Abstract

Mandarin Chinese, a widely spoken language globally, has abundant, regularly updated short news texts online. Generating precise summaries of these texts is vital for effective information transmission and comprehension. This article introduces DMSeqNet-mBART, an enhanced mBART-based model, as a state-of-the-art solution for Chinese short news text summarization. This model incorporates Adaptive-DropMessage technology, a novel approach that intelligently discards or retains information based on the attention mechanism’s output. This paper demonstrates that DMSeqNet-mBART excels across several benchmarks, including BERTScore, BLEU, and ROUGE metrics, surpassing other advanced models like GPT-4, T5, and MLC. The paper outlines the Adaptive-DropMessage mechanism, enhanced dynamic convolutional layers, gated residual connections, custom feed-forward networks with batch normalization, and improvements to self-attention and cross-attention. Results from comparative experiments on six recognized Chinese short news text summary datasets indicate that the model’s performance in terms of fluency, completeness, robustness, and accuracy significantly outperforms leading industry models. The DMSeqNet-mBART’s success is attributed to its unique combination of architectural and methodological enhancements, suggesting its suitability for various complex text data processing tasks. The model provides novel insights and methods for processing similar complex text data.

Kangjie Cao, Yiya Hao, Jueqiao Huang, Yichao Gan, Ruihuan Gao, Junxu Zhu, Jinyao Wu, Weijun Cheng

Abstract—Mandarin Chinese, a widely spoken language globally, has abundant, regularly updated short news texts online. Generating precise summaries of these texts is vital for effective information transmission and comprehension. This article introduces DMSeqNet-mBART, an enhanced mBART-based model, as a state-of-the-art solution for Chinese short news text summarization. This model incorporates Adaptive-DropMessage technology, a novel approach that intelligently discards or retains information based on the attention mechanism’s output. This paper demonstrates that DMSeqNet-mBART excels across several benchmarks, including BERTScore, BLEU, and ROUGE metrics, surpassing other advanced models like GPT-4, T5, and MLC. The paper outlines the Adaptive-DropMessage mechanism, enhanced dynamic convolutional layers, gated residual connections, custom feed-forward networks with batch normalization, and improvements to self-attention and cross-attention. Results from comparative experiments on six recognized Chinese short news text summary datasets indicate that the model’s performance in terms of fluency, completeness, robustness, and accuracy significantly outperforms leading industry models. The DMSeqNet-mBART’s success is attributed to its unique combination of architectural and methodological enhancements, suggesting its suitability for various complex text data processing tasks. The model provides novel insights and methods for processing similar complex text data.


I. Introduction

To contextualize the current advancements in Chinese text generation and the significance of optimizing models specifically for the Chinese language, it is important to understand the evolution and challenges in this domain. Over the past few years, there has been a surge in interest in natural language processing (NLP) with a specific focus on Chinese, driven by its status as the most spoken language globally. This interest is not only due to the linguistic and cultural richness of Chinese but also because of the unique challenges it presents in NLP tasks. Initially, NLP research heavily focused on English due to its widespread use and the availability of large datasets. Models developed during this period were primarily based on English language structures, which posed significant challenges when applied to languages with different grammatical and syntactic structures, like Chinese. Chinese, with its tonal nature, lack of clear word boundaries, and context-dependent meanings, presents a distinct set of challenges for text generation and summarization.

Since Chinese is the most widely spoken language in the world, many well-known companies and research laboratories are conducting extensive research on Chinese text generation, facing a series of challenges mentioned above, and achieving remarkable success using the above models. Google’s T5[3], a universal model, reimagines NLP tasks into a text-to-text format, handling translation, summarization, and question-answering with a unified, end-to-end approach. The mBART-Large-CC25[8] expands on mBART for 25 languages, while Pegasus-Large-Chinese-Chuecorpusmall[2] is dedicated to Chinese text summarization. Despite these advances, there is a growing recognition that models not only need to be tuned for a specific language, but also need to be optimized for specific types of Chinese text, such as short news articles, social media posts, and academic papers. At the same time, a comprehensive grasp of Chinese context and linguistic nuances is crucial, necessitating model and data optimizations tailored to Chinese characteristics.

In the field of Chinese news text summarization, researchers have tried various methods. Xi X et al.‘s sequence-to-sequence (Seq2Seq[14]) model based on the LCSPIRT corpus demonstrated competitive outcomes[13] but had limited success beyond that corpus. Xu et al. used a Chinese text summary algorithm based on Word2vec to represent the words in the article as vectors trained by Word2vec, and combined the word-sentence relationship and the graph-based ranking model to generate the final weight to generate the summary[28]. Experiments
on real data sets. Proven better summary quality than TF-IDF and TextRank, but due to the simplicity of the method, the processing of complex Chinese texts is less detailed. Ma, T. et al.’s Topic-based[16] method, combining topic words with key sentences using TF-IDF, excelled on Sina Weibo datasets focused on natural disasters and social hotspots but lacked generalizability. Innovations like Deng, Z.’s two-stage algorithm[20], Ma et al.’s summarization framework[21], Wei et al.’s regularization approach[22], and Liu et al.’s summarization method[23] advanced Chinese text summarization yet did not focus on optimizing the model structure itself. Sun et al. proposed an improved scheme based on automatic text summarization technology. By combining the characteristics of BERTSUM (BERT-based text summarization) and BART (bidirectional and autoregressive transformer) models, the processing results of the two models were merged into mixed data. Set, and then generated a summary through the BART model to obtain the final summary result.[29]. However, they still did not go deep into the architecture of the neural network model itself for targeted optimization in Chinese, but only conducted mixed screening for the output of the model. The complexity of Chinese, such as its tonal nature, lack of clear word boundaries, context-dependent meanings, and rich homophones and polysemy, requires a nuanced text summarization method for text summary generation, which general models may not be able to do. To fully address these issues, it is therefore crucial to enhance the model architecture specifically for Chinese. Inspired by the mBART model’s summary generation and BART-IT[9] model’s success in Italian summarization, our team has optimized the mBART model for Chinese short news (less than 500 Chinese characters) text summarization, aiming to develop a robust model for this specific domain.

