International Journal of Software Innovation
Volume 11 • Issue 1

A Two-Stage Long Text Summarization Method Based on Discourse Structure

Xin Zhang, Communication University of China, China*
Qiyi Wei, Communication University of China, China
Qing Song, Communication University of China, China
Pengzhou Zhang, Communication University of China, China

ABSTRACT

This paper proposes a two-stage automatic text summarization method based on discourse structure, aiming to improve the accuracy and coherence of the summary. In the extractive stage, a text encoder divides the long text into elementary discourse units (EDUs). Then a parse tree based on rhetorical structure theory is constructed for the whole discourse while annotating nuclearity information. The nuclearity terminal nodes are selected based on the summary length requirement, and the key EDU sequence is output. The authors use a pointer generator network and a coverage mechanism in the generation stage. The nuclearity information of EDUs is to update the word attention distribution in the pointer generator, which not only accurately reproduces the critical details of the text but also avoids self-repetition. Experiments on the standard text summarization dataset (CNN/DailyMail) show that the ROUGE score of the proposed two-stage model is better than that of the current best baseline model, and the summary achieves corresponding improvements in accuracy and coherence.

KEYWORDS

Attention Mechanism, Discourse Structure, Extraction, Generation, Long Text Summarization, Nuclearity Information

INTRODUCTION

The rapid development, active innovation, and widespread popularity of the internet have rapidly brought people from the era of information scarcity to the era of information explosion. According to the International Data Corporation (IDC) prediction, the global data volume may reach as high as 175ZB in 2025, with China's data volume likely to increase to 48.6ZB, accounting for 27.8% of the global data volume (IDC, 2018). Text data is an essential component of data, and although it can improve data search efficiency through keyword searches, it remains subject to the problem of information overload. In addition, with the popularization of mobile devices and the acceleration of work and life, people place increased demands on information browsing and reading methods, which have led to the new trends of digital reading and fragmentation reading. Text summarization, the simplification of text data to quickly extract adequate information, is an effective way to solve the above problems.

DOI: 10.4018/IJSI.331091 *Corresponding Author

Text summarization is one of the applications of natural language processing and one of the most challenging and exciting problems in natural language processing. From the information theory perspective, a text abstract is an information compression process that expresses the maximum amount of information in the original text with the minimum loss of information (Peyrard, 2019). Early text summaries were manually completed, which was time-consuming, labor-intensive, and inefficient. There was an urgent need for automated summarization methods to replace manual forms. In recent years, with the progress of research on unstructured text data, automatic text summarization has received widespread attention and research. Much research has emerged around algorithm technology, datasets, evaluation indicators, and systems. Various fields such as government affairs, finance, news, medicine, and media are applied rapidly. In particular, the recently released multimodal large model GPT-4 has shown strong processing power when performing various natural language tasks (OPENAI, 2023). It can analyze a large amount of text and obtain the required information faster. However, it requires a large training dataset and may result in incorrect summary results when processing text in specific fields (Dylan et al., 2023).

This study proposed a two-stage generative model for long text summarization. To avoid introducing a large amount of redundant information, the long text was first segmented into more fine-grained discourse units, known as EDUs. Then the discourse structure was analyzed based on the rhetorical structure theory and it was constructed based on understanding semantics. At the same time, annotated nuclearity information for EDUs and the terminal node extraction depth were set according to the abstract length requirement to output key EDU sequences. In order to improve coherence and readability, pointer generation networks and coverage mechanisms were adopted. EDU nuclearity information was utilized to update attention distribution at the word level, which solved the out-of-vocabulary problem and avoided self-repetition of crucial information. This model was mainly validated on the standard text summary dataset CNN/Daily Mail.

RELATED WORK

Text Summarization Method

According to the processing method, text summarization can be divided into extractive and abstractive summarization. According to the length of the input document, it can be further divided into short and long text summarization. This paper mainly focuses on single document long texts, and the processing method combines extractive and abstractive methods organically.

The extractive method can be seen as a binary classification problem. The mainstream approach in the early work was based on statistics. When Luhn (1958) first proposed the concept of automatic text summarization in 1958, he used the statistical feature of word frequency to solve the automatic text summarization task. The graph-based method was represented by TextRank algorithm. Mihalcea et al. (2004) use words in the document as vertices of the graph, then construct edges of the graph based on the co-occurrence relationship between words to calculate the importance of words and the corresponding sentence values, thereby obtaining a text summary. The vocabulary chain-based method was proposed by Barzilay et al. (1999), and it consists of three steps: text segmentation, vocabulary chain recognition, and searching for solid vocabulary chains for abstract sentence extraction. Subsequently, Chen et al.(2005) applied this method to Chinese texts for abstract extraction.

In contrast with extractive methods, the purpose of abstractive summarization methods is to fully understand the document content, reorganize the language, and generate a syntactically correct, coherent, and readable summary. Currently, the most widely studied model is based on the sequence-to-sequence framework (Li et al., 2021). In 2015, Rush et al. (2015) first used the models of attention-based encoders and neural network language model decoders for generative summarization. Subsequently, Chopra et al. (2016) introduced conditional recurrent neural networks to construct decoders based on Rush. Nallapati et al. (2016) proposed for the first time the seq2seq framework

combined with a recurrent neural network (RNN) process, extended text abstracts, and added generator pointers to solve out-of-vocabulary (OOV) and low-frequency word problems. However, due to the poor long-distance dependency of RNN, long short-term memory network (LSTM) is used to replace RNN in generating text abstracts. Tu et al. (2016) proposed using a coverage mechanism to handle the duplication problem during the decoding process of generating the abstract. See et al. (2017) proposed a pointer generator network, which automatically selects whether to copy the words needed for the abstract from the original text or generate new words from the vocabulary through pointers. The Google team proposed the Transformer model in 2017 (Vaswani et al., 2017), which uses only the attention mechanism and completely abandons traditional neural network units. Recently, OpenAI's GPT-4 and ChatGPT (OPENAI, 2022), Baidu's ERNIE Bot (Baidu, 2023), and Ali's Tongyi AliceMind (Wang et al., 2023), all based on Transformer.

Today, many models in the field of text summarization no longer rely on a single method and technology to achieve summarization tasks; rather, they combine multiple methods and models, and multiple technologies interact with each other. Zhang et al. (2019) were inspired by the pretrained Bidirectional Encoder Representation of Transformers (BERT) model proposed by Devlin et al. (2018) and proposed a hierarchical Bidirectional Encoder Representation of Transformers (HIBERT) model for document encoding, using unlabeled data to train the model. Liu et al. (2019) further simplified and generalized the usage of the BERT model and proposed a general framework for both extractive and abstractive models. Subsequently, researchers have continuously proposed various derivative models of BERT based on this framework.

Discourse Structure

A discourse refers to a language whole composed of consecutive paragraphs and sentences in a specific structural manner and order (Xu, 2010). The essence of long text summarization is the analysis and processing of the text. Improving the quality of the abstract begins with considering the structure of the text. Relationship, functionality, and hierarchy are the three characteristics of a text. According to the level of discourse, the discourse structure can be divided into macro and micro levels.

Van Dijk et al. (1980) proposed the macro discourse structure theory, and related theories include discourse patterns (Marsh, 1984) and hyper theme theory (Martin et al., 2003). The macro discourse structure refers to the structure and relationships between paragraphs and the above textual units, including paragraphs and paragraphs, chapters and chapters, manifested as the overall semantic coherence of the text. Micro discourse structure also refers to the structure and relationships between discourse units with sentences as the main body – these include clauses and clauses, sentences and sentences, and sentence groups and sentence groups. Based on the rhetorical structure theory (RST) of Mann et al. (1992), macro-discourse relationships can be classified into coordinate, progressive, complementary, comparative, causal, background, explanatory, and evaluative relationships.

