Analysis on the latest research hotspots of computer deep learning optimization algorithms

Xiaokang Guo*

Apollo Intelligent Mobility Technology (Guangzhou) Co., Ltd., Guangzhou, Guangdong 510660, China

*Corresponding author e-mail:guoxiaokang@gz.apollo.baidu.com

Abstract. Deep learning can effectively learn the characteristics of data, which is of great significance to the development of artificial intelligence. This article first outlines the development of deep learning, then describes four typical deep learning algorithms, and then presents the progress of the current deep learning in the optimization of learning algorithms After a review, it finally discusses the advantages and disadvantages of deep learning in data analysis, model building, and algorithm optimization, as well as the problems that need to be further studied and resolved.

Keywords: Deep Learning, Algorithms, Hot Spot Analysis.

1. Introduction

Deep learning belongs to an important category of machine learning in artificial intelligence, and it is also an important research field of machine learning. The third wave of artificial intelligence (AI) led by deep learning reflects its outstanding capabilities and enables the development of AI products to break through the ideal practice, thus truly reaching the application stage.

The concept of deep learning is inspired by traditional artificial neural networks, but its true meaning is still different from traditional neural networks. Deep learning is a structure of perceptrons with multiple layers of concealment. Compared with traditional artificial neural networks, deep learning algorithms use the most basic logical structure to create a neural network system that simulates brain behavior analysis and learning, and understand and analyze data by simulating the working mechanism of the human brain neural network. At present, deep learning has continuously set new records in audio recognition, image recognition, text recognition, NLP and other fields, and its accuracy rate has surpassed that of human beings. It can be said that deep learning is an improvement of traditional neural networks.

2. Typical deep learning algorithms currently being studied

2.1 Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) is mainly used in image processing. Image processing is similar to the working mechanism of human visual neural network.

After AI recognizes the image, it needs to be digitally processed. The image is composed of images, and each image is composed of colors. Now, every photo is at the level of 1000×1000 pixels, and each image has 3 parameters of RGB to display color information. If we need to process a photo with a resolution of 1000×1000 pixels, we must perform 3 million (1000×1000×3=3,000,000) parameters. And processing such a large amount of data requires a lot of time and powerful computing power, and this is just a simple picture. The problems that CNN can solve mainly include: simplifying the image to be processed with a large amount of data and retaining the original features of the image to be processed.

Among them, simplification refers to the dimensionality reduction processing of a large amount of original data to reduce the large amount of data originally required. The purpose is to save processing costs and resources. In CNN, in most processing modes, the dimensionality reduction of

the parameters will not affect the results. Compared with the original high-pixel image, the simplified low-pixel image does not affect the visual recognition of the content in the image.

The preservation of image features refers to the special digital processing of the characteristic part of the original image, which is equivalent to "marking" the characteristic part of the image, and this "marking" will not change with the change of the position of the feature in the image. If an image is digitally simplified in a traditional way, for example: the content of the image is a ping pong ball on a table, if the pixel of a ping pong ball is 1, and the pixel of no ping pong ball is 0, then the content of the ping pong ball is 0. Different locations will result in completely different processed data. But from the perspective of traditional human visual neural network, the original image content has not changed, only the position that can represent the original image features. Therefore, when the position of the object in the image changes, the processed data is very different from the result processed by the human visual neural network. CNN uses a processing method similar to human vision to "mark" the features of the original image. When the feature position of the original image changes, the image it recognizes also has a high similarity to the original image.

CNN consists of convolutional layer; pooling layer; fully connected layer as its basic structure. The working principle of each of its basic structures is: the convolutional layer is used to identify and "mark" the local features in the image for digital processing; the pooling layer performs dimensionality reduction after obtaining a large number of parameters; the fully connected layer is equivalent to the traditional Artificial neural network will further process the reduced parameters and output the result.

The practical applications of CNN mainly include:

The classification and retrieval of pictures. Among them, the classification and retrieval of images with obvious characteristics have a high accuracy rate. It is widely used in machine recognition of images and classification of similar pictures; target segmentation: a brief understanding is only a pixel-level Divide. It is used to make a pixel-level division of people or objects and the background in the picture. It is mainly used in PS, PR, etc.; face recognition: matching by identifying facial image features is mainly used in payment, life, security, etc.; bone recognition: matching and recognizing the behavior of bones by identifying the key bones of the human body. Mainly used in security, image generation, games, etc.

