

大连理工大学 图书馆

检索证明

大连科技学院 信息科学与技术学院 2022 年发表的部分论文被国外著名检索工具收录情况如下:

美国《科学引文索引》Science Citation Index Expanded (SCI-E) 收录 1 篇

1. Accession Number: WOS:000766932400009
Title: Intelligent Fuzzy Optimization Algorithm of Data Mining Based on BP Neural Network
Author(s): Li, Y (Li, Yao); Wei, HM (Wei, Huimei); Jiang, JJ (Jiang, Jingjing); Wang, LJ (Wang, Lijuan)
Source: WIRELESS COMMUNICATIONS & MOBILE COMPUTING, Volume: 2022; Article Number: 3084035; Published: FEB 9 2022
Document Type: Article
Addresses: [Li, Yao; Wei, Huimei; Jiang, Jingjing; Wang, Lijuan] Dalian Univ Sci & Technol, Sch Digital Technol, Dalian 116052, Liaoning, Peoples R China.
Corresponding Address: Wang, LJ (通讯作者), Dalian Univ Sci & Technol, Sch Digital Technol, Dalian 116052, Liaoning, Peoples R China.
E-mail Addresses: znhy.wang@163.com
ISSN: 1530-8669

上述检索结果经委托人本人核实, 确认无误

特此证明

证明人  大连理工大学图书馆
2022年4月14日 查收引专用章

Research Article

Intelligent Fuzzy Optimization Algorithm of Data Mining Based on BP Neural Network

Yao Li, Huimei Wei, Jingjing Jiang, and Lijuan Wang 

School of Digital Technology, Dalian University of Science and Technology, Dalian 116052, Liaoning, China

Correspondence should be addressed to Lijuan Wang; znhy.wang@163.com

Received 10 December 2021; Revised 5 January 2022; Accepted 18 January 2022; Published 9 February 2022

Academic Editor: Deepak Kumar Jain

Copyright © 2022 Yao Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article aims to explore the intelligent fuzzy optimization algorithm for data mining based on BP neural network. Although the database technology has been improved with the increase of the amount of data, facing the explosive growth of the amount of data, the previous database management methods have been unable to meet and analyze the hidden knowledge in this scale of data. Therefore, it is important to find better automated data processing methods to satisfy the classification and analysis of massive data. However, the current BP neural network is not yet perfect. This method has some problems, such as slow convergence speed. The problem is reflected in the problem of pattern recognition and insufficient generalization ability and stability. Based on the above description, the research content of this paper is an intelligent fuzzy optimization algorithm for data mining based on BP neural network. Considering that the training of the BP algorithm is based on the weight correction principle of error gradient descent, the genetic algorithm is good at global search, but it does not have accurate local searchability. Therefore, this paper uses the weight of the genetic algorithm. This paper improves BP neural network based on a genetic algorithm. The experimental simulation results of Iris show that the quantity of hidden nodes usually increases with the number of training samples. ACBP algorithm can construct a better network structure based on the number of training samples. And through the experimental comparison of the traditional BP neural network algorithm, it is concluded that the improved algorithm can allow data mining technology to mine relatively more ideal data from complex environments.

1. Introduction

With the development of machine intelligence, agricultural robots are becoming the protagonists. In the process of robot research, basic industrial technologies such as drive control of gearboxes and engines and machinery manufacturing are easy to solve, but intelligent control algorithms remain difficult. Owing to the increasing performance of computers, the drastic reduction in costs, and the successful application of data management technology, the degree of informatization within various departments in the enterprise has become higher and higher. In the face of massive data, it is difficult for decision-makers to directly extract valuable knowledge from it, which has led to strong demand for data analysis tools. The knowledge and information obtained using data analysis tools can be widely used in many fields.

Data mining is an advanced data analysis tool generated by the natural evolution of information technology [1, 2]. Since the beginning of the 21st century, with the popularization and application of various large-scale data systems for query and practical processing [3], data analysis and knowledge understanding have become new goals, and people are looking forward to finding valuable information from data and knowledge, discover the trend of things, and look forward to the emergence of automatic analysis tools [4, 5]. In recent years, the Internet and related technologies have promoted the consolidation of computers, networks, and communications, and the total amount of data processing has become larger and larger [6]. Nowadays, with the rise of the concept of “big data,” the trend of dataization in society is becoming more and more obvious. For industries such as finance and telecommunications, “data is the business itself” [7, 8]. Therefore, data mining and business

intelligence have become the most striking research topics [9, 10].

Guo proposed a two-phase method for online identification of dynamic signatures in power systems using PMU measurement. It usually only predicts the transient stable state to assist in the correction control, but under unstable conditions, it does not produce the dynamic behavior of power generation. The method proposed by Guo proved that the two-stage method of online recognition has high prediction accuracy [11, 12]. Ma proposed a method that can extract the influence of temperature on the load characteristic index and internal related characteristics. Considering the seasonal characteristics of temperature and load, each season was modeled and analyzed. First, a qualitative analysis is performed on the underlying physical relationship between the indicators, and a quantitative calculation is performed by using the Pearson correlation coefficient of historical data to obtain the correlation characteristics between the two factors [13, 14]. Although the security of web applications has been studied for more than ten years, the security of web applications is still a difficult problem. Medeiros tries to find vulnerabilities in the source code to reduce false positives [15, 16]. Mishra proposed a microgrid intelligent protection scheme combining wavelet transform and decision tree. Their scheme flow is roughly to obtain the current signal at the relay point. In addition, the fault classification task is performed by including wavelet-based features derived from sequence components and features derived from current signals, and a new data set is used to construct a DT for fault detection and classification [10]. Gamin introduced a new domain adaptive representation learning method. Gamin's method is mainly inspired by domain adaptation theory. The new domain adaptation method is implemented in the neural network architecture. The neural network formation trains the labeled data from the source domain and the unlabeled data from the target domain. Gamin's results show that adaptive behavior can be achieved by adding a few standard layers and new gradient inversion layers to almost all feedforward models [17, 18].