There are many micro-discourse structure theories, including RST, Pennsylvania discourse tree theory (Prasad et al., 2008), sentence group theory (Wu et al., 2000), complex sentence theory(Xing, 2001), and discourse structure theory based on the connective-driven dependency tree (CDT) (Li, 2014). RST was developed by Mann and Thompson in 1986, initially defining 23 structural relationships (Mann et al., 1986). Two or more discourse units can be connected through rhetorical relationships and constituted into an RST discourse tree. Li et al. (2014) proposed a discourse structure theory based on CDT and provided a complete definition and description of discourse structure. In the tree structure of CDT theory, terminal nodes represent elementary discourse units (EDUs), internal nodes are conjunctions, and arrows point to central discourse units. Each level of text unit forms a higher level of text units through conjunction and thus combines layers to form a complete discourse tree (Li, 2014).

In 2003, Wang et al. used statistical methods and heuristic rules to extract keywords and critical sentences under the guidance of discourse structure for Chinese web document summarization. Jia (2007), combined the discourse structure features, calculated sentence relevance based on topic

division, merged the overlapping content, and simplified it before producing a summarized text. In 2019, Zhang et al. (2019) utilized the primary and secondary relationships in discourse structure analysis to improve the summarization quality. Liu et al. (2019) extracted sentences based on RST and then modeled and evaluated the summarization coherence. In 2020, Xu et al. proposed a BERT-based neural discourse perception extractive summarization model and constructed a discourse graph based on the RST tree and co-referential relationships to generate low redundancy and rich content abstracts. Most of the above research focuses on the micro-discourse structure, mainly on the relationships and structures within or between sentences, with little research conducted at the macro level. Fu et al. (2021) constructed a hierarchical encoder based on a macroscopic structure using a graph method. They added an information fusion module to assist the decoder in generating abstracts. Although the evaluation indicators of ROUGE have improved significantly, the coherence and redundancy of the abstracts still need to be addressed.

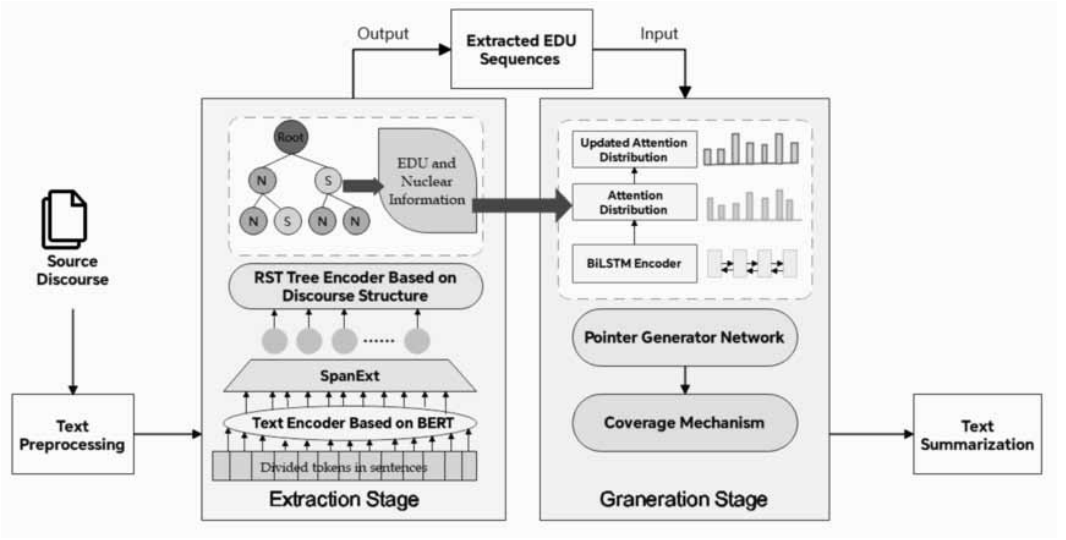
LONG TEXT SUMMARIZATION METHOD BASED ON DISCOURSE STRUCTURE

Overview

The extractive summarization method can maximize confidence that the summary content comes from the original text with smooth sentences and almost no need for grammar modification, but when the length of the summary is limited, there may be some critical sentences that should have been extracted but were not, resulting in missing summary content. Moreover, the extracted sentences could also introduce redundant information or uninformative phrases, resulting in weak sentence coherence and poor readability. On the other hand, the generative summarization method ensures coherence and readability but cannot extract semantic information accurately. Due to the lack of guidance on crucial information, the summarization accuracy could be higher.

In order to comprehensively utilize the advantages of the two methods, this study proposed a two-stage summarization method based on discourse structure, and the model framework is shown in Figure 1. In the first stage, critical sentences were extracted. In the second stage, sentence coherence and readability were improved and low accuracy in long text summarization was solved.

Figure 1. Framework diagram of long text summarization method based on discourse structure



Extraction Stage

The input text often needs to be preprocessed before the extraction stage. The Stanford CoreNLP toolkit for paragraph and sentence boundary detection was used, and the [PAR] identifier was inserted before each paragraph and the [/PAR] identifier was inserted at the end. The [CLS] identifier was inserted at the beginning of each sentence and the [SEP] identifier was inserted at the end. The sentences were then tokenized through space, punctuation, and specific segmentation rules in the extraction stage.

A critical information extraction method based on RST in the extraction stage was proposed. First, BERT for text encoding was applied, followed by RST-DT encoding based on the discourse structure. Finally, the terminal node extraction depth was set according to the required length of the abstract, and the critical information was produced.

Text Encoder Based on BERT

Most of the sentences in the input text were composed of two or more clauses. A clause refers to a language structure that can express an essential, independent, and complete semantic meaning. Based on text preprocessing, EDU was considered the smallest content selection unit in text summarization. The sentence S_i(i=1, 2, ..., n) of the text D was sequenced into EDUs, which are contiguous, adjacent, and nonoverlapping.

Assuming that the sentence S_i is divided into l_i-th EDUs, then:

$$\begin{aligned} S_1 &= \{E_1, E_2, ..., E_{l_1}\}, \\ S_2 &= \{E_{l_1+1}, E_{l_1+2}, ..., E_{l_1+l_2}\}, \ldots \\ S_n &= \{E_{L+1-l_n}, E_{L+2-l_n}, ..., E_L\}, \ (L = l_1 + l_2 + \ldots + l_n) \end{aligned} \tag{1}$$

Figure 2 shows the segmentation of a paragraph and the clause after each ordinal is an EDU. In traditional extractive summarization methods, the entire sentence is usually regarded as the smallest extraction unit, even if some sentences have redundant information. In this study, the EDU was used as the minimum extraction unit, because it is independent and has complete semantics and finer granularity than sentences, making it less likely to generate redundant information when generating abstracts.

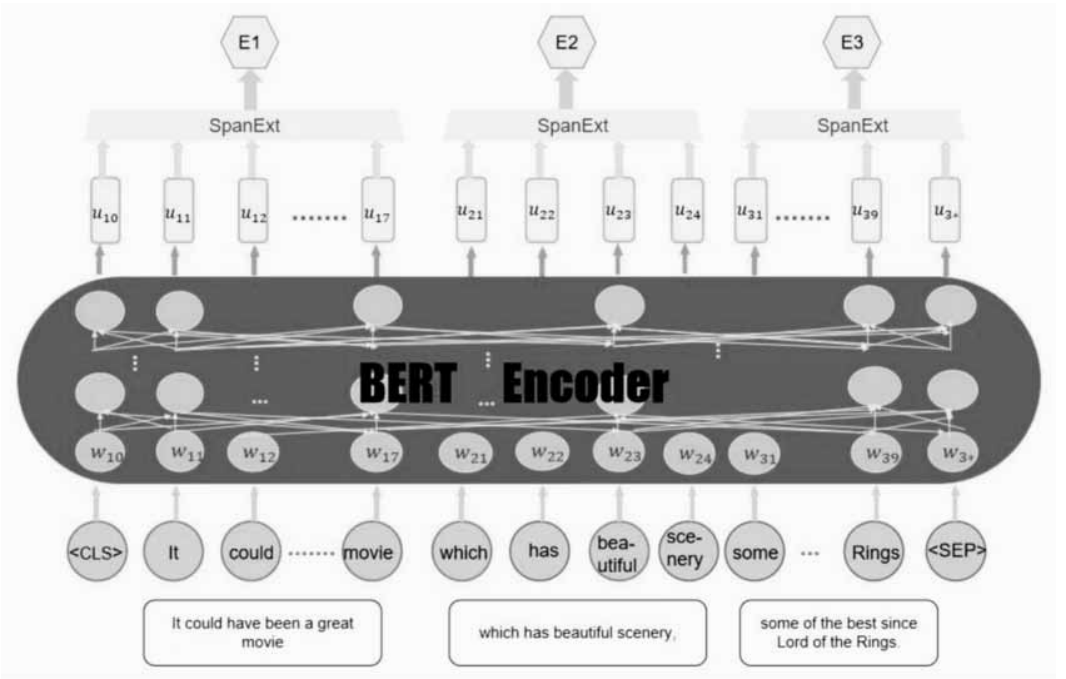
BERT is a pretrained model for deep bidirectional transformer that performs extremely well in text processing tasks (Vaswani, 2017; Zhang, 2019; Xu, 2020). Thus, the BERT model was used to construct a text encoder and encode the text D (Figure 3). BERT was initially trained to encode individuals or pairs of sentences, but it was used to encode tokens in this study. After text encoding, the text can be represented as:

