

Received March 5, 2020, accepted May 2, 2020, date of publication May 12, 2020, date of current version May 29, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2994092

# Generative Text Summary Based on Enhanced Semantic Attention and Gain-Benefit Gate

JIANLI DING<sup>1</sup>, YANG LI<sup>1</sup>, HUIYU NI<sup>1</sup>, AND ZHENGQUAN YANG<sup>2</sup>

<sup>1</sup>College of Computer Science and Technology, Civil Aviation University of China, Tianjin 300300, China

<sup>2</sup>College of Science, Civil Aviation University of China, Tianjin 300300, China

Corresponding author: Yang Li (liuhaifeiyu8@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61573199 and Grant 61571441, in part by the Civil Aviation Major Science and Technology Special Fund under Grant MHRD20160109, in part by the Civil Aviation Security Capacity Building Project under Grant TRSA201803, and in part by Fundamental Research Funds for the Central Universities of the Civil Aviation University of China under Grant 3122018C025.

**ABSTRACT** Generative text summary is an important branch of natural language processing. Aiming at the problems of insufficient use of semantic information, insufficient summary precision and the problem of semantics-loss in the current generated text summary method, an enhanced semantic model is proposed based on dual-encoder, which can provide richer semantic information for sequence-to-sequence architecture through dual-encoder. The enhanced attention architecture with dual-channel semantics is optimized, and the empirical distribution and Gain-Benefit gate are built for decoding. In addition, the position embedding and word embedding are merged into the word embedding technology, and the TF-IDF (term frequency-inverse document frequency), part of speech, key score are added to word's feature. Meanwhile, the optimal dimension of word embedding is optimized. This paper aims to optimize the traditional sequence mapping and word feature representation, enhance the model's semantic understanding, and improve the quality of the summary. The LCSTS and SOGOU datasets are used to validate proposed method. The experimental results show that the proposed method can improve the performance of the ROUGE evaluation system by 10-13 percentage points compared with other listed algorithms. We can observe that the semantic understanding of the text summaries is more accurate and the generation effect is better, which has a better application prospect.

**INDEX TERMS** Generative text summarization, encoder-decoder, enhanced semantic attention, empirical distribution, improved word feature.

## I. INTRODUCTION

Nowadays, with the rapid development of the internet, the network platform has become an important way for people to interact and communicate with each other, and that also makes it easier for people to browse and publish information. Information overload has become a serious problem with the explosive growth of online data. In the face of massive information, how to obtain useful data from it has become an urgent problem in the field of information processing.

Text summarization is an important direction in the field of natural language processing. Text summary refers to extract the key information from a large number of texts by computer. The text summarization is the significant technology of information extraction and compression. Methods about text

summary have appeared as early as the 1950s. According to the form of summary, it can be divided into extractive and generative forms. Extractive text summary is the way that the central idea of an article can be summarized by one or several sentences in the article. Generative text summary method is based on the understanding of the contextual semantics of the article. It can summarize the article like human beings. The summary does not need to appear in the original text, but should be appropriate to the full text semantics. By contrast, generative text summary is closer to human thinking and is more accurate in reflecting the artistic conception of the text. But it also relates to the natural language understanding and text remodeling, which makes it more difficult to understand the summary.

At present, deep learning technology has been widely used in the field of natural language processing, including reading comprehension [1], automatic question and answer [2],

The associate editor coordinating the review of this manuscript and approving it for publication was Imran Sarwar Bajwa<sup>1</sup>.

machine translation [3] and text remodeling [4] *et al.* The text summary generation method is inspired by the neural network model of machine translation. In 2015, Lopvrev [5] used recurrent neural networks and LSTM (Long Short-Term Memory) neural units to construct a summary generation model based on Encoder-Decoder framework, and combined with self-attention mechanism to generate text high-quality text summary. In 2016, Rush *et al.* [6] proposed an improved attention model for sentence summary and solved the problem of inconsistent sentence generation. Hu *et al.* [7] *et al.* designed an encoder-decoder structure based on RNN for Chinese text summary tasks, and proposed a large-scale text training corpus LCSTS. In 2017, Zhang *et al.* [8] used character-level features as input to the seq2seq framework based on LSTM. And the problem of too large dimension of feature space is solved while ensuring the performance of the model. See *et al.* [9] proposed a hybrid pointer generator network, which guaranteed the accuracy of generating summary for original semantic restatement and retained the ability to generate new words. In 2018, Chen [10] combined the hidden layer semantics of multi-layer neural networks to solve the problem of insufficient semantic understanding of the model in order to improve the quality of summary generating.

IBM Nallapati *et al.* [11] proposed a text summarization method combining multiple features based on sequence-to-sequence model with attention mechanism, which was validated by the Gigaword dataset and the DUC dataset, and got a great semantic result. Stanford University scholar See *et al.* [12] proposed a generative text summarization method, which can solve the problem of semantic repetition in the process of text organization through pointer network and coverage mechanism. In 2017, Google Brain Vaswani *et al.* [13] proposed a new network architecture, transformer, and based on this proposed a multi-head attention mechanism. This method completely eliminates repetition and convolution processes, has more parallelizable and less training time. Experimental results on machine translation tasks have shown the method to be superior in quality. In 2019, Guo *et al.* [14] proposed an ms-pointer network that based on the multi-head self-attention to enhance the semantic features by the combination of semantics, which makes the semantic structure of abstractive text summary more reasonable. Nowadays, the abstractive text summary generation of long and short texts is a hotspot in natural language research [15]–[19]. The multi-head attention mechanism is widely used in natural language processing tasks such as machine translation and text summarization, and has achieved good results.

In addition, in order to improve the training speed and expressiveness of language models and assist models to acquire better results, there are also many studies focusing on sentence embedding and document embedding. Artetxe and Schwenk *et al.* [20] proposed an architecture to learn representations in different languages, which can solve the problem of cross-language information utilization and multiple practical scenarios sentence representations for different

languages. In 2017, Arora *et al.* [21] proposed an embarrassingly simple but tough-to-beat sentence embedding method, which is well-suited for domain adaptation settings, so the sentence vectors trained on various corpora can be used on different testbeds. And, Chatzakou *et al.* [22] evaluated actual verification on doc2vec, and the results show that the document embedding method proposed by Google Mikolov performs well in various text tasks, and has better quality than the vector obtained by the word vector averaging method. Lau *et al.* [23] proposed the MultiSpot, a multilevel sentiment analysis approach that combines word and sentence level features to enhance emotional analysis. The experiment shows that the combination of word level and sentence level features can better understand the perceptual content of the document and better capture the emotion of text expression than the word level method.

