A Study on the Application and the Advancement of Deep Neural Network Algorithm

Jiulin Song\textsuperscript{1}, Yansheng Chen\textsuperscript{2,*}

\textsuperscript{1}RDFZ Xishan School, Beijing, China, 100193
\textsuperscript{2}Institute of Big Data and Artificial Intelligence Applications, Guangdong Industry Polytechnic, Guangzhou, China, 510300

*Corresponding author: 2005106056@gdip.edu.cn

Abstract. Deep neural network is a new type of learning algorithm, which has both global and local aspects and performs well in pattern recognition and computational speed. In recent years, deep neural network algorithm has been widely used in scientific research and real life, but its complexity, parallelism and other characteristics lead it to be a very challenging and innovative research area. This study briefly introduces the basic principles and theoretical knowledge of deep neural network algorithms, and mainly discusses their applications and Advancement of feature extraction in the field.

Keywords: Deep Neural Networks, Machine Learning, Feature Extraction, The Advancement, Algorithm

1. Introduction
Deep neural networks are composed of RBF algorithm and single or double layer parallel distributed units with multiple hidden layers, which have strong nonlinear mapping ability and can well solve the problems encountered in traditional methods that are difficult to handle and have obvious defects. Feature extraction is a key technology of artificial neural networks, which is important for learning algorithms. Traditional classification methods mainly rely on the human brain to recognize different colors, shapes and sizes in images. However, there are often many problems in the practical application. We study feature extraction based on deep neural networks, and explore the application of deep neural network algorithms in feature extraction and the latest progress.

2. Feature extraction and its deep neural network related model

2.1. Extraction of features
In pattern recognition, features are typically used to describe and represent the subject under study. The features extracted from the data by the algorithm have a direct impact on the performance of the model, and how to efficiently extract features that fit the structure of the data distribution is a hot research topic in this area. Most traditional feature extraction methods select features according to different tasks and data structures, such as Gabor features, SIFT and local binary models [1-3]. However, developing a good feature is not easy, especially for large datasets, and is not only time consuming but also requires...
heuristic expertise from the researcher, relying heavily on experience and luck, otherwise there is a risk of extracting features with low generality. With rapid advances in cognitive biology and neurobiology, researchers have begun to explore ways to automatically learn feature descriptions. Figure 1 shows a diagram of information processing in the visual system of the brain. Deep learning models have been developed to simulate the process of hierarchical iterative learning of abstract signals in the human brain. Deep learning is essentially a feature extraction process, and its main advantage over traditional pattern recognition methods is the independent learning of feature representations in a multi-step abstraction process. In addition, more complex features can be represented with fewer parameters and effective features for new applications can be learned quickly from training data, making this feature extraction perform better than shallow networks [4].

![Diagram of information processing in the brain vision system](image)

**Figure 1.** Schematic diagram of information processing in the brain vision system

2.2. **Autoencoder**

Deep neural network is developed on the basis of artificial neural network, which is a computing model built by imitating the structure and function of biological brain and consists of a large number of nodes (or neurons) interconnected with each other, and different connections can form different networks, referred to as neural networks or neural-like networks. Neuron is the smallest information processing unit of neural network, and its basic model is shown in Figure 2.

![Neuron structure diagram](image)

**Figure 2.** Neuron structure diagram

The output process of a neuron is generally:

1. Receiving input signals $x_i$, each connection represents an input signal, and each input letter $x_i$ multiplied by the connection weight $w_i$ is a stimulus quantity for the neuron.
2. Calculate the weighted sum of the input signal $x_i$ on the neuron's synapse, and calculate the input weighted sum $v_m$ from $w_i$ and $b$:

$$V_m = \sum_{i=1}^{n} w_i x_i + b$$

(1)
(3) After the nonlinear mapping function restricts the output of the neuron within a certain range, the output is:

\[ h_{w,b}(x) = f(v_m) = f(\sum_{i=1}^{n} w_i x_i + b_0) \] (2)

Where \( f(\cdot) \) is the activation function, to limit the amplitude of the output signal of the neuron. The commonly used activation functions \( f(\cdot) \) are sigmoid function and tanh function. When the mapping interval is \([0, 1]\), the continuously derivable sigmoid function is usually used; when the mapping function value is in the interval \([-1, 1]\), the symmetric tanh function is used as the activation function. Different choices of activation functions result in models with different functions and structures, and the activation function can be chosen flexibly according to the actual situation. The activation function determines whether a neuron node should be activated or not. If the activation function is linear, then the weights and biases are combined into linear equations in the network, which only have the ability of linear expression and cannot solve complex problems, even if it is a multilayer neural network.

