

PAPER • OPEN ACCESS

## Analysis of the Application of Machine Learning Algorithms in Data Mining

To cite this article: Chen Xiaojun and Bian Jiang 2020 *J. Phys.: Conf. Ser.* **1550** 032039

View the [article online](#) for updates and enhancements.

You may also like

- [Application Research of Big Data Technology in Engineering Cost](#)  
Songyang Yan
- [Multi-source Geographic Information Data Fitting for Transmission Line Design](#)  
Yu Ren, Linjie Chai, Bingjun Zhao et al.
- [Java Web Cloud Data Analysis and Application Based on Spark Machine Learning Algorithm](#)  
Pengfei Che



**ECS**  
The  
Electrochemical  
Society  
Advancing solid state &  
electrochemical science & technology

**DISCOVER**  
how sustainability  
intersects with  
electrochemistry & solid  
state science research

# Analysis of the Application of Machine Learning Algorithms in Data Mining

**Authors name:** Chen Xiaojun<sup>1</sup>, Bian Jiang<sup>2</sup>

**Author unit:** QingDao Institute Of Technology<sup>1,2</sup>

**postal code:** 266300<sup>1,2</sup>

**Abstract:** In the context of the era of big data, it is necessary to process information data more quickly to further develop various industries. At present, traditional data processing and calculation methods can no longer meet the requirements of modern big data calculations. With the increase of the amount of information data, the calculation and analysis methods of data are also growing, and machine learning algorithms are one of them. This article mainly expounds and analyzes the application of its learning algorithm in data mining.

## 1. Introduction

With the progress of science and technology, the application of data mining technology has become more and more widespread. Under the trend of informationization and dataization, people have paid more attention to the research and analysis of data management technology. Accurate judgment and scientific decision-making can also promote communication and exchange between departments. But with the growth of data, the difficulty of analysis is also increasing. To perform data analysis more accurately and comprehensively, it is necessary to adopt effective data mining techniques and innovate data calculation and analysis methods.

## 2. An overview of machine learning algorithms and data mining

### 2.1 Machine learning algorithm

With the development of modern science and technology, China has entered the wave of “Internet +”, and both data mining and data processing have shown a saturated state. People need to re-recognize the importance of machine learning algorithms, set the algorithm according to the needs, and then perform deep mining and analysis of the data through the machine. The cost of manual processing can be reduced, and the accuracy and efficiency of data processing can be improved. In today’s society, the application of this algorithm is more extensive, such as Internet search, social and business platforms. The network data is complex and complicated, especially information and shopping websites are mixed with various kinds of information. Therefore, most websites use machine learning algorithms to analyze user preferences and perform more accurate analysis based on their preferences<sup>[1]</sup>. For example, shopping websites such as Taobao and Jingdong use collaborative filtering algorithms to perform filtering on neighbors. Like will draw like. People who are alike attract each other. It is needed to analyze the user’s preferences, orders, and intersection users and more than 80% of the product’s coincidence rate in the order, etc.

In addition, when machine learning algorithms are applied in these fields, the main task is to identify various types of data in the information, including pictures, text, audio, etc. It can also output speech-



to-text, identify the article, and shield the bad. This technology has been widely used in many fields with a very broad development prospect.

## 2.2 Data mining

In data mining algorithms, there are two more common algorithms, one is a machine learning algorithm and the other is the statistical. The former mainly uses artificial intelligence technology, which can use samples for training and learning, and understand the rules and modes of operation. The latter mainly uses cluster analysis, probability and other calculation methods. Different calculation methods can be applied in different fields simultaneously or independently.

The most widely used in machine learning algorithms is the artificial neural network method with strong processing and learning capabilities. It can accurately identify, and is conducive to data classification. It is mainly implemented by building models. There are many types and forms of models, which can meet various needs. In general, the artificial neural network model has the characteristics of high accuracy and good robustness, which can be accurately described<sup>[2]</sup>. However, this method also has some shortcomings, such as long training data, low knowledge understanding ability, and insufficient openness. The traditional BP neural network has many shortcomings, such as poor learning, insufficient anti-interference ability. Genetic algorithms have problems such as precocity and insufficient local optimization ability.