However, there is a problem of Semantics-Loss in sequence-to-sequence text tasks. As we all know, the essence of the decoding process in the seq2seq model is a conditional probability. In the decoding, the generation of the latter word depends on the information of the previous generated word and the attention mechanism. But as the length of the generated text increases, the decoding accuracy will decrease, which reflects the asymmetry of the decoding. This phenomenon occurs in both machine translation and text generation tasks, and we call it Semantics-Loss. The performance of the Semantics-Loss problem is that as the length of the generated text increases, the accuracy of subsequent word generation gradually decreases. In 2019, Zhou *et al.* [24] proposed a synchronous bidirectional decoding mechanism, left-to-right and right-to-left decoding methods, and tried to solve the problem of Semantics-Loss in machine translation tasks, but the generated words may be confused. In this paper, a Gain-Benefits gate structure is designed in the decoder to solve this problem.

Facing different word-formation structures and complex structured texts, how to generate summaries with sufficient semantic understanding and high accuracy is an urgent problem to be solved. At present, the method of summaries generation and learning based on encoder-decoder is widely used, that is, sequence-to-sequence text learning model, to generate text abstracts. However, the ability of word-level or character-level semantic comprehension can not be improved by linear stacking of multiple network layers or increasing the ability of replication of related words in the original text. When the number of short text sentences is small and the length of the text is various, the linear superposition of the layers of the neural network can not make the model obtain good quality.

From the above discussions, in this paper, an enhanced semantic architecture is proposed based on sequence-to-sequence with dual-encoder, Gain-Benefit gate and keywords empirical probability distribution for text summary.

The innovations of this paper are mainly focused on as follows: 1) The dual-encoder is applied to the encoder, and the high-level encoder is used to obtain the global semantic

information of the text, while the low-level encoder focuses on the semantic representation of the original sequence in the encoding process. 2) By the internal alignment relationship between global and local semantic information, the attention mechanism is optimized to integrate the mixed semantic information of dual-encoder and the hidden state of the decoder. 3) The Gain-benefit gate structure is designed to solve the problem of Semantics-Loss along with the length increase of text summary length of in decoder. Meanwhile, The empirical probability distribution of keywords is added to the decoder to accelerate the convergence of the model and make the generated summary more accurate. 4) Integrating position embedding and word embedding methods, emerging POS (parts of speech), TF-IDF and Sco (key scores) to improve the word's feature representation to improve the model's understanding of word meaning. Specially, the optimal dimension of word vector generation is optimized according to the size of corpus.

The organization of this paper is described as follows: In Section 2, an enhanced semantic network model based on seq2seq is proposed. The model includes a dual encoder, an enhanced attention mechanism, and the decoder with an empirical probability distribution and Gain-Benefit gate structure. In Section 3, the improvement strategy of word features is proposed, which combines word vector and position vector, and adds new features such as POS, TF-IDF, Sco. Then, the optimal dimension of word embedding is optimized. In Section 4, the effectiveness of the method is verified by experiments. Conclusions are given in Section 5.

## II. AN ENHANCED SEMANTIC MODEL WITH DUAL-ENCODER AND GAIN-BENEFIT GATE

In this section, the enhanced semantic text summary model is designed, which consists of high-level and low-level dual-encoder structure, the decoder with an empirical probability distribution, and the Gain-Benefit gate structure. The detailed architecture is shown in Figure 1.

In Figure 1, The high-level encoder focuses on context semantic understanding, while the low-level encoder focuses on the aligned input of text features and hidden state, which forms a joint semantic vector representation. In the decoder, the multi-layer unidirectional neural network is designed. Meanwhile, high quality text summary can be output by combining the Gain-Benefits gate structure and the empirical probability distribution of vocabulary.

Where the global semantic information  $M$  is generated by the high-level encoder accompanied by self-attention mechanism, and the alignment vector of text features is generated by the low-level encoder. The two vectors are emerged into a joint semantic vector for the decoder.  $Q_o$  is the vocabulary empirical distribution to achieve fast convergence of the model.  $GB$  is the Gain-Benefit gated structure to solve the problem of Semantics-Loss due to the increasement in the summary length generated. The above model is described in detail in the following sections.

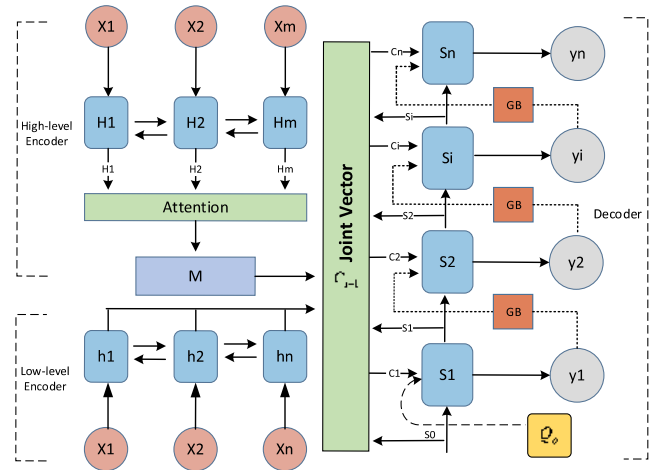


FIGURE 1. Enhanced semantic model based on seq2seq with dual-encoder and Gain-Benefit gate.

### A. DUAL-ENCODER BASED ON BIDIRECTIONAL RECURSIVE NEURAL NETWORK

In this section, the design of dual-encoder is mainly introduced. The abstractive generation of text summary requires not only semantic comprehension of the full text but also compression and reconstruction of the original text. It can not generate high-quality summary through parallel alignment of two languages as machine translation tasks. In order to solve the problem of incomplete and inadequate semantic information output by the encoder in the traditional Seq2Seq model, in this paper, a dual-encoder structure is designed to obtain global and local context semantic information of the original text. As shown in Figure 1, The high-level encoder maps the text input sequence  $(x_1, x_2, \dots, x_n)$  into a high-dimensional semantic vector  $M$  through a bidirectional recurrent neural network accompanied with a self-attention mechanism. The low-level encoder focuses on the consistent representation of local semantics of the original text. The text input sequence  $(x_1, x_2, \dots, x_n)$  is mapped to the hidden state vector  $(h_1, h_2, \dots, h_n)$  of the encoder through BiRNN. The specific calculation method of the dual-encoder is described as:

$$H = (H_1, H_2, \dots, H_m) = \text{BiRNN}(x_1, x_2, \dots, x_m) \quad (1)$$

$$S_i^H = V_H \tanh(W_H H_i) \quad (2)$$

$$a_i^H = \frac{\exp(S_i^H)}{\sum_i^n \exp(S_i^H)} \quad (3)$$

$$M = \sum_i^n a_i^H H_i \quad (4)$$

In equation (1), the original text is transformed into a hidden state representation of high-level encoder  $H$  by a bidirectional recurrent neural network. Then,  $H$  is mapped to the similarity score by non-linear activation function  $\tanh$ , where  $V_H$  and  $W_H$  means the optimization parameter matrix. Equation (3)

denotes that  $S_i^H$  is normalized by Softmax. In equation (4),  $\mathbf{M}$  is obtained by weighted summation of normalized score and encoder's hidden state information.  $\mathbf{M}$  represents the global semantic vector representation of the original text.

