# **Big Data Stratification and Feature Extraction Based on Affine Neural Network**

### Hong Chen

School of Information Science and Engineering, Xi'an Eurasia University, Xi'an, Shaanxi, China

Keywords: Affine nonlinear system, Neural network, Big data stratification, Data characteristics

**Abstract:** This paper analyzes the relationship between affine neural network and big data based on the definition of affine neural network and big data layering, and expounds autoencoder algorithm, a big data layering method based on affine neural network. At the same time, two big data feature extraction algorithms based on affine neural network are proposed, which are based on the principle of affine invariance to identify shape features and affine invariance to extract contour curves.

### **1. Introduction**

Under the background of big data era, it is of great practical value to carry out big data stratification and feature extraction, and it is also the main content of big data analysis. At this time, the traditional data hierarchical pattern and feature extraction method are no longer applicable and need to be further updated. In order to reduce the actual workload under the premise of ensuring the quality of big data stratification and feature extraction, the introduction of affine neural network is an inevitable choice.

## 2. Brief Description of Concepts

#### 2.1 Affine Neural Network

The relation between the differential xdtot and x of the state quantity is non-linear for affine terms, but the relation of the control quantity u is linear; For the neural network, the human brain neural network is abstracted from the information processing angle to establish a simple model, The different network[1] is made up of different connections. The affine neural network is an affine nonlinear neural network and has the characteristics of affine invariance.

#### 2.2 Big Data Tiering

Big data stratification is a form of big data processing, which mainly classifies different user data based on certain characteristics. In big data management, there is a need to have a clearer control of the data, so it is necessary to expand the big data hierarchy. Relying on big data stratification, we can construct a clearer data structure, realize the "blood tracing" of data, reduce the number of repeated development, and promote the transformation of complex problems to the direction of simplification.

## 3. Big Data Tiering Based on Affine Neural Network

## 3.1 The Relationship between the Affine Neural Network and the Large Data

In essence, the imitation neural network is not big data, and the neural network is a big data processing method. At present, big data and neural networks (and other artificial intelligence) are generally mentioned at the same time, and the two are extremely closely linked. Because there are many network big data about user behavior, the related personnel can use neural network, affine neural network and other methods to analyze big data, so as to simulate human behavior, and urge the computer to complete the work of recognizing graphics, recognizing sound, analyzing problems and finding the best solution to the problem. At the same time, the emergence and rise of big data, also promoted the neural network technology.

technology. In this context, in order to deal with a large number of search behavior data efficiently, Internet enterprises, including Google, Baidu, Alibaba and so on, have invested a large number of researchers to optimize artificial neural networks in order to improve efficiency and form a variety of research results into use.

For neural networks (affine neural networks) and big data, both can be likened to a person's brain and what this person looks at. in which, the affine neural network is a data processing method, and the method is often dependent on the computer program, and the large data is a plurality of objective data and information, which is not dependent on the computer program, but is stored in a physical device such as a hard disk, a cloud disk, and the like, It belongs to an objective existence.

### 3.2 Big Data Hierarchical Method Based on Affine Neural Network

It is difficult to complete the layering of big data by using the traditional algorithm. In order to complete big data layering under the consideration of all the data characteristics, the autoencoder algorithm has high practical value and can give full play to the advantages of affine neural network.

Autoencoder algorithm belongs to unsupervised learning algorithm, which mainly uses machine self-study pattern to complete feature extraction and stratification. Compared with supervised learning algorithm, the biggest difference of unsupervised learning algorithm is that there is no clear label [2]. For example, when the supervised learning algorithm is used to identify the digital picture, the final database will show the corresponding digital classification of the picture, and build a neural network on the basis of it. Combined with training and learning, the prediction results of neural network can be adjusted to correspond to the known tags. From this we can see that the supervision of learning mainly completes the corresponding clear label of the study. But in practice, Not all the data have clear labels, or the sample size of the tags is small, so the unsupervised learning algorithm is used to make the rules and methods of machine autonomous mining more practical. This goal can be achieved by relying on autoencoder algorithm.

The specific work flow of autoencoder algorithm is as follows: suppose, the current data is  $60000 \times 600$ , in which 60000 is the number of users, 60000 is the characteristics of each user; because autoencoder belongs to a special neural network architecture, it includes input layer, hidden layer and output layer, input 600 input values in the input layer, hidden layer temporarily, and extract 600 output values in the output layer. It can be seen that the input layer is the same as the output layer in autoencoder algorithm, and it is also the core of the algorithm. Stand at A In terms of the structure of the utopencumber's algorithm, this layering process is equivalent to a reduction in the process of amplification. When the data passes through the neural network, the first "compression" re"decompression" process is realized, and since the compressed data is another existence form of the original data, the input is consistent with the output data, and the compressed data also inherits the characteristic value of the original data.

For the original data, the number of features can be regarded as dimensional features, relying on Autoencoder algorithm, a large number of eigenvalues can be reduced to two-dimensional. As an example of 600 user features proposed above, 600 customer features can be "compressed" to two-dimensional features with the support of this algorithm. At the same time, because the original data is still obtained by compression in the decompressed mode, it can be inferred that the compressed state represents the data characteristics of the original data itself.