$$\{\boldsymbol{u}_{i1}, \boldsymbol{u}_{i2}, ..., \boldsymbol{u}_{il_i}\} = Bert\left\{\left(w_{i1}, w_{i2}, ..., w_{il_i}\right)\right\} (i = 1, 2, ..., L) \tag{2}$$

Figure 2. A schematic diagram of the segmentation of a paragraph

[a]①It could have been a great movie ②which has beautiful scenery,③ some of the best since Lord of the Rings.④The acting is well done,⑤and I really liked the son of the leader of the Samurai.⑥ He was a likable chap, ⑦and I hated to see him die. ⑧But, other than all that, this movie is nothing more than hidden rip-offs.

Figure 3. A text encoder framework based on BERT



$\boldsymbol{u}_{il_i}$ is the output obtained by text encoding the l_i-th word in the *i*-th EDU.

Extracting EDU is an essential step in the text encoding process. The self-attentive Span Extractor (SpanExt) proposed by Lee et al. (2017) was used to indicate the representation of EDU. For the *i*-th EDU with l_i words, BERT outputs $\{\boldsymbol{u}_{i1}, \boldsymbol{u}_{i2}, ..., \boldsymbol{u}_{il_i}\}$, so the importance of words $\boldsymbol{u}_{ij}$ (j=1,2,...,l_i) in the EDU can be expressed as:

$$s_{ij} = W_2 \cdot ReLU\left(W_1 \boldsymbol{u}_{ij} + b_1\right) + b_2$$

$$\boldsymbol{s}_{ij} = \frac{exp\left(s_{ij}\right)}{\sum_{k=1}^{l_0} exp(s_{ik}} \quad \boldsymbol{E}_i^s = \sum_{j=1}^{l_i} \boldsymbol{s}_{ij} \cdot \boldsymbol{u}_{ij} \ (i = 1,2,...,L) \tag{3}$$

Using matrix **W** and vector **b**, which are learnable parameters, the score of the *j*-th word of the *i*-th EDU could be calculated, and the Softmax function was used to normalize the result $\boldsymbol{s}_{ij}$. The variable $\boldsymbol{s}_{ij}$ could be used as the weight of the word u_{ij}, and then the entire sentence is weighted and summed, resulting in $\boldsymbol{E}_i^s$ being the representation of the *i*-th EDU.

After processing by SpanExt, E_i could be abstracted as SpanExt($u_{i1,}$ $u_{i2,\ ...,}$ u_{ij}), and the text D was converted into an EDU sequence, that is, D={E_1, E_2, ..., E_L}, which served as input to the RST tree encoder.

RST Tree Encoder Based on Discourse Structure

In RST, Mann and Thompson summarized 23 rhetorical relationships, generally divided into two categories: single-core and multi-core. Single-core is mainly used in discourse. In a single-nucleus relationship or a nucleus-satellite relationship, the connected units have a hierarchical structure. All EDUs in the text D were marked with Nuclei (N) or Satellite (S) to indicate their primary and secondary relationships, and their judgment criteria were directly related to the textual relationship. N represented the core content, while S provided supplementary or auxiliary explanations.

In traditional extractive summarization methods, the sentence is the smallest extraction unit, and each sentence is grammatically independent. For EDUs, the semantics are relatively independent, but the syntactic structure may still need to be completed. When extracting candidate EDUs, some limitations must be considered to ensure the syntactic structure. According to the definition of rhetorical relationships, as shown in Table 1, the rhetorical relationships between adjacent EDUs were identified and expressed as directed edges, with the arrow direction pointing from S to N (Mann et al., 1988).

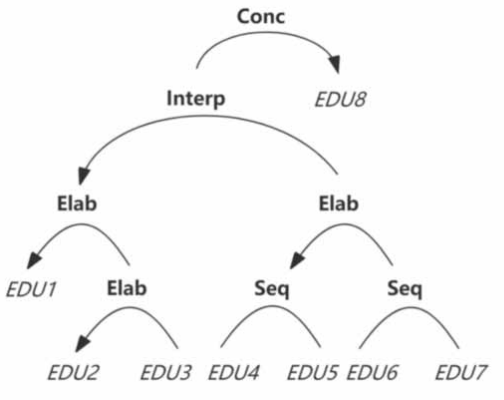
Figure 4 is a hierarchical tree based on the paragraphs shown in Figure 2. There is a semantic juxtaposition between EDU6 and EDU7 and between EDU4 and EDU5. In this case, the two EDUs are in a nucleus-nucleus-symmetric relationship. The whole composed of EDU6 and EDU7 has an elaboration relationship with the whole composed of EDU4 and EDU5. According to the nucleus-satellite relationship, the arrow points from the satellite to the nucleus. Two EDUs are combined to form a composite EDU (CEDU), which can form other semantic relationships with adjacent EDUs or CEDUs, thereby constructing an RST discourse parse tree.

The construction of RST-DT has always been based on a bottom-up approach (Yu et al., 2018), but this method could be more intuitive in practice. Bottom-up parsing can easily limit the construction of trees with local information while neglecting the macrostructure of the text. Therefore, Kobayashi et al.(2020) began to explore a top-down construction method, which uses recursion to segment text

Table 1. Organization of the relation definitions

Circumstances	Antithesis and Concession
Solutionhood	Antithesis
Elaboration	Concession
Background	Condition and Otherwise
Enablement and Motivation	Condition
Enablement	Otherwise
Motivation	Interpretation and Evaluation
Evidence and Justify	Interpretation
Evidence	Evaluation
Justify	Restatement and Summary
Relations of Cause	Restatement
Volitional Cause	Summary
Non-Volitional Cause	Other Relations
Volitional Result	Sequence
Non-Volitional Result	Contrast
Purpose	

Figure 4. An example of RST discourse parse tree, where {EDU1,EDU2,...,EDU8} are EDUs, Conc, Interp, Elab, and Seq are relation labels (Conc=concession, Interp=interpretation, Elab=elaboration, Seq=sequence)



[a]①It could have been a great movie ②which has beautiful scenery, ③some of the best since Lord of the Rings. ④The acting is well done, ⑤and I really liked the son of the leader of the Samurai. ⑥He was a likable chap, ⑦and I hated to see him die. ⑧But, other than all that, this movie is nothing more than hidden rip-offs.

composed of words across spans to construct a composition tree. RST discourse tree (RST-DT) is a constituent tree, so this study adopted a top-down approach to construct RST-DT.

The input to the text encoder is an EDU sequence of the text D. The subscripts are unified here for convenience expression.

Text D consists of m paragraphs, that is, $\boldsymbol{D = \{P_1, P_2, \ldots, P_m\}}$.

