Research on the Application of Deep Learning in Text Generation

Shuohua Zhou1,*

1School of Information, Renmin University of China, BeiJing, 100872, China
*Corresponding author’s e-mail: zshgrey@126.com

Abstract. In recent years, deep learning technology has made great progress, and text generation based on deep learning has received extensive attention. In the form of complex and diverse information, text information is a mainstream form of data, and its number is growing very fast. How to quickly and accurately locate and use effective information from massive text data has become an urgent research problem in the field of text information extraction. The deepening of deep learning technology has broken through the dependence of traditional natural language generation technology on templates. It can automatically learn the input to output mapping from the data to form an end-to-end solution and reduce the degree of human participation. It enables the generation system to generalize in a wider field, and can generate more free text under the given conditions according to the needs. This paper explores the text generation method based on deep neural network, conducts research work on the text summary task, designs a generative summary generation method based on improved cluster search, and conducts experiments. Experimental results show that this method is effective.

1. Introduction

At present, we are in the age of the Internet and generate a lot of text content every day. Every user of the Internet can be a content producer. They can publish and modify content on the Internet, so the content on the Internet can increase a lot every day. We can get what we want from the Internet at any time. At the same time, there is a lot of invalid content in the massive content on the Internet, which has caused great obstacles for people to quickly obtain effective content [1]. In the 1990s, with the continuous development of natural language processing technology, some scholars proposed a generative text summary method. This method imitates people's thinking based on neural networks, and then trains to generate text summaries. From the classification of processing objects, automatic summary technology can be divided into single document automatic summary and multi-document automatic summary. The continuous development of neural network technology has made deep learning a breakthrough in the field of text generation. Generative automatic summarization is an important part of the field of text generation. Research hotspots in the field of automatic summarization technology have gradually shifted from extractive to generative, and generative automatic summarization technology has received widespread attention [2].

2. Deep learning

2.1 Deep learning

One of the sub-fields of machine learning is deep learning. Deep learning originated from the study of neural networks and is an End2End learning structure that contains multiple layers of networks. Deep
learning algorithms improve the effectiveness of the algorithm by adding more layers. Deep learning allows researchers to focus on how to build a deeper network and let the network automatically learn effective combination features [3].

Deep learning algorithms are mainly divided into three categories: Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), and Deep Neural Network (DNN). CNN is mainly used in image processing. It learns the spatial features of images through different convolution kernels, and then uses pooling techniques to combine different image features to improve the effect of the algorithm [4]. DNN keeps stacking hidden layers of different structures to improve the algorithm's ability to understand abstract features, so it is suitable for search and advertising applications that require a lot of abstract features. RNN uses recursive form to learn the characteristics of time series, so it is suitable for natural language processing scenarios. The text summary generation method proposed in this paper is based on the RNN model and can learn the time series information in the text content, which is helpful for the generation of text summaries [5].

2.2 Recurrent Neural Network
In the summarization model based on deep learning, the input document is a sequence, and the document needs to be represented in vector. RNN is a model learned on sequence data, which can process input sequences with varying lengths [6].

Recurrent neural networks were introduced thirty years ago. The traditional neural network model is mainly based on the feedforward neural network. The input dimension information of the feedforward neural network is independent, and the dimension of the input information is fixed. In natural language processing tasks, each word depends on the previously entered words. Natural language input is a sequence of information, the length of the sequence is variable, and there is a connection between the sequence information. Therefore, the feedforward neural network is restricted in natural language processing tasks [7].

RNN has a memory function. It can remember the information input from the beginning to the current network. RNN has the feature of being able to memorize previously input information. And it will affect the output of subsequent networks. The function of the recurrent neural network is very powerful, and the sequence information it processes is not limited by the length. Recurrent neural network serial input sequence information. Only after the information at the previous time is processed, the information at the next time can be processed.

RNN is composed of three parts: output layer, hidden layer, and input layer. At each moment of the RNN, the structure of the neuron is similar to that of a normal feed-forward neural network. The difference between them is that the recurrent neural network shares the parameters and structure of a neuron at all times, and the current output of the network depends on the input of the previous time.