### B. ENHANCED ATTENTION MECHANISM THAT INTEGRATES GLOBAL AND LOCAL SEMANTICS

In this section, the design of enhanced attention is mainly introduced. The attention mechanism originates from the field of computer vision. The core idea of attention mechanism is to get the target area which needs to be focused after browsing the global image. Subsequently, the attention mechanism was introduced into the field of text processing, which achieved good results.

In the task of text summary, attention mechanism is a connection architecture between encoder and decoder to solve the problem of inadequate text semantic understanding in the Seq2Seq model. The traditional attention mechanism integrates the hidden state of encoder and decoder into high-dimensional semantic representation by weighted summation. The purpose is to indicate which word semantics in the sequence should be paid more attention to by decoder.

On this foundation, the method proposed in this paper integrates the global and local semantic information obtained by dual-encoder into the attention mechanism, to enrich the semantic information representation of the original text, and enhance the model's reading comprehension about the text. The detailed design of the enhanced attention mechanism is expressed as follows:

$$\mathbf{h} = (\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_n) = \text{BiRNN}(x_1, x_2, \dots, x_n) \quad (5)$$

$$\mathbf{S}_i^L = S(\mathbf{h}_i, \mathbf{M}, \mathbf{S}_{t-1}) = \mathbf{V}_e^T \tanh(\mathbf{W}_h[\mathbf{h}_i, \mathbf{M}] + \mathbf{V}_s \mathbf{S}_{t-1}) \quad (6)$$

$$\mathbf{a}_{t_i}^L = \frac{\exp(\mathbf{S}_{t_i}^L)}{\sum_i^n \exp(\mathbf{S}_{t_i}^L)} \quad (7)$$

$$\mathbf{C}_{t-1} = \sum_{i=1}^n \mathbf{a}_{t_i}^L \mathbf{h}_i \quad (8)$$

In equation (5), the sequence  $(x_1, x_2, \dots, x_n)$  is mapped to the hidden state vector  $(\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_n)$  by the low-level encoder. In equation (6), the global semantic vector  $\mathbf{M}$  of the original text is cascaded with the hidden state  $\mathbf{h}_i$  of the low-level coder, and transformed into the semantic representation of the current state by optimizing parameter  $\mathbf{W}_h$ . The vector representation of the decoder's current state information is obtained by multiplying of hidden state  $\mathbf{S}_{t-1}$  and parameter  $\mathbf{V}_s$ . Then, the two parts are emerged by the alignment of the corresponding elements and mapped to a fusion state vector by the non-linear activation function  $\tanh$ . The fusion state vector combines the semantic information of the  $i$  step at the encoder and the decoder's hidden state information at the  $t - 1$  time. Where  $\mathbf{V}_e^T$  is optimization parameter. The essence of the process is to map the semantic representation of attention mechanism and the hidden layer state of

encoder-decoder to similarity score by the activation function. In equation (7), the scores are mapped into probability distributions through Softmax. Finally, the joint context semantic representation  $\mathbf{C}_{t-1}$  is calculated with the low-level state vector and the result of equation (7), as expressed in equation (8). Enhanced attention mechanism with dual-encoder structure combines global and local dual-channel semantics, it can be better focused on the abstract semantic features of the original text and acquired richer semantic understanding.

### C. DECODER WITH EMPIRICAL PROBABILITY DISTRIBUTION AND GAIN-BENEFIT GATE

In this section, a decoder is designed with the empirical probability distribution and Gain-Benefit gate. The traditional decoder maps the hidden vector at the current time  $t$ , the high-level semantics of the context, and the output of the decoder at time  $t-1$  to the candidate probability result at time  $t$  through the conditional probability function.

However, there are some problems in the decoding process. The essence of decoding is to maximize the conditional probability. The decoding of text summary accords with the rule of generation from left to right. The accuracy of the first few words of the generated summary is higher than that of the last few words. And the accuracy of the generated text summary is reduced from left to right. This phenomenon is called Semantics-Loss. In order to solve the Semantics-Loss problem in the process of text summary generation, the Gain-Benefit gate structure is designed to solve the semantic loss of decoding with the increase in length of text summary.

In addition, although the summary generated by the decoder is not compulsorily derived from the original text, the relevant part comes from the original text. On this basis, the empirical probability distribution of keywords is introduced to increase the tendency of the original words to be selected when the summary sequence is generated by the decoder, and to accelerate the convergence speed of the decoder.

$$p(y_t | x) = \text{Soft max}(\mathbf{W}_c \mathbf{C}_t + \mathbf{W}_o \mathbf{S}_t + \mathbf{Q}_o^S + \mathbf{P}_t + \mathbf{b}_o) \quad (9)$$

$$\mathbf{Q}_o^S = \mathbf{W}_Q \circ \mathbf{V} + \lambda_o \quad (10)$$

In equation (9), the SoftMax activation is used to normalize the probability of each predicted word at time  $t$  by the fully connected layer with the inputs. Where  $\mathbf{W}_c$ ,  $\mathbf{W}_o$ ,  $\mathbf{b}_o$  are the optimization parameters,  $\mathbf{C}_t$  is a contextual semantic representation of integration global and local semantic information,  $\mathbf{S}_t$  is the hidden state of the current time decoder.  $\mathbf{Q}_o^S$  is the empirical probability distribution of the candidate words of the output sequence, and it is designed in equation (10).  $\mathbf{W}_Q$  and  $\lambda_o$  are optimized parameter, and the  $\circ$  means the operation of Hadamard.  $\mathbf{V} = (v_1, v_2, \dots, v_m)$  represents 0/1 matrix composed of vocabularies, and  $v_i = 1$  when the word has appeared in the article, otherwise  $v_i = 0$ .