This paper first analyzes the current status of BP network structure optimization, then proposes an adaptive organization BP network algorithm, and finally combines the idea of growing hidden nodes into the genetic BP algorithm. The results are verified through simulation experiments of Iris flowers. Compared with the traditional BP neural network algorithm, experiments show that the improved algorithm can make the data mining technology mine more useful and accurate data from the nonlinear and complex large amount of data.

2. Proposed Method

2.1. Data Mining Technology

2.1.1. Data Mining Technology. Nowadays, especially with the development of the Internet and the surge in the amount of information, "information" is becoming more and more important to us [16, 19]. Therefore, the goal of data mining has become how to effectively acquire more valuable

Internet knowledge and information [20]. So far, there is no strict standard for the discipline of data mining [21, 22]. In a broad sense, data mining (DM) is to find things that are valuable to users from vast databases or data systems and to extract and analyze relationships that users will not easily detect or assert from huge observation data sets and then give a valuable conclusion that the user can fully understand [23]. Figure 1 shows the process of how to get useful patterns from the original data and then get further knowledge.

As shown in Figure 1, the basic process of data mining is as follows: the proportion of the work that collects the original data in the entire data mining process is not less than other tasks, and the original data collected must be abundant, so that they make the performance of data mining better meet our requirements; sampling and cleaning can then be carried out. Then, you can get the data sample set, which can be used for training and learning.

Then, we propose a data mining system. The composition of the system is shown in Figure 2.

Figure 2 shows the typical structure of a data mining system, where the data sources that the database and data warehouse need to be mined can be various types of information bases, such as databases and spreadsheets. The types of data mining systems include centralized data mining systems and distributed data mining systems. You can perform data cleaning and integration on this data source; the knowledge inventory puts the existing knowledge in the field, which is used to guide the search or evaluate the interest degree of the pattern. This kind of knowledge may include knowledge about concept layering and user confidence; a set of functional modules of the data mining engine implement different types of mining, such as association, classification, or clustering according to user requirements; pattern evaluation modules generally use interest to evaluate whether the pattern is useful; the graphical user interface provides interaction between the user and the data mining system. It functions as a data mining task, provides knowledge, helps focus the search, and is based on the intermediate results of the mining process explored and the mining results are displayed with a visually friendly interface.

2.1.2. Common Algorithms for Data Mining. There are two types of data mining methods: statistics and machine learning. Each method has its own advantages and disadvantages, and choosing different methods for data mining will also produce different results. The commonly used data mining algorithms are as follows:

(1) Decision tree

It mainly conducts an inductive analysis of the attributes of the data. The commonly used method is the "if-then" algorithm. The biggest advantage of decision trees is that they are simple, intuitive, and highly readable. However, due to the characteristics of the algorithm, its branches will be quite complicated and difficult to manage when faced with complex and changeable problems. There is also the

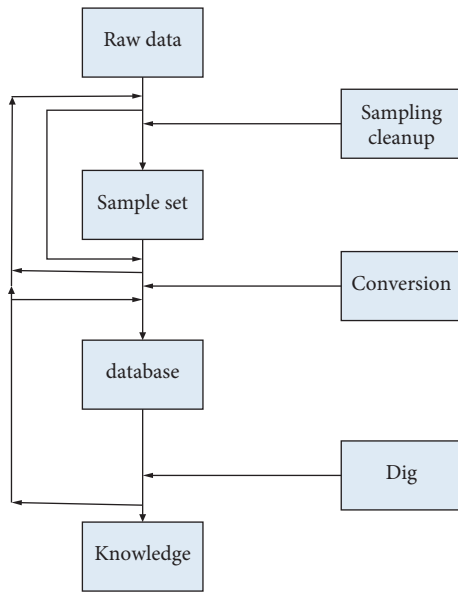


FIGURE 1: The basic process of data mining.

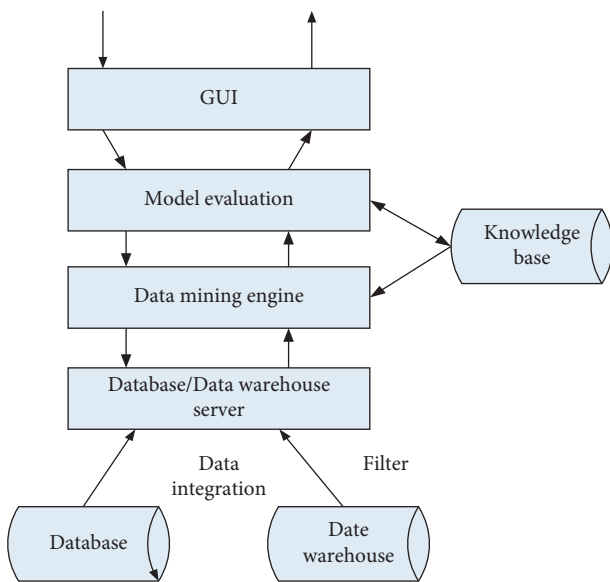


FIGURE 2: The typical structure of a data mining system.

problem of missing data processing. There are many variants of decision tree algorithms, including ID3, C4.5, C5.0, and CART.

(2) Genetic algorithm

This method is a global search algorithm. The genetic algorithm generally searches the data globally through selection and mutation operations to obtain the optimal solution. Generally, the tasks in data mining are treated as the search target.

(3) Bayesian network

Uncertain problems are linked together through the Internet, and other problems are predicted. The network can be hidden or visible. This method mainly has functions such as clustering, classification, and

prediction. Its main advantages are easy to understand and good prediction effect, but it has poor prediction effect on small probability events. Uncertain problems are linked together through the Internet, and other problems are predicted. The network can be hidden or visible. This method mainly has functions such as clustering, classification, and prediction. Its main advantages are easy to understand and good prediction effect, but it has poor prediction effect on small probability events.

