Deep learning review and discussion of its future development

Zhiying Hao[*]

University of Electronic Science and Technology of China, No.4, Block 2, North Jianshe Road, Chenghua District, Chengdu, Sichuan, China

Abstract. This paper is a summary of the algorithms for deep learning and a brief discussion of its future development. In the first part, the concept of deep learning and the advantages and disadvantages of deep learning are introduced. The second part demonstrates several algorithms for deep learning. The third part introduces the application areas of deep learning. Then combines the above algorithms and applications to explore the subsequent development of deep learning. The last part makes a summary of the full paper.

1 Introduction

As early as 1952, IBM's Arthur Samuel designed a program for learning checkers. It can build new models by observing the moves of the pieces and use them to improve their playing skills. In 1959, the concept of machine learning was proposed as a field of study that could give a machine a certain skill without the need for deterministic programming. In the process of machine learning development, various machine learning models have been proposed, including deep learning. Due to its complicated structure and the need for a large amount of calculation, the computing cost is very high, so it had not been paid attention to at the beginning. However, with the great improvement in computer performance, the excellent performance of deep learning makes it rose rapidly and has become one of the hottest research areas. In this paper, the main deep learning models will be briefly summarized and the development prospects of deep learning will be analyzed and discussed at the end.

2 Introduction to deep learning

2.1 What is Deep Learning

Deep learning is a branch of machine learning [1]. It is an algorithm that attempts to use the high-level abstraction of data using multiple processing layers consisting of complex structures or multiple nonlinear transforms. In machine learning, deep learning is an algorithm based on characterizing learning data. The concept of deep learning is relative to shallow learning. Shallow machine learning models such as Support Vector Machines and

---

[*] Corresponding author: zhiyinghao@std.uestc.edu.cn

Logistic Regression were introduced in the 1990s. These shallow machine learning models have only one layer or no hidden layer nodes, as shown in Fig 1. Deep learning is based on multiple hidden layer nodes. The essence of deep learning is multi-layer neural network. Deep learning uses the input of the previous layer as the output of the next layer to learn highly abstract data features.
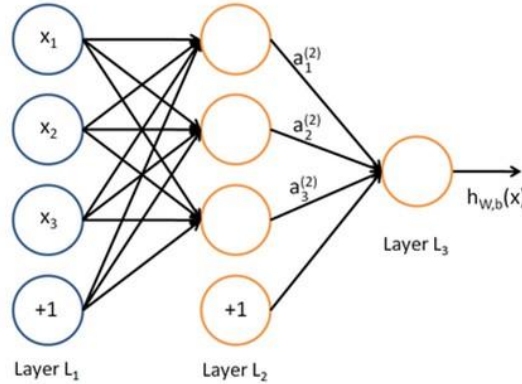

Fig. 1. A single-layer neural network

Like machine learning, deep learning can be categorized into supervised learning, semi-supervised learning, and unsupervised learning. At present, the classical deep learning framework includes Convolutional Neural Networks, Restricted Boltzmann Machines [2], Deep Belief Networks [3], and Generative Adversarial Networks [4]. In the next section, these algorithms will be introduced briefly.

2.2 Advantages and disadvantages of deep learning

Deep learning has shown better performance than traditional neural networks. After a deep neural network is trained and properly adjusted for certain task like image classification, it saves a lot of calculations, and can complete a lot of work in a short time. . Deep learning is also malleable. Usually, for traditional algorithms, if you need to adjust the model, you may need to make copious changes to the code. For the determined network framework used for deep learning, if you need to adjust the model, you only need to adjust the parameters, thus deep learning has great flexibility. The deep learning framework can be continuously improved and then reached the almost perfect state. Deep learning is also more general, it can be modelled based on problems, not limited to a fixed problem.

Deep learning has some shortcomings as well. First of all, its training cost is relatively high. Now, the performance of computer hardware has been improved a lot, and some simple neural networks can be trained on some of the common computing modules. However, the training of some more complex neural networks still requires relatively expensive high-performance computing modules. Although the price of such modules has been greatly reduced compared with the previous ones, the demand for such hardware still makes the training cost of deep learning relatively high. At the same time, not only the economic cost, the training of neural networks requires a large amount of data to be trained to achieve a satisfactory level, but it is often difficult to obtain a sufficient amount of data. Secondly, deep learning can't directly learn knowledge. Although models such as AlphaGo Zero can learn without prior knowledge have emerged, most deep learning frameworks still need to rely on manual feature marking for training. The workload is enormous for marking large-scale datasets, which also increases the training cost of deep learning. Another point is that deep learning lacks sufficient theoretical support. Although deep learning has achieved good

results in various application fields, there is still no complete and rigorous theoretical derivation to explain the deep learning model at this stage, which limits the follow-up study and the improvement of deep learning.

