

# Model establishment and risk management model of industrial big data based on neural network

Zhihan Jiang

College of Applied Science and Technology, Hainan University, 450007, China

jiangzhihan@wut.edu.cn

**Abstract.** With the maturity of big data, the data hierarchy and data processing methods of big data are changing with each passing day. Big data has begun to occupy an important position in all walks of life, and data is an asset. Industrial risk has always been the focus of data generation and application. In this paper, we use TensorFlow to implement a big data classification model, and add different optimization methods and convolutional neural networks. In the research, the usage scenarios of different optimization methods and the characteristics of self-sufficiency were compared, so that the final classification accuracy of the model reached 99.67%. Using TensorFlow to build different types of neural networks to fit the Fourier transform, and use the trained model for ultrasonic detection of time-domain signals, and obtain better signal processing results. On this basis and analyze the characteristics of different models, as well as the advantages and disadvantages of signal processing.

**Keywords:** Industrial big data, neural network, risk management, model building, computational analysis.

## 1. Background of the subject research

### 1.1 The development of big data and the rise of industrial big data

In recent years, with the rapid development of computer technology and data storage capabilities, a large amount of data has been continuously accumulated in all walks of life. How to obtain valuable information from massive data and enhance enterprise value has become a current research hotspot[1]. In the current Internet industry, the Internet of Things (IOT) technology and computing technology continue to develop and mature, and big data has become a well-deserved hot spot in this era. Big data technology provides opportunities and challenges for the development of modern enterprises. For some traditional industries, the emergence of big data has brought huge challenges, forcing the industry to re-integrate, and some companies in the industry have to start changing. For modern enterprises, the scale and activity of data assets, as well as their ability to interpret and use data, begin to be directly linked to enterprise value[2-3].

### 1.2 Application and importance of big data

With the development of big data, its importance has become increasingly apparent, not only in the IT industry where it was originally produced, but also in applications in other industries, such as big data medical treatment, big data investment, and recently very popular Autonomous driving technology is all closely related to big data. The investment field is originally a data center, and big data can show its talents even more here. For example, in the popular global quantitative investment, through the analysis and modeling of historical data, the best investment portfolio is selected to obtain the maximum return and the minimum risk. Traditional investment cannot do this, because on the basis of big data, the statistical results will be closer to the mathematical expectation and standard error of the actual situation, which is the conclusion that the law of large numbers tells us. Big data can help investors lock in better investment performance[4].

## 2. Research Methods and Neural Network Algorithms

In this paper, data modeling technology is used to process related gear design and calibration data. First, big data modeling technology is used to model the data. Commonly used modeling methods are: cluster analysis, association analysis, regression and classification analysis, deep learning[5]. For industrial big data, neural network deep learning is used for classification processing. Common algorithms are: support vector machine, logistic regression, K-means, and random forest. Similar algorithms are also needed in subsequent chapters to fit the data processing method—Fourier transform. Therefore, this chapter sets out to discuss the rationale from these aspects[6-8].

### 2.1 Big data modeling methods

#### 2.1.1 K-means clustering

The K-means clustering algorithm mainly uses the Euclidean distance between data points, which is called the k-nearest neighbor-based (k-NN) algorithm. In this method, each data point  $x$  that exists in an  $n$ -dimensional set of data points  $X$  needs to be evaluated against the  $k$  closest points to  $x$  with respect to a subset  $N_k(x)$  of  $X$ . In this process, we use a fixed distance function, such as Euclidean distance. This score based on the average distance of the nearest neighbors is called the k-NN Global Anomaly Score (GAS)

$$GAS_{kNN}(x) = \frac{\sum_{o \in N_k(x)} d(x, o)}{|N_k(x)|}$$

The instance  $x$  with the largest GAS value is considered as a potential anomaly. The local outlier of an instance is defined as the ratio of the average of the reciprocal GAS values of all its adjacent points to the inverse of its own GAS value[9].