## 3. The Application principle of machine learning algorithm and data mining

### 3.1 The application principle of machine learning algorithm

With the continuous development of science and technology, the amount of data is constantly increasing, and many software and computing methods are gradually unable to meet the needs of modern data processing and analysis. Therefore, it is necessary to optimize the data analysis method. According to the characteristics of big data, the machine learning algorithms are integrated to comprehensively optimize the indifferent data transmission. Although the application of this method on the basis of big data has problems such as complexity and lack of experience, there is still a huge research space and it can be widely applied. The main research areas include three aspects, as follows:

First, neural networks. Machine learning algorithm data simulates the internal nerves of the human brain and analyzes based on the role and location of nerves, and simulates the working of the brain using nerves. To improve the performance of traditional BP neural network, it can be improved to GA-BP. It is mainly optimized and improved from the aspects of chromosome structure design, fitness function, operator selection, and cross mutation operator<sup>[3]</sup>. The specific expression of the fitness function is as follows:

$$f = a * f_{mse} + b * f_{com}, 0 < a, b < 1$$

$$f_{mse} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

$$f_{com} = N(1) / (N(0) + N(1))$$

Among them,  $n$  is the number of training samples;  $f_{mse}$  is the root mean square error of the training data, the range is 0-1;  $y_i$  is the actual value;  $\hat{y}_i$  is the recognition value. Through machine algorithms to study neural modeling, different neural tissues can build different theoretical models. The study of neural network units has complex characteristics, which can also be called deep algorithm learning. Deepening the research on this algorithm will not only benefit data analysis, but also promote the progress of China's medical cause.

Second, vector machine. In-depth exploration of neural network algorithms, biological species as the research basis are used on machine algorithm. According to the research results, machine algorithms can be researched not only on the basis of biology, but also on vector machines. Studying the levels of vector machines and using mathematical thinking for analysis can use regression algorithms, which is

conducive to the guessing of unknown results. First, for example, in the film “Wandering Earth”, the earth is affected by huge attraction. The rescue team will ignite flint to complete the rescue. The space station predicted that the rescue method will not work within hours after the dangerous explosion. At this time, the algorithm used is vector machine calculation. Using the combination of arithmetic and analysis of three-dimensional space from multiple dimensions, the results are predicted through algorithmic deduction.

Third, recommendation algorithm. First, the algorithm is widely used in e-commerce, news, and other platforms. It can recommend products that are of interest to users based on their browsing. For example, in headline platform, machine learning algorithms are used to analyze and optimize data. In the process of news reading, the algorithm can analyze the reading of the user, understand the direction and the reading, and classify the user according to the calculation and analysis results. The user is given more information that meets the user’s interests. In the process of online shopping and browsing, people often receive product and consulting pushes from different platforms, which can arouse the interests of users. This is because the platform website uses a recommendation algorithm to perform calculations based on the calculation results. Reasonable message push makes the message more accurate.

### 3.2 The application principle of data mining technology

In-depth research and analysis of data mining technology use the specific data, deduction and calculation from different angles and levels, and grasp the data change rules. Data mining technology can enhance the efficiency of information search, and grasp the changing rules of data faster and more accurately. Using the technology can analyze data changes in multiple fields such as agriculture, commerce, and medicine, and promote the progress of various industries.

## 4. A practical application of machine learning algorithm in data mining

### 4.1 Fault diagnosis system for air compressor in coal mine

The fault diagnosis system of coal mine air compressor is only used as an analysis case, and the improved GA-BP neural network model is used for data mining and analysis. Before the diagnosis model is established, the operating instructions of various air compressors, the operator’s field experience, fault descriptions, and implementation monitoring data should be consulted and collected, and the types and causes of air compressor failures should be analyzed and summarized.