Even a multilayer neural network, which is actually equivalent to a multilayer perceptron, has only one hidden layer, and it is impossible to fit any function at all. Therefore, some scholars propose to introduce, a nonlinear function to perform a nonlinear transformation on the input to approximate the data distribution, thus giving a real meaning to deep neural networks. Also, the activation function allows back propagation, and the gradient error of the activation function can be used to adjust the weights and biases.

Autoencoder (AE) is a conceptually simple deep neural network used to obtain useful data representations through unsupervised training. It consists of an encoder that outputs a hidden (latent) representation and a decoder that tries to reconstruct the input using the hidden representation as input. Where the number of nodes in the input layer is the same as the number of nodes in the output layer; the autoencoder can be used to implement many dimensionality reduction techniques, which can be seen as dimensionality reduction of the data when the number of nodes in the hidden layer is lower than the input and output layers; when the number of nodes in the hidden layer is much higher than the input and output layers and some constraints are added, such as sparse autoencoder adding sparsity restrictions, sparsity can be obtained data [5].

3. **Deep belief network model**

The restricted Boltzmann machine is a deformed structure of Boltzmann machine, and its essence is to maximize the probability of eligible samples generated by RBM. BM model is an energy model based on statistical mechanics, named BM because the sample distribution obeys Boltzmann machine distribution. According to physics, every thing that exists in nature has its own corresponding steady state, i.e., the lowest state of corresponding energy, and thus it is possible to The energy function of the network in the steady state is defined, and then the steady state of the network is solved by an optimization algorithm. In fact, the energy model describes that each situation in the parameter space corresponds to the energy in scalar form through a mapping, which is called the energy function [6].

Boltzmann machines are stochastic neural networks defined by an energy function, consisting of a visible layer and an implicit layer, with connected weights for both intra-layer and inter-layer nodes, and with only two states for the output nodes: active and inactive, with 1 indicating activation and 0 indicating inactivity, and with unsupervised learning. Because the structure of the RBM model is symmetric, and each layer of neurons is independent of each other, Gibbs sampling can be used to obtain random data consistent with the model distribution, as shown in Figure 3.
4. Feature homogenization phenomenon

In the pre-training phase of DBN, each RBM is trained in turn to optimize the network, so RBM is also an unsupervised learning process, and its goal is to use the CD algorithm to maximize the likelihood function to find the appropriate parameters to fit the distribution of the training data. Therefore, the KL distance can be used to represent the difference between the input sample distribution and the Gibbs distribution represented by the RBM, and maximizing the likelihood logarithm function is equivalent to minimizing the KL distance between the two conditional distributions. In general, the value of the KL distance is greater than or equal to 0, and it is 0 only when the two distributions are identical. In addition, the value of the KL distance decreases as the difference between the two distributions is reduced. For RBM, the entropy of the input sample in the calculation formula is a fixed value and the other one cannot be found. If the KL distance is to be minimized, then the one that cannot be found has to be maximized, so from the CD algorithm point of view, instead of maximizing the log-likelihood function, the difference between the KL distances of the two distributions is calculated, which allows the network to reconstruct the distribution of the input sample to the maximum [7].

The current solution to the feature homogeneity phenomenon is to adjust the sparsity of the nodes in the hidden layer to reduce the similarity between the columns of connected weights, i.e., by adding a sparse penalty factor to the network for sparsification. According to a related study, the human visual system has only a few neurons activated for targeted things. Inspired by this study, We proposed a sparse coding theory in order to model the sparse representation of the visual system. If one is to have a fixed size representation, then a sparse representation is more efficient (than a non-sparse representation) in an information-theoretic sense, which allows each example to vary in the number of effective bits [8]. The sparse representation has properties in it that can be used to interpret the learned features, i.e., it has factors that correspond to meaningful aspects of the input as well as capturing changes in the data, so that the features are usually represented by a sparse representation, in other words only a few parts of a large number of features are activated.

5. Non-linear modified deep belief network classification method

The DBN model is composed of multiple RBM modules cascaded, so the activation function also determines the feature extraction ability during the training process of RBM. The RBM performs one step Gibbs sampling through the CD algorithm, first the visible layer is mapped by the activation function to obtain the output of the implicit layer, and then the output is used as the input of the visible layer, and again the mapping of the activation function reconstructs the input data distribution, and this The closer the reconstructed distribution is to the original data distribution, the better the model is fitted. Therefore, how to choose a suitable activation function to make the reconstructed distribution of the model as close as possible to the original data distribution becomes a current issue. In addition, because finite weights are important for the representation of network features, there are more stable gradient-based optimization methods when the output value after nonlinear mapping is finite; when the output is infinite, the model will have more efficient training as well as require a lower learning rate.