#### 2.1.2 Logistic regression

In statistics, Logistic Regression is a classification model based on probability statistics, which uses a linear function to fit logarithmic likelihood ratios. For example, for a classification problem with only two categories, logistic regression is:

$$\ln \left( \frac{p(y = 1 | x)}{p(y = 2 | x)} \right) = w^T x$$

The value of parameter  $w$  can be estimated by moment estimation or maximum likelihood method. The log-likelihood function in the maximum likelihood method is:

$$L(w) = \sum_{j=1}^2 \sum_{i=1}^{N_j} \ln p(x_i^{(j)} | y = j; w)$$

Through the above formula, we can obtain the parameter  $w$  of the regression, that is, to complete the mathematical model of the entire logistic regression.

### 2.2 Regression analysis

Logistic regression is a model used to estimate the probability of an event. It can be seen as a classifier, predicting the category with the highest probability. In logistic regression, the input variables can be continuous or discrete. It is the preferred method when dealing with some binary classification problems. Logistic regression is shown below:

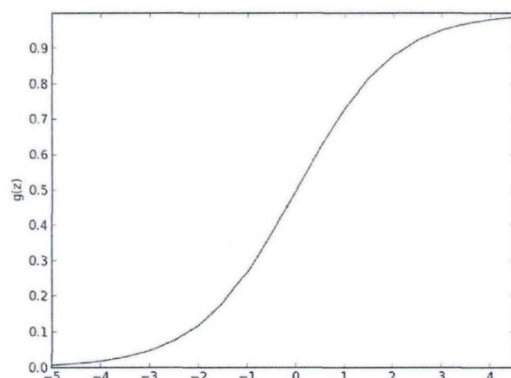


Figure 1 Logistic regression

Logistic regression has interpretable output values, and it is easy to determine the variables that affect the results, which makes it more complex than linear regression. Also, it is stable to redundant variables and has an accurate representation of the output. However, it cannot handle missing very well, and it still defaults that the variables affect the results in a linear way, so it will have certain limitations when it is used in the question of nonlinear relationships[10].

### 2.3 Physical structure and algorithm of neural network

Humanity's pursuit of artificial intelligence has never stopped, and there were similar ideas hundreds of years ago. Since the 20th century, some theories and algorithms have been proposed to make great progress, but the full realization is still far away. From the proposal of neuron structure and the establishment of Hebb's rule in the last century, to the Hodgkin-Hexley equation, perceptron model and self-adapting filter in the 1950s, to the SOM (Self OrganizingMap) network and Neocognitron in the 1960s The concept of ART, the ART model (Adaptive Res-onance Theory), the above models have all developed into classic models in the fields of DSP, computer vision, NLP (natural language processing) and optimization computing, and become milestones. There are no less than a hundred models of neural network models, and there are many excellent application cases in research fields such as Handwriting Recognition, image annotation, semantic understanding and speech recognition[11].

The neuron model was proposed earlier, but its structure has not changed substantially. Its output is only single and the input can have multiple. The following figure is the structure model of the neuron:

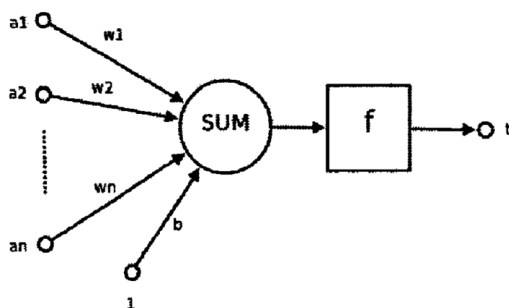


Figure 2 Neuron Model

### 2.4 TensorFlow introduction and its neural network implementation

TensorFlow is the second-generation artificial intelligence open source learning system developed by Google. It is a built-in framework learning software library used to implement the network. The nodes (Nodes) in the calculation graph represent mathematical operations, and the lines in the graph represent the propagation of different data groups between different nodes, that is, tensors. Its advantage is that the architecture is more flexible and can be operated on a variety of platforms, such as a single or several CPUs (or GPUs) in a personal PC, also used in servers, mobile devices, etc.