Through experimental research, after collecting and analyzing field data, the working conditions of 5 coal mine air compressors can be summarized, and different working states are represented by Y1-Y5, the output of the neural network fault diagnosis model the way. Y1-Y5 indicates different contents, as shown in the following table:

Table 1. Working state of air compressor in coal mine

Code symbol	Y <sub>1</sub>	Y <sub>2</sub>	Y <sub>3</sub>	Y <sub>4</sub>	Y <sub>5</sub>
Working condition	Trouble-free working condition	Failure of the cooling water system	Failure of the lubrication system	The working condition of a bearing in fault	The working condition of a faulty circuit system

Before the various fault conditions described in the table, there will be different fault symptoms. According to application experience and various reference materials, it can be divided into different fault symptoms, which can be expressed using X1-X10. The specific is as follows:

Table 2. 10 warning signs before failure

Symbol	Sign
X <sub>1</sub>	Low discharge
X <sub>2</sub>	Insufficient exhaust pressure
X <sub>3</sub>	Exhaust temperature over limit
X <sub>4</sub>	Cooling water temperature over limit

X <sub>5</sub>	Insufficient cooling water pressure
X <sub>6</sub>	Low speed limit of main engine
X <sub>7</sub>	Vibration transfinite
X <sub>8</sub>	Oil temperature overshoot in lubricating system
X <sub>9</sub>	Insufficient oil pressure
X <sub>10</sub>	Bearing temperature exceeding limit

There is a certain correlation between the fault signs and phenomena of air compressor in coal mine, but it is not completely corresponding. The relationship between the two is non-linear. Through specific arrangement and analysis, corresponding internal correlation data can be obtained, as shown in the table below

Table 3. The relationship between compressor fault state and fault phenomenon

malfunction	Fault phenomenon									
	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>	X <sub>9</sub>	X <sub>10</sub>
Y <sub>1</sub>	0	0	0	0	0	0	0	0	0	0
Y <sub>2</sub>	0	0	1	0	1	0	0	0	1	0
Y <sub>3</sub>	0	0	1	0	0	0	1	1	1	1
Y <sub>4</sub>	1	1	0	0	1	1	1	0	0	1
Y <sub>5</sub>	1	0	0	1	0	1	0	1	0	0

During the long-term monitoring, the corresponding working data can be obtained, processed and selected. The fault diagnosis system of the data mining bureau can apply 100 sets of data for training, and use the 100 sets of data as training and performance samples. The specific data samples and data are as follows:

Order numbe	Discharge pressure /MPa	Discharge rate/ (M <sup>3</sup> /min)	Exhaust temperature/°C	Cooling temperature/°C	Cooling water pressure/MPa	Engine speed/ (r/min)	Mainframe vibration/mm/s <sup>2</sup>	Lubricating oil temperature/°C	Lubricating oil pressure/MPa	Bearing temperature/°C	Fault type
1	0.80	30.41	38.91	19.92	0.33	1 502.01	8.22	77.80	0.56	120.70	1,0,0,0,0
2	0.65	25.21	40.21	20.31	0.28	1 434.00	7.70	82.52	0.71	114.21	1,0,0,0,0
3	0.72	28.21	36.40	28.92	0.05	1 477.02	7.91	79.92	0.48	100.92	0,1,0,0,0
4	0.69	30.22	40.11	21.60	0.18	1 479.01	5.91	81.82	0.67	119.51	0,1,0,0,0
5	0.67	28.11	42.51	36.20	0.09	1 520.00	8.90	81.91	0.71	117.22	0,1,0,0,0
6	0.67	29.21	43.71	33.41	0.30	1 388.01	7.21	96.50	0.51	102.12	0,0,1,0,0
7	0.41	17.21	35.92	26.41	0.27	1 321.01	9.41	79.22	0.63	115.21	0,0,0,1,0
8	0.40	19.71	36.21	27.21	0.38	1 206.01	15.12	86.21	0.59	119.72	0,0,0,1,0
9	0.41	16.51	36.42	21.62	0.33	1 362.01	6.72	78.90	0.63	120.91	0,0,0,1,0
10	0.45	19.61	34.62	21.21	0.22	1 501.02	16.71	94.21	0.40	109.21	0,0,0,0,1

Figure 1. Sample chart of training and testing data

In this paper, data samples are used to normalize different dimensions of sensor signal data, which can eliminate the influence.