To tackle the intricacies of Chinese short news texts (less than 500 Chinese characters), which are typically information-dense with diverse topics, and to distill critical features from constrained data, we integrated DropMessage technology[24]—originally from graph neural networks—into the sequence-to-sequence (Seq2Seq[14]) neural network framework, marking the first application of Adaptive-DropMessage technology in the realm of short news Chinese text summarization tasks. This innovative approach has significantly bolstered the performance of our model in summarization tasks, propelling it to industry-leading status as measured by benchmarks such as BERTScore[10], fluency, and robustness. Our Adaptive-DropMessage technology stands out with its adaptive mechanism, which dynamically modulates the fraction of information discarded, based on input data characteristics, allowing for more nuanced operations that directly influence the message-passing process within the model. Consequently, the model is empowered to selectively retain messages, guided by the output of the attention mechanism and the message’s relevance, thereby managing the information flow with greater precision. This technique is particularly beneficial in short news text summarization tasks as it enhances the model’s capacity to discern subtle nuances and complex relationships within the news content, bolsters generalization ability, mitigates overfitting, and more accurately pinpoints the essence of each article to generate coherent summaries. By integrating Adaptive-DropMessage into the self-attention and cross-attention layers, we have significantly improved the model’s prowess in identifying and handling salient information, particularly in complicated sequence tasks. We further refined the dynamic convolution layers, gated residual connections, and feedforward networks within the mBART model, augmenting its performance and extendibility. The use of the nlpaug library for data augmentation enriched the diversity of our training corpus, thereby enhancing the model’s generalization capacity. Additionally, we employed the RAdam optimizer[11], CosineAnnealingLR learning rate scheduler[12], and mixed-precision training among other state-of-the-art techniques, which collectively contributed to the enhanced performance and stability of our model. In summary, the contributions of this study are threefold:

1) We specifically and innovatively apply Adaptive-DropMessage technology to the field of Chinese short news text summarization, which significantly improves the model’s ability to identify and prioritize key information in news texts, effectively solving common challenges of information overload and redundancy.

2) We customize a powerful model for the Chinese short news text summarization task by integrating specific improvements to the mBART model structure and demonstrate the contribution of each component.

3) By combining advanced neural network techniques with language-specific optimizations, our approach sheds new light on the challenges inherent in processing and summarizing complex and nuanced text data, especially in a language as complex in structure as Chinese.

II. Related Work

1) Advancements in Neural Networks for Text Generation. In recent years, the swift advancements in information technology and artificial intelligence have significantly propelled the capabilities of neural networks within the sphere of natural language processing (NLP), particularly in text generation. This era has witnessed the emergence of numerous potent models and methods. Pioneering models such as mBART[8], Pegasus[2], and GPT-3[1] have recorded substantial triumphs in tasks spanning multi-language text processing, text summarization, and dialogue systems, thereby profoundly enhancing the proficiency and efficiency of NLP practices. mBART is a distinguished pre-trained language model designed with a focus on multilingual text generation and translation. It boasts the ability to
process text in a diverse array of languages, thus offering robust support for cross-language information processing and translation. Pegasus, on the other hand, is a specialized pre-trained model dedicated to generative text summarization. It adeptly manages summarization tasks through counterfactual generation, addressing challenges associated with long texts, information distillation, and content summarization to produce precise, coherent, and substantial summaries. GPT-3 is an expansive pre-trained language model predicated on the transformative Transformers[4] architecture, and it encompasses an extensive array of text generation capabilities. With a staggering count of 175 billion parameters, it underpins a wide spectrum of text generation initiatives. The influence and value of these models in text generation primarily lie in their ability to amplify the potential and throughput of NLP. They render efficient and accurate solutions for multi-language text processing, particularly excelling in English text processing, text summarization, and dialogue systems. As a result, these cutting-edge models play a pivotal role in propelling the progression and practical implementation of text generation technologies.

2) Attention mechanism in deep learning The Attention mechanism [15] is a widely applied technique in the field of deep learning, particularly achieving significant accomplishments in natural language processing tasks. Through the Attention mechanism, the model can learn and focus on the most relevant parts of the input sequence, thereby enhancing its ability to process sequential information. Self-attention mechanism [4] is a special type of Attention mechanism, allowing the model to dynamically allocate attention weights at different positions while processing sequence data. This mechanism is particularly useful when dealing with variable-length sequences, such as sentences or documents, as the model can adaptively adjust attention based on the actual content of the input. Self-attention has been successfully applied to various NLP tasks, including machine translation, text classification, and dialogue systems.

3) Application of regularization methods Using regularization methods in self-attention networks can improve the robustness and generalization of the model. Dropout [5] is a regularization technique used in deep learning models to reduce overfitting and enhance the generalization ability of the model. In recent years, a lot of optimization research on Dropout has emerged [25][26][27], which performs better than the traditional Dropout method, but they do not fundamentally solve a problem, that is, the Dropout method usually fails every time. The entire neuron is randomly discarded in the output of the layer. It will not be selectively discarded based on the performance of the model or task requirements. The discarding method is simple and rough, which may cause some important nodes to be randomly discarded. This also means that in short news Chinese text The training process in the summary task may lead to the loss of key semantic information and semantic incoherence in the Chinese environment. Therefore, in complex tasks such as short news Chinese text summarization, a new method is needed to base the output of each layer on the message. Depending on factors such as importance, some nodes are selectively discarded, making the control of neuron activity more flexible and fine-grained, and enhancing the decision-making and generalization capabilities of the model. DropConnect [6] is another regularization technique, similar to Dropout but targets the weights of the neural network rather than the neuron’s outputs. DropConnect is applied to the weight matrices in fully connected layers or convolutional neural networks, reducing model overfitting and enhancing generalization by randomly dropping weights. During each training iteration, weights are dropped at a certain probability, making the model more robust to specific weights and reducing overfitting to noisy data. Unlike Dropout, DropConnect randomly drops weights rather than neuron outputs, making it more flexible for network architecture and applicable to more types of neural networks. The introduction of this method helps deep learning models better address overfitting during training while improving generalizability. However, a drawback is that, since DropConnect randomly drops weights in the neural network layers, it may not handle long-distance dependencies in the Chinese context well. Chinese sentences often have longer dependencies, and DropConnect may not effectively capture and maintain these long-distance contextual relationships. Another emerging regularization method is DropAttention [7], which uses attention mechanisms to arrange the attention scores of hidden representations to estimate the mutual information (MI) related to the model’s output. By randomly dropping elements in the attention weight matrix, this method, applied in the text generation field, reduces model overfitting and improves test performance. Compared to traditional Dropout, this method focuses more on element dropout in self-attention layers rather than random dropout of entire neurons. However, in the Chinese context, where semantic combinations between words are complex, the DropAttention mechanism might not capture these complex semantic information combinations well, leading to the neglect or loss of some important semantic information.

