Review of Deep Neural Network Based on Auto-encoder

Xinlin Zhang, Yuanmeng Hu, Li Zhang, Yajing Kong, Xiang Gao and Huajing Wei

ABSTRACT

Deep Learning gets a new research direction of machine learning. After years of deep learning development, researchers have put forward several types of neural network built on the Auto-encoder. In this article, firstly, the origins and basic concepts of deep learning, automatic encoders, deep belief networks, and convolutional neural networks are introduced. The principle of deep neural networks based on Auto-encoders is described, and the application of hybrid neural networks in various types is introduced. Finally, the problems existing in the current stage of deep neural network based on Auto-encoders and the future prospects of it are described.

KEYWORDS

Deep Learning; Auto-Encoder; Deep Brief Network; Convolutional Neural Network; Hybrid Neural Network.

INTRO TO DEEP LEARNING

As a brand-new research area in Machine Learning, Deep learning has made some breakthrough progress in some fields such as Natural Language Processing and Computer Vision. The principle of Deep Learning is establishing a model to simulate the structure of the connections of the neurons in human brains. To

---

1Xinlin Zhang, Yuanmeng Hu, Li Zhang, Yajing Kong, Xiang Gao, School of Mathematical Sciences of OUC, Qingdao, China
Huajing Wei, School of Earth Sciences and Engineering of Nanjing University, Nanjing, China
interpret the data, the model which was set up by deep learning counts describing for
the features of data by many transforming stages when processing the signal such as
images, audio and text.

The concept of Deep Learning was put forward by Hinton et al. from the
university of Toronto in 2006 [1]. The main idea is to extract the more abstract and
distinct high-layer presentations from lower-layer features and finally discover the
distributed traits of data to deal with the complicated problems. As deep learning
contributes to simplifying the data and keeping the input information simultaneoulsy, the input is almost equal to the output [2].

The deep neural network based on the auto-encoder is a special architecture of
deep learning, which aims to combine the ability to debug parameters in pre-training
of the auto-encoder with the generalization of the deep neural network to obtain
better simulation results. As the most efficient learning method in the field of
machine learning, deep learning is widely used in the fields of natural language
processing, computer vision, intelligent manufacturing, etc., and promotes the
concept of artificial intelligence to become a reality.

AUTO-ENCODER

With the progress of scientific research, the researchers have made some
adjustments and improvements to the auto-encoder, producing a variety of auto-
encoders [3]. However, according to the adjustment, improvement and pre-training,
the existing auto-encoder can be divided into the following types [4].

A. Sparse Auto-Encoder

Sparse auto-encoder (SAE) is the concept of new auto-encoder proposed by
Bengio in 2007 [5]. This algorithm adds an extra penalty factor in the reconstruction
error function to achieve the sparse limit.

After increasing the sparse constraint conditions, the average value of the hidden
layer neuron output should be as low as possible, and the expression of SAE's loss
function is following.

\[
J_{sparse}(W,b) = J(W,b) + \beta \sum_{j=1}^{s_2} KL(\rho_j^{\hat{\cdot}} \| \rho_j)
\]  

(1)

In (1), \(\beta\) is the weight that controls the sparsity penalty factor. \(\rho\) is the sparse
parameter. \(s_2\) is the number of hidden layer neurons.

\[
\hat{\rho}_j = \frac{1}{m} \sum_{i=1}^{m} [a_j^{(2)}(x^{(i)})] \]  

(2)
Equation (2) is the average output of hidden layer neurons, in which \( a_j^{(2)}(x^{(i)}) \) is the activation degree of the hidden neuron in the neural network of the encoder for the given input as \( x^{(i)} \). Minimizing the loss function can make \( \hat{\rho} \) close to \( \check{\rho}_j \) for sparse representation.

SAE can automatically extract the sparse high-dimensional data variables explanatory factor, retains the original input non-zero characteristics, and learn from without annotation data characteristics. It can give better feature description than raw data and automatically extract edge feature. However, due to the inconsistency of the density of the original data distribution, the sparse variables are difficult to control after the information is unlocked, and the problem of overfitting often appears.

**B. De-noising Auto-Encoder**

The De-noising Auto-Encoder (DAE) was proposed by Vincent et al. in 2008 [6]. The main idea of DAE is to first perform a corruption on the input vector, then encode and decode the input vector based on the interference, and the input vector after decoding is required to keep the original information as much as possible. If the output is able to reconstruct the original input that has been added interference, the network is very robust to the input data. In another word, it is required that the expression of the hidden layer neurons has certain robustness to the noise interference on the input data, which is called the noise robustness constraint. And the auto-encoder that satisfies the noise robustness condition is called the de-noising auto-encoder.

