Overview of Deep Learning Research

Yanmei Liu[1,a] and Yuda Chen [2,b]

[1] Wuhan Institute of Design and Sciences, Wuhan, Hubei, 430205, China

[2] Wuhan Center of Geological Survey, China Geological Survey Wuhan, 430205, China

[a] 61704088@qq.com, [b] chyd555@163.com

Abstract. Deep learning is the research hotspot and direction in all major fields at present, its motivation lies in the establishment and simulation neural network of human brain's analysis and learning, and to interpret data by imitating the mechanism of human brain. This paper mainly overviewed the related research of deep learning. Firstly, it introduced its concept, and introduced several important neural network models in detail, compared and analyzed the current mainstream deep learning framework, finally introduced the main application field, and prospected the future research directions.

Introduction

As an important branch of machine learning, deep learning not only profoundly affects the direction of machine learning, but also is a powerful way to realize artificial intelligence. It is a deep machine learning model whose depth is reflected in the multiple transformation of features. Its concept was first proposed by G. E. Hinton et al. in 2006[2,5]. This algorithm integrates many neurons, which are independent of each other. They receive input from the bottom neurons by connecting with the upper neurons, and obtain more semantic information with abstract identifiers in the form of combination and transformation, and achieves the purpose of learning the inherent distributed characteristics of the target, which belongs to a deeper neural network.

In view of the popularity of deep learning in recent years, this paper will give a systematic introduction to deep learning from the following aspects: network structure, mainstream framework, application situation, summary and prospect

Network Structure of Deep Leaning

The "depth" of deep learning refers to the number of layers experienced from "input layer" to "output layer", that is, the number of layers of the "hidden layer", and the more layers, the deeper the depth. So, the more complex the choice problem is, the more hierarchical it needs. With the rapid development of deep learning, a variety of deep network structures have emerged. This paper mainly introduces the following three network structures: convolutional neural network, recurrent neural network, and deep belief network.

Convolutional Neural Network

Convolutional Neural Network (CNN) [6]is a structure that can successfully use BP for training deep networks. It is proposed by Yan Lecun et al. based on previous work. It combines local receptive fields, weight sharing and subsampling in space or time to ensure translation and deformation invariance. CNN usually contains convolutional layer, subsampled layer full Connection layer and output layer, there can be multiple convolution layers and subsampled layers. The specific process is as follows:

- Convolution: the data is convoluted in the convolutional layer and output to the next layer after the activation function. Through convolution operations, the neural network can extract local features of each part of the data and can record the positional relationship between the features;

- Pooling: divide the features of the output of the previous layer into regions, and reduce the number of features by sampling operation, thereby avoiding the phenomenon of over-fitting and better synthesizing the characteristics of the data;
- Full connection: the data feature after the pooling operation is multiple sets of signal data, and the full connection combines multiple sets of signal data into one set of signal data;
- Classified output: after integrating the data of feature signals into a group, output the features according to the requirements of data analysis to achieve classification or regression problems.

The following is a typical CNN network, as shown in Figure 1. The basic structure of CNN consists of two layers: 1. Feature extraction layer (S layer), the input of each neuron is connected with the local acceptance field of the previous layer, and extract the local features. Once the local feature is extracted, its location relationship with other features is also determined. 2. Feature mapping layer (C layer). Each computing layer of the network consists of multiple feature maps. Each feature map is a plane, and the weights of all neurons on the plane are equal. The sigmoid function is used as the activation function of the convolution network in the feature mapping structure, which makes the feature mapping displacement invariant. In addition, since the neurons on one mapping surface share weights, the number of network free parameters is reduced. Each extraction layer (S layer) is followed by a computing layer (C layer) which is used to calculate the local average and the secondary extraction. This unique structure of two feature extraction reduces the feature resolution and makes the network have a high distortion tolerance to the input samples during the identification.
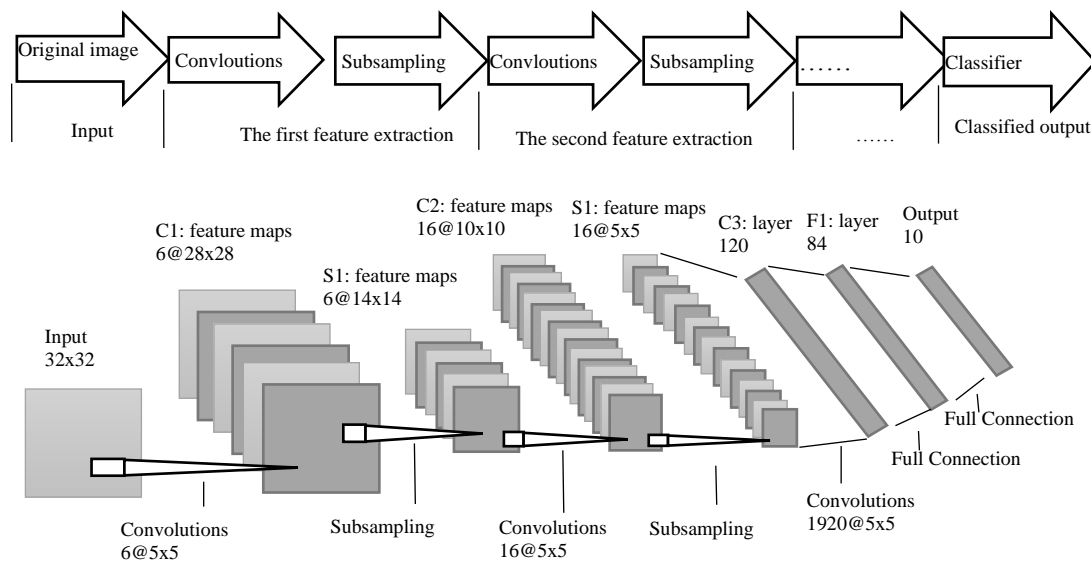
Fig. 1 A classical convolutional neural network