#### 4.2 Making fault diagnosis and analysis according to data mining

The policy technology is used to study the fault diagnosis system described above. To reflect the superiority of GA-BP neural network announcement on the basis of GA optimization, a corresponding system can be established using BP neural network. The same training data samples can be used in performance testing. After collecting and processing 100 sets of training samples and performing model training, the error approximation curve can be obtained, as follows:

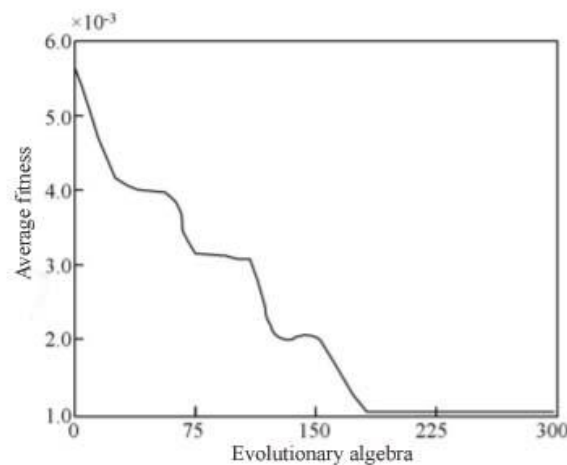


Figure 2. Average fitness changes

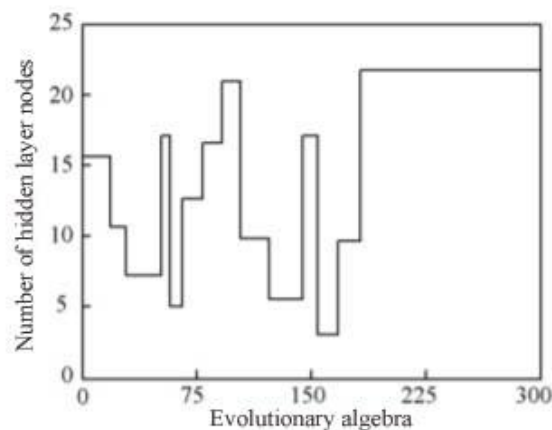


Figure 3. Update process of hidden layer node data

According to the comparison of the pictures, it can be seen that when using the GA-BP neural network algorithm for calculation and analysis, the error has entered the set range after 569 iterations. If the original calculation method is used, it can meet error requirements after 2779 iterations. Therefore, the GA-BP neural network algorithm is superior in both training and convergence speed or accuracy. After testing and analyzing 100 sets of test data samples, the results show that the original neural network diagnosis method has a correct rate of about 87.5%, which requires 564 seconds of diagnosis time, and the output value is not stable. After technical improvements, the accuracy rate has increased to 98.2%, and the diagnosis time has been greatly reduced. It takes about 246 seconds and the output value is relatively stable.

## 5. Conclusion

In summary, the artificial neural network algorithms in machine learning algorithms are researched and analyzed. It is known that the data mining technology using GA-BP neural network algorithms is more efficient and accurate. Therefore, technical research must be actively carried out to comprehensively improve the efficiency and accuracy of data mining.

## References:

- [1] Lv Jianchi. 2019. Application of Machine Learning Algorithm in Data Mining [J]. Electronic World, (13)

- [2] Zhang Lei, Chen Dong, Wang Jianxin, et al. 2017. Machine Learning Algorithm And Application [J]. Journal of Beijing Institute of Electronic Science And Technology, (4): 51-56
- [3] Sun Zhiyi, Ji Zhiwei, Zhang Pei, et al. The Application of Data Mining in The Psychological Needs Assessment of Mobile App Users [C] // The 20th National Psychological Academic Conference Psychology And National Mental Health.