4) Model research on Chinese abstract environment In deep learning tasks facing the Chinese context, traditional random dropout methods have some shortcomings. They might cause the loss of important semantic information in Chinese text processing, affecting the model’s performance and generaliz-
ability. At the same time, although the mBART model has achieved significant success in processing multilingual tasks, current research still needs to be strengthened in adapting to the Chinese context, including the specificity of the Chinese context, such as grammatical structure, lexical expression, and semantic complexity. The adaptability improvements of the mBART model to the Chinese context remain a topic worthy of in-depth research and discussion. At the same time, many scholars have done valuable research on improving the model’s ability to generate text summaries. Hongseok Kwon and others proposed a novel generative text summarization architecture: dynamic-convolutional-encoder-decoder architectures [17], which uses gating mechanisms. By controlling and maintaining important contextual information as dynamic convolutions through encoder-decoder layers and incorporating part-of-speech information as external knowledge into the model, it better predicts the filters of dynamic convolutions. Yunsheng Shi and others proposed a normalized encoder-decoder structure [18], addressing the issue of semantic structure disruption of pre-trained word embeddings during the training process. However, directly integrating the aforementioned techniques into the mBART model might not capture the word order information well in Chinese due to the significant differences in word order between Chinese and languages like English, affecting its accuracy and effectiveness in Chinese summary generation. Moreover, the Chinese language has complex features at the word level, including homophones, ambiguities, etc. These features might make it challenging for unmodified gating mechanisms to accurately learn and process language information in Chinese summarization tasks. Additionally, the presence of a large amount of sparse data or low-frequency words in Chinese texts might affect the performance of gating mechanisms, leading to information loss or incoherence in summary generation. The aforementioned gating mechanisms, dynamic convolution layers, and attention layers have been individually proven to have unique advantages in text summarization generation. However, current research has not demonstrated whether these methods can be specifically optimized and uniformly applied to enhance the mBART model’s performance in adapting to complex Chinese contexts.

5) Conclusion and our research direction In summary, further research is needed on how to use deep learning technology to solve the problem of prioritizing key semantic information and complex syntactic structures in Chinese summary tasks. At the same time, the adaptability of the advanced mBART model in China requires more attention and exploration. Our research will focus on exploring improvements and optimizations to random dropout methods for the Chinese context and how to integrate and optimize gating mechanisms, dynamic convolution layers, and attention layers to better adapt the advanced mBART model to the Chinese context.

III. Methodology

The overall improvement method we use for the model is shown in Figure 1. Next, we will dismantle and analyze each one:

A. First application of Adaptive-DropMessage

In this study, we realize for the first time the extension of the DropMessage technique from graph neural networks to sequence-to-sequence neural network architectures, and innovatively propose the Adaptive-DropMessage method. The core objectives of the method are to reduce the sample variance, enhance the stability of the training process, and maintain the diversity of information. DropMessage is essentially a fine-grained sampling process that produces a post-perturbation matrix by generating an independent Bernoulli distribution mask for each element that determines its probability of retention or discard. Mathematically, the process can be expressed as: For each element of the matrix $M_{i,j}$, we generate an independent mask $\epsilon_{i,j}$ whose value is determined by the Bernoulli distribution:

$$\epsilon_{i,j} \sim \text{Bernoulli}(1 - \delta)$$

(1)

This distribution represents the probability of keeping or discarding. Then, by multiplying each element with its corresponding mask, we obtain the perturbed matrix $M_f$. Finally, we adjust the scaling factor by multiplying $\frac{1}{1 - \delta}$ to adjust the $M_f$, to ensure that the perturbed message matrix is equal to the original message matrix in expectation. Taken together, the whole process can be expressed mathematically as

$$M_{f_{i,j}} = \frac{1}{1 - \delta} \epsilon_{i,j} M_{i,j}$$

(2)

which $\epsilon_{i,j}$ obeys the Bernoulli distribution

$$\text{Bernoulli}(1 - \delta)$$

(3)

Inspired by the DropMessage technique from graph neural networks, we pioneered the Adaptive-DropMessage technique for sequence-to-sequence neural networks. As shown in Fig. 2, the Adaptive-DropMessage technique is a new algorithm for adaptively adjusting the dropout rate according to the training loss, which centers on its ability to dynamically adjust the dropout rate according to the input characteristics in each layer of the neural network. Unlike traditional dropout, which typically randomly discards nodes or connections at the network level, the Adaptive-DropMessage approach operates at a finer granularity, acting directly on the model’s message passing process. This means that the model is able to manage the information flow at a finer level by deciding...
whether to keep each message based on its importance. Mathematically, this process can be precisely expressed and quantified. The core mathematical expression is:

\[ \text{Dropout Rate}_{i,j} = \delta \times \text{Attn Adjustment}_i \times \text{Message Adjustment}_i \]  

(4)

Where, the \( \delta \) is the base dropout rate, this expression combines the base dropout rate with the attention weights and message importance to realize the dynamic adjustment of the dropout rate for each element. The attention weight adjustment reflects the distribution of the model’s attention on different inputs, while the message importance adjustment is based on how much importance the model attaches to different messages in a given task. The mask is then generated through the Bernoulli distribution:

\[ \text{mask} = \text{bernoulli}(1 - \text{Dropout Rate}_{i,j}) \]  

(5)

This mask determines which elements in the network are kept and which are dropped. The generation of the mask depends on the dynamically adjusted dropout rate for each element. At the same time, based on the model’s performance on the validation set, the model adjusts the base dropout rate by a preset threshold, thus reducing the risk of overfitting while maintaining high-quality summary generation. In the short news Chinese text summarization task, Adaptive-DropMessage optimizes the information processing strategy by integrating the attention mechanism and the dynamic convolutional layer to improve the flexibility and robustness of the model. The application of this method enhances the regularization effect of the feed-forward network, optimizes the gated residual connections, and improves the model’s flexibility and robustness when processing information. And it can dynamically adjust the dropout rate according to the performance of the model in a specific task, effectively reducing the risk of overfitting while maintaining high-quality summary generation, which makes the model more flexible to adapt to the characteristics of different data, thus improving the model generalization ability.