2.2 Recurrent Neural Network (RNN)

Convolutional neural networks and ordinary machine calculations are mostly one-to-one correspondence between entry and output, that is, a certain entry can get a certain output. The different inputs and outputs are completely unrelated to each other. It's just that in a specific situation, a connection between a certain entry and an output is far from enough. And RNN is a calculation that can efficiently manage sequential data information. Including: text information content, audio and video, stock market quotes, etc. The reason why it can manage the sequence data information is because the entry before the sequence will directly affect the output afterwards, which is equivalent to the "remember function". However, because RNN has a great short-term memory problem, it has little effect on long-term data. And because of RNN, variant calculations such as LSTM and GRU are produced. The most important advantage of this variant algorithm is that long-term messages can be stored more efficiently and important information can be selected for storage, and unimportant messages will be "forgotten".

2.3 Generative Adversarial Networks (GANs)

Among them, the most unique and powerful field of deep learning is through refining the characteristics of one's own learning. Powerful supercomputing capabilities help to deal with many problems that humans cannot handle at all, but the process of human judgment and generating results is very costly and very inefficient. And GAN can automatically judge and optimize, and continuously optimize, which is a very efficient and low-cost way. After automation, machine learning is stronger and more adaptable.

3. Optimization in deep learning

Deep learning is essentially optimization, and the most common problems in deep learning are also optimization problems, because the optimization algorithm involves the processing efficiency and computing power of the neural network of deep learning, and it is also an important criterion for evaluating the results of deep learning. One of the most basic methods of deep learning algorithms is to create an optimization model first, and then use the optimization algorithm to optimize the objective function in order to train the best neural network. To establish an optimization model and obtain an optimization problem algorithm, it is necessary to understand the classification and general form of the optimization problem. Optimization problems can usually be divided into three levels: the simplest is unconstrained optimization, followed by optimization with equality constraints, and the most difficult is optimization with inequality constraints and equality constraints. One of the most common optimization problems is the unconstrained optimization problem. At present, this problem is usually solved by multiple iterations of the derivative of the objective function. Common methods are as follows.

3.1 Gradient Descent

The gradient descent method is currently the most concise optimization method for the processing process with the longest use cycle. It is also the most common algorithm to deal with unconstrained optimization problems. When the objective function is a convex function, since its solution is a global solution, it only needs to solve the derivative of the objective function through multiple iterations to obtain the optimal solution. Under normal circumstances, the convergence rate of the solution process is obviously slower when it is close to the global optimal solution, and it needs a lot of iterations, which is undoubtedly slow.

3.2 Newton's method and quasi-Newton's method

3.2.1 Newton's method.

Newton's method is a method that can similarly calculate equations in the entire real number field and the complex number field. The specific algorithm steps of Newton's method are to approximate the quadratic function through the quadratic term of the Taylor series of the function f(x), and then use the minimum point of the constructed quadratic optimization model as the new iteration point, and repeat this continuously A process until the approximate minimum point that satisfies the accuracy is obtained. The biggest advantage of this algorithm lies in its extremely fast convergence rate.

3.2.2 Quasi-Newton Methods.

The quasi-Newton method is usually used to deal with nonlinear optimization problems, and the positive definite matrix is used to approximate the inverse of the Hessian matrix. This method simplifies the calculation amount of the Newton method in the algorithm to a certain extent. The most basic idea of the quasi-Newton method is to optimize the problem that the inverse matrix of the more complex Hessian matrix must be calculated every time in the Newton method, and the calculation efficiency is improved. Compared with the gradient descent method, the quasi-Newton method is far superior to the gradient descent method for solving complex optimization problems, because it only needs to know the gradient of the objective function at each iteration, and determine the change law of the gradient, Construct an optimization model of the objective function problem, so that it can generate the superlinear convergence problem.

3.3 Conjugate Gradient

The conjugate gradient method has the characteristics of the gradient descent method and its efficiency is better than that of the gradient descent method, but it is not as efficient as the Newton method. However, the conjugate gradient method makes up for the shortcomings of the two methods: the conjugate gradient method modifies the gradient descent method repeatedly in The iteration of

the same optimization direction makes up for the slow convergence rate of the gradient descent method and the construction of the quadratic function avoids the defects of calculating the complex Hessian matrix and finding the inverse in the Newton method. In various optimization calculations, the conjugate gradient method performs more prominently. The advantage is that the required storage volume is small, the stability is high, and other external parameters are not required.