![Figure 1. Recurrent neural network structure](image)

The basic principle is shown in formulas (1) to (6).

\[ S_t = U \cdot h_{t-1} + W \cdot x_t \]  \hspace{1cm} (1)

\[ h_t = \tanh(s_t) \]  \hspace{1cm} (2)

\[ z_t = V \cdot h_t \]  \hspace{1cm} (3)
\[ y_t = \text{soft max} (z_t) \]  
\[ E_t = -y_t \cdot \log (y_t) \]  
\[ E = \sum_{t} E_t \]  

In the above formula, \( x_t \) represents the input at the current time; \( s_t \) represents the input at the hidden layer at the current time; \( h_t \) represents the output at the hidden layer at the current time; \( z_t \) represents the input at the current layer; \( y_t \) represents the output at the current time layer; \( E_t \) represents the loss value at the current moment, and \( E \) represents the sum of the loss values at all moments. The goal of RNN is to make the value of \( E \) the lowest through data training. At this time, the prediction effect of the RNN model reaches the best [8].

3. Automatic summary model based on deep learning

Related technologies of deep learning are widely used in NLP. Attention mechanism is introduced into the structure of self-encoder. The attention mechanism is a focused thought, which makes the neural network have the ability to reorganize the input information. Analyze the problem, make each item of the original data a zoom-in or zoom-out transformation, and enlarge the part related to the problem, and vice versa. This design also introduces this attention mechanism. The task processing flow is shown in Figure 2 and consists of four steps. The four steps are described in detail below [9].

1) Text preprocessing stage. After word segmentation of the original information, the word vectorization process is carried out. In this process, including the calculation of part of speech, word frequency, and inverse text frequency, a sequence of word vectors is finally formed as input for the next stage. Then the neighboring words of high frequency words in the corpus are counted, and a neighboring word list is formed to assist the decoder vocabulary generation.

2) Semantic understanding stage. The recurrent neural network has a memory function, and the word vector sequence of the previous stage is sequentially input into the encoder. The encoder will generate a semantic vector of the current time step at each time step, and finally merge these semantic vectors together to form the semantic vector of the full text, and pass it to the next stage.

3) Information reorganization. The attention mechanism reorganizes the full-text semantic information most suitable for the current time step according to the intermediate state fed back by the decoder, and sends the reorganized intermediate semantic information back to the decoder for word prediction at the current time step.

4) Abstract generation. The RNN predicts a word at each time step, and predicts the next word based on the previously predicted word and the intermediate semantics summarizing the full text, and finally forms a summary sentence.

4. Text generation research and experimental analysis based on deep learning

4.1 Hybrid neural network encoder

This paper proposes a new form of encoder. The encoder combines the structural characteristics of the CNN and the deep RNN. It explicitly uses the convolutional layer to capture the contextual relationship...
between the target vocabulary unit and its neighboring words, and strengthens the role of the context. It is a good supplement to the traditional encoder structure based on a single RNN. Based on this, the encoder can not only learn the timing information and long-distance dependence that the recurrent neural network is good at, but also detect local timing-independent features, so as to obtain a high-quality original text representation, which lays the foundation for summary generation [10].

This article uses bidirectional LSTM as the basic calculation unit of the encoder, as shown in Figure 3. The two-way modeling method can more completely capture the association between the units in the sequence. Specifically, the input original text sequence \( (x_1, x_2, \cdots, x_m) \) is mapped to the forward hidden state vector \( (\overrightarrow{h_1}, \overrightarrow{h_2}, \cdots, \overrightarrow{h_m}) \) and the backward hidden state vector \( (\overleftarrow{h_1}, \overleftarrow{h_2}, \cdots, \overleftarrow{h_m}) \), the forward and backward hidden states at each moment are concatenated according to formula (7) as the overall hidden state at that moment.

\[
h_i = [\overrightarrow{h_i}, \overleftarrow{h_i}] \tag{7}
\]

In order to achieve gradient transfer between multi-layer neural networks, we use a residual connection between the two-layer recurrent neural network to facilitate its training. Residual connections can help build deeper networks and alleviate problems such as gradient disappearance. At each moment, the input of the bottom LSTM is added to its output, and its sum is fed as an input to the second layer LSTM. On top of the RNN, we cascaded a layer of convolutional neural network without pooling operation. Its purpose is to use the convolutional layer to extract local features between adjacent states, as shown in formula (8). The hidden state \( h_i \) of the i-th input is determined jointly by the adjacent hidden states in the convolutional network window.