B. Enhanced dynamic convolutional layers

In order to improve the expressive and generalization capabilities of the model and better adapt to the complexity of the generative summarization task for Chinese short news text, we have improved the dynamic convolutional layer of mBART. We use a dynamic weight computation that maps input features to the shape of dynamic weights through a linear layer. This computation allows the model to dynamically adjust the convolution kernel when processing different inputs, thus capturing local features and patterns in the text data more accurately. A regularization mechanism is introduced to perform Adaptive-DropMessage operations on the dynamic weights and outputs during the training process. Regularization helps the model’s generalization ability and improves the processing of news content with large temporal variations. We process the input features in groups, each containing...
an equal number of input channels, using multi-head computation, where we define that for a given input feature matrix $X$ we can divide it into $h$ heads $X_1, X_2, X_3, \ldots, X_h$. Performing an independent linear transformation on each head yields $Q_i, K_i, V_i$, where $i$ denotes the first $i$ header, the multi-header output can be expressed as

$$\text{head\_output}_i = \text{Attention}(Q_i, K_i, V_i) \quad (6)$$

Weighted summation of outputs from different heads after using multi-head computation

$$\text{output} = \sum_{i=1}^{h} W_i \cdot \text{head\_output}_i \quad (7)$$

Thus, the information of the input text is expressed more comprehensively. By improving the dynamic convolutional layer of mBART, we improve the stability of the model when dealing with news content with large temporal variations, which helps the model to deal with Chinese syntactic structure and semantics more efficiently, and improves the performance of the generative summarization task.

Building on the existing improvements to the dynamic convolutional layer in the mBART model, we further enhance the model’s capability by implementing character-based embeddings. This method of embedding focuses on the granularity of individual Chinese characters, which is crucial due to the idiosyncrasies and complexities of the Chinese language. Character-based embeddings allow for a more nuanced understanding of the language, capturing the intricate meanings conveyed by each character and addressing the challenges posed by homophones and polysemy common in Chinese.

In addition to the modifications in the convolutional layers, the model’s architecture, which can be seen in Figure 3, employs an advanced embedding layer that maps each Chinese character to a high-dimensional space. This process helps in capturing the semantic and syntactic properties of each character, crucial for understanding and generating coherent and contextually accurate summaries.

Moreover, the use of character-based embeddings complements the existing multi-head computation structure. By processing character-level features, the model gains a finer understanding of the textual nuances. This is particularly beneficial when dealing with Chinese text, where the meaning can drastically change with slight variations in characters. The multi-head attention mechanism, applied on these character embeddings, facilitates the model to focus on the most relevant characters in a given context, enhancing the model’s ability to generate more accurate and coherent summaries.

The enhanced dynamic convolutional layer, combined with character-based embeddings, leads to a more robust model. This robustness is particularly evident in handling the complexity of Chinese syntactic structures and semantics. The model becomes better equipped to process and summarize Chinese short news text, which often contains nuanced information and requires high precision in understanding and representing the content.

Furthermore, the integration of character-based embeddings with the Adaptive-DropMessage mechanism and multi-head computation ensures that the model not only captures the essential information in the text but also maintains a balance between retaining critical information and preventing overfitting. This balance is key in achieving high-quality summarization results that are both accurate and generalizable across various texts and contexts.

In conclusion, the incorporation of character-based embeddings significantly enhances the mBART model’s ability to process and summarize Chinese short news text. This approach, combined with the improvements in the dynamic convolutional layer and the sophisticated attention mechanism, positions the model as a highly effective tool for Chinese text summarization tasks.

C. Innovative gated residual connection design

Continuing from the previously discussed enhancements, the incorporation of gated residual connections into the mBART model architecture plays a pivotal role in refining the model’s capability to process and synthesize complex information inherent in Chinese short news texts. The essence of the gated residual connection lies in its ability to modulate the flow of information through the network. This modulation is achieved by implementing a gate mechanism that calculates gating weights, which determine the proportion of information to be passed through the residual connections.

In order to introduce gated residual connections, we first introduce a gating mechanism that determines how
much information should pass in the residual connections by learning the obtained gating weights. This mechanism allows the model to selectively pass and filter the input information, which improves the learning ability of the network when dealing with complex features. The Gate network is a sequential network containing linear layers, GELU activation functions and Sigmoid activation functions. Through this network, the model is able to learn to get the gating weights, which effectively regulates the information flow in the residual connections and enhances the flexibility of the model. LayerNorm is introduced to layer normalize the inputs, which helps to stabilize the training process. In addition, the generalization ability of the model is further improved by applying Adaptive-DropMessage for random drop operations, which makes it more robust. The gating weights obtained through Gate network $G$ is multiplied with the input features $X$ are multiplied to obtain the gating term, i.e.

$$ Gate\_Item = G \cdot X $$  \hspace{1cm} (8)

The terms outside the information flow are represented by $1 - G$ denoted by the scaled residual term

$$ Outside\_Item = (1 - G) \cdot Residual $$  \hspace{1cm} (9)

The gated term is added to the residual term to get the final gated residual connection output

$$ Y = Gate\_Item + Outside\_Item $$  \hspace{1cm} (10)

This structure enhances the model’s learning capabilities by effectively balancing the new features learned at each layer with the residual information from previous layers. Layer Normalization (LayerNorm) is employed to normalize the inputs before they enter the gate network, which aids in stabilizing the training process by ensuring consistent scales of data across different layers.

Furthermore, the integration of Adaptive-DropMessage within this structure serves to improve the model’s robustness and generalization. By dynamically adjusting the dropout rate based on the input characteristics and training performance, Adaptive-DropMessage aids in preventing overfitting and ensuring that the model remains adaptable to various textual nuances found in Chinese short news text.

In summary, the addition of gated residual connections to the mBART model introduces a sophisticated mechanism for managing the flow of information, enhancing the model’s ability to learn complex features and relationships within the text. This leads to more accurate and coherent summarizations, bolstering the model’s effectiveness in handling the intricate task of generating summaries for Chinese short news texts.

E. Enhanced self-attention and cross-attention mechanisms

Building upon the enhancements in the self-attention and cross-attention mechanisms within the DMSeqNet-mBART framework, let’s delve deeper into the modifications and their implications for the model’s performance in Chinese short news text summarization.

1) Enhanced Self-Attention Mechanism The self-attention mechanism, a cornerstone in modern NLP
models, has been further refined in our architecture. By integrating the Adaptive-DropMessage module, we address a critical challenge in attention mechanisms: the management of information significance. In conventional self-attention frameworks, while the attention weights are dynamically allocated, there is still a risk of diluting critical information, especially in complex text sequences typical of Chinese language.