#### 4. Big Data Feature Extraction Algorithm Based on Affine Neural Network

#### 4.1 Identification of Shape Features on the Basis of the Principle of Affine Invariance

#### **4.1.1 Affine Invariance Feature Extraction System Structure**

When the affine invariance principle is used for feature extraction, the main analysis object is the image. In this study, the collected image information is de-noised and filtered in advance to obtain a relatively single image, which is easy to be recognized by the computer system. In this process, the K-means algorithm is used to segment the image information to be processed, and the binarization processing is carried out to transform it into binarization image. At the same time, combined with

expansion corrosion treatment, the black area with moderate size and relatively uniform can be obtained. By analyzing these black areas, we can get some areas with affine non deformable features, and combine the algorithm to complete the feature extraction.

#### 4.1.2 Determination of Data Processing Software

At present, VC ++ and matlab are two more common image data processing software, and have different application advantages. For VC ++, it has an integrated development environment that provides a variety of programming languages for the associated personnel and allows users to develop remote debugging [3]. For matlab, it combines the functions of matrix calculation, numerical analysis, data visualization and nonlinear dynamic system modeling and simulation, so it has higher advancement. For comparison, the use of matlab is better because the amount of programming used in image processing is more complex. The effect is to realize the reduction of the workload in a large number of image data processing.

#### 4.1.3 Selection of Feature Extraction Scheme

The affine geometric invariance or the local affine invariance can be used when the image feature extraction is carried out using the affine invariance principle. In this paper, when the affine geometric invariance is used to carry out feature extraction, the centroid is extracted as the feature point, and the affine invariant model is constructed; and when the local affine invariance is used for extracting feature extraction, the gray-scale extreme point needs to be used as the "Key points" to complete the feature extraction. In comparison, the process of the feature extraction based on the local affine invariance is simpler and relatively practical, and the effect is more obvious, so the study uses the local affine invariance to finish the extraction of the image feature.

## 4.1.4 Algorithm Design

In this study, road sign images are selected for feature extraction, and eight-direction recognition, chain code recognition and Hu moment recognition are used to complete feature extraction and recognition. Among them, the eight-direction recognition method mainly determines the arrow direction and type of road sign image, and completes the judgment of missing and noisy image; chain code recognition mainly completes stable and clear image judgment and determines the important geometric shape features in the image; Hu moment recognition also mainly realizes stable and clear image judgment and realizes the distinction of different road signs.

#### 4.1.5 Analysis of Recognition Results

In order to determine the effectiveness of the feature extraction method, six road sign images are selected for recognition. The results show that although the stretch image is deformed in the geometric transformation of stretching, the recognition results are obviously affected. Combined with geometric change and noise processing, it can be concluded that the feature extraction method (algorithm) has high accuracy and can complete the image recognition and feature extraction, which is difficult to recognize by the ordinary algorithm.

## 4.2 Local Description of Profile Curves Based on Affine Invariance

#### 4.2.1 Extraction of Chord Heights

In an open curve, the two ends are connected in a straight line, and the chord high point A1 is marked in the curve. If there are n points with the largest distance, it is necessary to take  $(n \ 1) / 2$  (n odd) points or  $n \le 2$  (n even) points as chord high points according to the direction of the curve; connect A1 and the two ends of the curve, use the above operation to find out the chord high points in the separate two ends of the curve, and repeat n operations. Among them, the value of n should be combined with the calculation needs and accuracy requirements to be determined.

#### 4.2.2 Algorithm Steps

The concrete steps of the feature extraction algorithm are as follows: using gaussian function to complete denoising and obtaining smooth outline curve; extracting corner points to obtain corner

coordinates arranged counterclockwise along the outline curve; dividing the curve between the two corners to construct the subcurve segment; extracting the chord point and constructing affine invariants; measuring the difference between the two subcurves; judging the concavity and convexity of the curve.

## 4.2.3 Results Analysis

In this study, the effectiveness of the algorithm is verified by using the 70 images in the gallery. The result shows that the method can complete the judgment of the asperity of the sub-curve on the basis of the recognition vector matching, and has the advantages of validity.

## 5. Conclusion

In summary, the affine neural network plays an important role in the large data layering and the feature extraction of large data. Among them, in the large data hierarchy, the use of the autoencumber algorithm has a higher real value and can play an advantage of the affine neural network. In the feature extraction of large data, based on the affine invariance principle to identify the shape features, the two methods have good effect on the local description of the contour curve based on the affine invariance, and the recognition experiments confirm the practicability and effectiveness of the two methods.

## References

[1] Judy, Zhu., Ningdan., Yang, Ninghua. (2019). And the like. A review of the method of image matching. China Image and Graphics, vol. 24, no. 05, pp. 677-699.

[2] Li, Hao. (2019). Feature extraction of human face image based on convolution neural network. Journal of Henan Institute of Engineering (Natural Science Edition), vol. 31, no. 01, pp. 66-70-76.

[3] Zhou, Hongho., Yi, Weining., Du, Lili. (2019). And the like. SIFT feature description subdimension reduction method based on convolution neural network. Progress in laser and photoelectronics, vol. 56, no. 14, pp. 121-128.