(4) Rough set

This method also plays an important role in DM. It is mainly used to deal with ambiguity or uncertainty. It can also be used for feature reduction and correlation analysis. The advantage of this method is that it can predict the problem well without any initial information, so it is widely used in uncertain problems. In all existing knowledge, find the two most similar to him as the lower approximation and one as the upper approximation. Lower approximate set = the set of distinguishable objects in the subset. Upper approximate set = the set of all objects in which the objects in the lower approximate set are indistinguishable.

(5) Neural network

This method was first proposed by biologists and psychologists. It is mainly a simulation of human brain nerves. Through the strong academic interest rate of the human brain, a stable network is obtained. Artificial neural networks have four basic characteristics: nonlinearity, nonlimitation, very qualitative, and nonconvexity. The network predicts other samples. The application of neural network methods in data mining is also quite common. The disadvantage of this method is that it is poorly interpretable and not easy to understand, because the network structure is complicated, and there are no clear steps to explain the results. However, this method can make good predictions for complex problems and has a good tolerance for noise data. Neural networks can be generally divided into feedforward, feedback, and self-organizing neural networks. They have good prediction capabilities for the prediction of complex data and have extensive applications in various aspects.

(6) Statistical analysis

This method is based on the principles of probability and statistics and uses known models to accurately analyze and mine the data, including factor analysis, discriminant analysis, and regression analysis. Because of its accurate description and ease to understand, this method has a wide range of applications in practice, and its products also have a certain market position.

2.2. Artificial Neural Network Technology

2.2.1. Artificial Neural Network. The human thinking process has always troubled many people, and many scholars have also joined the study of the human brain. A large

number of experiments show that the human brain is connected by a large number of neurons through complex nonlinearities, and this powerful network can handle very complex and changeable problems. An artificial neural network is a simulation of the human brain. It constructs a large number of neurons and combines these neurons to form a network. A neural network consists of many neurons interconnected to form a complex network system. One of them is a BP neural network, which adds a hidden layer to the neural network. This network will have better classification and memory. The learning process of the BP neural network can generally be divided into two directions, namely, forward propagation and backpropagation. First, the network forwards the input data. Secondly, when the system enters the error backpropagation, the connection weights of each layer will be modified successively according to the corresponding algorithm to achieve the network learning process.

BP network is a network structure composed of three layers: input, output, and middle layers. The middle layer generally has only one hidden layer. Each neuron in the input layer corresponds to a variable feature, the neuron in the input layer is equivalent to a container with numbers, the output layer can return to the problem as one neuron, the classification problem is multiple neurons, and all the parameters in the network are the middle layer, that is, the weights and biases of neurons. The basic three-layer BP network structure is shown in Figure 3.

As shown in Figure 3, the learning process of the BP model is basically as follows: providing training samples to the network model and propagating the processed data forward; setting the value of the error, that is, the difference between the actual output and the expected output; after learning, changing its weights, in turn, to make the network gradually stable, the error will be within the specified range.

The input information of the network first passes through the input layer, and each feature of the data corresponds to a neuron in the input layer. Since the BP neural network is a fully connected method, the information of the input layer and its connection weights are combined to obtain their weighted sum, and then, they can reach the hidden layer. This layer is the core processing part of the network. The received data is further processed and then passed to the output layer. The same is true at the output layer. After processing the data, the corresponding actual output information can be obtained. The data processing method of each layer is as follows:

(1) Input layer

After the input layer obtains the data, it can be directly forwarded to the hidden layer. Because this layer does not need special processing of the data, the output vector of this layer is the received sample data set. So the input vector expression for this layer looks like this:

$$\vec{X} = [x_0, x_1, \dots, x_m]. \quad (1)$$

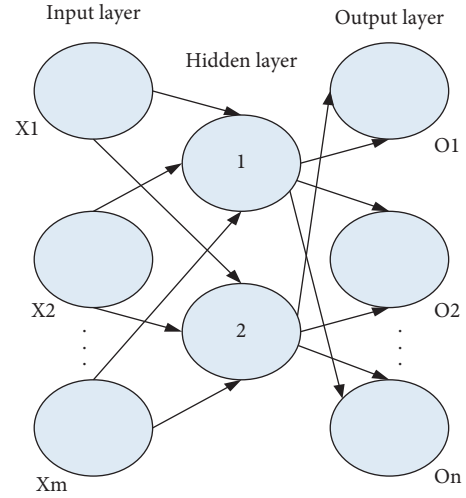


FIGURE 3: BP neural network structure diagram.

(2) Hidden layer

This layer is an effective processing layer for the network. Generally, the weighted data is first received from the input layer, and then, the data is regulated by the activation function so that it falls in the specified area, and it is smoothed through the threshold to get better effect data. Therefore, the expression of the neuron i input information in the hidden layer is as follows:

$$x_i = \vec{X} \vec{W}_i, \quad (2)$$

where \vec{W}_i represents the connection weight between the hidden layer neuron i and the input layer, as shown as follows:

$$\vec{W}_i = \begin{bmatrix} w_{i0} \\ w_{i1} \\ w_{i2} \\ \vdots \\ w_{im} \end{bmatrix}. \quad (3)$$

The specific expression of neuron i is

$$x_i = [x_0 w_{i0}, x_1 w_{i1}, \dots, x_n w_{in}]. \quad (4)$$

When i receives the input data, it will constrain its value to the expected value through the activation function. The S-type activation function is usually used, as follows:

$$f(x) = \frac{1}{1 + e^{-ax}} \quad (0 < f(x) < 1). \quad (5)$$

Therefore, the output value of i after being bound by activation function $f(\cdot)$ is

$$o_i = f(x_i) = f(\vec{X} \vec{W}_i). \quad (6)$$

So the output vector of the entire hidden layer is as follows (where k is the am, amount of hidden layer neurons):

$$\vec{O} = [o_1, o_2, \dots, o_k]. \quad (7)$$

(3) Output layer

Then, the input expression of neuron j in the output layer is as follows, where \vec{W}_j represents the connection weight vector between neuron j and the hidden layer.