\[
h_i = \sigma \left( \theta \ast h_{(i+j-1) \frac{m-1}{2}} + b \right) \tag{8}
\]
4.2 Seq-to-Seq abstract model combined with attention mechanism
The automatic summarization model based on improved cluster search proposed in this paper is shown in Figure 4. Given a piece of original text as input, we first map the words in the sequence to a continuous word vector space to get its vector representation, and encode it using a hybrid neural network encoder combining CNN and RNN to obtain its hidden state. After the original text is encoded, we use a decoder based on coverage and combined with an attention mechanism to generate a verbatim summary. During the decoding process, this paper constrains the generated candidate abstract sentences from the perspective of grammar enhancement and text diversification, and obtains its initial score. Then a reordering module based on key phrases is introduced to reprocess these candidate sentences. At the same time, its original score and the importance score of the included key phrases are considered, and finally the candidate sentence with the highest score is selected as the generated abstract.

![Figure 4. Structure of an automatic summary model based on improved cluster search](image)

4.3 Experimental analysis
The purpose of this experiment is to test the types of encoders and the dimensions of word vectors to select the best combination. We set the dimension of the word vector to 128, 200 or 300 respectively. The LSTM used in the encoder selects one-way or two-way respectively, and combines them in pairs. The influence of the two factors of word vector dimension and encoder direction on the quality of the generated sentences was tested under the sub-word model of sub-word unit extraction of the vocabulary of complex sentences and simple sentences in the corpus.

| Encoder type | Word vector dimension | BLEU | SARI |
|--------------|-----------------------|------|------|
| Uni-LSTM     | 128                   | 17.07| 15.96|

Table 1. Model word vector dimension and encoder directionality test
The experimental results of the sub-word model on the WPKP data set are shown in Table 1. It can be found that when the same type of encoder is used, the increase in the dimension of the word vector can contribute to the improvement of the model performance, which is consistent with intuition. When using neural networks to learn sequence information, higher dimensional vectors can present richer text feature information. At the same time, when using the same word vector dimension, the effect of bidirectional LSTM encoder is better than unidirectional LSTM. Therefore, when generating sentences, using bidirectional LSTM can obtain more context information, which is conducive to improving text quality. In subsequent experiments, we will use a bidirectional LSTM encoder and set the word vector dimension to 300 in order to obtain the best experimental results.

5. Conclusion
This article makes an in-depth study on generative summary technology and proposes a solution to the problem. This scheme builds a new self-encoder model. The model's encoder uses a new type of combination, and the encoder uses a double-layer recurrent neural network. By comparing and analyzing the experimental results, it can be concluded that this combination has a better effect. This paper believes that deep learning methods have many valuable attempts in the field of NLP, and will achieve greater success in the near future. But the future is still full of challenges, and it is worthy of more researchers to conduct extensive and in-depth research.

Reference
[1] Ronan C, Jason W, Michael K, et al. Natural Language Processing (Almost) from Scratch. Journal of Machine Learning Research, 2011, 12(1):2493-2537.
[2] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks. Advances in Neural Information Processing Systems, 2014, 3: 2672-2680.
[3] Kucukelbir A, Tran D, Ranganath R, et al. Automatic differentiation variational inference. Journal of Machine Learning Research, 2017, 18(1):430-474.
[4] LIU Zhi-yuan, SUN Mao-song, LIN Yan-kai, et al. Knowledge representation learning: a review. Journal of Computer Research and Development, 2016, 2: 247-261.
[5] LI P, ZHOU G. Joint argument inference in Chinese event extraction with argument consistency and event relevance. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2016, 24(4): 612-622.
[6] Greff K, Srivastava R K, KOUTNIK J, et al. LSTM: a search space odyssey. IEEE Transactions on Neural Networks and Learning Systems, 2016, 99: 1-11.
[7] Huang J J, Li P W, Peng M, et al. Review of Deep Learning-Based Topic Mode. Chinese Journal of Computers, 2020, 43(5): 827-855.
[8] Dong C X. Research on Short Text Automatic Summary Method Based on Deep Learning[D]. Beijing: Beijing University of Posts and Telecommunications, 2019.
[9] Lin Y O, Lei H, Li X Y, et al. Deep Learning in NLP: Methods and Applications. Journal of University of Electronic Science and Technology of China, 2017, 46(6): 913-919.
[10] Guo T Z, Sun B S. Research on Application of Deep Learning in Text Generation. Electronic Instrumentation Customers, 2020, 27(2): 110-112+42.