Each paragraph is split into n sentences, that is:

$$P_i^s = \left\{\left(S_{i1},\ldots\ldots,S_{in_i}\right)\right\}(i\ldots,2,\ldots,n)$$

Each sentence S is split into l EDUs, that is:

$$S_{ij} = \{E_{ij1},\ldots,\ldots,\ldots(j=1,2,\ldots,l).$$

In order to determine the nucleus and rhetorical relationship labels of two adjacent spans at the sublevel of a specific span (usually text D, paragraph P, or sentence S), Kobayashi's (2020) scoring method was used, and $s_{label}\left(k,p\right)$ was defined as a scoring function to set the nucleus and rhetorical relationship labels for spans. Below is an example of two adjacent EDUs (E_{ijk} and E_{ijk+1}) under the j-th sentence of the i-th paragraph:

$$s_{label}\ k,q\ = W_q MLP\left(\left[E_{ijk};E_{ijk+1};\sum\nolimits_{k=1}^{k}E_{ijk};\sum\nolimits_{k=k+1}^{k=l}E_{ijk}\right]\right),(0<k<l) \tag{4}$$

W_q is the projection layer of the nucleus and rhetorical relationship labels, *MLP* is a multilayer perceptron and a single feedforward network, and the ReLU function is used as the activation function. $\sum_{k=1}^{k}E_{ijk}$ and $\sum_{k=k+1}^{k=l}E_{ijk}$ are the current E_{ijk} left and right spans. A label for the maximum value of the above equation was selected and is defined as follows.

$$\hat{q} = \underset{q \in Q}{argmax}\left[s_{label}\left(k, q\right)\right] \tag{5}$$

Q represents a set of effective nucleus label combinations {N-S, S-N, N-N} for predicting the nucleus and a set of rhetorical structure labels {Circulances, Solutionhood, Elaboration,...} for predicting rhetorical relationships (Mann et al., 1988).

Similarly, to determine the nucleus and rhetorical relationship labels of two adjacent sentences (S_{ij} and S_{ij+1}) under the i-th paragraph, then:

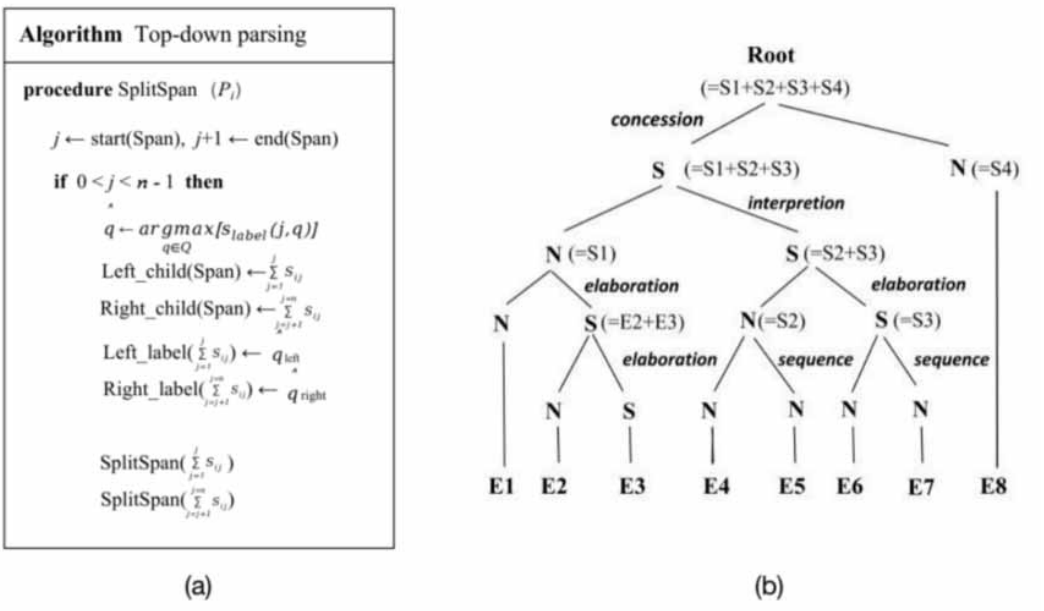
$$s_{label}\ j, q\ = W_q MLP\left(\left[S_{ij}; S_{ij+1}; \sum_{j=1}^{j} S_{ij}; \sum_{j=j+1}^{j=n} S_{ij}\right]\right) \tag{6}$$

If there are two adjacent paragraphs (P_i and P_{i+1}) under the text D, then:

$$s_{label}\ i, q\ = W_q MLP\left(\left[P_i; P_{i+1}; \sum_{i=1}^{i} P_i; \sum_{i=i+1}^{i=m} P_i\right]\right) \tag{7}$$

Therefore, a paragraph tree for text D was first constructed based on the macro discourse structure, where the terminal nodes of the tree are paragraphs. Subsequently, a sentence tree for paragraphs was constructed based on the micro discourse structure, where the leaf nodes of the tree are sentences; then an EDU tree for the sentence was constructed, where the leaf nodes of the tree are EDUs. Finally, the terminal nodes of the paragraph tree were replaced with a sentence tree, and the terminal nodes of the sentence tree were replaced with an EDU tree. The construction process of the RST tree is shown in Figure 5. At this point, the RST discourse tree corresponding to the entire text D was constructed, and

Figure 5. Schematic diagram of RST tree construction process (a) shows a top-down parsing algorithm; (b) shows an example which uses (a) to build a subtree of the RST tree corresponding to Figure 2



the terminal nodes (EDUs) marked as the nucleus set the extraction depth according to the summary length requirement. Assuming that the minimum depth of the RST tree for text D is d_{min}, the extraction depth is $d_{ext}(d_{ext} > d_{min})$. After extraction, the EDU sequence of text D was further reduced to obtain $D'=\{E_1', E_2', ..., E_l'\}$ (l represents the number of extracted EDUs, l<L), which was used as input for the generation stage. Taking the RST tree in Figure 5 as an example, if the minimum depth is 3 and the extraction depth is set to 5, the extracted values are E1 and E8.

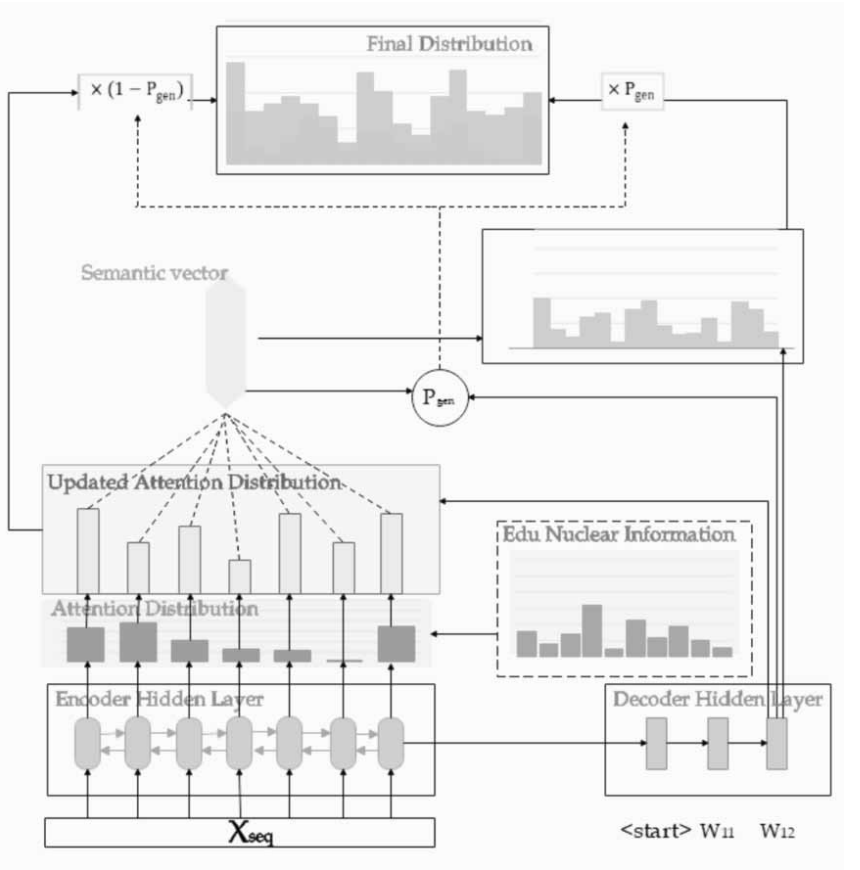
Generation Stage

In the generation stage, the extracted EDU sequences were mainly rewritten into concise sentences to comply with manual summarization habits and enhance coherence and readability. Here, the pointer generation network proposed by See et al.(2017) was used as the generation stage network model and the word level attention distribution in the pointer generator was updated using the EDU nucleus information obtained during the extraction stage. The overall architecture of the generation stage model is shown in Figure 6, which includes a pointer generator and a coverage mechanism, which can effectively solve the problem of unknown words and duplicate summaries.