The Adaptive-DropMessage module’s incorporation post-attention score calculation enables a dynamic adjustment of the dropout rate based on the calculated attention weights. This innovative approach allows the model to emphasize more critical pieces of information within the text while reducing the emphasis on less relevant details. It significantly enhances the model’s capability to discern and prioritize key semantic relationships within a text sequence. The mathematical representation of this process can be expressed as:

\[
\text{attn\_output, attn\_weights} = \text{multihead\_attn}(x, x, x) \tag{11}
\]

\[
\text{attn\_output} = \text{Adaptive-DropMessage}(\text{attn\_output, attn\_weights}) \tag{12}
\]

2) Enhanced Cross-Attention Mechanism The cross-attention mechanism is pivotal in understanding the relationship between the input sequence and external context, such as encoder outputs in sequence-to-sequence tasks. In our improved model, the cross-attention mechanism, similar to the enhanced self-attention, employs Adaptive-DropMessage. This enhancement is particularly beneficial for tasks requiring a deep understanding of contextual relationships, a common characteristic in summarizing Chinese short news texts.

The use of Adaptive-DropMessage in the cross-attention mechanism ensures that the model dynamically adapts its focus between the input sequence and the external memory. This capability is crucial in tasks dealing with long or complex text sequences, where the context and nuances play a significant role in generating accurate summaries. The mathematical representation of this process is:

\[
\text{attn\_output, attn\_weights} = \text{multihead\_attn}(x, \text{memory, memory}) \tag{13}
\]

\[
\text{attn\_output} = \text{Adaptive-DropMessage}(\text{attn\_output, attn\_weights}) \tag{14}
\]

3) Overall Impact on Model Performance By integrating these enhanced attention mechanisms, the DMSeqNet-mBART model not only improves its self-attention mechanism but also introduces a sophisticated cross-attention layer. This dual enhancement substantially boosts the model’s capability to process and summarize text sequences more effectively. The adaptive nature of the DropMessage process right after the attention score computation accentuates the model’s proficiency in identifying and emphasizing the most salient information, thereby ensuring that the generated summaries are both coherent and contextually rich.

F. Hyperparameter Selection and Design Decision Rationale

In the development of the DMSeqNet-mBART model, specific choices regarding hyperparameters and model configurations were made to optimize performance for Chinese short text summarization. This subsection discusses these choices and the rationale behind them, further illustrating the feasibility of the proposed method.

1) Adaptive-DropMessage Parameters: The Adaptive-DropMessage technique’s effectiveness hinges on the careful tuning of its parameters, particularly the base dropout rate (\(\delta\)). This rate was set after extensive experimentation to balance the model’s need for information retention and the avoidance of overfitting. The dynamic adjustment mechanism for the dropout rate, governed by the attention and message importance adjustments, was calibrated to ensure that the model could adaptively respond to varying degrees of information significance across different text segments.

2) Dynamic Convolutional Layer Configuration: The dynamic convolutional layers were configured to efficiently capture the local contextual features of Chinese text. The decision to use a linear layer for dynamic weight computation was driven by the need for a flexible adaptation of the convolution kernel to different inputs. This design choice enhances the model’s ability to understand nuanced Chinese linguistic patterns, crucial for accurate summarization.

3) Gated Residual Connection Design: Implementing gated residual connections was a strategic decision to improve information flow within the network. The gating mechanism, comprising linear layers, GELU, and Sigmoid activation functions, was optimized to control information passage effectively. The addition of LayerNorm within the gate network was crucial for maintaining training stability and consistency.

4) Character-Based Embedding: The choice to implement character-based embedding stemmed from the unique characteristics of the Chinese language, where meaning is heavily dependent on individual characters. This embedding approach ensures a deeper semantic understanding and more accurate representation of the input text.

5) Multi-Head Computation in Dynamic Layers: The multi-head computation approach in dynamic convolutional layers was adopted to enhance the model’s ability to process information comprehensively. By dividing the input feature matrix into multiple heads and processing them independently, the model can capture a broader range of linguistic features, essential for handling the complexities of Chinese text.
6) Feedforward Network with Batch Normalization:
The integration of batch normalization in the feedforward network was a critical decision to accelerate the model's convergence and improve training stability. This normalization, combined with LayerNorm and GELU activation, optimizes the network's feature extraction capabilities, making it more effective in summarizing Chinese news text.

These design decisions and hyperparameter selections were grounded in a deep understanding of the challenges inherent in Chinese text summarization. The resulting DMSeqNet-mBART model showcases a bespoke architecture fine-tuned for the task at hand, demonstrating the proposed method's practicality and effectiveness.

G. Summary of overall model improvements

We have applied the Adaptive-DropMessage technology to Chinese short news text summarization, enhancing the mBART model to better suit the unique linguistic structure of Chinese. The dynamic convolutional layers handle nuances in Chinese syntax, gated residual connections optimize information flow for complex sentences, the custom feedforward network manages Chinese semantics efficiently, and the enhanced attention mechanism improves the model's ability to prioritize key semantic relationships in text sequences. Together, these enhancements create a powerful model tailored for Chinese short news text summarization.

IV. Experiments

A. Experimental setup

1) Datasets
   * LCSTS dataset [19]: A large dataset for the task of generating Chinese short text summaries. It contains many different types of short text.
   * THUCNews: A Chinese news corpus created by the Natural Language Processing and Social Humanities Computing Laboratory of Tsinghua University, which contains real news texts in different fields and topics.
   * Sogou News Data and Sohu News Data: Media news data with a wide audience in China, including real news texts on various topics and themes.
   * WeChat public account news summary data and Weibo data: summary data generated by social communication software with a wide audience in China.

The six authoritative and widely used data sets used in this experiment cover a variety of different types of Chinese short texts and have high practical application value.

2) Introduction to comparative modeling

In this study, we selected several industry-leading models for comparative experiments to verify the performance of the DMSeqNet-mBART model. These models are:

- T5 (Text-to-Text Transfer Transformer): T5 is an advanced sequence-to-sequence model based on Transformer, designed to handle various NLP tasks such as text summarization.
- mBART-Large-CC25: mBART is a powerful multilingual sequence-to-sequence pre-trained model for large text datasets in 25 different languages, especially suitable for processing text generation tasks such as text summarization.
- Pegasus-Large-Chinese-Cluecorpussmall: Pegasus is a powerful model specially designed for Chinese text summarization tasks.
- GPT-4: GPT-4 is the latest generation of autoregressive language model released by OpenAI. It has a huge parameter scale and extremely powerful text generation capabilities. It is recognized as one of the most powerful models in the field of text generation.

By comparing the full range of indicators with these state-of-the-art models, we can have a more comprehensive understanding of the performance and advantages of DMSeqNet-mBART in the field of Chinese short news text summarization.