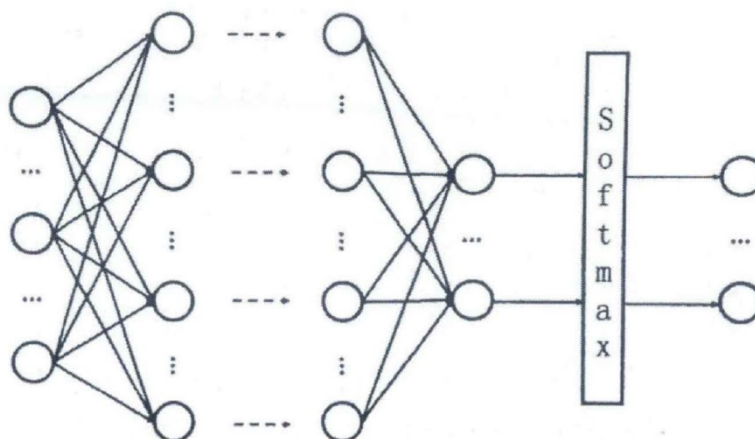


Figure 4 Schematic diagram of softma layer

#### 4. Industrial data fitting model and analysis based on neural network

The neural network algorithm has its unique advantages in industrial big data classification. It can establish the distinguishing marks between different classes through the model output, which can reasonably and naturally represent each classification. Although most of the applications of neural networks appear as classifiers, classification is by no means the only application of neural networks. Another important application is fitting, which is used to fit some complex, nonlinear function body structures through training; generalized fitting is not only limited to the fitting of functions, but can even be used to fit certain functions. processing method, and fitting the processing method of the operation can make the method more adaptable[16].

##### 4.1 Introduction to Fuzzy Fourier Transform

Fuzzy Fourier transform is not an academic term, it is just to distinguish it from the traditional Fourier transform. It is a term used by the author to distinguish a problem solution used in this article. Original, if the research in this direction is valuable, it may also become a special academic term. Fourier transform is a common integral transform[17]. Its formula is as follows. It can transform the signal in the time domain into the frequency domain, from the corresponding relationship between time and amplitude to the corresponding relationship between frequency and amplitude. It occupies an important position in processing and is a common and effective analysis method.

$$F(\omega) = F[f(t)] = \int_{-\infty}^{\infty} f(t)e^{-i\omega t} dt$$

In the ultrasonic detection signal processing of this case, the time domain signal is firstly graphed and observed.

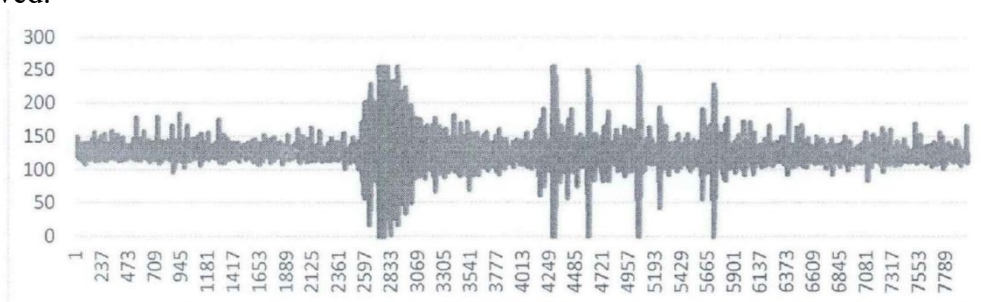


Figure 5 Time City Borrowing Number Waveform

##### 4.2 Fuzzy Fourier Transform Network Training

The data obtained in the previous section is divided into 2 parts, 30% test data and 70% training data. All the optimization methods used in the previous chapter are used in the training process:

softmax regression and cross-drop combination, regularization method and regularization loss , the moving average model. In the training process, the optimization function is still the regularization loss of the cross and the weights of each layer, and the correct rate is detected by the test sample[18].

```
10000 steps,accuracy 0.8832
10500 steps,accuracy 0.8895
11000 steps,accuracy 0.8934
11500 steps,accuracy 0.897
12000 steps,accuracy 0.9019
12500 steps,accuracy 0.8994
13000 steps,accuracy 0.8985
13500 steps,accuracy 0.9058
14000 steps,accuracy 0.9041
14500 steps,accuracy 0.9052
15000 steps,accuracy 0.9072
15500 steps,accuracy 0.9056
16000 steps,accuracy 0.9069
16500 steps,accuracy 0.908
17000 steps,accuracy 0.9083
17500 steps,accuracy 0.9075
18000 steps,accuracy 0.9086
18500 steps,accuracy 0.9084
19000 steps,accuracy 0.9084
19500 steps,accuracy 0.9084
```

Figure 6. Specific results of single-hidden-layer fully-connected network training

```
10000 steps,accuracy 0.9126
10500 steps,accuracy 0.9187
11000 steps,accuracy 0.9234
11500 steps,accuracy 0.9267
12000 steps,accuracy 0.9311
12500 steps,accuracy 0.9308
13000 steps,accuracy 0.931
13500 steps,accuracy 0.9366
14000 steps,accuracy 0.9352
14500 steps,accuracy 0.9358
15000 steps,accuracy 0.9371
15500 steps,accuracy 0.9363
16000 steps,accuracy 0.9377
16500 steps,accuracy 0.9389
17000 steps,accuracy 0.9379
17500 steps,accuracy 0.9383
18000 steps,accuracy 0.9392
18500 steps,accuracy 0.9386
19000 steps,accuracy 0.9395
19500 steps,accuracy 0.9391
```

Figure 7. Specific results of double-hidden-layer fully-connected network training

It can be seen from the results that the expressive ability of the convolutional neural network is better than that of the fully connected neural network. Its accuracy rate reached 96.77%.However, compared with the case of handwritten digit recognition in the previous chapter, the accuracy still cannot reach a high standard. Now analyze from the dataset.

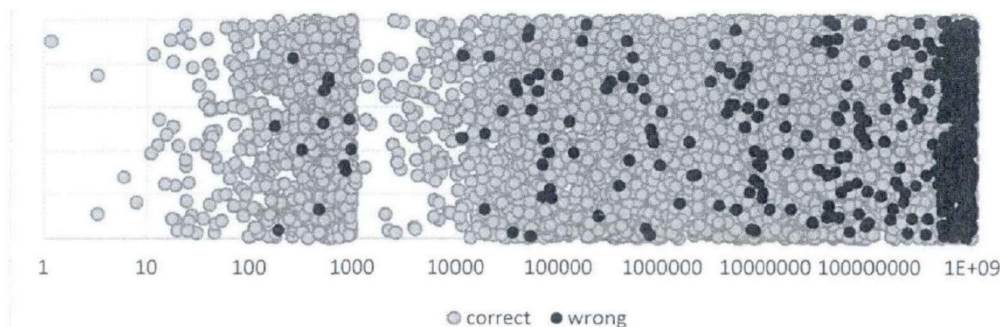


Figure 8. Fuzzy Fourier Transform Network Division Results

In the above figure, the heavy-colored points are the wrong output results, and the light-colored points are the correct output results. It can be clearly seen that the errors are concentrated in the high frequency band, and the upper limit of the high frequency band is set to 1 GHz, which is much higher than the sampling rate. Therefore, this is also an important cause of error, and the sampling frequency is insufficient. However, this sample is not the most authoritative training sample, it is just a sample created by the author for preliminary training and testing. It can also be seen from the distribution of the data in the above figure that the distribution of the data set is not uniform on the logarithmic axis, and is biased towards the upper limit. uneven[19-20].

## 5. Conclusion

This article stems from the current popular big data technology and neural network technology. Big data technology is used for data information mining, and neural network technology is used for data classification and fitting. It is this kind of problem that is studied in this paper. The neural network can use the labeled data to directly train the model, and by continuously updating the network parameters, the network structure can be adapted to the classification method of the labeled data. Different models also need to be optimized in different ways. In general, this paper uses an innovative method to study industrial big data processing. The research results have high use value and scientific research value. At the same time, they are highly innovative and have a certain accuracy, but there are also certain errors. and optimization space. Complete all research objectives as a whole.