As the main biometric recognition technology, face recognition has important applications in artificial intelligence, pattern recognition and other fields, and has become a hot research topic for many scholars because of its advantages of good interactivity, friendliness and convenience [9]. In face
recognition systems, how to extract the main features of face images becomes a key issue. Local binary pattern method is a feature extraction method used for texture classification, and it has wide applications in expression analysis and face recognition because it can well describe the local micropatterns as well as the distribution of face images. Although LBP has advantages such as rotational invariance, it has a high dimensionality in the representation of face images, which leads to problems such as low recognition efficiency and unsatisfactory classification results.

According to the latest research results shown in, the face features extracted by deep neural networks have advantages such as moderate sparsity, good robustness, and also have strong selectivity for face attributes [10]. Currently, DBNs have been widely used in the field of face recognition, and the traditional DBNs ignore the local structure for low-level features of face images because the input pixel features are in vector form, making it difficult for the network to learn the local main features of face images, resulting in poor robustness. We imagined a MEDBN model with two implicit layers to extract features from face images and finally classify them by a classifier. The MEDBN-based face recognition framework constructed in this study consists of three main stages: preprocessing, pretraining, and fine-tuning. First, we input the face image, chunk the face image into n regions, and then compute each region using the unified model LBP, and then count the histogram of local texture in each region, and then use the texture feature h obtained in the preprocessing stage as the input of MEDBN, and finally fine-tune the whole MEDBN globally using the gradient descent method to get the better network. Then, the test samples pre-processed by the uniform mode LBP are used as the input of MEDBN to extract the features, and a Softmax classifier is used at the top layer of the network to classify them to obtain the final classification accuracy.

6. Conclusion

Feature extraction algorithms are a fast method of information retrieval that can be used to classify different objects in the original image as needed, and we often use artificial neural networks when processing images. Feature extraction has an important position in the field of computer intelligence, it is the processing of the original image to express the object to be measured into a specific form, and then use this data to build a function expression that can be understood by the recognizer and can distinguish between different categories and attributes.

Acknowledgments

This research was funded by Natural Fund Project of Guangdong Industry Polytechnic (Project # X20190132), Guangdong Provincial Social Science Planning Foundation (Project # GD20XGL49/GD21CGL13), and the Base and Basic Applied Study Foundation of Guangzhou City (Project# 202102080475).

References

[1] Chen Y S, Zhu L, Ghamisi P, et al. Hyperspectral Images Classification With Gabor Filtering and Convolutional Neural Network [J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14(12):2355-2359.

[2] Rashid M, Khan M A, Sharif M, et al. Object detection and classification: a joint selection and fusion strategy of deep convolutional neural network and SIFT point features[J]. Multimedia Tools & Applications, 2019, 78(12):15751-15777.

[3] Meng X, Liang C, D Polajnar, et al. Local binary pattern network: A deep learning approach for face recognition[C] IEEE International Conference on Image Processing. 2016 IEEE International Conference on Image Processing (ICIP), 2016, 3224-3228

[4] Zhang M, Li X, Wang R. Incipient Fault Diagnosis of Batch Process Based on Deep Time Series Feature Extraction [J]. Arabian Journal for Science and Engineering, 2021, 46 (10): 10125-101362021.

[5] Singh M, Mishra D, Vanidevi M. Sparse Autoencoder for Sparse Code Multiple Access [C] 2021 International Conference on Artificial Intelligence in Information and Communication (IEEE
[6] Wang R, Zhang Z, Chen G. Energy Function and Energy Evolution on Neuronal Populations [J]. IEEE Transactions on Neural Networks, 2008, 19(3):535-538.

[7] He Z, Liu J, Dang K, et al. Leveraging maximum entropy and correlation on latent factors for learning representations [J]. Neural Networks, 2020, 131:312-323.

[8] Schleif F M, Raab C, Tino P. Sparsification of core set models in non-metric supervised learning [J]. Pattern Recognition Letters, 2020, 129: 1-7.

[9] Imaoka H, Hashimoto H, Takahashi K, et al. The future of biometrics technology: from face recognition to related applications [J]. APSIPA Transactions on Signal and Information Processing, 2021, 10:10. 1017/ATSIP.2021.8.

[10] Singh A, Prakash S, Kumar A, Kumar D. A proficient approach for face detection and recognition using machine learning and high-performance computing [J]. Concurrency Computat Pract Exper, 2021, e6582: 10.1002/cpe.6582