Pointer Generator

The pointer generator finds and preprocesses the corresponding EDU based on the EDU number obtained during the extraction stage. Since the generator network is based on words as the minimum

Figure 6. Generation stage model framework diagram



unit, it cannot be represented using EDU vectors trained during the extraction stage. The input of the generator is the vector representation of the extracted EDU word segmentation. This model used a standard encoder-decoder structure and an encoder with an attention mechanism in the generator. The encoder encoded the input document into a vector representation. It sequentially input the original words w_{ij} from the input sequence into the encoder f_{en} (single-layer bidirectional LSTM), generating a series of hidden states h_{ij} of the encoders.

$$h_{ij} = f_{en}\left(w_{ij}\right) \tag{8}$$

Generative summarization is the output in words. At each step t, the decoder f_{de} (single layer unidirectional LSTM) received the word embedding of the previous word and had the decoder state s_t. According to the calculation method for attention distribution proposed by Bahdanau et al. (2015), there are:

$$e_{ij}^{t} = v^{T} tanh\left(W_h h_{ij} + W_s s_t + b_{attn}\right)$$

$$\alpha_{ij}^{t} = softmax\left(e_{ij}^{t}\right) \tag{9}$$

where v^T, W_h, W_s and b_{attn} are all learnable parameters and the attention distribution can be seen as the probability distribution of words w_{ij} input into the original text sequence at the current moment. When generating the target summarization, the nucleus information of the EDU was used to update the word attention distribution in the pointer generator.

$$\widehat{\alpha_{ij}^{t}} = \frac{\alpha_{ij}^{t}\left(1 + P_{edu_i} s_{ij}\right)}{\sum \alpha_{ij}^{t}\left(1 + P_{edu_i} s_{ij}\right)} \tag{10}$$

s_{ij} was the score of the j-th word of the i-th EDU, and P_{edu_i} was the degree of influence of i-th EDU, which could be expressed as the reciprocal of the depth of the EDU in the RST tree. The smaller the depth, the greater the impact. It could be represented as $P_{edu_i} = d_{edu_i}^{-1}$. At this point, $\widehat{\alpha_{ij}^{t}}$ could be seen as the probability distribution of words in the input text sequence at the moment, combined with the nucleus information of the EDU where they were located. Words with a more extensive probability distribution were the core words decoded and output. The attention distribution $\widehat{\alpha_{ij}^{t}}$ and encoder implicit state h_{ij} were weighted and operated on to generate semantic vectors h_t^*. At decoding step t, vocabulary distribution P_{vocab} could be generated based on the decoder state and semantic vector h_t^*.

$$h_t^* = \sum_i^0 \sum_j^0 \widehat{\alpha_{ij}^{t}} h_{ij}$$

$$P_{vocab} = softmax\left(V'\left(V\left[s_t, h_t^*\right] + b\right) + b'\right) \tag{11}$$

where V,V', b, and b' were all learnable parameters. P_{vocab} was the probability distribution of all words in the glossary.

Due to the word overflow issue in the generator, a copy mechanism needed to be introduced. Using the pointer network to calculate the probability helped determine whether to generate words from the glossary according to the vocabulary distribution P_{vocab} or directly copy the words in the input sequence according to the attention distribution $\widehat{\alpha_{ij}^t}$.

The calculation formula for generating probability p_{gen} was:

$$p_{gen} = \sigma\left(W_h * h_t^* + W_s * s_t + b_{gen}\right) \tag{12}$$

using updated attention distribution $\hat{a}^t$ and generation probability p_{gen} to calculate the final distribution.

$$P\left(W\right) = p_{gen} P_{vocab}\left(w\right) + \left(1 - p_{gen}\right) \sum_{j:w_j=w}^{0} \hat{a}_j^t \tag{13}$$

Coverage Mechanism

Because the attention mechanism repeatedly focused on certain words in the input sequence, a coverage mechanism was introduced during the generation stage aiming to prevent words that already gained excessive weight from being given high weight once again, thereby reducing self-repetition.

The concrete method used the attention weights previously obtained to influence the current word's calculation. First, the coverage vector c^{t} was summarized and calculated according to the attention distribution $\hat{a}^t$. c^{t} could be represented the past attention information, and used to calculate the current word attention. define the coverage loss, and participate in calculating the primary loss function.

$$\begin{aligned}
c^t &= \sum_{t'=0}^{t-1} \hat{a}_{ij}^{t'} \\
e_{ij}^t &= v^T tanh\left(W_h\left[h_{ij}, \theta\right] + W_s s_t + w c_{ij}^t + b_{attn}\right) \\
covloss_t &= \sum_i^0 \sum_j^0 min\left(\widehat{\alpha_{ij}^t}, c_{ij}^t\right) \\
loss &= -lnP\left(w_t^*\right) + \lambda \sum_i^0 \sum_j^0 min\left(\alpha_{ij}^t, c_{ij}^t\right)
\end{aligned} \tag{14}$$

According to the formula, if a word had previously gained high weight, then its past attention information c^{t} was more extensive, and $covloss_t$ was equal to $\sum_i \sum_j \widehat{\alpha_{ij}^t}$. In order to reduce the loss, it was necessary to reduce the attention to the word again, thereby solving the problem of repetition.

EXPERIMENTS AND RESULTS

Dataset

The original dataset was CNN/Daily Mail proposed by Ramesh Nallapati (2016), which is currently one of the benchmark datasets in English text summarization. The dataset size is shown in Table 2.

Table 2. CNN/DM dataset scale

Dataset	*Size*
Training set	286817
Validation set	13368
Testing set	11487

Evaluation Method

This paper used the ROUGE evaluation method, which Chin Yew Lin et al. (2004) proposed in 2004 and widely used in the Document Understanding Conference (DUC) summary evaluation task. This method is based on the co-occurrence information of n-grams in the abstract to evaluate the abstract and is a method for evaluating the recall rate of n-grams. This study mainly used Rouge-1, Rouge-2, and Rouge-3 to evaluate the proposed two-stage model and the current mainstream summary model.

Advanced Text Summarization Model

Extractive Text Summarization Models

SummaRuNNer is a Recurrent Neural Network (RNN) based sequence model for extractive summarization of documents. It proposes a new training mechanism that uses a generative summary pattern to train extracted tasks.

Refresh is a system based on reinforcement learning, trained through global optimization using ROUGE indicators.

Sumo is an end-to-end extractive model. It defines the abstract problem as a tree induction problem, utilizing structured attention and iterative structure improvement methods to learn document representation.

Neusum is a novel end-to-end neural network framework for extractive document summarization by jointly learning to score and select sentences.

Banditsum is a novel method for training neural networks to perform single-document extractive summarization without heuristically generated extractive labels.

Discobert is a discourse-aware neural summarization model that uses a discourse unit as the minimal selection basis to reduce summarization redundancy and leverages two types of discourse graphs as inductive bias to capture long-range dependencies among discourse units.