B. Experimental results

1) accuracy of different models tested on different datasets (BERTScore)

The DMSeqNet-mBART model demonstrates its excellence in Chinese short news text summarization through an impressive array of BERTScore metrics across multiple datasets. This model, leveraging the novel Adaptive-DropMessage technique, excels in synthesizing concise and contextually accurate summaries by dynamically modulating information retention based on the significance derived from the text’s semantic structure.

a) Synthesis of Results

As shown in Table 1, across the datasets, DMSeqNet-mBART not only shows superior performance but also maintains a consistent lead in the BERTScore F1 metrics, particularly highlighted by its scores in the LCSTS (0.8595) and Weibo (0.8368) datasets. These results exemplify the model's robustness and adaptability to both formal news content and socially driven text data, respectively.

LCSTS: The model achieved an F1 score of 0.8595, surpassing other models like T5 (0.8316) and GPT-4 (0.8468), which underscores its enhanced capability to handle formal news text efficiently.
THUCNews: Here, the model recorded an F1 of 0.8005, reflecting its effectiveness across diverse journalistic content, despite the inherent challenges posed by varied thematic elements within this dataset.

Sougou and Souhu: The scores of 0.8146 and 0.8245, respectively, indicate a strong performance, underscoring the model’s ability to engage with a broader spectrum of news styles and formats.

WeChat and Weibo: These platforms, where informal interaction predominates, presented unique challenges. The model’s scores of 0.7698 (WeChat) and 0.8368 (Weibo) demonstrate its adeptness at navigating less formal content and extracting salient information effectively.

Technological Underpinnings and Impact

The Adaptive-DropMessage technology underpins the model’s superior performance. This feature allows DMSeqNet-mBART to intelligently manage the trade-off between precision and recall, optimizing for content relevance and conciseness. By assessing the significance of each data point within the attention mechanism, the model ensures that essential information is retained, thus enhancing the overall quality of the summaries.

b) Comparative Analysis

As shown in Table 1, when juxtaposed with leading models such as T5 and GPT-4, DMSeqNet-mBART not only achieves higher BERTScores but also demonstrates enhanced stability across varied content types. This stability is critical in applications like real-time news summarization, where the ability to deliver accurate and immediate summaries of events is of paramount importance. The experimental results validate the DMSeqNet-mBART model’s design philosophy and its practical effectiveness in the realm of NLP. By setting new benchmarks on widely recognized datasets, the model not only advances the field of Chinese text summarization but also opens avenues for further research into language-specific NLP applications.

2) Comparison of robustness of different models

In Experiment 2, the DMSeqNet-mBART model was rigorously evaluated under adversarial testing, revealing its robustness and adaptability to complex linguistic challenges. The model’s architecture, specifically its innovative Adaptive-DropMessage technique, enabled it to maintain high BERTScore-F1 scores despite adversarial manipulations, including synonym replacements, information distortions,
Model Adversarial Testing Dataset BERTScore-F1
proposed Adversarial Testing 0.8595
T5 Adversarial Testing 0.8316
mlc Adversarial Testing 0.8312
plcc Adversarial Testing 0.8273
GPT-4 Adversarial Testing 0.8468
proposed Synonym Replacement 0.8378
T5 Synonym Replacement 0.8045
mlc Synonym Replacement 0.8063
plcc Synonym Replacement 0.7994
GPT-4 Synonym Replacement 0.8295
proposed Distortion Information Injection 0.8277
T5 Distortion Information Injection 0.7976
mlc Distortion Information Injection 0.7908
plcc Distortion Information Injection 0.7864
GPT-4 Distortion Information Injection 0.8177
proposed Extreme Text Testing 0.8329
T5 Extreme Text Testing 0.7905
mlc Extreme Text Testing 0.7947
plcc Extreme Text Testing 0.7875
GPT-4 Extreme Text Testing 0.8226
proposed Logical Contradictions and Antonyms 0.7985
T5 Logical Contradictions and Antonyms 0.7712
mlc Logical Contradictions and Antonyms 0.7754
plcc Logical Contradictions and Antonyms 0.7598
GPT-4 Logical Contradictions and Antonyms 0.7982
proposed Spelling and Grammar Errors 0.8165
T5 Spelling and Grammar Errors 0.7843
mlc Spelling and Grammar Errors 0.7801
plcc Spelling and Grammar Errors 0.7703
GPT-4 Spelling and Grammar Errors 0.8031

TABLE II
Mean BERTScore-F1 scores after adversarial testing for different models and scenarios

<table>
<thead>
<tr>
<th>Model</th>
<th>Euclidean Distance of F1 Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>proposed</td>
<td>0.0368</td>
</tr>
<tr>
<td>T5</td>
<td>0.0419</td>
</tr>
<tr>
<td>mlc</td>
<td>0.0417</td>
</tr>
<tr>
<td>plcc</td>
<td>0.0466</td>
</tr>
<tr>
<td>GPT-4</td>
<td>0.0926</td>
</tr>
</tbody>
</table>

TABLE III
Comparison of the Euclidean distance of the average F1 scores of each model before and after adversarial testing

- extreme text variations, logical contradictions, and grammatical errors.
  - a) Profound Resilience and Architectural Sophistication
    - i) Synonym Replacement: As shown in Table 2, the model exhibited a slight reduction in F1 scores to 0.8378, a predictable outcome given the semantic complexities introduced. Nevertheless, this performance outstripped the competition, underscoring the model's nuanced understanding of context and robust lexicon.
    - ii) Distortion Information Injection: As shown in Table 2, an F1 score of 0.8277 in the face of distorted inputs underscores the
model's exceptional ability to discern and retain critical information, a testament to its practical application in real-world summarization tasks.
  - iii) Extreme Text Testing: As shown in Table 2, the model proved its mettle with an F1 of 0.8329, adeptly handling hyperbolic and unconventional language use, thus safeguarding the original intent of the texts.
  - iv) Logical Contradictions and Antonyms: As shown in Table 2, despite the adversarial challenge, the model scored an F1 of 0.7985, showcasing its cognitive sophistication in navigating complex syntactic and semantic structures and in ensuring logical consistency in its summaries.
  - v) Spelling and Grammar Errors: As shown in Table 2, with an F1 of 0.8165, the DMSeqNet-mBART model demonstrated commendable resilience to linguistic inaccuracies, emphasizing its capacity to capture underlying semantic meanings—a crucial attribute for summarization in variable-quality text environments.

b) Stability Assessed Through Euclidean Distance
As shown in Table 3, the stability of the DMSeqNet-mBART model's performance pre- and post-adversarial testing was measured using the Euclidean distance metric. Here, the model showcased a distance of 0.0368, asserting itself as one of the front runners in performance consistency, only marginally outpaced by GPT-4. The model outperformed other high-caliber models, including its predecessor mlc, solidifying the efficacy of the DMSeqNet-mBART enhancements.

c) Stability Assessed Through Euclidean Distance
The DMSeqNet-mBART model has demonstrated exceptional performance across a suite of adversarial challenges, affirming its design strengths and its status as a leading tool for Chinese short news text summarization. Its sophisticated architecture enables it to deliver consistent, high-quality summaries, establishing it as a resilient and reliable model suitable for the nuanced demands of NLP applications. The results from Experiment 2 not only validate the model's operational excellence but also its pioneering role in advancing text summarization under adversarial conditions.

