

# Research on the Application of Deep Learning in Natural Language Processing

Fangxu Yan, Jianhui Wang, Wei Li

School of Mathematics and Systems Science, Shenyang Normal University, Shenyang Liaoning, 110034, China

**Abstract:** Deep learning is a kind of machine learning, which is the necessary path to realize artificial intelligence. The application of deep learning in the field of natural language processing has gone far beyond the limitations of traditional methods, showing unparalleled advantages such as automatically learning abstract features from the original data to form preciser representation. This paper aims to analyze the application of deep learning in natural language processing, and provides reference for the follow-up research and development in this field.

**Keywords:** Deep Learning; Natural Language Processing; Neural Network.

## 1. Introduction

In recent years, as a hot field of artificial intelligence research, deep learning (DL) has had an increasingly important impact on many tasks due to its excellent feature extraction and learning capabilities. The concept of DL was first proposed by Hinton in 2006[1]. It aims to learn the inherent laws and representation levels of sample data, so that machines can have analytical learning capabilities like humans. At the same time, DL has been widely used in natural language processing (NLP), computer vision (CV), speech recognition (SR) and other fields, which has attracted much attention.

NLP is regarded as the pearl on the crown of artificial intelligence. The goal of NLP is to enable computers to understand, process and generate natural language, so as to provide users with better interactive experience and more efficient output. Therefore, how to use DL to promote the development of NLP is the hot spot and difficulty in current research. At present, the breakthrough of deep learning technology has brought great changes to the development of NLP, which has greatly promoted the frontier research and practical application in this field. DL has dominated especially in key applications such as machine translation, text sentiment analysis, and question answering systems. In addition to these main application fields, DL also shows strong potential and development prospects in many directions such as text generation, named entity recognition and relation extraction. These not only promote the rapid development of NLP technology, but also have great economic value in many practical application scenarios such as search engine optimization, knowledge graph construction, intelligent customer service and so on.

In summary, In-depth study of the application development and problems of DL in the field of NLP is of great practical significance for promoting the deep integration of artificial intelligence and various fields of economic society, stimulating high-quality economic development, and motivating the high-level application of intelligent manufacturing [2].

## 2. Principles of Deep Learning

The concept of deep learning originates from the research

of artificial neural network. DL combines low-level features to form more abstract high-level representation attribute categories or features to discover distributed feature representations of data. The motivation of studying deep learning is to establish a neural network that simulates the human brain for analytical learning. It imitates the mechanism of the human brain to interpret data, such as images, sounds and texts[3]. A basic neural network includes input layer, hidden layer and output layer. The input layer is responsible for receiving the original data set, and the output layer delivers the prediction results of the model. The hidden layer is located between the input layer and the output layer, which usually appears as multiple layers and contains multiple neurons. Each neuron and the others in the anterior and posterior layers transmit information through weight. By non-linearly converting the input signals of neurons, the activation function enables the network to learn and perform more complex tasks. Common activation functions contain *sigmoid*, *tanh*, *ReLU*, *ELU* and so on. In addition, the propagation of data is divided into forward propagation and back propagation. In forward propagation, the input data is linearly transformed by the weight and bias of each layer, and non-linearly transformed by the activation function, and then output to the next layer until reaching the output layer. In the back propagation, the error of the output layer is first calculated, and then propagated from the output layer to the hidden layer until propagated to the input layer. Through back propagation, the gradient information of each parameter on the loss function can be obtained, so as to optimize and update the parameters. Different from traditional shallow learning, DL emphasizes the depth of the model structure, which can be added to obtain deep meaning. Secondly, DL clarifies the importance of feature learning. By layer-by-layer feature transformation, the feature representation of the sample in the original space is transformed into a new feature space, which makes classification or prediction easier. The most classic deep learning networks include convolutional neural network (CNN) and recurrent neural network (RNN).

### 2.1. Convolutional Neural Network

CNN is a kind of feedforward neural networks with convolution calculation and deep structure, which is one of the representative algorithms of deep learning. CNN adopts the method of local connection and weight sharing. On the

one hand, it reduces the number of weights and makes the network easy to optimize. On the other hand, it reduces the complexity of the model and reduces the risk of overfitting[4]. When the input of the network is an image, these advantages will be more obvious. Each neuron only needs to perceive some elements in the image or text, and the final neuron integrates the perceived local information to obtain the comprehensive representation information of the image or text.

