency. Use the GPT2 model as the basic model for automatic text summary generation, adopt a pre-training-fine-tuning two-stage training strategy, use a large amount of corpus of Chinese Wikipedia for pre-training, and finally enter the major disasters and unexpected accidents in the emergency field to be studied in this study. Data, fine-tuned to get the final automatic text summary model;

(2) Improve the optimizer of the model training process. The basic model uses Adam [14] (Adaptive Moment Estimation) optimization method for training, and Adam designs different adaptive learning rates for different parameters. Although the Adam optimization algorithm is widely used in the field of natural language processing, its generalization performance is too poor, and sometimes the model cannot converge due to unstable and extreme learning rates. We have improved the optimizer of the model training process to the Adabound [15] optimization algorithm, which uses a dynamic range of learning rate to achieve a gradual smooth transition from adaptive methods to SGD (Stochastic Gradient Descent), avoiding the impact of extreme learning rates. This study uses two optimizers for training, compares the summary results of the two training methods, and confirms the effectiveness of this study to replace the optimizer;

(3) Based on the Jaya evolutionary algorithm [16], optimize the hyper-parameters of the Adabound optimizer, improve the performance of the optimizer, and train a better model. This paper selects the optimizer that sets the default parameters and the optimizer that sets the parameters after evolutionary algorithm tuning respectively for training compares the parameters of the two and analyzes the results, verifying the effectiveness of the hyper-parameter tuning using an evolutionary algorithm.

The overall steps of the proposed method are shown in figure 1. First, the input of the model is the Chinese accident news report after we crawled and cleaned it. The news reports in the emergency field in this study are all Chinese text, so before inputting to the model. For the Chinese word segmentation, this research uses BERT Tokenizer as a word segmentation tool. Then, input the segmented Chinese news sequence into the GPT2 pre-training model based on the unsupervised training of the Chinese

![Figure 1. Working steps of the proposed method.](image-url)
Wikipedia data set, and then perform fine-tuning training on the emergency data set to obtain the Chinese text automatic summary model for the emergency field. By improving the optimization method of model training and optimizing the hyperparameters of the optimization method based on the Jaya evolutionary algorithm, the results of model training are improved, and a better Chinese automatic text summary model is obtained, and the corresponding emergency accident news summary is output.

3. Results & Discussion

To evaluate the pros and cons of the proposed summary model, this paper chooses ROUGE [17] (Recall-Oriented Understudy for Gisting Evaluation) as the evaluation index. According to the different optimizers of the parameters during training, this study has trained a total of three models, the GPT2 model trained using the Adam optimization method, the GPT2 model trained using the Adabound optimization method, and the GPT2 model trained by the Adabound optimization method with Jaya evolutionary algorithm.

To prove the effectiveness of using the Adabound optimizer to replace the Adam optimizer and the effectiveness of using the evolutionary algorithm to optimize the hyper-parameters of the Adabound algorithm, this paper did a comparative experiment on the three models, using the ROUGE evaluation index mentioned in the first section of this chapter. The model is evaluated on the test set, and the experimental results are as follows:

| Model               | ROUGE-1 | ROUGE-2 | ROUGE-L | ROUGE-W | ROUGE-S | Sum   |
|---------------------|---------|---------|---------|---------|---------|-------|
| Adam+GPT2           | 49.95   | 34.12   | 47.94   | 22.50   | 29.50   | 184.01|
| Adabound+GPT2       | 50.33   | 34.36   | 48.27   | 22.63   | 29.31   | 184.90|
| Proposed method     | 51.69   | 34.60   | 49.70   | 23.66   | 30.80   | 190.45|

The model in the first row of table 1 is the original model trained using Adam's algorithm. The ROUGE-1, ROUGE-2, ROUGE-L, ROUGE-W, and ROUGE-S are 49.95, 34.12, 47.94, 22.50, and 29.50 respectively. After using the Adabound algorithm to train the model, although the ROUGE-S score has been reduced by 0.19, the ROUGE-1, ROUGE-2, ROUGE-L, and ROUGE-W scores have increased by 0.38, 0.24, 0.33, and 0.13 respectively. The indicator increased by 0.9. This shows that after the Adam algorithm may start training quickly, it is affected by the extreme learning rate and enters a stagnation period, only converging to the local minimum. While the Adabound optimizer dynamically shrinks the learning rate boundary during training, it is finally transformed into the SGD algorithm. Make the model converge to the global optimal solution.