The design of  $Q_o^s$  can increase the output probability of candidate words that appear in the original text, and accelerate the convergence process. Importantly,  $P_t$  is the output vector of the Gain-Benefit gate, which is described as:

$$a = \text{sigmoid}(W_1 C_t + W_2 S_{t-1} + y_{t-1}) \quad (11)$$

$$P_t = (1 - a) \odot M + a \odot C_{t-1} \quad (12)$$

As mentioned above, there is a problem of Semantics-Loss in the decoding process, which means the prediction accuracy of words will gradually decrease with the increase of the length of the text summary. The reason is that the word  $y_{t-1}$  generated at time  $t-1$  needs to become the basis for word prediction at time  $t$ , but each prediction will lose part of the semantics, which will gradually reduce the accuracy. In order to solve this problem, a Gain-Benefit gate structure is designed. In equation (11), the semantic supplementary weight coefficient  $a$  is obtained through the activation function Sigmoid whose inputs are context semantic vector  $C_t$ , state vector  $S_{t-1}$  and  $y_{t-1}$ . In equation (12), the beneficial vector  $P_t$  is obtained by emerging between high-level semantic vector  $M$  multiplied by supplemental weight coefficient  $(1 - a)$  and context semantic vector  $C_{t-1}$  multiplied by supplemental weight coefficient  $a$ . The semantic supplementary weight coefficient  $a$  can control the contribution ratio between high-level semantic vector and context semantic vector to the decoder. This strategy can effectively supplement the semantic information lost in the previous step.

### III. IMPROVED WORD EMBEDDING

In this section, the improved word embedding is designed. Word2vec or Glove are usually used to generate word vectors in natural language processing experiments. The former uses the vector representation of current words to infer the vector representation of surrounding words. The latter uses global semantic information to make training faster. In principle, both approaches focus on the relative position of words in vector representation space, which is semantic similarity. In order to enrich semantics, position vectors are emerged into word vectors and new word features are added to the model.

#### A. ADD NEW FEATURES FOR WORDS

The generated text summary should express the comprehensible semantics containing the whole passage in a limited number of words. So the order of words has a great influence on the quality of the summary, such as “I love you” and “you love me”, the reversal of the position of ‘you’ and ‘me’ changed the subject-object relationship of the sentence and also changed the original expression of the summary. It has a significant impact on the summary results and quality.

This paper combines position embedding [25] and word embedding to form a new word vector representation. In addition, POS (Part of speech), TF-IDF and Sco (key score) are added to enrich the semantic features of words, which makes the characterization of words more abundant and sufficient,

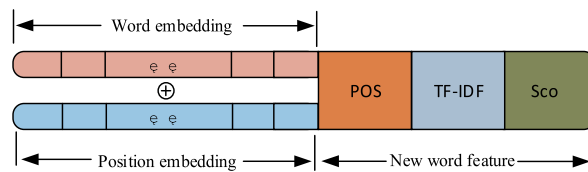


FIGURE 2. Improved word embedding model.

and provides better feature expression for model training. The improved word embedding model is shown in Figure 2.

In this paper, all elements of words’ dictionary  $\{x_1, x_2, \dots, x_m\}$  are mapping into word vectors  $W = (w_1, \dots, w_m)$ ,  $w_i \in R^f$ . Meanwhile, the position of the word in the text is mapped to the position vector  $P = (p_1, \dots, p_m)$ ,  $p_j \in R^f$ . Combining position embedding and word embedding into a fusion vector  $U = (w_1 + p_1, \dots, w_m + p_m)$ . The generating algorithm of position embedding is described as follows.

$$PE_{(pos, 2i)} = \sin(pos/10000^{2i/d_{model}})$$

$$PE_{(pos, 2i+1)} = \cos(pos/10000^{2i/d_{model}}) \quad (13)$$

Note: in the text that is segmented into length  $n$ , the position  $POS$  of the word is mapped to a  $d_{model}$  dimensional position vector. And the value of the  $i - th$  dimension of the position vector is  $PE_{(pos, i)}$ . The position vectors seem to be the absolute positions generated by formulas, but the vector of position  $(\alpha + \beta)$  can be expressed as the linear transformation of position  $\alpha$  vector with the help of the properties of triangular functions of  $\sin(\alpha + \beta)$  and  $\cos(\alpha + \beta)$ , and it also provides the possibility of the expression of relative position, which means that there are similarities with the essence of word vector representation.

Most of the generated text abstracts are based on declarative summary sentences, focusing more on nouns and verbs, and paying less attention to adjectives with emotional color. The TF-IDF index can reflect the comprehensive features of words in the corpus. The POS (part of speech), TF-IDF and Sco (key score) of words are simply connected at the back end of the fusion vector  $U$  to form a new word vector. The new vector as inputs of the encoder. The definition of Sco is shown in equation (14).

$$Sco(w_i) = \frac{\beta}{p(w_i) + \beta} \quad (14)$$

where  $P(w_i)$  is the word frequency and  $\beta$  is the smooth inverse frequency coefficient. In texts, the key degree of words is negatively correlated with their frequency because of words’ number limitation, and the degree of correlation varies with the corpus. The parameter  $\beta$  was introduced to balance the effect of frequency on the criticality of words. Less frequent occurrences may be more critical.

#### B. OPTIMIZING THE DIMENSION OF WORD EMBEDDING

In this section, we mainly introduce the dimension optimization algorithm of word vector. Word2vec and Glove are the most popular word embedding methods at present, but the

choice of word embedding dimension is quite arbitrary at present. There is no certain theoretical basis. The use of arbitrary embedding dimension in different corpus size can not guarantee the optimality of the model, which will inevitably affect the quality of semantic representation. A large number of experimental studies have shown that too small selection of word embedding learning dimension will result in semantic loss, and too large dimension will pull in too much noise.

A good quality semantic high-dimensional matrix cannot change the relative position of words in space no matter how the matrix is transformed. In this paper, PIP (Pairwise Inner Product) [26] is designed based on matrix perturbation theory to design the optimal embedding dimension for the current corpus and improve the accuracy of word vector representation. The loss function PIP is described as:

$$PIP(\mathbf{E}) = \mathbf{E}\mathbf{E}^T \quad (15)$$

$$E[\|\hat{\mathbf{E}}\hat{\mathbf{E}}^T - \mathbf{E}^*\mathbf{E}^{*T}\|] \leq \sqrt{\sum_{i=k+1}^d \lambda_i^{4\alpha} + 2\sqrt{2n\alpha}\sigma \sqrt{\sum_{i=1}^k \lambda_i^{4\alpha-2}}} \\ + \sqrt{2} \sum_{i=1}^k (\lambda_i^{2\alpha} - \lambda_{i+1}^{2\alpha})\sigma \\ \times \sqrt{\sum_{r \leq i < s} (\lambda_r - \lambda_s)^{-2}} \quad (16)$$