$$x_j = \vec{O} \vec{W}_j. \quad (8)$$

The output layer is also an S-type activation function that constrains the data in the data output layer. Then, the output of neuron j is

$$o_j = f(x) = f(\vec{O} \vec{W}_j). \quad (9)$$

So, the output vector of the entire network is

$$\vec{OO} = [o_0, o_1, \dots, o_n] = F_2\left(F_1\left(\vec{X} \vec{W}_1\right) \vec{W}_2\right), \quad (10)$$

where m is the number of neuron nodes in the output layer, F_2 is the activation function of the output layer, F_1 is the activation function of the hidden layer, \vec{W}_2 is the connection weight vector between the output layer and the hidden layer, and \vec{W}_1 is the hidden layer. The weight vector of the connection with the input layer \vec{X} represents the input vector of the input layer.

2.2.2. Algorithm of BP Neural Network

- (1) Initialization network: the input and output of the neural network form a sequence, expressed as (X, Y) . According to this sequence, we can determine the number of nodes in the input layer, hidden layer, and output layer of the network, which are n nodes, l nodes, and m nodes. Then, initialize the connection weight and threshold, and set the connection weight between the input layer and hidden layer neurons to ω_{ij} , the connection weight between the hidden layer and output layer neurons to ω_{jk} , and the hidden layer and output layer. The thresholds are a and b , respectively. Finally, the neuron excitation function $f(x)$ and learning function η of the neural network are given.
- (2) Calculate the output of the hidden layer: according to the input variable X of the network, the connection weight ω_{if} between the input layer and the hidden layer, and the threshold a of the hidden layer, the output H of the hidden layer can be obtained:

$$H_j = f\left(\sum_{i=1}^n \omega_{ij} x_i - a_j\right), \quad j = 1, 2, \dots, l, \quad (11)$$

Here, f is the excitation function of the hidden layer; l is the number of nodes of the hidden greedy layer. There are many forms of expression of the excitation function. The ones selected in this paper are

$$f(x) = \frac{1}{1 + e^{-x}}. \quad (12)$$

- (3) Calculate the output of the output layer: according to step 2, calculate the output H of the hidden layer, and connect the connection weight ω_{jk} of the hidden layer and the output layer and the output layer threshold b to obtain the BP neural network predicted output:

$$O_k = \sum_{i=1}^l H_i \omega_{ik} - b_k, \quad k = 1, 2, \dots, n. \quad (13)$$

- (4) Calculation error: according to the expected output Y of the network and the predicted output O of the network obtained in step 3, we can calculate the prediction error e of the network:

$$e_k = Y_k - O_k, \quad k = 1, 2, \dots, n. \quad (14)$$

- (5) Update weight: update the connection weights ω_{if} and ω_{jk} of the network according to the prediction error e of the neural network:

$$\omega_{ij} = \omega_{ij} + \eta H_i (1 - H_j) x(i) \sum_{k=1}^n \omega_{jk} e_k, \quad (15)$$

$$i = 1, 2, \dots, n; j = 1, 2, \dots, l,$$

$$\omega_{jk} = \omega_{jk} + \eta H_j e_k, \quad j = 1, 2, \dots, l; k = 1, 2, \dots, n,$$

where η is the learning rate of the neural network.

- (6) Update the threshold: update the thresholds a and b of the hidden layer and output layer of the network according to the prediction error e of the network:

$$a_j = a_j + \eta H_j (1 - H_j) \sum_{k=1}^n \omega_{jk} e_k, \quad j = 1, 2, \dots, l, \quad (16)$$

$$b_k = b_k + e_k, \quad k = 1, 2, \dots, n.$$

- (7) Determine whether the iteration end condition of the algorithm is satisfied, and if not, return to step 2.

2.3. Combined BP Neural Network. The structure of the combined BP neural network is shown in Figure 4.

As shown in Figure 4, the combined BP neural network is composed of two levels: the first level is used as a classifier and consists of a BP network; the second level is used as a predictor and consists of n BP networks, and the value of n is the total number of classes. It was decided that each class trained the corresponding BP network with samples belonging to this class. Among them, fuzzy clustering uses the method of calculating the fuzzy similarity matrix between samples; the first-level classifier, the classification rule of this BP network, is the rule generated by fuzzy clustering and has

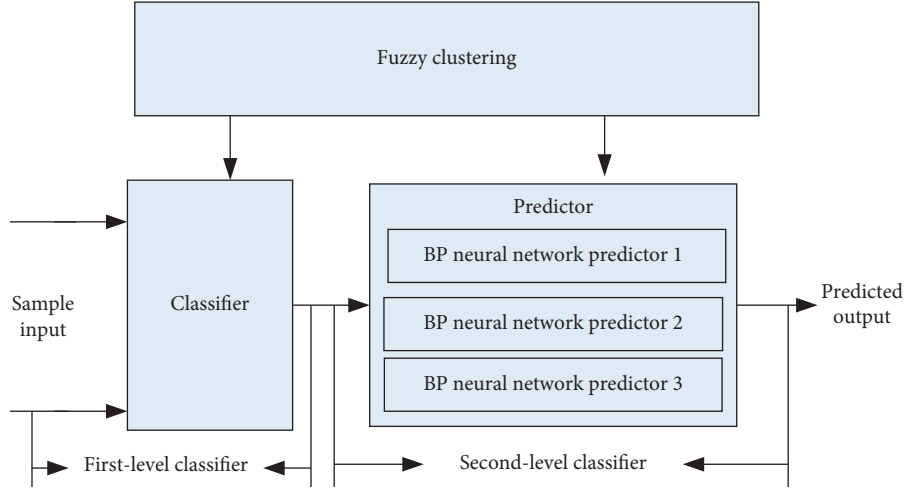


FIGURE 4: Combined BP neural network model.

n output units. The function of the classifier is to judge the category to which a new observation sample belongs based on the training data of the marked category. The second-level predictor determines the use of n BP networks based on the n values obtained by fuzzy clustering, and each BP network has l output units. All BP networks use a three-layer structure (input layer, hidden layer, and output layer), and in order to improve the efficiency and accuracy of the overall combined BP network, heuristic BP improvement algorithm Heuristic BP is used; that is, the momentum method and the learning rate are used. Adapt to adjustment strategies.