As shown in Figure 1, CNN is usually composed of input layer, convolutional layer, activation function, pooling layer, fully connected layer and output layer. The convolutional layer is an important part of CNN. Each node input in the convolutional layer is part of the previous neural network layer, which aims to extract different features of the input picture or text. When dealing with the text sequence problem, the convolution layer usually uses filters of different sizes to extract different features in the text sequence. Then the non-linearity of CNN is increased by the activation function. The pooling layer is to reduce the input dimension of the model, thereby reducing the complexity and parameters, improving robustness, and preventing model from overfitting to a certain extent. The most common pooling methods are maximum pooling and average pooling. The fully connected layer is mainly responsible for dimensionality reduction while retaining truly effective information. Usually, the components from the convolutional layer to the pooling layer are regarded as the process of automatically extracting features. After the feature extraction is completed, the output layer needs to be used to complete the classification or prediction tasks. Generally, the learned high-dimensional feature representation is fed to the output layer, and the probability that the current sample belongs to different categories is calculated by the Softmax function.

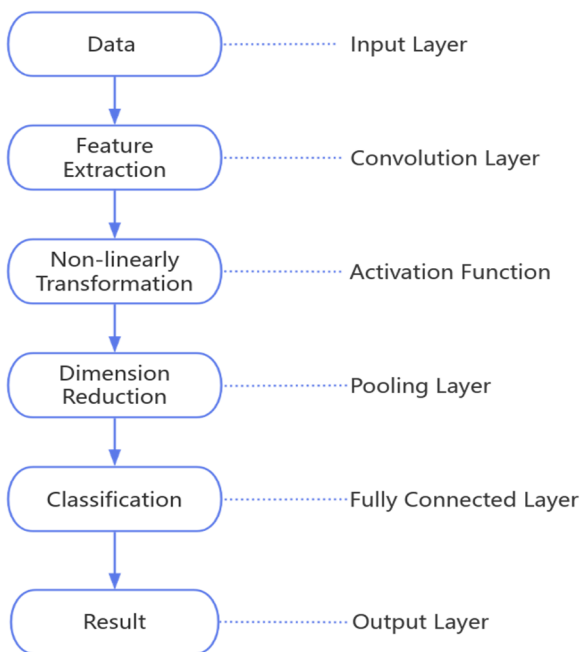


Fig 1. Overall structure of CNN

## 2.2. Recurrent Neural Network

Recurrent neural network has the structure of tree layer, which is an artificial neural network in which the network nodes recursive the input information according to their connection order. RNN has a variable topology and shares

weights. It is mostly used in machine learning tasks containing structural relationships and has attracted the attention of researchers in the field of NLP. The basic structure of RNN includes input layer, hidden layer and output layer. The biggest difference between RNN and traditional neural network is that RNN will send the output of the previous word into the hidden layer of the next word to train together, and finally only output the calculation result of the last word.

At present, RNN has a great influence on short-term memory, but has little effect on long-term memory, which decides the predicament of the incompetence in dealing with input sequences with large length. And in the process of back propagation, the gradient multiplication is involved in the process of finding the parameter gradient of the bottom layer, which is prone to the problem of gradient disappearance or gradient explosion. Long short-term memory (LSTM) network and gated recurrent unit (GRU) can solve this problem to a certain extent.

## 3. Application of Deep Learning in NLP

### 3.1. Machine Translation

Machine Translation (MT) refers to the process of using a machine to translate a natural language in a written form or sound form into a natural language in another written form or sound form through a specific computer program. With the development of DL, the quality of MT has been significantly improved, and it has been able to handle more complex and natural language expressions. The sequence-to-sequence (Seq2Seq) model is one of the applications of DL in MT. It is designed to directly establish a mapping between the input and output sequences, without the need for complex feature engineering or intermediate representation. Seq2Seq includes an encoder responsible for understanding the input language and a decoder responsible for translating the target language. This end-to-end training method enables the model to capture rich language patterns and context information on a large number of bilingual data. Different from the Seq2Seq model, Transformer relies entirely on the self-attention mechanism, which allows the model to process input data in parallel and better capture long-distance dependencies in the sequence. This brings higher efficiency and better performance to the model and becomes the core of many leading translation systems today.

### 3.2. Sentiment Analysis

Sentiment analysis, also known as tendency analysis, aims to analyze and process the text with subjective emotion [5]. Different from the traditional sentiment analysis methods that have the problem of semantic loss, over-reliance on prior background knowledge, and more requirement on manual intervention, DL provides a method to learn and extract features directly from data. CNN 's multiple convolutional layers are able to extract more complex text features layer by layer to form a hierarchical text representation, which is achieved by capturing the N-gram features in the text through a sliding window and identifying keywords or phrases related to specific emotions. Collobert et al. proposed a general framework based on CNN to solve a large number of NLP tasks in 2011[6]. Then in 2014, Kalchbrenner et al. successfully used CNN to extract significant N-gram features from input sentences [7]. RNN and its extension, such as

LSTM and GRU, are designed to solve the problem of long-distance dependency in text. The emotion in the text often depends not only on the local vocabulary selection, but also on the context, the information in the front and back sentences. In 2011, Mikolov et al. successfully used RNN for language modeling [8]. Sutskever et al. proposed a general deep LSTM encoder-decoder framework that enables mapping between sequences in 2014[9].