3.4 Heuristic optimization method

The heuristic method is the way of thinking that people use when solving problems based on the law of experience. Its advantage is that when dealing with optimization problems, you can use the laws and basic ideas of traditional optimization algorithms to obtain the optimal solution without having to deal with the optimization problem through certain algorithm steps. You can directly select the feasible solution and refer to the previous optimization of processing algorithm.

4. Conclusion

Deep learning can enable the machine to automatically complete the extraction and representation of feature data, and its strong learning ability also makes it the most important research direction for AI development and reaching a high level. Its huge range of adaptation and the processing of a large number of levels of data also enable the development of deep learning to save a lot of manpower and material resources to a large extent, and its accuracy in data extraction and analysis far exceeds that of traditional human neural networks. deal with. Once the deep learning algorithm is trained as a model, its preservation is relatively high. The extraction and analysis of feature data under the model can also improve the efficiency of data processing to a certain extent.

However, while deep learning processes a large amount of data, it also requires the machine to have high hardware facilities to support deep learning's ultra-high computing power requirements. The defects shown in the hardware facilities are mainly high cost requirements and portability. Poor. At the same time, a lot of resources and time are needed to develop the optimization algorithm model, and the optimization model that is often developed can only be applied to a fixed scene, which shows the defects of deep learning, such as poor flexibility and undifferentiated recognition.

In order to reach the climax of development in the AI field again, it is necessary to solve the shortcomings exposed by deep learning, improve its large-scale data extraction requirements, and improve the portability of deep learning so that it can be applied to a wider range of practical scenarios. For the processing of some complex problems, a combination of multiple algorithms is used, which requires a certain amount of research in the development of combinatorial algorithms. And its deep learning should also be changed in terms of undifferentiated extraction. When training its model, it also sets certain restrictions to fundamentally improve the shortcomings of deep learning.

References

[1] Fan Yaqin, Wang Binghao, Wang Wei, Tang Yewei. Domestic research on deep learning [J]. China Distance Education, 2015(06):27-33+79.DOI:10.13541/j.cnki.chinade.2015.06. 007.

[2] Hu Yue, Luo Dongyang, Hua Kui, Lu Haiming, Zhang Xuegong. A review and discussion on deep learning [J]. Journal of Intelligent Systems, 2019, 14(01):1-19.

[3] Tian Qichuan, Wang Manli. Research progress of deep learning algorithms [J]. Computer Engineering and Applications, 2019, 55(22): 25-33.

[4] Su Fu, Lu Qin, Luo Renze. A review of research on image classification based on deep learning [J]. Telecommunications Science, 2019, 35(11): 58-74.

[5] Yang Li, Wu Yuqian, Wang Junli, Liu Yili. Overview of cyclic neural network research [J]. Computer Applications, 2018, 38(S2): 1-6+26.

[6] Tang Meili, Hu Qiong, Ma Tinghuai. Research on speech recognition based on recurrent neural network [J]. Modern Electronic Technology, 2019, 42(14):152-156. DOI: 10.16652/j.issn.1004-373x. 2019.14.035.

[7] Zou Xiufang, Zhu Dingju. A review of generative adversarial networks [J]. Computer System Applications, 2019, 28(11):1-9. DOI:10.15888/j.cnki.csa.007156.

[8] Guo Tiande, Han Congying. From numerical optimization method to learning optimization method [J]. Journal of Operations Research, 2019, 23(04):1-12. DOI:10.15960/j.cnki.issn. 1007-6093. 2019. 04. 001.

[9] Jing Hongxia. Algorithm research and realization of unconstrained optimization problem [D]. Beijing University of Posts and Telecommunications, 2013.

[10] Liu Jinkui. Research on Unconstrained Optimization Problems and Several Solutions of Nonlinear Equations [D]. Chongqing University, 2016.

[11] Chen Yuquan. Research on the basic theory of fractional descent method[D]. University of Science and Technology of China, 2020.DOI:10.27517/d.cnki.gzkju.2020.000018.

[12] Chen Yu. Research on Convergence of Conjugate Gradient Method [D]. Yangtze University, 2012.

[13] Cao Bangxing. Application of Quasi-Newton Method in Solving the Extreme Value of Unconstrained Multi-dimensional Function [J]. Journal of Dali University, 2019, 4(06): 1-4.

[14] Hou Yuqingyang, Quan Jicheng, Wang Hongwei. Overview of the development of deep learning [J]. Ship Electronic Engineering, 2017, 37(04): 5-9+111.

[15] Zhang Haodong, Li Linzong. Discussion on the challenges and prospects of the development of deep learning [J]. China New Telecommunications, 2020, 22(11):217.