3 Main Deep Learning Algorithm Introduction

3.1 Convolutional Neural Network

The convolutional neural network, as seen in Fig 2, is a feedforward neural network whose convolution operation allows its neurons to cover peripheral units within the convolution kernel and has excellent performance in large image processing. A convolutional neural network typically consists of one or more convolutional layers and a fully connected layer, which also includes a pooling layer for integration. Convolutional neural networks give better results in terms of image and speech recognition. It requires fewer parameters to consider than other deep neural networks. The advantages of convolutional neural networks make it one of the most commonly used deep learning models. The basic structure of the convolutional neural network is briefly introduced below.
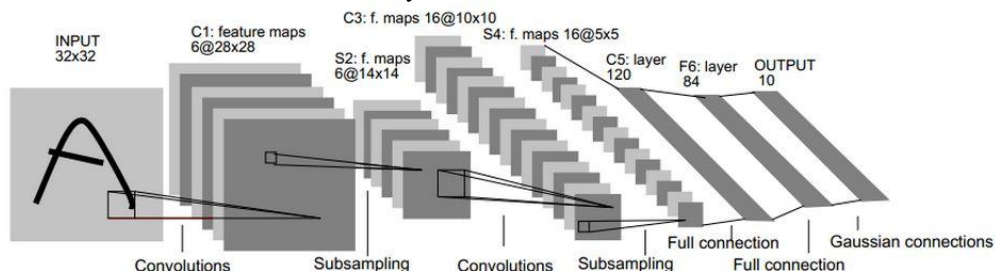

Fig. 2. Convolutional Neural Network, LeNet-5[5]

3.1.1 Convolutional layer.

The convolutional neural network convolves data using multiple convolution kernels in the convolutional layer to generate a plurality of feature maps corresponding to the convolution kernel.

The convolution operation has the following advantages:

1. The weight sharing mechanism on the same feature map reduces the number of parameters;

2. Local connectivity enables convolutional neural networks to take into account the characteristics of adjacent pixels when processing images;

3. There is no object in the image recognition due to the position of the object on the image.

These advantages also make it possible to use a convolutional layer instead of a fully connected layer in some models to speed up the training process.

3.1.2 Pooling layer

After obtaining the features by convolution, we hope to use these features to do the classification. However, the amount of data that is often obtained is very large, and it is prone to over-fitting. Therefore, we aggregate statistics on features at different locations. This aggregation operation is called pooling. In the convolutional neural network, the pooling

layer is used for feature filtering after image convolution to improve the operability of the classification.

3.1.3 Fully connected layer

After pooling layer is the fully connected layer, its role is to pull the feature map into a one-dimensional vector. The working mode of the fully connected layer is similar to that of a traditional neural network. The fully connected layer contains parameters in approximately 90% of the convolutional neural network, which allows us to map the neural network forward into a vector of fixed length. We can grant this vector to a particular image class or use it as a feature vector in subsequent processes.

3.2 Deep Belief Network

The deep belief network is a probability generation model. Compared with the neural network which is a traditional discriminative model, the generated model is to establish a joint distribution between observation data and labels, and to evaluate both P (Observation|Label) and P (Label|Observation) while the discriminative model has only evaluated the latter, that is, P (Label|Observation).

The deep confidence network consists of multiple restricted Boltzmann layers, a typical neural network type as shown. These networks are "restricted" to a visible layer and a hidden layer, with connections between the layers, but there are no connections between the cells within the layer. The hidden layer unit is trained to capture the correlation of higher order data represented in the visible layer.