Equation (15) is defined as the PIP matrix, and equation (16) is the expected estimator of the loss function. Where  $\hat{\mathbf{E}} \in R^{n \times k}$  means the obtained word vector matrix,  $\mathbf{E}^* \in R^{n \times d}$  is the optimal word vector matrix not obtained,  $k, d$  represent dimension  $k < d$ ,  $\sqrt{\sum_{i=k+1}^d \lambda_i^{4\alpha}}$  is the deviation, which represents after  $k+1$  dimension will be lost when the word with the limited  $k$  dimension of word embedding.  $2\sqrt{2n\alpha}\sigma \sqrt{\sum_{i=1}^k \lambda_i^{4\alpha-2}}$  and  $\sqrt{2} \sum_{i=1}^k (\lambda_i^{2\alpha} - \lambda_{i+1}^{2\alpha})\sigma \sqrt{\sum_{r \leq i < s} (\lambda_r - \lambda_s)^{-2}}$  is variance.  $2\sqrt{2n\alpha}\sigma \sqrt{\sum_{i=1}^k \lambda_i^{4\alpha-2}}$  denotes that the existence of noise leads to errors in estimating the size of the semantic matrix, which increases with the increase of  $k$ .  $\sqrt{2} \sum_{i=1}^k (\lambda_i^{2\alpha} - \lambda_{i+1}^{2\alpha})\sigma \sqrt{\sum_{r \leq i < s} (\lambda_r - \lambda_s)^{-2}}$  represents the estimation error in the direction of semantic matrix with the existence of noise, which increases with the increase of  $k$ .  $\alpha \in (0, 1]$ ,  $\sigma$  is the standard deviation for noise.  $\lambda_i$  is the  $i$ -th empirical singular value parameter. In order to obtain the best dimension of the word embedding matrix, the essence is to find the balance point between deviation and variance, that is, the minimum value of the PIP loss function.

## IV. EXPERIMENTS AND ANALYSIS

### A. EXPERIMENTAL DATA SELECTION

In this section, we mainly introduce experimental design and experimental analysis. The experimental data sets of this paper adopt the LCSTS dataset, which was proposed by Hu et al. [7], and the SOGOU dataset, which was proposed

by Sogou smart-Lab with multiple categories text summary. These two data sets contain news-headline data pairs from different categories in entertainment, culture, education, military, social, financial and other fields. The LCSTS data set contains three parts, and the first part is selected of this with 2400591 pairs of text abstracts in this paper. The SOGOU data set contains 1359956 pairs of news-headline data. Besides, there are 174 pairs of long text summary data are included in SOGOU data set, which are used to verify the effect of the proposed method on long text data. Note: All datasets used in this article have been translated into English by Google Translate.

For the task of text summary generation, the quality of corpus data also affects the final experimental results. Data sets need to be filtered to extract high-quality text summary data. Firstly, the datasets were preprocessed to remove the text with length less than 25 and replaces the chaotic characters, special characters and emoticons. Secondly, text abstracts are highly compressed and reconstructed based on language understanding. Comparing with the number of text and summary, the number of text is reduced, but it still has a certain degree of correlation.

According to the semantic similarity between the abstract and the original text, the data are divided into three levels to select high-quality experimental data pairs. 1 means the least relevant, 3 means the most relevant. Text-summary semantic similarity has a correlation of 1 in the interval (0,0.15), 2 in the interval (0.15,0.65), and 3 in the interval (0.65,1). A semantic similarity algorithm is designed as follows.

$$\mathbf{v}_s = \frac{1}{|s|} \sum_{w \in s} IDF_w \mathbf{v}_w \quad (17)$$

$$\cos \theta = \frac{a \cdot b}{\|a\| \|b\|} = \frac{\sum_{i=1}^n \mathbf{A}_i \mathbf{B}_i}{\sqrt{\sum_{i=1}^n \mathbf{A}_i^2} \sqrt{\sum_{i=1}^n \mathbf{B}_i^2}} \quad (18)$$

Equation (17) focuses on the generation of sentence vectors.  $|s|$  is the number of words contained in the current sentence,  $IDF_w$  means the inverse document frequency index of word documents, and  $\mathbf{v}_w$  represents the word vector. The similarity between text-summary pairs is calculated by cosine distance as Equation (18). The partition of data sets is described as shown in Table 1.

As shown in Table 1, we divide the data into three levels of correlation according to the semantic similarity. On the LCSTS dataset, as shown in Data Set I, the total number of data is 2400591, which is used as the training set. Data Set II is a validation set of 11504 randomly selected data from the data of correlation degree 2 and 3 of Data Set I. Data Set III contains 2535 pairs data, which is selected from Data Set I of correlation degree 3. Data Set III data sets are used as test sets, and these data will not be included in the training set I. Similarly, the partition results of the SOGOU dataset are shown in the table above.

TABLE 1. Division of experimental data set.

Datasets	Class	Correlation	Numbers
LCSTS	I	1	521805
		2	1788071
		3	90715
	II	2	7143
		3	4361
		III	3
SOGOU	I	1	30804
		2	1103296
		3	225856
	II	2	6283
		3	5912
		III	3

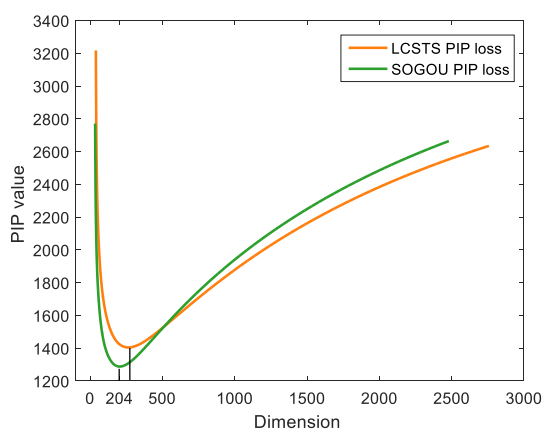


FIGURE 3. Dimension optimal choice of word embedding.

B. EXPERIMENTAL PARAMETERS

In this paper, we use the Jieba word segmentation tool to segment the corpus. We select 20,000 high frequency words as the Dictionary Vocabulary of the coder. The optimal word embedding dimension of the LCSTS dataset is 263, and the optimal dimension of the SOGOU dataset is 204, which are calculated by the paired internal product loss function PIP as shown in Figure 3.

The high-level encoder uses 200 GRU neurons with a 3-layer BiRNN structure, and the low-level encoder uses 200 GRU neurons with a 3-layer BiRNN structure. The decoder uses the beam search method with a beam size of 5 and batch\_size of 64,  $\beta = 10^{-3}$  is smoothing parameters. The proposed En-semantic model with dual encoders runs on the server PowerEdge R210 II for nearly six days to get the final result.