3) Comparison of fluency and completeness of summaries generated by different models
Experiment 3 employs a blind evaluation conducted by professional Chinese language experts, offering an impartial assessment of the quality of summaries generated by different models. In this assessment, three critical dimensions are examined: Fluency,
**TABLE IV**
Comparison of fluency and completeness of summaries generated by different models

<table>
<thead>
<tr>
<th>Model</th>
<th>Fluency</th>
<th>Completeness</th>
<th>Pithiness</th>
</tr>
</thead>
<tbody>
<tr>
<td>proposed</td>
<td>96.9</td>
<td>98.6</td>
<td>96.4</td>
</tr>
<tr>
<td>t5</td>
<td>95.3</td>
<td>96.7</td>
<td>95.9</td>
</tr>
<tr>
<td>mlc</td>
<td>95.4</td>
<td>96.4</td>
<td>96.1</td>
</tr>
<tr>
<td>plcc</td>
<td>95.6</td>
<td>96.7</td>
<td>95.8</td>
</tr>
<tr>
<td>GPT-4</td>
<td>96.6</td>
<td>98.2</td>
<td>96.3</td>
</tr>
</tbody>
</table>

**TABLE V**
Aggregate BLEU and ROUGE Scores Across Datasets for Summarization Models

<table>
<thead>
<tr>
<th>Model</th>
<th>BLEU Score</th>
<th>ROUGE Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>proposed</td>
<td>0.576</td>
<td>0.542</td>
</tr>
<tr>
<td>t5</td>
<td>0.559</td>
<td>0.528</td>
</tr>
<tr>
<td>mlc</td>
<td>0.561</td>
<td>0.531</td>
</tr>
<tr>
<td>plcc</td>
<td>0.562</td>
<td>0.529</td>
</tr>
<tr>
<td>GPT-4</td>
<td>0.571</td>
<td>0.536</td>
</tr>
</tbody>
</table>

Completeness, and Pithiness.

a) Performance Evaluation of the DMSeqNet-mBART Model

i) Fluency: As shown in Table 4, DMSeqNet-mBART achieves the highest score of 98.6, reflecting its exceptional ability to generate summaries that read naturally and smoothly. This suggests that the summaries are grammatically coherent and stylistically polished, indicative of the model’s adept language processing capabilities.

ii) Completeness: As shown in Table 4, with a score of 96.4, DMSeqNet-mBART tops the completeness category as well. This illustrates the model’s proficiency in encapsulating all necessary information from the source material into a comprehensive summary.

iii) Pithiness: As shown in Table 4, the model also leads in pithiness with a score of 96.4, which speaks to its efficiency in distilling the essence of the text. This means the summaries are not only succinct but also devoid of superfluous content, maintaining the original text’s core message without unnecessary elaboration.

b) Implications of the Results

As shown in Table 4, the DMSeqNet-mBART’s performance in Experiment 3, as indicated by the highest sentence score of 96.9, showcases its superior design and the successful realization of its intended capabilities in generating summaries. Each score in the evaluation criteria stands as a testament to the model’s nuanced approach to language modeling and summarization tasks. By achieving the highest marks across all metrics, DMSeqNet-mBART solidifies its place as a leader in Chinese short news text summarization.

The consistency of high scores across fluency, completeness, and pithiness reaffirms that the model’s architectural innovations—particularly the Adaptive-DropMessage technique—effectively address the challenges of summarization in terms of linguistic fidelity, information retention, and concise expression.

c) Conclusion

The findings of Experiment 3 confirm that the DMSeqNet-mBART model not only meets but exceeds the benchmarks set by other advanced models in summarization tasks. Its unmatched scores in a blind test by language experts underscore its effectiveness and the potential to be a transformative tool in the field of NLP, especially for applications involving Chinese text. This evaluation strongly supports DMSeqNet-mBART’s deployment in scenarios where high-quality, nuanced, and concise summaries are crucial.

4) Evaluating Summarization Efficacy of DMSeqNet-mBART Using BLEU and ROUGE Metrics

For Experiment 4, the evaluation focuses on the BLEU and ROUGE metrics.

a) Detailed Analysis

Proposed Model (DMSeqNet-mBART): With an aggregate BLEU score of 0.576 and a ROUGE score of 0.542, the proposed DMSeqNet-mBART model outperforms all other models across the six datasets. This high BLEU score suggests that our model is particularly precise in generating summaries that align closely with reference summaries in terms of word and phrase usage. Similarly, the ROUGE score indicates a superior ability to recall the salient points from the original texts, ensuring comprehensive coverage in the summaries.

i) T5: As shown in Table 5, the T5 model, while showing strong performance, falls slightly behind the proposed model with a BLEU score of 0.559 and a ROUGE score of 0.528. These scores indicate that while T5 is effective at capturing the essence of the source texts, it may miss certain nuances that the proposed model captures.

ii) MLC: As shown in Table 5, the MLC model’s performance is competitive with a BLEU score of 0.561 and a ROUGE score of 0.531. However, it still does not quite match the precision and recall exhibited by
5) Analysis of the results of ablation experiments
To further understand the impact of each component within the DMSeqNet-mBART model, we conducted a series of ablation studies. These experiments involved systematically removing key components from the model and observing the resultant effects on performance. The components subjected to ablation included the Adaptive-DropMessage, the enhanced dynamic convolution layer, the gated residual connection, the custom feedforward network with normalization, and the enhanced self-attention and cross-attention layers.