In terms of network architecture, the network is deepening. the deeper the network in theory, the more abstract the features in the image can be captured, and the more learning ability it has, the more difficult the training will be. In 2012, Krizhevsky et al. proposed AlextNet, which consists of a 5-layer convolutional layer and a 3-layer fully-connected layer. The new activation function ReLU is used in the network to help the model converge, and the Dropout method is proposed to mitigate the over-fitting phenomenon. VGGNet emerged in 2014. Through smaller convolution kernel and reduced sampling size, the network was improved to 19 layers at most, which verified that network layer deepening could obtain better performance. In the same year, GoogleLetNet inherited the idea of "network in network" and adopted Inception structure as the basic unit, which improved the training speed and promotion ability. In 2015, ResNet introduced a direct connection from input to output, and the network depth was increased to 152 layers at most.

Recurrent Neural Network

The Recurrent Neural Network (RNN) [7,16] is proposed to solve the problem that convolutional neural network cannot solve and retain time series. It is a recursive neural network, in which sequence data is input, recursion is carried out in the direction of sequence evolution and all nodes (cyclic units) are linked in a chain to form a closed loop recursive neural network. The closed loop connection shown in Figure 2 is the core of RNN. They perform the same task for each element in the sequence, and the output depends on the previous calculation. That is, RNN has the memory function, and the memory can capture the information that has been calculated so far.
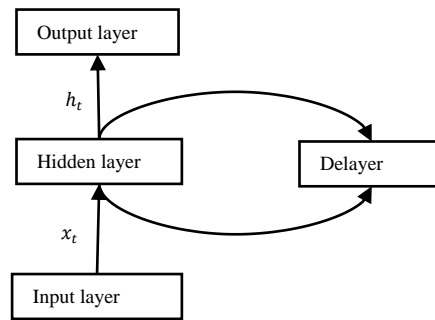
Fig. 2 RNN structure

Deep Belief Network

The system DBN [12,14] is structurally composed of a self-encoder and a restricted Boltzmann machine (RBM [8]). The DBN consists of several layers of neurons, and the components are RBM-restricted Boltzmann machines. The process of training DBN is carried out layer by layer. In each layer, data vectors are used to infer the hidden layer, and then the hidden layer is regarded as the data vector of the next layer. DBN network structure is limited to two layers: visual layer and hidden layer. There are links between layers, but there are no links between layers. Hidden layer units are trained to capture the relativity of higher-order data in the visual layer. The specific DBN network structure is shown in Figure 3.
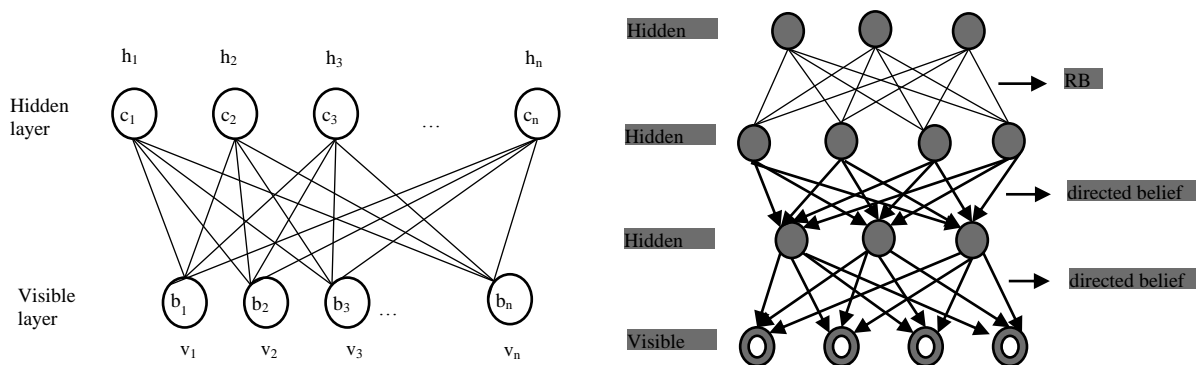
Fig. 3 DBN structure