Abstractive Text Summarization Models

BOTTOMUP is a generative summarization model that uses a data-efficient content selector to overdetermine phrases in a source document that should be part of the summary and uses the selector as a bottom-up attention step to constrain the model to likely phrases.

Seq2Seq+Attention is a generative summarization model based on RNN, which includes an encoder and decoder, typically using LSTM or bidirectional LSTM to reduce gradient vanishing and explosion issues.

Pointer Generator Network is a viable approach for abstractive summarization, which can copy words from the source text via pointing, using a coverage mechanism to solve duplicate text generation. This network was adopted in the generation stage of this study.

Hybrid Text Summarization Models

SFExt-PGAbs is a summarization model that consists of a submodular function extraction abstract SFExt and a pointer generator generated abstract PGAbs.

TASTE is a hybrid summarization model that combines extraction and generation based on topic awareness. The topic model first obtains the potential topic representation of the text, then adds it to the hybrid model to assist the generation of long text summaries in obtaining summaries that fit the topic.

Experiment Process

This study adopted a non-anonymous version of the CNN/DM dataset, extracting standard abstracts from the original dataset and then using the coreNLP tool to detect sentence boundaries and tokenize the dataset.

In the extraction stage, a pretrained BERT model was used. Due to the limitations of the BERT encoder, the maximum sequence length that BERT can handle is 512, which also needs to include [CLS] and [SEP]. The actual available length was 510. The input sequence length was extended from 512 to 1024 by using a hierarchical position encoding mechanism and fine adjustments were made to all experiments.

In the EDU preprocessing stage, the Self Attention Span Model (SpanExt) was used to learn the representation of EDU. According to statistics, the average number of EDUs in the CNN/DM dataset was 66. Based on the top-down construction of RST discourse trees proposed by Naoki et al. (2020), an EDU parsing tree was constructed and the extraction depth was set according to the abstract length. Additionally, the EDU sequences marked as N were extracted to obtain the key EDU sequences.

In the generation stage, the Seq2Seq+Attention model was adopted. Essential information from the extraction stage was used as input for the generation stage, and the point generator network was used to solve the OOV problem in the summary task. In addition, the coverage mechanism could reduce repeated word generation. A total of 10 epochs were trained in the experiment, with a batch size of 64, and 4488*10 updates were made. The model was stored every 500 times. From this, a summary of the entire text was obtained.

Experiment Results

Results and Analysis

Since the data in the CNN/Daily Mail dataset are all news genres, and most of the news's key content appears at the beginning, Lead-3 has become a robust baseline model. In addition, due to the two-stage model of "extractive+generative" used in this study, the more advanced models in the current extraction and generation algorithms were selected for comparative research in the experiment.

Compared with the experiment results (table 3), the model achieved better summary performance on Rouge-1 and Rouge-L than other baseline models. Compared with DiscoBERT, which performed better in the extraction model, the model in this study achieved progress of 0.5%, 1.5%, and 1.3% in routing 1, routing 2, and routing L. Compared to Bottomup, which achieved the best results in the generative summarization model, progress of 6.8%, 13.3%, and 7.5% in Rouge-1, Rouge-2, and Rouge-L was made. Compared with the SFExt-PGAbs, the study's model made progress of 10.2%, 20.7%, and 13.7% in Rouge-1, Rouge-2, and Rouge-L. Combining micro-basic discourse units and macro-discourse structure analysis could clarify the generated abstract. The seq2seq model combined with the attention mechanism and pointer network solved the problem of duplicate text generation and reduced redundancy.

In addition, ablation experiments on the settings of the TLTSum model generation module were conducted, including the TLTSum model with coverage mechanism, pointer mechanism, and both pointer and coverage mechanism. TLTSum with coverage mechanism increased by 0.14, 0.05, and 0.07 in Rouge-1, Rouge-2, and Rouge-L. TLTSum pointer mechanism significantly improved the rating of Rouge-1, Rouge-2, and Rouge-L, with increases of 0.39, 0.24, and 0.13. The results indicated the pointer mechanism could effectively solve the problem of unregistered words. TLTSum, which simultaneously added a pointer mechanism and coverage mechanism, achieved the best rouge scoring results, with improvements of 0.33, 0.25, and 0.19 compared to the TLTSum model.

Table 3. Experiment results

Model	Rouge-1	Rouge-2	Rouge-L
Lead-3	40.42	17.62	36.67
Extractive			
SUMMARUNNER(Nallapati et al., 2017)	39.60	16.20	35.30
SUMO (Liu et al., 2019)	41.00	18.40	37.20
REFRESH (Narayan et al., 2018b)	40.00	18.20	36.60
NEUSUM (Zhou et al., 2018)	41.59	19.01	37.98
DiscoBert	43.77	20.85	40.67
Abstractive			
BOTTOMUP (Gehrmann et al., 2018)	41.22	18.68	38.34
Seq2Seq+Attention	31.10	11.54	28.56
PointerGeneratorNetwork(See et al., 2017)	36.44	15.66	33.42
PointerGeneratorNetwork+coverage	39.53	17.28	36.38
Extractive + Abstractive			
SFExt-PGAbs(zhou et al, 2021)	39.97	17.54	36.24
TASTE(Yang et al, 2022)	39.86	17.15	35.48
TLTSum	44.03	21.17	41.20
TLTSum + Coverage	44.17	21.23	41.27
TLTSum + Pointer	44.42	21.41	41.33
TLTSum + Pointer&Coverage	44.50	21.48	42.39

Comparison of Generated Summarization

Through comparison and analysis (shown in Figure 7), the difference between the study's model and the baseline model PGN could be seen. Due to the smaller granularity of extracting discourse units, the model designed for the study had lower redundancy. Moreover, with the introduction of discourse structure analysis, the understanding of text semantics was more profound, and the abstract was more hierarchical. The pointer generator considered contextual information, comprehensively solved the OOV problem, and also prevented self-repetition of crucial information. In summary, the model significantly improved in accuracy and coherence.

An automatic grammar check was performed based on the approach proposed by Xu and Durrett (2019), and the grammar check results are shown in Table 4. The average number of errors per 10000 characters on the CNN/DM dataset is displayed, with a comparison of the summary errors of the TLTSum and BERT models.

Where CR represents the correctness check of grammar, PV represents the passive voice check of grammar, PT represents the punctuation check of compound sentences, and O represents the check of other syntax errors. The results in Table 4 show the summary generated by the TLTSum model was consistent with the BERT model in terms of the number of syntax errors. There were many errors in the punctuation of compound statements, which may be related to the division errors when building the RST tree from top to bottom.

Amazon Mechanical Turk (MTurk) was used for manual evaluation, scoring the generated abstracts in coherence, conciseness, and information, with a score range of 1 to 5. The results are shown in Table 5.