The traditional BP algorithm uses the approximate steepest descent method to update the weight and offset values and has the following formula:

$$\begin{aligned}\Delta W(k) &= -aS^n(a^{n-1})T, \\ \Delta b(k) &= aS^n.\end{aligned}\quad (17)$$

The momentum improvement method is based on the theory that if the oscillations in the trajectory can be smoothed, the convergence performance will be improved. When the momentum coefficient γ is added to the update of the parameter value, the momentum improvement formula of the backpropagation is obtained:

$$\begin{aligned}\Delta W(k) &= \gamma\Delta W(k-1) - (1-\gamma)aS^n(a^{n-1})T, \\ \Delta b(k) &= \gamma\Delta b(k-1)(1-\gamma)aS^n.\end{aligned}\quad (18)$$

The adaptive learning speed improvement method is based on the theory that increasing the learning speed on flatter surfaces and reducing the learning speed on steeper surfaces can increase the convergence speed. Its rules are as follows. If the mean square error is over the entire training set, weights are increased after the update and exceed the preset parameter ζ . If γ is set to 0, it returns to the previous value.

The flowchart is shown in Figure 5.

Fuzzy clustering uses the method of calculating the fuzzy similarity matrix between samples. The specific steps are as follows:

- (1) Take n samples, and obtain the original matrix $X(m*n)$ for each sample X_i according to the determined value X_{ji} ($j = 1, 2, 3, \dots, m; i = 1, 2, 3, \dots, n$) of m feature attribute values X_j .
- (2) Calculate the fuzzy "similar" relationship matrix ($R = r_{ij}$), $i, j = 1, 2, 3, \dots, n$. between each sample. Each element r_{ij} represents a fuzzy similarity relationship between samples X_i and X_j , and the calculation formula is

$$\begin{aligned}r_{ii} &= 1, \\ r_{ij} &= \frac{\sum_{k=1}^m X_{ki}X_{kj}}{M},\end{aligned}\quad (19)$$

$$(i \neq j), i, j = 1, 2, \dots, n,$$

where M is a parameter and its role is to ensure that the value of r_{ii} does not exceed 1.

- (3) According to R obtained above, R_2, R_3, \dots are calculated sequentially until a certain $R^m = R^{m+1}$ is obtained, and the R^m obtained at this time is a fuzzy equivalent relationship, which is the basis for clustering.
- (4) Determine a threshold $\lambda, 0 \leq \lambda \leq 1$. If $r_{ij} \geq \lambda$, the samples X_i and X_j can be grouped into one class. According to the clustering results, the total number of classes n is worth clustering $C_i (i = 1, \dots, n)$.

3. Experiments

3.1. Experimental Setup and Environment. The well-known Iris database is widely used as instance data for pattern classification, and it is also a benchmark question for testing the effectiveness of learning algorithms. Iris contains 4 attributes, sepal length, sepal width, petal length, and petal width; Iris contains 3 classes, each of which has 50 elements. Each category represents a type of Iris, and 200 samples are evenly distributed