The training rule of de-noising automatic encoder is to reconstruct log-function. And the function is:

\[
- \log P(x | c(\tilde{x}))
\]  

(3)

In (3), \( x \) is the input data that is not interfered by noise. \( \tilde{x} \) represents the noise. And \( c(\tilde{x}) \) is the data code obtained from \( \tilde{x} \). The log-likelihood function is used for the training of DAEs of this kind. And this method can utilize the non-classification data to the maximum extent. The original data can be expressed by using data that has not received noise interference. De-noising auto-encoders learn more robust representations of input signals, and their regularization ability is stronger than that of general auto-encoders, which lays a solid foundation for the development of depth theory [4]. However, when the original data are a complex high-dimension function, the computation is so large that de-noising auto-encoders have less efficiency in computation and the processing time is longer.
DEEP NEURAL NETWORK

A. Deep Brief Network

1) BOLTZMANN MACHINE

The Boltzmann machine (BM) is a random recursive neural network, whose neurons determine their state values through probability distribution. In a Boltzmann machine, any neurons in the visible layer can find its corresponding neuron in the hidden layer. Furthermore, the neurons in the visible or hidden layers of the Boltzmann machine are connected to each other, which is the biggest difference with the restricted Boltzmann machine. Although the neurons in the visible and hidden layers of the Boltzmann machine are connected, the ones in the layer are not connected. That is the Boltzmann machine with special structure proposed by G. E. Hinton in 2006 [7]. The method of getting a restricted Boltzmann machine is to find the optimal parameter values by using the probability method. When the parameter value reaches the maximum value of the maximum likelihood function, the parameter value can also minimize the energy function.

\[
E(v, h, \theta) = \sum_{i=1}^{n} \sum_{j=1}^{m} W_{ij} v_i h_j - \sum_{j=1}^{m} b_j v_j - \sum_{i=1}^{n} c_i h_i
\]

In (4), \( \vec{b} = (b_1, b_2, ..., b_m) \) stands for the offset of the visible node; \( \vec{c} = (c_1, c_2, ..., c_n) \) stands for the offset of the hidden node. \( \theta (\vec{W}, \vec{b}, \vec{c}) \) Means the parameters of the model.

At the same time, Hinton proposed a training RBM algorithm called the contrast divergence (CD) algorithm [8]. The CD algorithm uses the reconstruction error as the index for changing its weight. That is, the input error between the hidden layer output and the visual layer serves as the index of the update weight. The CD algorithm is generally applied to initialize the weight of the model, because the initial weight of a good whole network can greatly reduce the time of experiment and improve the efficiency of experiment.

2) DEEP BELIEF NETWORKS

Deep belief networks (DBN), also known as the deep Boltzmann machine, is a probability generation model. It consists of several restricted Boltzmann planes, which could be divided into visible layers, hidden layers and a BP neural network. All the neurons between the visible and hidden layers are interlinked and are bidirectional. However, due to the special structure of the restricted Boltzmann machine, the deep belief network has no interconnecting neurons in the layer. The
training algorithm of the deep belief network contains two steps: unsupervised pre-training and supervised fine-tuning [9]. The method commonly used in the process of the unsupervised training step is greedy algorithm. But some problems still exist in deep belief networks, although the possibility of convergence to local minimum values is reduced compared to BP network [10]. Moreover, the learning speed of it is too slow. Hence, these are what deep hybrid neural networks need to solve.

B. Convolutional Neural Network

Convolutional Neural Network (CNN) is one of the deep neural network models. This neural network uses the local correlation of image data, reduces the number of parameters in the neural network, and facilitates the solution of parameters in the network. Convolutional neural networks have been widely used in the image field [11].

Traditional neural networks often involve a large number of parameters when processing image data, which gives rise to a large amount of calculation during the training process. However, convolutional neural networks can reduce parameters without affecting the accuracy of model training [12].

Sparse connection, weight sharing and pooling are three important concepts in convolutional neural network. Sparse Connectivity is to model the local areas in the data to discover local features. Sparse-connections reduce the number of edges in a neural network compared to full-connections. In the convolutional neural network, the parameters of each set of receptive fields are shared with each other, which is called Shared Weights. The weights between the nodes of the upper layer connected to the nodes of the next layer are the same. The sharing of weights can greatly reduce the number of parameters that need to be learned and improve learning efficiency. Pooling refers to dividing an input image into a series of non-overlapping square matrices, extracting its eigenvalues and outputting the eigenvalues of each region. The eigenvalues can be either average or maximum. According to different characteristics, pooling can be divided into average pooling and maximum pooling. Using the pooling strategy can eliminate non-feature values and further reduce the amount of computation.

The convolution neural network consists of three parts: convolution layer, subsampling layer and full connection layer. The Convolutional Layer applies several filters to the input data for filtering, and different filters extract different features of the input parameters to obtain different feature maps (FM). The number of feature maps and the number of filters are same. Subsample Layer, also known as the Pooling Layer, is a layer that implements pooling operations and can reduce the size of input parameters. Full Connected Layer is the same as the traditional neural network. Each neuron on this layer is connected to each neuron of the previous layer, and the full connected layer corresponds to the sparse connecting layer. The last sub-sampling layer or convolution layer is usually connected to one or more full-connection layers. The full-connection layer is a classifier, and the scores
calculated obtain a one-dimensional matrix, that is, a vector. The output of the fully connected layer is the final output of the entire convolutional neural network.