The model in the third row is a model trained using the Adabound algorithm after optimizing the hyper-parameters of Adabound using the Jaya evolutionary algorithm, which is the overall model of this study. Compared with the original model, its ROUGE score has increased by 1.75 and 0.48 respectively. 1.76, 1.66, and 1.30, compared with the improved version, increased by 1.74, 0.24, 1.43, 1.03, and 1.49 respectively. After using the Jaya evolutionary algorithm to optimize the hyper-parameters of the Adabound optimizer, the ROUGE score of the model trained by the Adabound algorithm has been comprehensively improved on the test set, which shows that the Jaya evolutionary algorithm can find the global optimal solution that maximizes the fitness function, and finds the most Better hyper-parameters enable the Adabound algorithm to function better. Therefore, it shows that the overall model we proposed helps to improve the training effect of the model and improve the performance of the model.

4. Conclusions

To form a text plan in the field of emergency and improve the response speed of disaster prevention and rescue of relevant emergency departments, this paper studies an automatic Chinese text summarization system for the field of emergency. This system can extract key information from online disaster accidents and emergency accident data, and it significantly improves the utilization of emergency text
data. Therefore, the proposed method has an important significance and social value for emergency management research, disaster prevention, and emergency rescue.

Generating tasks are generally sequence-to-sequence tasks. The relationship between the generated words and the current words to be generated is often considered. This is a relatively difficult task in the field of natural language processing. In response to this problem, this paper uses the GPT2 model that has an excellent performance in the generation task. As a basic framework of the proposed method, GPT2 guarantees the effect of text summarization. To further improve the performance of the proposed method, the Adam algorithm is employed. The model enters a stagnation period due to the extreme and unstable learning rate and finally converges to a local minimum. For this problem, this paper uses the Adabound algorithm instead of the Adam algorithm to train the model, so that the model is not subject to the extreme learning rate that interferes and eventually converges to the global minimum. At the same time, given the problem that the Adabound optimizer has a lot of hyper-parameters and the hyper-parameters have a huge impact on the results, the evolutionary algorithm is used to optimize the hyper-parameters to obtain the best hyper-parameters, which makes the Adabound algorithm better perform model training and improve the quality of the generated summary.

Acknowledgments
This work was supported by the National Key R&D Program of China under Grant 2018YFC0810601.

References
[1] Huang C, Wu T and Renn O (2016) A Risk Radar driven by Internet of intelligences serving for emergency management in community Environmental Research 148 550–9
[2] Panagiotopoulos P, Barnett J, Bigdeli A Z and Sams S (2016) Social media in emergency management: Twitter as a tool for communicating risks to the public Technological Forecasting and Social Change 111 86–96
[3] Nazari N and Mahdavi M A (2019) A survey on Automatic Text Summarization Journal of AI and Data Mining 7 121–35
[4] Rush A M, Chopra S and Weston J (2015) A Neural Attention Model for Abstractive Sentence Summarization arXiv:1509.00685 [cs]
[5] Lopyrev K (2015) Generating News Headlines with Recurrent Neural Networks arXiv:1512.01712 [cs]
[6] Chopra S, Auli M and Rush A M (2016) Abstractive Sentence Summarization with Attentive Recurrent Neural Networks Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (San Diego, California: Association for Computational Linguistics) pp 93–8
[7] Nallapati R, Zhou B, santos C N dos, Gulcehre C and Xiang B (2016) Abstractive Text Summarization Using Sequence-to-Sequence RNNs and Beyond arXiv:1602.06023 [cs]
[8] Hu B, Chen Q and Zhu F (2016) LCSTS: A Large Scale Chinese Short Text Summarization Dataset arXiv:1506.05865 [cs]
[9] Tan J, Wan X and Xiao J (2017) Abstractive Document Summarization with a Graph-Based Attentional Neural Model Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) ACL 2017 (Vancouver, Canada: Association for Computational Linguistics) pp 1171–1181
[10] Zhou Q, Yang N, Wei F and Zhou M (2017) Selective Encoding for Abstractive Sentence Summarization Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) 1095–104
[11] Celikyilmaz A, Bosselut A, He X and Choi Y (2018) Deep Communicating Agents for Abstractive Summarization arXiv:1803.10357 [cs]
[12] Devlin J, Chang M-W, Lee K and Toutanova K (2019) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding arXiv:1810.04805 [cs]
[13] Radford A, Wu J, Child R, Luan D, Amodei D and Sutskever I (2019) Language Models are Unsupervised Multitask Learners OpenAI blog 8 9
[14] Kingma D P and Ba J (2017) Adam: A Method for Stochastic Optimization arXiv:1412.6980 [cs]
[15] Luo L, Xiong Y, Liu Y and Sun X (2019) Adaptive Gradient Methods with Dynamic Bound of Learning Rate arXiv:1902.09843 [cs, stat]
[16] Wang L, Huang C and Huang L (2018) Parameter estimation of the soil water retention curve model with Jaya algorithm Computers and Electronics in Agriculture 151 349–53
[17] Lin C-Y (2004) ROUGE: A Package for Automatic Evaluation of Summaries Text Summarization Branches Out (Barcelona, Spain: Association for Computational Linguistics) pp 74–81