For these experiments, each variant of the model was trained on a corpus of 5 million Chinese short news summary samples to ensure full convergence. To test the models, 40% of the data was randomly selected from each of the six Chinese short news text datasets, including LCSTS. We employed Average-BERTScore F1 and Average-BLEU as evaluation metrics, which accurately reflect the quality and accuracy of the summaries generated by the models. The results of these ablation experiments, as shown in Table 6, reveal the significant contribution of each component to the overall performance of the DMSeqNet-mBART model. The findings from these experiments are summarized as follows:

- Adaptive-DropMessage: Removing this component will cause a significant decrease in the model’s ability to process critical information. This shows that it plays a vital role in effectively managing the flow of information. Specifically, Adaptive-DropMessage dynamically adjusts the drop rate based on the information’s relevance and message importance, which is critical to maintaining attention to salient details during summarization. Without it, the model will have difficulty distinguishing between critical and non-critical information, resulting in less accurate summaries.
- Enhanced Dynamic Convolutional Layer: The exclusion of this component slightly impaired the model’s capability to capture local features in the text. This highlights its importance in enhancing the model’s feature extraction proficiency. Dynamic convolutional layers adapt to the variable contexts in text, enabling the model to better understand and represent the nuances of language structure, especially in complex Chinese syntax.
- Omitting these connections slightly affected the model’s efficiency in filtering and processing information. This underscores their role in balancing information flow within the model. Gated residual connections regulate the contribution of each layer’s output to the next, ensuring that only relevant information is propagated forward. This helps in maintaining the coherence and relevance of the generated summaries.
- Custom Feedforward Network: Removing this component had a noticeable impact on the model’s overall robustness. The custom design of the feedforward network, tailored for the specific demands of Chinese text summarization, plays a significant role in capturing and processing the intricate semantic patterns in the data.
- Enhanced Self-Attention and Cross-Attention Layers: The lack of these layers significantly reduced the model’s efficiency in processing complex text structures. This emphasizes their critical role in the model’s performance. Enhanced attention mechanisms are pivotal for understanding the contextual relationships within the

iii) PLCC: As shown in Table 5, with a BLEU score of 0.562 and a ROUGE score of 0.529, the PLCC model demonstrates capabilities closely aligned with MLC. It suggests that while the model is adept at summarization tasks, it may not handle the language variety as effectively as the proposed model.

iv) GPT-4: As shown in Table 5, GPT-4 presents a formidable challenge with a BLEU score of 0.571 and a ROUGE score of 0.536, coming closest to the proposed model’s performance. This shows that GPT-4’s vast training data and sophisticated architecture allow it to generate high-quality summaries. Nonetheless, it still slightly lags behind DMSeqNet-mBART, especially in capturing the nuanced details of the source text as reflected in the ROUGE score.

b) Conclusion
The data from Experiment 4 clearly positions the DMSeqNet-mBART model at the forefront in the context of BLEU and ROUGE performance metrics. It confirms that the advancements embedded within DMSeqNet-mBART, such as its Adaptive-DropMessage and dynamic attention mechanisms, not only contribute to its linguistic precision but also to a more faithful encapsulation of the source content. The model’s leading scores are indicative of its superior summarization capabilities, making it an excellent choice for applications requiring high-fidelity translations and detailed content retention. This experiment adds to the growing body of evidence supporting DMSeqNet’s suitability as a top contender in the realm of NLP summarization tasks.
text, enabling the model to generate summaries that are not only accurate but also contextually rich and coherent.

In conclusion, the ablation study demonstrates that each component of the DMSeqNet-mBART model contributes uniquely to its robustness and high performance. Particularly, the Adaptive-DropMessage and the enhanced attention layers (self-attention and cross-attention) are crucial, significantly impacting the model’s ability to generate high-quality, accurate summaries.

V. Conclusion

A. Summary of results

This research marks a significant advancement in the field of Chinese short news text summarization through the innovative application of Adaptive-DropMessage technology within the mBART model framework. Our pioneering work in integrating this technology has proven instrumental in enhancing the model’s ability to process and summarize Chinese short news texts, characterized by their complexity and information density.

The core of our contribution lies in the tailored optimization of the mBART model to suit the intricacies of the Chinese language, a challenge that has been inadequately addressed in previous models. By incorporating Adaptive-DropMessage, we have successfully augmented the model’s capacity to handle complex sequences and contexts that are typically constrained in terms of data availability and diversity. This enhancement significantly elevates the model’s efficiency in discerning and retaining key information, thus improving the quality of the generated summaries.

Furthermore, the enhancements made to the dynamic convolutional layers, gated residual connections, and the feedforward network, along with the refinement of the self-attention and cross-attention mechanisms, collectively contribute to the robustness and adaptability of the model. These improvements enable the DMSeqNet-mBART model to effectively tackle the nuances and subtleties of Chinese syntactic and semantic structures, setting a new standard for text summarization in this domain.

The achievements of this study extend beyond the realm of Chinese text summarization. The methodologies and insights gleaned from this research offer valuable paradigms for future explorations in processing complex texts in various languages. The adaptability and effectiveness of the DMSeqNet-mBART model in handling diverse linguistic challenges open avenues for its application in a broad spectrum of natural language processing tasks.

B. Practical Implications

The DMSeqNet-mBART model’s capabilities have significant practical implications, especially in applications where efficient and accurate summarization of Chinese short news texts is essential. For instance, in the context of rapidly evolving news environments, this model can be employed to generate concise and accurate summaries for quick dissemination, aiding journalists and news aggregators in staying abreast of breaking news and key developments.

In the corporate world, the model can be leveraged for efficient meeting summarization, where it can process and condense meeting contents into actionable summaries. This application is particularly relevant in the era of video conferencing and remote work, where capturing the essence of discussions and decisions is crucial.

Another promising application is in the realm of AI personal assistants. The DMSeqNet-mBART model can be integrated into these systems to provide users with brief, relevant summaries of news articles, reports, or documents, enhancing the efficiency and utility of these assistants in daily tasks and decision-making processes.

Furthermore, the model’s robustness and adaptability make it suitable for educational purposes, such as creating study guides or summarizing academic papers and research articles. This could significantly aid students and researchers in managing the vast amounts of information they encounter.

In conclusion, this paper not only sets a new benchmark in the domain of Chinese short news text summarization but also paves the way for innovative approaches in natural language processing. The integration of Adaptive-DropMessage technology and the comprehensive optimization of the mBART model have the potential to inspire further research and development in this rapidly evolving field.

Acknowledgment

This work was supported by the National Natural Science Foundation of China under Grant 52374169. (Corresponding author: Weijun Cheng)
References