Figure 7. Comparison of different model summary generation examples

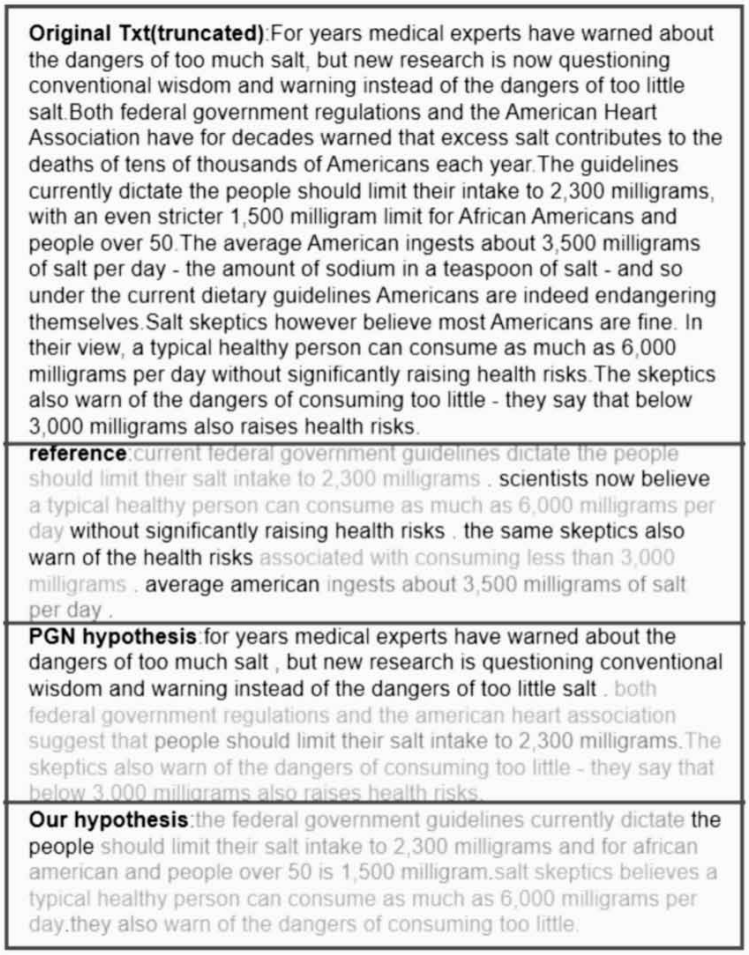


Table 4. Grammar check results

Model	CR	PV	PT	O
BERT	18.0	2.9	2.3	3.0
TLTSum	18.2	2.9	2.5	3.0

Table 5. Manual evaluation summary scoring

Model	Coherence	Cociseness	Informedness
BERT	3.30 ± 0.90	3.12 ± 0.73	3.26 ± 0.80
PGN	3.28 ± 0.82	3.28 ± 0.75	3.17 ± 0.61
TLTSum	3.29 ± 0.71	3.24 ± 0.81	3.30 ± 0.56

The results of the manual evaluation show that the TLTSum model proposed in this study achieved high scores in the informative evaluation, which was related to the extraction of kernel relationships between EDUs in this study, allowing critical information to be selected for output. In the simplicity evaluation, the model outperformed the BERT model with a slight advantage but was slightly inferior to the pointer network model, which may be related to the depth of the EDU extraction. In coherence assessment, the model was almost consistent with the baseline model. From the above analysis, the TLTSum model met the requirements of abstract coherence, conciseness, and information.

CONCLUSION

This study proposed a long text summarization model based on discourse structure, effectively solving the accuracy and coherence issues of long text summarization by extraction and generation. In the extraction stage, the long text was first divided into EDUs to reduce the granularity of crucial information extraction and avoid introducing too much redundant information. Then, a parse tree for the entire text was constructed and the critical information was hierarchically labeled based on rhetorical structure theory. Finally, the corresponding deep core leaf nodes were extracted according to the abstract's length requirements. In the generation stage, a pointer generation network and coverage mechanism were used to solve the problem of unknown words and duplicate abstracts. At the same time, the nuclearity information of the extraction stage EDU was used to update the attention distribution at the word level, better assisting the pointer generator in extracting critical text information. The experiments on the standard text abstract dataset (CNN/Daily Mail) showed that the ROUGE indicators of the proposed abstract model were superior to the current best benchmark model, and the abstract achieved corresponding improvements in accuracy and coherence.

REFERENCES

Bahdanau, D., Cho, K., & Bengio, Y. (2015). *Neural machine translation by jointly learning to align and translate.* International Conference on Learning Representations (ICLR), San Diego, CA. https://arxiv.org/pdf/1409.0473

Baidu. (2023). *The report of Ernie bot*. https://aistudio.baidu.com/aistudio/projectdetail/5748979

Barzilay, R., & Elhadad, M. (1997). *Using lexical chains for text summarization*. Intelligent Scalable Text Summarization Workshop, Madrid, Spain. https://aclanthology.org/W97-0703.pdf

Chen, Y., Wang, X., & Guan, Y. (2005). *Automatic text summarization based on lexical chains* [Conference paper]. International Conference on Natural Computation (ICNC), Berlin, Germany. doi:10.1007/11539087_125

Chopra, S., Auli, M., & Rush, A. M. (2016). Abstractive sentence summarization with attentive recurrent neural networks. *Conference of the North American Chapter of the Association for Computational Linguistics (NAACL): Human Language Technologies*. 10.18653/v1/N16-1012

CNN/Daily Mail. (2018). *Process-Data-of-CNN-DailyMail*. https://github.com/JafferWilson/Process-Data-of-CNN-DailyMail10.18653/v1/N16-1012

Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *Conference of the North American Chapter of the Association for Computational Linguistics*, 4171- 4186. https://doi.org/10.48550/arXiv.1810.04805

Dylan, P., & Gerald, W. (2023). *GPT-4 Architecture, Infrastructure, Training Dataset, Costs, Vision, MoE*. https://www.semianalysis.com/p/gpt-4-architecture-infrastructure

Fu, Y., Wang, H., & Wang, Z. (2021). Scientific paper summarization model using macro discourse structure. *Jisuanji Yingyong*, *41*(10), 2864–2870. https://kd.nsfc.gov.cn/achievement-system/isisn/detailsPage/7073376

IDC. (2018). *IDC: Expect 175 zettabytes of data worldwide by 2025.* https://www.networkworld.com/article/3325397/idc-expect-175-zettabytes-of-data-worldwide-by-2025.html

Jia, G. (2007). Study on automatic abstraction based on text structure. *Computer and Engineering Institute, 38*(6), 10-13. https://xueshu.baidu.com/usercenter/paper/show?paperid=5918a739e4629aee47439649f74f7d45&site=xueshu_se&hitarticle=1

Jiacheng, X., Zhe, G., Yu, C., & Jingjing, L. (2020). *Discourse-Aware Neural Extractive Text Summarization.* The 58th Annual Meeting of the Association for Computational Linguistics. doi:10.18653/v1/2020.acl-main.451

Jiu, J. X. (2010). Introduction: Text linguistics in contemporary Chinese. The Commercial Press, 2-4.

Kobayashi, N., Hirao, T., Kamigaito, H., Okumura, M., & Nagata, M. (2020). Top-down RST parsing utilizing granularity levels in documents. *The Thirty-Fourth AAAI Conference on Artificial Intelligence*, New York, NY. doi:10.1609/aaai.v34i05.6321

Lee, K., He, L., Lewis, M., & Zettlemoyer, L. (2017). *End-to-end neural coreference resolution.* The 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark. doi:10.18653/v1/D17-1018

Li, J., Zhang, C., Chen, X., Hu, Y., & Liao, P. (2021). Survey on Automatic Text Summarization. *Journal of Computer Research and Development.*, *58*(1), 1–21.

Li, Y., Feng, W., Sun, J., Kong, F., & Zhou, G. (2014). Building Chinese discourse corpus with connective-driven dependency tree structure. *Conference on Empirical Methods in Natural Language Processing*, Doha, Qatar. doi:10.3115/v1/D14-1224

Lin, C. Y. (2004). *ROUGE: A package for automatic evaluation of summaries*. Information Sciences Institute, University of Southern California. http://anthology.aclweb.org/W/W04/W04-1013.pdf

Liu, K., & Wang, H. (2019). Research Automatics Summarization Coherence Based on Discourse Rhetoric Structure. *Journal of Chinese Information Processing*, *33*(01), 77–84.

Liu, Y., & Lapata, M. (2019). *Text Summarization with Pretrained Encoders.* The 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, Hong Kong, China. doi:10.18653/v1/D19-1387

Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM Journal of Research and Development*, *2*(2), 159–165. doi:10.1147/rd.22.0159

Mann, W. C., Matthiessen, C., & Thompson, S. A. (1989). *Rhetorical structure theory and text analysis.* Information Sciences Institute. https://apps.dtic.mil/dtic/tr/fulltext/u2/a222655.pdf

Mann, W. C., & Thompson, S. A. (1986). Relational propositions in discourse. *Discourse Processes*, *9*(1), 57–90. doi:10.1080/01638538609544632

Mann, W. C., & Thompson, S. A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text*, *8*(3), 243–281. doi:10.1515/text.1.1988.8.3.243

Marsh, D. (1984). *Michael Hoey: On the Surface of Discourse*. George Allen and Unwin, London. 211 pp. *Nordic Journal of Linguistics*, *7*(1), 77–79. doi:10.1017/S033258650000113X

Martin, J. R., & Rose, D. (2003). *Working with Discourse: Meaning Beyond the Clause*. Continuum.