The optimal dimension of word embedding is related to the size of the corpus. Different size corpus is accompanied by different word number, dictionary size and word structure. The essentials of finding the optimal dimension of word embedding is to calculate the best balance between deviation and variance of the word vector matrix. As shown in Figure 3,

the PIP value of word embedding loss function for LCSTS dataset decreases with the increase of dimension in 0-263 dimension, and increases with the increase of dimension after 263, and reaches the minimum at 263 dimensions. Consequently, 263 is the optimal dimension of the LCSTS dataset. Similarly, the optimal dimension of the SOGOU dataset calculated by the PIP function is 204.

The quality evaluation methods of generative text summary can be divided into two categories: external evaluation method and internal evaluation method. The external evaluation method is to use the generated summary to perform the related tasks of this paper, such as document retrieval, document classification and so on. The quality of the summary can be judged according to its application performance level. The internal evaluation method needs to provide reference summary. The higher the coincidence between the reference abstract and the generated abstract, the better the quality of the generated abstract. The most commonly used method is internal evaluation. EDMUNDSON and ROUGE [27] are the most frequently used internal methods for text summary tasks in China and abroad. ROUGE is widely used in text summary task evaluation. This paper, the ROUGE evaluation system is used to evaluate the quality of the text summary. ROUGE is a method for evaluating the quality of abstracts based on the co-occurrence information of n-gram in abstracts, which is oriented to the recall rate of the n-gram. In this paper, ROUGE-1, ROUGE-2 and ROUGE-L are used to evaluate summary quality and model performance.

C. ANALYSIS OF EXPERIMENTAL RESULTS

In this paper, we use the above data sets to carry out experiments, including comparison experiments with references [7], [11], [13], [24] and a self-contrastive experiment was conducted on whether the fusion position was embedded or not, and whether the dimension of word embedding was optimized or not. Details of the experiment are expressed in Table 2.

Seq2Seq+attention [7]: represents a neural network model based on sequence-to-sequence with the self-attention mechanism.

Abstractive model [11]: stands for the text summary model combining multiple features based on the sequence-to-sequence model with attention mechanism.

Multi-head attention [13]: means the model of text summary based on a multi-head attention mechanism.

Sy-Bi+multi-head atten [24]: represents the synchronous bidirectional mechanism, which contains left-to-right and right-to-left decoding strategies in the decoder.

En-semantic-model+w2cPro(our): represents the enhanced semantic model proposed in this paper, which includes new word features without position embedding.

En-semantic-model+pos-w2cPro(our): is the enhanced semantic model proposed in this paper with the word embedding optimization algorithm that integrates position embedding, word embedding and new word features.

**TABLE 2. Comparison of Rouge evaluation indicators between different models on LCSTS dataset.**

Model	Rouge-1	Rouge-2	Rouge-L
Seq2Seq+attention	0.2770	0.1524	0.2432
Abstractive model	0.3454	0.1657	0.2804
Multi-head attention	0.3522	0.1719	0.3315
Sy Bidierct+multi-head atten	0.3758	0.1732	0.3443
En-semantic-model+w2cPro	0.3807	0.1912	0.3628
En-semantic-model+pos-w2cPro	0.3929	0.2123	0.3632
En-semantic-mode+pos-w2cPro+dimen	<b>0.4139</b>	<b>0.2231</b>	<b>0.3722</b>

**TABLE 3. Comparison of Rouge evaluation indicators between different models on SOGOU dataset.**

Model	Rouge-1	Rouge-2	Rouge-L
Seq2Seq+attention	0.2657	0.1543	0.2441
Abstractive model	0.3532	0.1664	0.2804
Multi-head attention	0.3533	0.1725	0.3347
Sy-Bi+multi-head atten	0.3779	0.1737	0.3426
En-semantic-model+w2cPro	0.3835	0.1813	0.3626
En-semantic-model+pos-w2cPro	0.3941	0.2132	0.3731
En-semantic-mode+pos-w2cPro+dimen	<b>0.4054</b>	<b>0.2442</b>	<b>0.3849</b>

En-semantic-mode+pos-w2cPro+dimen(our): represents the optimal dimension of word embedding based on the last experiment.

From Table 2 and Table 3, we can observe that the results of the model in this paper are better than those of other listed models in the ROUGE evaluation system. The ROUGE value on LCSTS dataset is increased by 0.1-0.137, and the ROUGE value on the SOGOU dataset is increased by 0.1-0.139. which shows that the model with dual-encoder and Enhanced-attention designed in this paper has richer context semantics and better language understanding. And through the self-contrastive experiments of the last three groups, we can see that the improved word embedding technology, including new features and optimization dimensions, has played a good role in assisting the model. On the LCSTS dataset, the ROUGE value is 0.01-0.03 higher in combining with position embedding than that without position embedding, and the ROUGE value with dimension optimization is 0.01-0.03 higher than that without dimension optimization. Similarly, on the SOGOU dataset, the ROUGE value is increased by 0.02-0.06 compared to the method without dimensional optimization and position embedding.

Table 4 is an example of several text summary tests, including the original text and summaries generated using different models. The models are Multi-head attention, Sy-Bi+multi-head atten,and En-semantic-model+pos-w2cPro+dimen (our).

**TABLE 4. Text summary comparison generated by different models on SOGOU dataset.**

<p><b>Source text:</b> There are many ways to improve memory. In daily life, we must develop good habits, especially pay attention to two aspects of eating and sleeping. Good eating habits can provide the body with sufficient energy and physical strength, and secondly, ensure adequate sleep and high quality sleep. Good sleep can give the brain a good rest and help improve memory. Some people have poor memory because of poor sleep. Most of the problems occur during work or study. Because of poor sleep, they can't focus on it, making the brain's memory poor. In addition, there are also ways to strengthen brain memory, such as reviewing knowledge to a certain extent when getting up in the morning and before going to bed at night can also effectively improve memory.</p> <p><b>Reference summary:</b> There are many ways to improve memory, such as developing good eating and sleeping habits, and enough rest in the brain can improve memory.</p>
<p><b>Sy-Bi+multi-head atten:</b> In order to improve memory. We should pay special attention to diet and sleep in daily life. The brain is fully rested, attention can be improved, and memory can be good.</p>
<p><b>Multi-head attention:</b> To improve memory, you must develop good habits, pay attention to diet and sleep, and then use some methods to strengthen brain memory.</p>
<p><b>En-semantic-model+pos-w2cPro+dimen(our):</b> There are some method to improve memory by developing good eating and sleeping habits in daily life, and high quality sleep can rest the brain and improve memory.</p>

Table 4 is an example of experiment generation summary, which contains the original text and the summary generated by each method.