Deep Learning Framework

Deep learning framework is a set of independent architecture, unified style template and reusable solution specially developed for the field of deep learning. It generally has the characteristics of high cohesion, strict specification, expansibility, maintainability, and high versatility, and has a unified code style and template structure, which can reduce the writing of many repeated codes. Due to different requirements, more and more deep learning frameworks are developed. At present, there are mainly five mainstream deep learning frameworks, as shown in table I.

Table 1. Current mainstream deep learning framework

Framework	Support Language	Support System	Characteristic	Maintenance Team
TensorFlow	Python, C++, Java, Go	Linux, Window, ios, MacOS	High flexibility, strong expansibility, automatic differentiation and visualization	Google
pytorch	Python, C++	Linux, Window, ios, MacOS	Numpy with GPU operation, dynamic graph, can quickly prototype	Facebook
paddepaddle	Python, C++	Windows,Ubuntu, CentOS,MacOS	Support dynamic and static graphs, integrated design of inference engine, seamless integration of training to multi-ended reasoning, and deep learning platform providing systematic technical services and support	Baidu
keras	Python	Linux,macOS,Windows	Written by an object-oriented approach, fully modular and extensible, and trying to simplify the implementation of complex algorithms.	François Chollet etc.
Deeplearning4j	Java	Linux,macOS,Windows	Driven by its own open source numerical computing library ND4J, it can be run using CPU or GPU	Adam Gibson etc.

The Application of Deep Learning

As one of the important technologies in the field of artificial intelligence, deep learning has excellent performance in many application fields. This paper mainly discusses the application of deep learning in image recognition, speech recognition, natural speech processing, data analysis and other fields.

Image Recognition

Image recognition [13] is the earliest application field of deep learning. As early as 1989, LeCun et al. applied convolutional neural networks to handwritten digit recognition tasks [3], achieved the best results in the world at the time. and widely applying it to handwritten numeral recognition tasks of checks of major Banks. In 2012, deep learning technology was first applied to the ImageNet competition. It can be seen that compared with the best traditional recognition error rate in 2011, the error rate of image recognition based on deep learning technology has greatly decreased by 41.1%. In 2015, the error rate of image recognition has surpassed that of human beings. In 2016, the latest recognition error rate of ImageNet has reached 2.991% [4].

At present, deep learning has been widely used in image recognition and computer vision. It is mainly used in face recognition, face tracking, object scene recognition, etc. For example, the popular smart phones on the market have the functions of pattern, fingerprint and face unlocking. Its speed and accuracy have achieved ideal results. With the continuous development of deep learning technology, the accuracy rate of image recognition will be higher and higher, which will provide greater convenience for our intelligent life.

Speech Recognition

Speech recognition technology [15-16,20] mainly includes keyword recognition and continuous speech recognition. Traditional language recognition technology is difficult to obtain better results in large speech data. The initial research focused on Hidden Markov Model-based algorithms, and Hinton et al. improved the traditional GMM-HMM and replaced the GMM with the Deep Neural Network (DNN) in the model, thus generated a new model DNN-HMM. Compared with the original model, the lexical error rate of the improved model is greatly reduced. Since deep learning was introduced into speech recognition in 2009, a series of breakthroughs have been made. Nowadays,

speech recognition technology has been widely used in daily life. Google's Google Now, Apple's Siri, and Microsoft's skype are all based on deep learning algorithms, which bring great convenience to human-computer interaction.

Natural Language Processing

Natural language processing [17] mainly studies various theories and methods of processing, understanding and using human language (also known as natural language) by computer, including Chinese word segmentation, part of speech tagging, language translation, information retrieval, recommendation system, etc. With the rapid development of the Internet, the explosive growth of network text, especially user-generated text, has brought tremendous application requirements for natural language processing. Before the deep learning technology, the research of natural language processing mainly used dictionary matching, regular expression, conditional random field and hidden Markov model, and achieved certain results, but it has not yet reached a satisfactory level. Since the birth of RNN, because of its advantages in time series processing, its algorithm was quickly applied to the field of natural language processing and achieved great breakthrough. At present, deep learning algorithms play an important role, such as Google translation, youdao translation and other language translation tools, Google, Baidu and other search engines etc.