### 3.3. Question Answering System

The question answering system is designed to provide users with accurate answers. Driven by DL technology, the ability of question answering system has been significantly enhanced, especially when dealing with large and complex text databases. Traditional question answering methods usually rely on manually formulated rules or shallow text matching techniques, but these methods lack the ability to deal with diverse and complex problems. In contrast, DL provides a way to learn directly from data, and to understand text content and contextual relationships more deeply. BERT, a bidirectional deep language model based on Transformer, is proposed by Devlin et al. in 2018[10]. As shown in Figure 2, The model is composed of multi-layer bidirectional Transformer decoder. It learns the internal structure of language by pre-training a large number of unlabeled texts and semantics, and then optimize the results through label data for question answering tasks. In this case, BERT can capture bidirectional context semantics. Another model named text-to-text transfer transformer(T5) adopts a new methods[11]. It treats all NLP tasks as text-to-text conversion problems. T5 emphasizes the combination of task-independent pre-training and task-specific fine-tuning, which enables the model to show a high degree of flexibility and accuracy in question and answer scenarios.

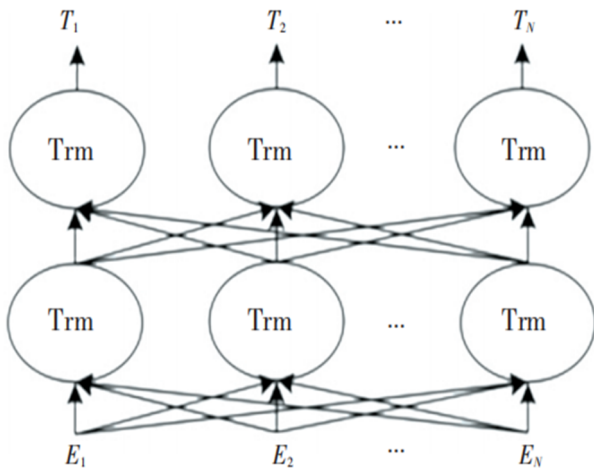


Fig 2. BERT model

## 4. Conclusion

This paper mainly introduces CNN and RNN in DL, and

expounds the research progress of each task in the field of NLP. Although DL has achieved great success in various tasks of NLP, there are still difficulties that need to be overcome. The larger the deep neural network model is, the longer the training time needs. How to reduce the model volume while maintaining the performance is a direction for future research. Looking ahead, DL may also be combined with other cutting-edge technologies such as quantum computing and edge computing to achieve higher computational efficiency and wider application scenarios. The application of DL in NLP will be more diverse and extensive.

## Acknowledgments

The authors gratefully acknowledge the financial support from Liaoning Provincial Department of Education Scientific Research Funding Project funds (LFW202004) and Postgraduate Research Project of Shenyang Normal University (SYNUXJ2024060).

## References

- [1] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. science, 2006, 313(5786): 504-507.
- [2] Chen W. Research on Artificial Intelligence in Natural Language Processing[J]. Information Recording Materials, 2023, 24 (10) : 92-94.
- [3] Wei J, Liu A, Tang J. Research on alarm model of digital TV monitoring platform based on deep learning neural network technology[J]. Cable TV technology, 2017, (07) : 78-82.
- [4] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [5] Yan F, Wang J. Research on Sentiment Analysis of Micro-blog based on Attention-BiLSTM[J]. Frontiers in Computing and Intelligent Systems, 2024, 7(3): 49-51.
- [6] Collobert R, Weston J, Bottou L, et al. Natural language processing (almost) from scratch[J]. Journal of machine learning research, 2011, 12: 2493– 2537.
- [7] Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences[J]. arXiv:1404.2188v1, 2014.
- [8] Mikolov T, Kombrink S, Burget L, et al. Extensions of recurrent neural network language model[C]//2011 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, 2011: 5528-5531.
- [9] Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks[J]. Advances in neural information processing systems, 2014, 27.
- [10] Devlin J. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv:1810.04805, 2018.
- [11] Raffel C, Shazeer N, Roberts A, et al. Exploring the limits of transfer learning with a unified text-to-text transformer[J]. Journal of machine learning research, 2020, 21(140): 1-67.