From example, we can see that the method designed in this paper has the characteristics of dual-channel semantic association and new word feature, and the accuracy of abstract is higher. The noun “eating and sleeping” is clearly extracted, and the generated word “methods” is synonymous with the original word “ways”. Comparing with the core word “diet and sleep” generated by other methods, the accuracy is not enough, and the accuracy rate gradually decreases as the sentence length increases. The sentence generated by the method proposed in this paper performs well overall. The experimental comparison expresses that the advanced semantic model in this paper has a good performance on the ability of text compression and remodeling, and has a better understanding of semantics and a higher quality of summary.

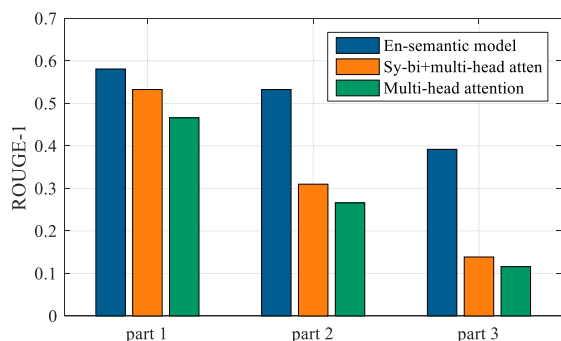
In addition, in order to verify the effectiveness of the proposed method in long text, we selected 174 pairs of long text summary data samples from the SOGOU dataset for experimental verification. The original text of this data sample contains 400-500 words, and the summary is that the central idea of the original text contains 60-80 words. We use this part of the long text dataset to verify the effectiveness of the method proposed in this article. Details are shown in Table 5.

From Table 5, we can observe that in the long text experiment, the text generation quality of each model in the ROUGE evaluation system is between 0.25-0.35. However, the method proposed in this paper reached 0.39 in the ROUGE evaluation system, which improved the ROUGE value by 0.07-0.14 compared with other methods. The results show



**TABLE 5. Comparison of Rouge evaluation indicators between different models on long text from SOGOU dataset.**

Model	Rouge-1	Rouge-2	Rouge-L
Seq2Seq+attention	0.2573	0.1451	0.2265
Abstractive model	0.3326	0.1624	0.3143
Multi-head attention	0.3452	0.1725	0.3216
Sy-Bi+multi-head atten	0.3534	0.1706	0.3341
En-semantic-model+w2cPro	0.3726	0.1963	0.3501
En-semantic-model+pos-w2cPro	0.3831	0.2123	0.3621
En-semantic-mode+pos-w2cPro+dimen	<b>0.3913</b>	<b>0.2223</b>	<b>0.3709</b>



**FIGURE 4. The comparison of semantics-loss of different models.**

that the Gain-Benefit gate structure designed in this paper plays a positive role, which makes the method proposed in this paper have better results.

Meanwhile, we also verify the performance of different methods in the face of the problem of Semantics-Loss. We have selected two methods with better performance, Sy-Bi+multi-head atten and Multi-head attention, to compare with the method proposed in this article. The sentence of summary generated by the test set is divided into three parts, part 1, part 2, and part 3 from left to right. Part 1 represents the front part of the abstract, part 2 represents the middle part of the summary, and part 3 represents the end part. Because ROUGE-1 is a 1-gram unit word structure, it can accurately evaluate the accuracy of generating words. So, the ROUGE-1 is used to verify the scoring and semantic loss of different parts as the sentence length increases. Details are shown in Figure 4.

As shown in Figure 4, the ROUGE-1 value of the Multi-head attention mechanism method is reduced by 0.3-0.35 from part 1 to part 3, which indicates that the semantic loss is large. The ROUGE-1 value of the Synchronous Bidirectional decoding model was reduced by 0.23-0.39 from part 1 to part 3. The enhanced semantics method proposed in this paper is only reduced by 0.048-0.18 from part 1 to part 3. Experiments show that the Gain-Benefit gate structure designed on the decoder side of this method plays a positive role in the summary generation process. The Gain-Benefit gate weakens the problem of Semantics-Loss as the summary

length increases. It can be seen that the method proposed in this paper has a precise understanding of text semantics, and the generated summary has high accuracy.

**V. CONCLUSION**

Through studying and researching the task of generative text summary, this paper proposes a text summary generation model (En-semantic model) with dual-encoder and Gain-Benefit gate to solve the problems of and Semantics-Loss and insufficient use of context semantic information, insufficient semantics understanding of traditional attention mechanism and low accuracy of generating summary in current text summary generation model. Our model integrating global and local semantic information to improve language comprehension. Meanwhile, new word vectors are synthesized by combining position embedding and word embedding methods, and parts of word, word's index inverse document frequency and word's key-score are integrated into word vectors to improve the understanding of words in the model. Secondly, for the skip-gram model of word2vec, the word embedding matrix is optimized by the pairwise inner product loss function (PIP) with unitary invariant properties, and the best word embedding dimension is selected for the current corpus, and the best performance of word vector representation is achieved. Finally, through the ROUGE evaluation system, the enhanced semantic model proposed in this paper has better performance than other listed models.

The text summary generation task is an important branch of natural language processing, and has important value in practical application tasks such as document retrieval, text classification, and text clustering. The performance of this model needs to be improved when dealing with texts with special names and unfamiliar nouns, so further research will be done in this area in future research.

**REFERENCES**

- [1] A. Wei Yu, D. Dohan, M.-T. Luong, R. Zhao, K. Chen, M. Norouzi, and Q. V. Le, "QANet: Combining local convolution with global self-attention for reading comprehension," 2018, *arXiv:1804.09541*. [Online]. Available: <http://arxiv.org/abs/1804.09541>
- [2] L. Xiao, N. Wang, and G. Yang, "A reading comprehension style question answering model based on attention mechanism," in *Proc. IEEE 29th Int. Conf. Appl.-Specific Syst., Archit. Processors (ASAP)*, Milan, Italy, Jul. 2018, pp. 1-4.
- [3] K. Chen, T. Zhao, M. Yang, L. Liu, A. Tamura, R. Wang, M. Utiyama, and E. Sumita, "A neural approach to source dependence based context model for statistical machine translation," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 2, pp. 266-280, Feb. 2018.
- [4] M. Gambhir and V. Gupta, "Recent automatic text summarization techniques," *Artif. Intell. Rev.*, vol. 47, no. 1, pp. 1-66, Mar. 2017.
- [5] K. Lopyrev, "Generating news headlines with recurrent neural networks," 2015, *arXiv:1512.01712*. [Online]. Available: <http://arxiv.org/abs/1512.01712>
- [6] A. M. Rush, S. Chopra, and J. Weston, "A neural attention model for abstractive sentence summarization," 2015, *arXiv:1509.00685*. [Online]. Available: <http://arxiv.org/abs/1509.00685>
- [7] B. Hu, Q. Chen, and F. Zhu, "LCSTS: A large scale chinese short text summarization dataset," 2015, *arXiv:1506.05865*. [Online]. Available: <http://arxiv.org/abs/1506.05865>