Data Analysis

In the era of big data, the new technological revolution represented by mobile Internet makes the data volume of the whole human society show an explosive growth trend. In-depth data analysis is not only a daily business requirement, but also a necessary condition to accurately grasp the current situation and trend of the industry. At present, it has been applied in parameter estimation, data diagnosis, modeling and prediction. for example Lei Yaguo et al [9] accurately identified the system health status under different fault scheme settings after analyzing the big data of the mechanical monitoring system; Tan Juan et al. [10] mainly selected traffic flow parameters, environment and time period to carry out traffic congestion prediction research, and the accuracy rate of morning rush hour congestion prediction was 86.7% and evening rush hour congestion prediction was 84.2%.

Summary and Prospect

With the development and changes in the field of deep learning research, this paper systematically introduces the concept of deep learning, neural network model, mainstream deep learning framework, and finally expounds the current application fields. Although it is excellent in the fields of image and speech, it also faces some problems that need to be solved urgently. At present, most of the deep learning models are based on the simplest neural network structure superposition or mixture of convolutional neural networks and restricted Boltzmann machines. When solving some problems of data types, complex data structures and complicated data relationships, the effect is not ideal. In addition, most of the research on deep learning is at the scientific research stage, and there are few products that are scaled into life. However, with the continuous development, more people will participate in it in the future, and the above problems and challenges will be solved and completed. Secondly, industry big data will also provide a broader stage. Future deep learning development will present a multi-platform, deep-level, and all-network trend.

Acknowledgments

References

[1]   Ram Deepak Gottapu, Cihan Dagli, Bharami Ali, Entity Resolution Using Convolutional Neural Network.Procedia Computer Science,2016,95:153-158.

[2]   Hinton G E, Osindero S, The Y W, A fast learning algorithm for deep belief nets.Neural Computationg, vol. 18, 2006, pp. 1527-1554.

[3]   LeCun Y, Boser B,Denker J S, et al. Backpropagation applied to handwritten zip code recognition. Neural Computation, 1989, 1(4):541-551.

[4]   Zhang Junyang,Wang Huili, Guo Yang, Hu Xiao. Review of deep learning. Application Research of Computers, 2018,7(35):1921-1936.

[5]   B Yin,W wang, L Wang, Review of Deep Learning, Journal of Beijing University of Technology,vol. 41, 2015, pp. 48-59.

[6]   Ram Deepak Gottapu, Cihan Dagli,Bharami Ali. Entity Resolution Using Convolutional Neural Network.Procedia Computer Science, 2016,95:153-158.

[7]   https://blog.csdn.net/heyongluoyao8/article/details/48636251

[8]   HINTON   G   E.   A   practical   guide   to   training   restricted   Boltzmann machines[M]//MONTAVON G, ORR G B, MULLER K R. Neural Networks: Tricks of the Trade. 2nd ed. Berlin, Heidelberg: Springer, 2012:599-619.

[9]   Lei Yaguo, Jia Feng, Zhou Xin, Lin Jing. A Deep Learning-based Method for Machinery Health Monitoring with Big Data. Journal of mechanical engineering.2015,51(21):49-55.

[10] Tan Juan, Wang Shengchun. Research on prediction model for traffic congestion based on deep learning. Application Research of Computers.2015,32(10):2951-2954.

[11] HINTON G E, SALAKHUTDINOV R R. Reducing the Dimensionality of Data with Neural Networks.Science,2006,313(5786):504-507.

[12] SILVER D, HUANG A, MADDISON C, et al. Mastering the game of Go With deep neural networks and tree search. Nature,2016,529(7587):484-489.

[13] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing System. Lake Tahoe: MIT Press,2012:1097-1105.

[14] MOHAMED A, DAHL G E, HINTON G. Acoustic modelling using deep belief networks. IEEE Transaction on Audio, Speech, and Language Processing,2012,20(1):14-22.

[15] FENG X, ZHANG Y, GLASS J. Speech feature denoising and dereverberation via deep autoencoders for noisy reverberant speech recognition. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Florence: IEEE,2014:1759-763.

[16] GRAVERS A, MOHAMED A, HINTON G. Speech recognition with deep recurrent neural network. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Vancouver: IEEE,2013,6645-6649.

[17] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE,1998,86(11):2278-2324.

[18] Yann Lecun, Yoshua Bengio, Geoffrey Hinton. Deep Learning. Nature,2015,521:436-444.

[19] Matthew D Zeiler, Rob Fergus. Visualizing and Understanding Convolutional NetWorks.ar Xiv:1311.2901,2013.

[20] ABDEL-HAMID O, MOHAMED A, JIANG H, et al. Convolutional neural networks for speech recognition. IEEE/ACM Transactions on Audio, Speech, and Language Processing,2014,22(10):1533-1545.