**DEEP NEURAL NETWORK BASED ON AUTO-ENCODER**

**A. Deep Belief Network Based on Auto-Encoder**

Because the restricted Boltzmann machine initial weights and bias in the process of the training are set in a random number between 0 and 1, which may cause the big reconstruction error (i.e., the difference between reconstructed data and input). A good pre-training network parameter can improve this phenomenon and make the model more stable [9]. So the hybrid neural network is to attach the deep belief networks to a sparse auto-encoder. Hence, the output of SAE is the input of a deep belief network and also the pre-training initial value of the first restricted Boltzmann machine in the deep belief networks. This combination makes the efficiency of training and efficiency of the classifying improved.

**B. Convolutional Neural Network Based on Stacked Auto-Encoder**

Convolutional neural network based on stack self-encoder (SA+CNN) by Zhang Chunyu is to use the deep convolutional neural networks to extract high-dimensional global feature of images, to use a stacked automatic encoder through unsupervised learning of features to yields a binary hash code, and to use the semantic similarity of image tags to finely tune the parameters of stacked automatic encoders [13]. Compared with the traditional hash algorithm, this method achieves better results in image retrieval.

**APPLICATION OF DEEP NEURAL NETWORK BASED ON AUTO-ENCODER**

**A. Application of Deep Belief Network Based on Auto-Encoder**

Due to the good feature of data extraction in sparse auto-encoder and the deep belief network’s great ability to classify data, the combination of them, a hybrid neural network, has the strengths of both, the great ability to extract data and also the excellent ability to classify the data. Moreover, when the volume of data is relatively large, the classification accuracy of the general hybrid neural network is higher than that of the traditional classifier, such as SVM. In this age of big data, this age requires hybrid neural networks, because the applications of them are very practical. Manual extractions of data are not only a waste of time but also a waste of money. The learning of text by machines makes society more efficient. Qin Shengjun et al. and Zhou Chao proposed the combination of the sparse auto-encoder and the deep
belief network, and they applied it to the study and the classification of texts or words [14] [15]. The experimental results show that the hybrid neural network has higher accuracy of classification than that of the traditional classifier when the data are set in high-dimensional case. Wu et al. put forward the hybrid model of stacked sparse auto-encoder and deep belief network, and applied it to the video monitoring to avoid the monitoring obscured [16]. It also avoids the artificial extraction, because the manual data extractions are complex and inaccurate.

B. Application of Convolutional Neural Network Based on Auto-Encoder

The convolutional neural network has an efficient generalization ability for image data, but the first few layers cannot be effectively trained. However, the automatic encoder has a strong preprocessing capability. Zhang Wen-da et al. used this feature to propose the MS-CNN algorithm, which uses multi-scale input images to increase local invariant information [17]. And it also uses different size filter convolutions to collaborate with different down-sampling intervals to obtain feature invariance without losing the details of the target. This algorithm effectively improves the recognition rate and robustness. In terms of the problem that the traditional neural network does not have a high recognition rate for complex background images, the method of sparse self-encoder unsupervised pre-training filters on convolutional neural networks solves the problem that the first few layers of traditional convolutional neural networks cannot be effectively trained. Li Hui et al. used a convolutional neural network based on sparse auto-encoder and used SVM as a classifier to identify target images [18]. The algorithm used has higher efficiency and recognition rate than traditional PCA+SVM and traditional convolutional neural network algorithms. Yu Tao also used this architecture to conduct image recognition experiments, and the result of network classification was ideal. And he also found that compared to the traditional image recognition artificial neural network, this architecture has the advantages that it has less network parameters and a certain tolerance for the translation of the image, moreover it is not easy to overfitting [19].

PROBLEM AND DEVELOPMENT TENDENCY OF DEEP NEURAL NETWORK BASED ON AUTO-ENCODER

Compared with traditional Auto-encoders or traditional classifiers, deep neural networks based on Auto-encoders have higher accuracy of classification and higher efficiency of learning in the case of high-dimensional data and large amounts of data. However, if the data are set in remarkably high dimensions, the classification accuracy of the deep neural network based on the Auto-encoder will decrease, and the experiment time will increase as well. The performance of the deep neural network based on the Auto-encoder is not as good as that of the traditional neural network under the same conditions. Even in the low-dimensional case, the neural
network based on the Auto-encoder will also produce over-fitting phenomenon, and the performance is inferior to the traditional classifier. In the cases of high dimensions or low dimensions, how to ensure that deep neural networks based on automatic encoders still maintain high classification accuracy and high learning efficiency is the goal of future research on deep neural networks based on the Auto-encoders.

REFERENCES