- [8] H. Zhang, J. Li, Y. Ji, and H. Yue, "Understanding subtitles by character-level sequence-to-sequence learning," *IEEE Trans. Ind. Informat.*, vol. 13, no. 2, pp. 616–624, Apr. 2017.
- [9] A. See, P. J. Liu, and C. D. Manning, "Get to the point: Summarization with pointer-generator networks," 2017, *arXiv:1704.04368*. [Online]. Available: <http://arxiv.org/abs/1704.04368>
- [10] G. Chen, "Chinese short text summary generation model integrating multi-level semantic information," in *Proc. Int. Conf. Netw., Commun., Comput. Eng. (NCCE)*, Beijing, China, 2018, pp. 1–12.
- [11] R. Nallapati, B. Zhou, C. dos Santos, C. Gulcehre, and B. Xiang, "Abstractive text summarization using sequence-to-sequence RNNs and beyond," in *Proc. 20th SIGNLL Conf. Comput. Natural Lang. Learn.*, Berlin, Germany, 2016, pp. 1–12.
- [12] A. See, P. J. Liu, and C. D. Manning, "Get to the point: Summarization with pointer-generator networks," in *Proc. 55th Annu. Meeting Assoc. Comput. Linguistics*, vol. 1, 2017, pp. 1073–1083.
- [13] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Conf. Neural Inf. Process. Syst.*, New York, NY, USA, 2017, pp. 1–15.
- [14] Q. Guo, J. Huang, N. Xiong, and P. Wang, "MS-pointer network: Abstractive text summary based on multi-head self-attention," *IEEE Access*, vol. 7, pp. 138603–138613, 2019.
- [15] Z. Tan, M. Wang, J. Xie, Y. Chen, and X. Shi, "Deep semantic role labeling with self-attention," in *Proc. AAAI*, New York, NY, USA, 2018, pp. 1–8.
- [16] Z. Tu, Z. Lu, Y. Liu, X. Liu, and H. Li, "Modeling coverage for neural machine translation," in *Proc. 54th Annu. Meeting Assoc. Comput. Linguistics*, Berlin, Germany, 2016, pp. 1–11.
- [17] O. Levy and Y. Goldberg, "Linguistic regularities in sparse and explicit word representations," in *Proc. 18th Conf. Comput. Natural Lang. Learn.*, 2014, pp. 746–751.
- [18] B. Rekadbar, C. Mousas, and B. Gupta, "Generative adversarial network with policy gradient for text summarization," in *Proc. IEEE 13th Int. Conf. Semantic Comput. (ICSC)*, New York, NY, USA, Jan. 2019, pp. 204–207.
- [19] H. Zhang, J. Xu, and J. Wang, "Pretraining-based natural language generation for text summarization," 2019, *arXiv:1902.09243*. [Online]. Available: <http://arxiv.org/abs/1902.09243>
- [20] M. Artetxe and H. Schwenk, "Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond," *Trans. Assoc. Comput. Linguistics*, vol. 7, pp. 597–610, Mar. 2019.
- [21] S. Arora, Y. Liang, and T. Ma, "A simple but tough-to-beat baseline for sentence embeddings," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Paris, France, Apr. 2016, pp. 1–16.
- [22] D. Chatzakou, N. Passalis, and A. Vakali, "Multispot: Spotting sentiments with semantic aware multilevel cascaded analysis," in *Proc. Int. Conf. Big Data Analytics Knowl. Discovery*. Madrid Spain: Springer, Sep. 2015, pp. 337–350.
- [23] J. Han Lau and T. Baldwin, "An empirical evaluation of doc2vec with practical insights into document embedding generation," 2016, *arXiv:1607.05368*. [Online]. Available: <http://arxiv.org/abs/1607.05368>
- [24] L. Zhou, J. Zhang, and C. Zong, "Synchronous bidirectional neural machine translation," *Trans. Assoc. Comput. Linguistics*, vol. 7, pp. 91–105, Mar. 2019.
- [25] J. Gehring, M. Auli, and D. Grangier, "Convolutional sequence to sequence learning," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Sydney, NSW, Australia, Aug. 2017, pp. 1243–1252.
- [26] Z. Yin and Y. Shen, "On the dimensionality of word embedding," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Toronto, CA, Canada, May 2018, pp. 887–898.
- [27] C.-Y. Lin and E. Hovy, "Automatic evaluation of summaries using N-gram co-occurrence statistics," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics Hum. Lang. Technol. (NAACL)*, Toronto, CA, Canada, Sep. 2003, pp. 150–157.



**JIANLI DING** was born in Henan, China, in 1963. He received the M.S. degree in management science and engineering from the University of Science and Technology of China, in 1995, and the Ph.D. degree in operations research and control theory from Nankai University, in 2004.

Since August 2004, he has been working with the Civil Aviation University of China. He has hosted and participated in several national natural science foundation projects. He is currently a Professor with the College of Computer Science and Technology. His research interests include civil aviation intelligent information processing, the aviation Internet of Things, and intelligent bionic algorithm.



**YANG LI** was born in Shandong, China, in 1995. He received the B.Sc. degree in software engineering from Qufu Normal University, in 2017. He is currently pursuing the master's degree with the Civil Aviation University of China.

He has published many articles in domestic and foreign journals. He is also involved in research on nonlinear control theory, focus on constraint consensus problem, agents formation control, and flocking control with time-varying delay. From 2017 to 2019, his research interests include natural language processing, machine learning, deep learning, and multidimensional trajectory anomaly detection.



**HUIYU NI** was born in Shandong, China, in 1993. She received the B.Sc. degree in computer science and technology from Qufu Normal University, in 2017. She is currently pursuing the master's degree with the Civil Aviation University of China.

From 2017 to 2019, her research interests include network security, information security, and deep learning in information security monitoring and evaluation. She is also interested in civil aviation information system security and impacts of system security on business. She is also studying on civil aviation business importance and the risks and impacts of business continuity.



**ZHENGQUAN YANG** received the B.S. degree in mathematics from the Institute of Mathematics, Qufu Normal University, Qufu, China, in 2003, and the Ph.D. degree in operational research and control theory from Nankai University, Tianjin, China, in 2009.

Since January 2009, he has been working with the Civil Aviation University of China. He is involved in several national natural science foundation projects. He is currently an Associate Professor with the College of Science. His research interests include synchronization of complex networks and flocking of multiagents.

• • •