Mihalcea, R., & Tarau, P. (2004). *TextRank: Bringing order into text.* The 2004 Conference on Empirical Methods in Natural Language Processing, Barcelona, Spain. https://digital.library.unt.edu/ark:/67531/metadc30962/m2/1/high_res_d/Mihalcea-2004-TextRank-Bringing_Order_into_Texts.pdf

Nallapati, R., Zhou, B., Gulcehre, C., & Xiang, B. (2016). *Abstractive text summarization using sequence-to-sequence RNNs and beyond.* 20th SIGNLL Conference on Computational Natural Language Learning, Berlin, Germany. doi:10.18653/v1/K16-1028

Naoki, K., Tsutomu, H., Hidetaka, K., Manabu, O., & Masaaki, N. (2020). Top-down RST parsing utilizing granularity levels in documents. *The Thirty-Fourth AAAI Conference on Artificial Intelligence*.

OPENAI. (2022). *Introducing ChatGPT.* https://openai.com/blog/chatgpt

OPENAI. (2023). *GPT-4 Technical Report.* https://arxiv.org/abs/2303.08774

Peng, W., Yang, A., & Rui, M. (2022). OFA: Unifying architectures, tasks, and modalities through a simple sequence-to-sequence learning framework. *The International Conference on Machine Learning*. https://arxiv.org/abs/2202.03052

Peyrard, M. (2019). *A simple theoretical model of importance for summarization.* The 57th Annual Meeting of the Association for Computational Linguistics, Firenze, Italy. doi:10.18653/v1/P19-1101

Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A. K., & Webber, B. L. (2008). The Penn discourse treebank 2.0. In The Language Resources and Evaluation Conference. Berlin: Springer-Verlag.

Rush, A. M., Chopra, S., & Weston, J. (2015). *A neural attention model for abstractive sentence summarization.* The 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal. doi:10.18653/v1/D15-1044

See, A., Liu, P. J., & Manning, C. D. (2017). *Get to the point: Summarization with pointer-generator networks.* The 55th Annual Meeting of the ACL, Vancouver, Canada. doi:10.18653/v1/P17-1099

Tu, Z., Lu, Z., Liu, Y., Liu, X., & Li, H. (2016). *Modeling coverage for neural machine translation.* The 54th Annual Meeting of the ACL, Berlin, Germany. doi:10.18653/v1/P16-1008

Van Dijk, T. A. (1980). *Macrostructures: An interdisciplinary study of global structures in discourse, interaction, and cognition*. Lawrence Erlbaum Associates, Inc.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., & Kaiser, L. (2017). *Attention is all you need*. The Advances in Neural Information Processing Systems., https://arxiv.org/pdf/1706.03762v5

Wang, J., Wu, G., Zhou, Y., & Zhang, F. (2003). Research on automatic summarization of web document guided by discourse. *Journal of Computer Research and Development*, *16*, 115–123. https://xueshu.baidu.com/usercenter/paper/show?paperid=cb7fd2235483af14becfbbc795151d88&site=xueshu_se

Wu, W., & Tian, X. (2000). *The Chinese Sentence Group*. The Commercial Press.

Xing, F. (2001). *Research on Chinese Complex Sentence*. The Commercial Press.

Xu, J., & Durrett, G. (2019). *Neural Extractive Text Summarization with Syntactic Compression.* The 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, Hong Kong, China. doi:10.18653/v1/D19-1324

Yu, N., Zhang, M., & Fu, G. (2018). *Transition-based neural RST parsing with implicit syntax features.* The 27th International Conference on Computational Linguistics, Santa Fe, NM. https://www.aclweb.org/anthology/C18-1047.pdf

Zhang, X., Wie, F., & Zhou, M. (2019). *HIBERT: Document level pre-training of hierarchical bidirectional transformers for document summarization.* The 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy. doi:10.18653/v1/P19-1499

Zhang, Y., Wang, Z., & Wang, H. (2019). Single Document Extractive Summarization with Satellite and Nuclear Relations. *Journal of Chinese Information Processing*, *33*(08), 67–76.

Zhang Xin, M.D., is a teacher at the School of Computer and Cyber Sciences at the Communication University of China. She is also a senior engineer and a member of the Media Information Technology Institute research team. Her research interests include computer vision, deep learning, natural language processing, blockchain, and more. Currently, she is involved in the "Research and Application of Key Technologies for Human-Machine Interaction in Government Affairs Based on Personification" project, which is part of the "New Generation of Artificial Intelligence" major project under the Science and Technology Innovation 2030 initiative. She has led and participated in dozens of scientific research and teaching reform project, and has published over 10 academic papers and 2 academic works.

Wei Qiyi is an undergraduate student of Communication University of China, majoring in computer science and technology of 2020 in the School of Computer and Cyberspace Security of Communication University of China and once served as the monitor of this major. 1. Learning experience 2020.09 to now. 2. Research interests and directions: She focuses on research in natural language text abstracts, emotion mining, and other fields. Currently, she is participating in the "New Generation Artificial Intelligence" major project "Research and Application of Key Technologies for Anthropomorphic Human Machine Interaction in Government Affairs" at the Media Information Technology Institute for Technological Innovation 2030. 3. Awards in school: She was awarded the title of Outstanding Class Cadre and Outstanding League Cadre of Communication University of China, the third prize scholarship and single scholarship of Communication University of China. She has won the second prize in the United States (International) Mathematical Modeling Competition, the second prize in the Huashu Cup Mathematical Modeling Competition, and she twice won the municipal level approval of the College Student Innovation and Entrepreneurship Plan.

Song Qing, PhD, is the Vice Director of the Convergence Media Center at the Communication University of China. He is also a senior engineer and a master's supervisor. Currently, his research focuses on intelligent media information processing, artificial intelligence, natural language processing, and information extraction. He is currently participating in the "Research and Application of Key Technologies for Human-Machine Interaction with Human-Like Characteristics in the Government Sector" project of the "New Generation of Artificial Intelligence" major project of the Science and Technology Innovation 2030 program. He has hosted and participated in multiple national and provincial-level scientific research projects, published more than 20 academic papers, and obtained 3 authorized invention patents and more than 10 software copyrights.

Zhang Pengzhou, PhD, is the Dean of the Internet Information Research Institute and a professor and doctoral supervisor at the State Key Laboratory of Media Convergence and Communication at the Communication University of China. He graduated from Beijing Institute of Technology in September 1997 and worked as a postdoctoral researcher at the Institute of Computing at the Chinese Academy of Sciences from 1997 to January 2000. He was introduced to Communication University of China in October 2002 and has been working there ever since. He was selected for the "New Century Excellent Talents Support Program" by the Ministry of Education in 2009. He has long been engaged in research on artificial intelligence and its application in digital multimedia, data journalism, etc., including construction of brain-like knowledge graphs, knowledge representation and reasoning, unified representation and retrieval of multimodal digital media, intelligent machine writing, and international communication. As a project leader or core member, he has undertaken more than 10 national-level projects, and he has won the Second Prize of Beijing Science and Technology Progress Award, the Second Prize of Science and Technology Innovation Award of the State Administration of Radio, Film and Television, the Third Prize of Chengdu Science and Technology Progress Award, and the Third Prize of China Film and Television Technology Society. In the past five years, he has completed the national project "R&D and Application Demonstration of Rich Media Public Contribution Platform under Complex Network Environment" and the Beijing project "R&D and Demonstration Application of Chinese News Automatic Writing System Based on Brain-Like Knowledge Graph", hosted two national standards, participated in the formulation of one international standard, applied for two patents, and published more than 20 papers in domestic and foreign academic journals and conferences.