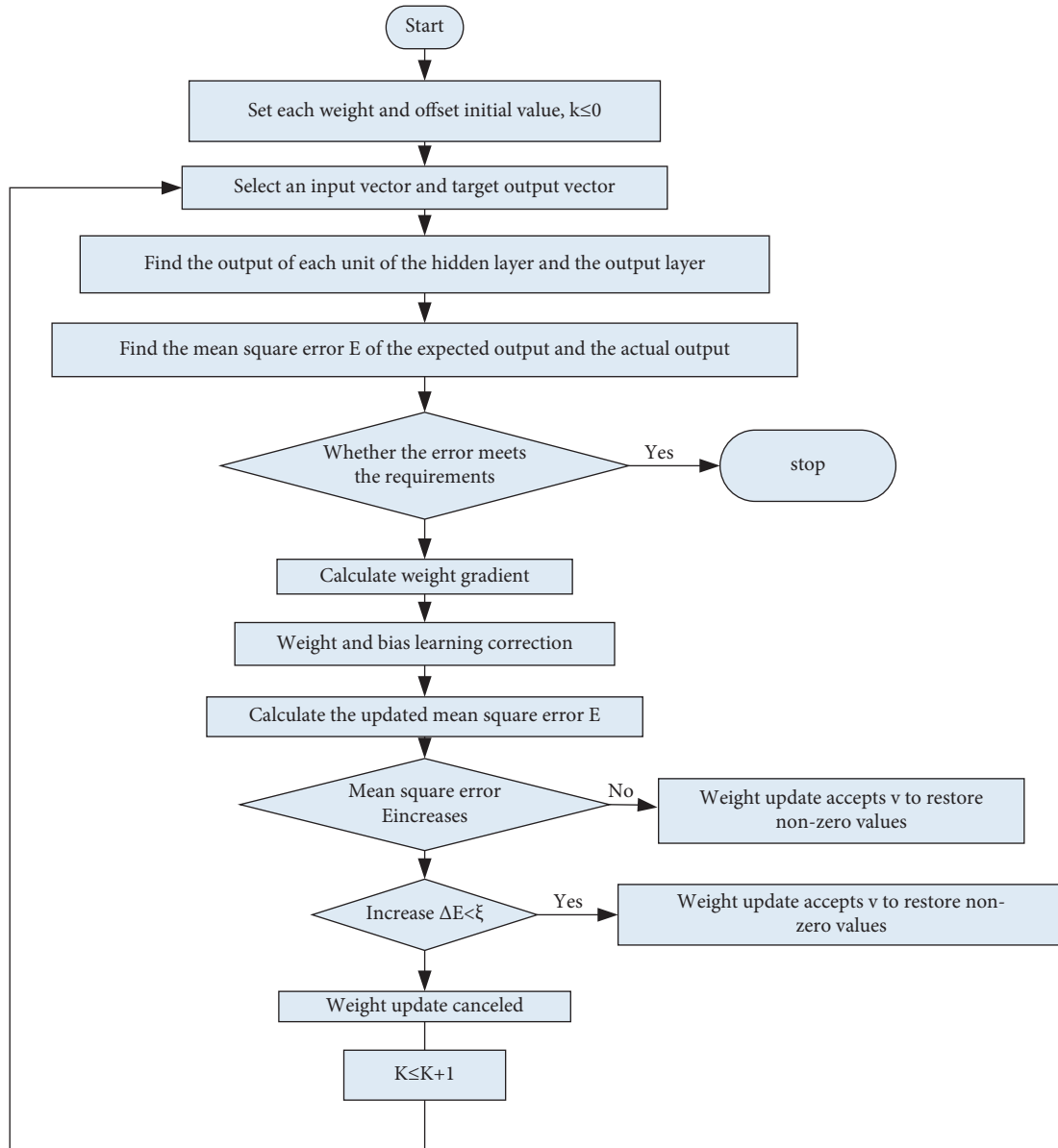


FIGURE 5: Heuristic BP algorithm flowchart.

among the 3 clusters; one is linearly separable with the other two, and the other two are partially overlapping.

This experiment was performed on the Matlab platform. Matlab is a set of very powerful visualization software for engineering calculations and mathematical analysis. Considering that Matlab has a highly reusable neural network function library and is better at operating matrix operations, this experiment is performed on Matlab. The specific experimental environment is as follows: CPU: Intel Pentium Dual T3200 2.00 GHz; memory: 1.87 GB; operating system: Microsoft Windows 10 Professional ServicePack3; operating platform: Matlab7.5.

3.2. Data Set. Experiments are carried out using the momentum, adaptive learning rate method of the fixed node and the ACBP algorithm based on the momentum-adaptive

learning rate proposed in this paper. In addition, two sets of samples were selected for training and testing: 120 training samples. There were 50 test samples, 100 training samples, and 100 test samples. Based on the data set with 4 attributes, the network's input nodes are set to 4; based on formula (16) in the second chapter of this article, because there are not many training samples, in order to avoid increasing the sum of the network weights and thresholds, we choose there is one output node, and three types of flowers are represented by 0.1 and 1, respectively. As for the number of hidden nodes, it is set to one in the ACBP algorithm: then the number of hidden nodes and its adjacent numbers are applied to the momentum-adaptive learning rate method. The initial learning rate is set to 0.1, the growth rate of the learning rate is set to 1.2, the decay rate is set to 0.8, the target accuracy is set to 0.01, the constant limit in ACBP is set to 0.0025, and the maximum number of training samples is 20,000.

TABLE 1: Results of 10 ACBP experiments.

Experiment number	Training times	Number of hidden nodes	Correct rate (%)
1	1917	4	100
2	1840	4	100
3	3179	6	100
5	1879	3	100
6	2749	4	100
7	1793	3	100
8	1469	5	100
9	3579	6	100
10	1724	4	100

4. Discussion

4.1. Initial Weight Exploration. The results of 10 experiments using 150 samples for training and 50 samples for testing ACBP are shown in Table 1.

As shown in Table 1, the accuracy of different hidden node numbers is 100%, so the momentum-adaptive learning rate algorithm with 3, 4, 5, and 6 hidden node numbers was tested 10 times each, and the average value is shown as follows.

As shown in Tables 2–5, when the number of hidden nodes is 3, the conditions for achieving the target accuracy are already available; when the number of hidden nodes is 4, it can basically meet the requirements. In addition, it can be known from the sample test results that when the number of hidden nodes is 2, the conditions for achieving the target accuracy are already met; when the number of hidden nodes is 3, the requirements can be basically met.

The change of the number of hidden nodes in training is shown in Figure 6. It can be seen that the ACBP algorithm can meet the principle of selecting the number of hidden nodes: on the premise of solving the problem, add 1 or 2 hidden nodes to speed up the error reduction. For different initial weights, after the corresponding hidden layer structure is constructed, its accuracy is generally higher than that of the network with a fixed hidden layer structure. In addition, the above experiments also reflect the difference in the number of suitable hidden nodes for different training samples.

4.2. Error Change Analysis. The typical error variation during ACBP training is shown in Figure 7.

It can be seen from Figure 7 that the number of hidden nodes that have just started training is l . As the training progresses, the “spikes” appear in the graph when new hidden nodes are added. Whenever there are new hidden nodes, training will appear jitter situation. This may cause the network to be unstable and may even cause the hidden nodes of the network to increase continuously. But it is also a more effective method to jump out of the dilemma when the training falls into the flat area of the error surface or reaches the minimum point. The figure also reflects a more obvious problem of the algorithm; that is, the number of training samples is large. This is because the initial training is equal to the number of hidden nodes of 1, 2, 3, ..., N included.