## References

- [1] Application Research of Equipment Quality Consumption Abnormal Analysis Model Based on Big Data Analysis [J]. Zhang Lixiu, Li Xiaoke, Li Ruirong. *Modern Industrial Economy and Informatization*. 2017(20)
- [2] Intelligent Technology in Cutting Process [J]. Liu Xianli, Liu Qiang, Yue Caixu, Wang Lihui, LIANG S Y, Ji Wei, Gao Haining. *Chinese Journal of Mechanical Engineering*. 2018(16)
- [3] Google TensorFlow Machine Learning Framework and Application [J]. Zhang Minmin, Xu Heping, Wang Xiaojie, Zhou Mengyun, Hong Shuyue. *Microcomputers and Applications*. 2017(10)
- [4] MapReduce training method of BP neural network based on input fragmentation disturbance [J]. Chen Wanghu, Yu Maoyi, Ma Shengjun. *Computer Engineering and Applications*. 2018(02)
- [5] Neural network method for big data analysis [J]. Zhang Yi, Guo Quan, Wang Jianyong. *Engineering Science and Technology*. 2017(01)
- [6] Research on mining analysis and application prospect of industrial big data [J]. Xu Yujie. *Electronic Technology and Software Engineering*. 2016(22)
- [7] Research on industrial big data mining analysis and application prospect [J]. Zhang Hongbo. *Science and Technology Innovation and Application*. 2016(24)
- [8] Improvement of model by big data preprocessing in hot-rolled C-Mn steel industry [J]. Wu Siwei, Zhou Xiaoguang, Cao Guangming, Shi Naian, Liu Zhenyu, Wang Guodong. *Steel*. 2016(05)

- [9] Real-time feature extraction method for complex industrial big data [J]. Kong Xianguang, Zhang Xiong, Ma Hongbo, Chang Jiantao, Niu Meng. Journal of Xidian University. 2016(05)
- [10] Big data modeling method and practice [J]. Chen Liqiang. Financial Electronics. 2015(11)
- [11] The Core Competitiveness of Made in China 2025: Mining and Using Data[J]. Li Jie, Liu Zongchang. Boao Observation. 2015(04)
- [12] Logistic regression method for class imbalance [J]. Guo Huaping, Dong Yadong, Wu Changan, Fan Ming. Pattern Recognition and Artificial Intelligence. 2015(08)
- [13] Research on big data technology and application in power industry [J]. Yan Mingliang. Journal of Nanjing Institute of Technology. 2015(02)
- [14] Application of big data analysis method in plant-level load distribution [J]. Wang Ningling, Fu Peng, Chen Degang, Yang Zhiping, Yang Yongping. Chinese Journal of Electrical Engineering. 2015(01)
- [15] Industry 4.0: From Automatic Production to Intelligent Manufacturing [J]. Peng Xunwen. Large Aircraft. 2014(07)
- [16] Research on Multivariate System Modeling Method Based on Big Data [J]. Yuan Shitong, Han Pu, Sun Ming. Journal of System Simulation. 2014(07)
- [17] The yesterday, today and tomorrow of deep learning [J]. Yu Kai, Jia Lei, Chen Yuqiang, Xu Wei. Computer Research and Development. 2013(09)
- [18] Review of Support Vector Machine Theory and Algorithm Research [J]. Ding Shifei, Qi Bingjuan, Tan Hongyan. Journal of University of Electronic Science and Technology of China. 2011(01)
- [19] The classic algorithm of decision tree: ID3 and C4.5 [J]. Huang Wen. Journal of Sichuan University of Arts and Sciences. 2007(05)
- [20] The influence of activation function on the performance of BP network and its simulation research [J]. Wang Xueguang, Guo Yanbing, Qi Zhanqing. Automation Technology and Application. 2002(04)