3.3 Restricted Boltzmann Machine

A Restricted Boltzmann Machine is a randomly generated neural network that can learn the probability distribution through the input data set. It is a Boltzmann Machine's problem, but the qualified model must be a bipartite graph. The model contains visible cells corresponding to the input parameters and hidden cells corresponding to the training results. Each edge of the figure must be connected to a visible unit and a hidden unit. In contrast, the Boltzmann machine (unrestricted) contains the edges between hidden cells, making it a recurrent neural network. This limitation of the constrained Boltzmann machine makes it possible to have a more efficient training algorithm than the general Boltzmann machine, especially for the gradient divergence algorithm.

The Boltzmann machine and its model have been successfully applied to tasks such as collaborative filtering, classification, dimensionality reduction, image retrieval, information retrieval, language processing, automatic speech recognition, time series modeling, and information processing. Restricted Boltzmann machines have been used in dimensionality reduction, classification, collaborative filtering, feature learning, and topic modeling. Depending on the task, the restricted Boltzmann machine can be trained using supervised learning or unsupervised learning.

3.4 Generative Adversarial Network

The Generated Adversarial Network was proposed in 2014. The Generative Adversarial Network uses two models, a generative model and a discriminative model. The discriminative model determines whether the given picture is a real picture, and the generative model creates a picture as close to the ground truth as possible. The generated model is designed to generate a picture that can spoof the discriminative model, and the discriminative model distinguishes the picture generated by the generated model from the

real picture. The two models are trained at the same time, and the performance of the two models becomes stronger and stronger in the confrontation process between the two models, and will eventually reach a steady state.

The use of generating a network is very versatile, not only for the generation and discrimination of images, but also for other kinds of data.

4 Deep Learning Application

4.1 Image processing

Manually selecting features is a very laborious approach. Its adjustment takes a lot of time. Due to the instability of manual selection, we consider letting the computer automatically learn the features. The automatic learning of the computer can be realized by deep learning.

In image recognition, deep learning utilizes patterns of multi-layer neural networks to pre-process, feature extract, and feature processing images.

Taking the convolutional neural network as an example, the convolutional neural network establishes a multi-layer neural network, and uses the convolutional layer to perform convolution operations to extract feature values, and then performs data processing and training through the pooling layer and the fully connected layer. The detailed process is explained in detail in the Technical Implementation section of 2.2 Neural Network below.

Although neural network image recognition can't reach the accuracy of the human eye at present, the neural network can process a large amount of image data, and the efficiency is much better than manual recognition. Facing huge amount of data that cannot be processed manually, using the neural network method will lead to magnificent improvement.

In addition, deep learning provides an idea for face recognition technology. Face recognition is a biometric recognition technology based on human facial feature information for identification. Face recognition products have been widely used in finance, justice, military, public security, border inspection, government, aerospace, electric power, factories, education, medical care and many enterprises and institutions. And with the further maturity of technology and the improvement of social recognition, face recognition technology will be applied in more fields and has an expectable development prospects. The characteristics of the neural network make it possible to avoid overly complex feature extraction when applied to face recognition, which is beneficial to hardware implementation.

4.2 Audio data processing

Deep learning has a profound impact on speech processing. Almost every solution in the field of speech recognition may contain one or more embedding algorithms based on neural models.

Speech recognition is basically divided into three main parts, namely signal level, noise level and language level. The signal level extracts the speech signal and enhances the signal, or performs appropriate pre-processing, cleaning, and feature extraction. The noise level divides the different features into different sounds. The language level combines the sounds into words and then combines them into sentences.

In the signal level, there are different techniques for extracting and enhancing the speech itself from the signal based on the neural model. At the same time, it is able to replace the classical feature extraction method with a more complex and efficient neural network-based method, which greatly improves the efficiency and accuracy. Noise and language levels also include a variety of different depth learning techniques, and different types of neural model-based architectures are used in both sound level classification and language level classification.

5 Discussion of the future development of deep learning

5.1 Representation Learning

The core of deep learning is the abstraction and understanding of features. Therefore, feature learning plays a very important role in deep learning. Since the essence of deep learning is a multi-layer neural network, some useful information is lost in the process of extracting features and transmitting to the lower layer. However, if too much image features are extracted, it may lead to over-fitting. Therefore, the study of representational learning may be one of the core research issues in deep learning research studying how to accurately extract the required features while avoiding over-fitting. The research progress on this problem will be of great help to the classification and generalization task of neural networks.