4.3. Improved Experimental Results of BP Neural Network.

Through learning, it is found that applying the improved neural network method can achieve good results, because neural networks can highlight their powerful functions when it is in a nonlinear and complex problem, and data mining is performed on a large amount of data analysis to obtain the patterns required by users, so it is of great research significance to combine neural network distribution with DM. However, the BP neural network also has some shortcomings, such as slow convergence speed and ease to fall into local minimums, which makes the BP algorithm fail to fully reflect its advantages in DM applications. Therefore, after studying the genetic algorithm, it was found that we can use GA to search for the characteristics of the global optimal solution, use the advantages of GA to make up for the shortcomings of BP algorithm, and then apply it in DM. Genetic algorithm has good global searchability and can quickly search out all solutions in the solution space, without falling into the trap of rapid decline of local optimal solutions; and using its inherent parallelism, it can easily perform distributed computing to speed up the solution. The improved experimental results are shown in Figure 8.

It can be seen from Figure 8 that the network structure obtained through experiments is a better model. Therefore, the improvement of the genetic BP algorithm does have a good effect. It can be applied to data mining so as to obtain more valuable knowledge in the information age.

4.4. Comparison of BP Algorithm before and after Improvement and Analysis of Experimental Results.

As the error decreases, the learning rate also decreases. In actual training, there will be cases where the training enters a flat area and the error keeps decreasing slightly. In this way, the learning rate will always decrease and cannot meet the requirement of increasing hidden nodes. The experimental results of the traditional BP neural network algorithm are shown in Figure 9.

The experimental results before BP improvement are shown in Figure 9. By comparing it with Figure 8, it can be found that the improved algorithm is slightly faster and the convergence effect is better than before, but these advantages are not obvious. Among the traditional methods of trial and error. Because the accuracy requirements may not be achieved when there are fewer hidden nodes, and even if the number of hidden nodes is sufficient, training may fall into the minimum point. So occasionally it will cause training

TABLE 2: Experimental results of five hidden nodes with two.

Experiment number	Training times	Correct rate (%)
1	843	100
2	287	97.3
3	1823	100
4	284	97.3
5	3497	100

TABLE 3: Experimental results of five hidden nodes with three.

Experiment number	Training times	Correct rate (%)
1	4829	100
2	Unable to reach within 20,000 times	100
3	Unable to reach within 20,000 times	33.5
4	4982	100
5	6928	97.2

TABLE 4: Experimental results of five hidden nodes with four.

Experiment number	Training times	Correct rate (%)
1	4273	100
2	497	100
3	5972	100
4	2441	100
5	619	100

TABLE 5: Experimental results of five hidden nodes with five.

Experiment number	Training times	Correct rate (%)
1	593	97.3
2	1074	100
3	1093	97.3
4	2698	100
5	1092	97.3

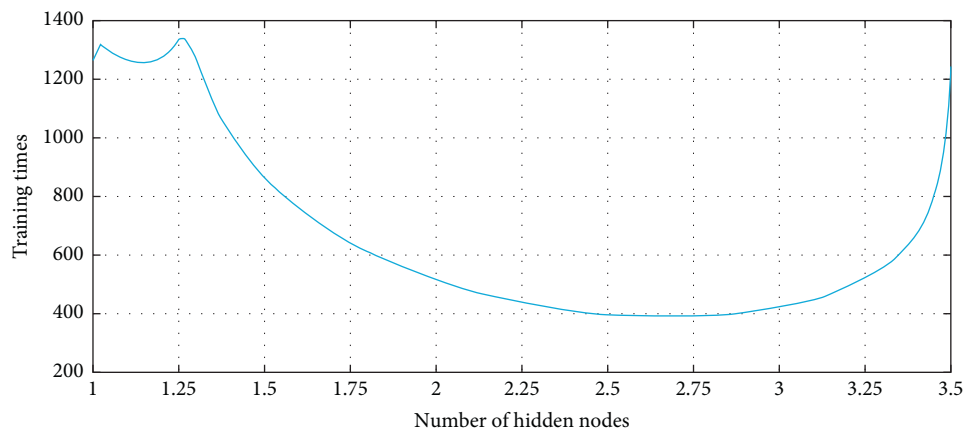


FIGURE 6: Hidden node number change graph.

failure. This makes the average error larger. Aiming at the difficulty of selecting hidden nodes, this paper proposes an optimization method for the hidden layer structure: increasing the number of hidden nodes in a growing way according to the gradient change. The algorithm of growing hidden nodes is used for Iris classification. Through

experiments, it can be proved that the improved algorithm has good advantages in data mining, which can enable data mining technology to mine better ideal data from a large amount of complex, nonlinear data. Therefore, it has great research value for the application of the improved BP algorithm in data mining.

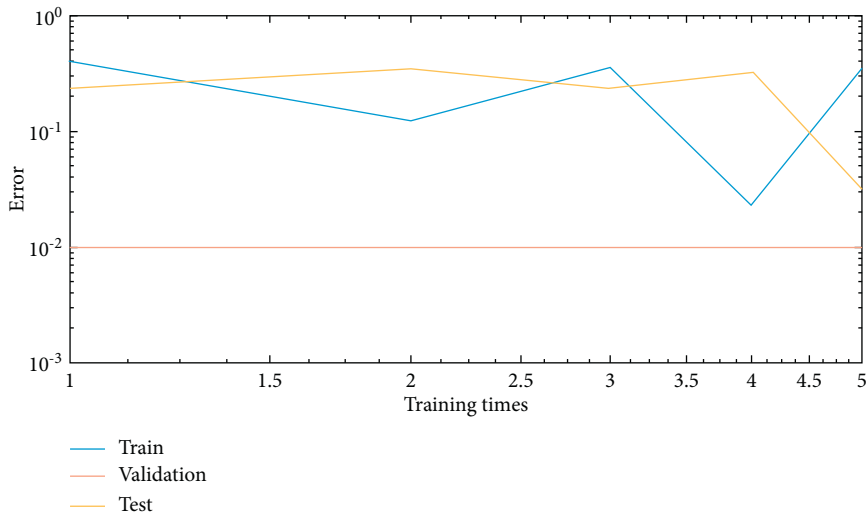


FIGURE 7: Error graph.