5.2 Unsupervised Learning

As mentioned above, training a supervised neural network requires a large amount of labelled data. The workload is very large, adding a lot of extra cost to the training of the neural network. Therefore, if the machine completes the work instead of human, the cost of network training will be greatly reduced. Unsupervised learning can also be used not only for the classification of markers, but also for the Go evaluation program such as AlphaGo Zero. The emergence of AlphaGo Zero proves that in some applications, even without the foundation of human prior knowledge, machines can achieve excellent training results. In this case, the application of unsupervised learning in some areas to automate the learning of the machine from the human knowledge base without the limitations of current state of the art, may contribute to the updating and breakthrough of technology in these fields. Moreover, the current research on unsupervised learning is not intensive now while most people's research focuses on supervised learning. Therefore, unsupervised learning has a rich research potential. In fact, unsupervised learning has recently become one of the hottest research areas. In my opinion, unsupervised learning is also one of the most valuable directions for deep learning in the future.

5.3 Theory Complement

One of the shortcomings of deep learning is that there is no complete theoretical support, which brings a lot of controversy to deep learning and the lack of theory does hinder the development of deep learning. With the increasing attention of deep learning this year, the research on deep learning has become more and more in-depth, and the theory of deep learning is constantly improving.

Nevertheless, the theory behind it is still not enough to rigorously prove the inner principles of deep learning. At present, the research relies on part of the theory combined with the actual test and the experiment-based research method. While the theory can't get further breakthroughs, it only depends on adjusting parameters to improve the models performance, which may easily lead to the bottleneck of the research.

Therefore, it seems that for the future development of deep learning, it is very important to obtain complete theoretical support. In the process of research, the theory of deep learning needs to be continuously improved, and finally reach a level sufficient to explain the structure of the inner principle.

5.4 Perspective of Deep Learning Application

The fourth part of this paper referred to two application areas of deep learning, image recognition and speech processing, which are the two main application areas of deep learning. In addition, recent deep learning has also been used in natural language processing.

Specific applications include, for example, autonomous driving, intelligent dialogue robots such as Siri, image classification, medical image processing, etc. Different deep learning frameworks tend to have slightly different application scenarios, such as convolutional neural networks, which are mainly used for image processing. In the work of medical image processing, for example, the use of convolutional neural networks for brain tumour segmentation has achieved an accuracy of more than 90%. Also, in medical applications, convolutional neural networks can be used to recognize Alzheimer's disease brain image, and more accurate diagnostic results can be obtained combined with manual judgement. Medicine is only a part of deep learning applications. In the application of deep learning, well-trained machines can often calculate some details that are hard to be obtained by human, saving people's workload and improving the quality of results. For example, in order to prevent the traffic light violation at the intersection, the usual way is to take picture by the camera and then manually recognize the license plate to give subsequent punishment. Manually viewing the images and recording them is a very boring job and is not so efficient. If the captured image is recognized by the deep neural network, and the license plate number entry system is automatically extracted, it not only saves manpower, but also improves the efficiency greatly.

And so on, I believe that deep learning applications will be developed in the future in transportation, medical, language, automation, etc., not to mention examples here. Although it seems that it can't completely replace human work now, deep learning and artificial combination can greatly improve the work efficiency.

6 Conclusion

In this paper, we introduced some of the main algorithms and some conjectures for future development of deep learning. Deep learning has already had in-depth research and a wide range of application scenarios, and has been put into practical use in real life with excellent performance. However, there's still a lot to exploit in the area of deep learning and neural network and it has great follow-up research space and considerable application potential.

References

1. Guo Y , Liu Y , Oerlemans A , et al. Deep learning for visual understanding: A review[J]. Neurocomputing, 2016, 187(C):27-48.

2. Chun-Xia Z , Nan-Nan J I , Guan-Wei W . Restricted Boltzmann Machines[J]. Chinese Journal of Engineering Mathematics, 2015.

3. Hinton G E , Osindero S , Teh Y W . A Fast Learning Algorithm for Deep Belief Nets[J]. Neural Computation, 2014, 18(7):1527-1554.

4. Goodfellow I J , Pouget-Abadie J , Mirza M , et al. Generative Adversarial Networks[J]. Advances in Neural Information Processing Systems, 2014, 3:2672-2680.

5. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. (1998). "Gradient based learning applied to document recognition". Proceedings of the IEEE, 86(11):2278–2324.