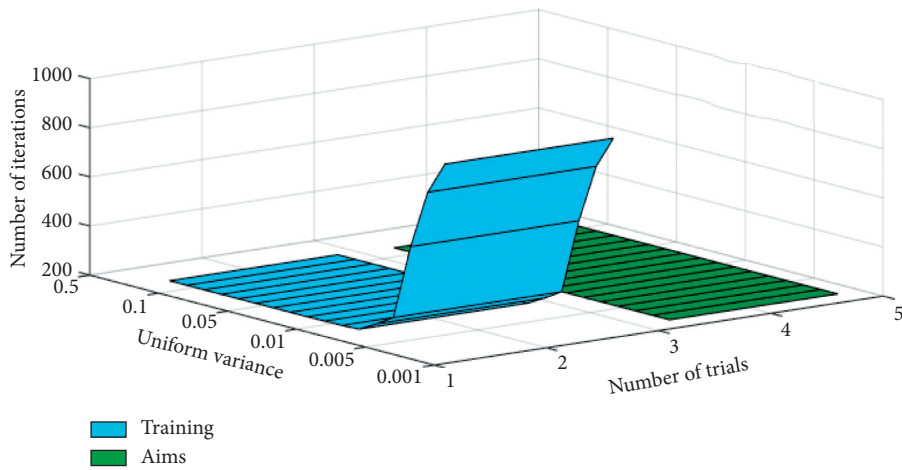


FIGURE 8: BP improved experimental results.

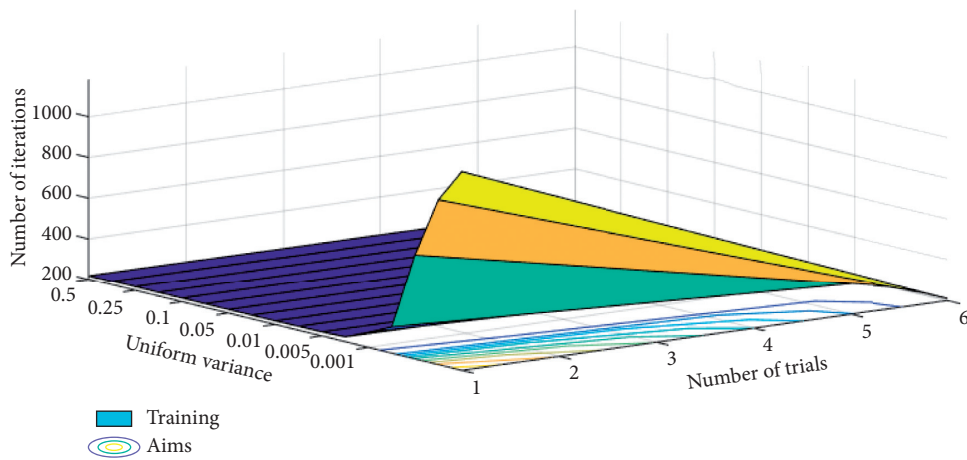


FIGURE 9: Experimental results before BP improvement.

5. Conclusions

This paper proposes improvements to the shortcomings of the BP algorithm. There are some shortcomings in the BP algorithm: slow convergence, minimal values, and lack of theoretical guidance in the selection of hidden nodes in the network. The objective is to improve the network structure selection method. If too few, the network will be too complex. Although this can improve the accuracy of the network and obtain the ideal training model, this also takes a lot of time to make it worth it.

The main purpose of this algorithm is to first obtain a small hidden layer node according to the designed method and then, based on the actual error and the set error of the network output, determine whether the network adds nodes. Finally, experiments show that the algorithm has a good effect. Because the BP algorithm still has the problem of local minimum, in order to overcome this shortcoming, according to the characteristics of global search optimal solution of genetic algorithm, this paper combines GA algorithm and BP algorithm, so that the advantages of GA can make up for the shortcomings of BP. Finally, the validity of the combination can be verified through experiments. Applying the above-improved algorithm to data mining can make DM's application range wider and can solve many large and complex problems and obtain better model mechanisms.

Because the improvement of the BP algorithm is only for certain aspects, although the ideal results can be obtained through experiments, no clear processing method is given in solving the convergence speed. Although the improvement of this paper can appropriately increase some speed under certain conditions, it does not solve it fundamentally. I hope to continue to improve this in subsequent studies and research. Although the disadvantages of the local minimum of the BP network can also be avoided through the combination of genetic algorithms, the GA algorithm itself has some shortcomings, so how can it overcome the shortcomings of the BP algorithm in the situation of the GA algorithm itself. The local searchability of genetic algorithms is poor, which leads to the time-consuming of pure genetic algorithms, and the search efficiency is low in the later stage of evolution. In practical applications, genetic algorithms are prone to premature convergence. In the process of optimization research, we learned that its optimization methods go far beyond changing the number of hidden nodes. You can also optimize the network structure by adjusting the information flow direction of the network. Just considering that the optimization method in this regard is not clear, a less scientific approach may be drawn. It is hoped that after this article is completed, a more systematic study in this area can be conducted.

Data Availability

This article does not cover data research. No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the 2020 Scientific Research Fund Project of Education Department of Liaoning Province: Research on Key Technologies of Media Big Data Depth Analysis System under 5g Platform (Project no. L2020004), Research on Prevention and Analysis of New Charging Attacks in Wireless Charging Sensor Networks (Project no. L2020006), and Research on Public Opinion Struggle and Guidance Mode of Public Emergencies in Virtual Community under Big Data Environment (Project no. L2020007).

References

- [1] Y. Zheng, "Trajectory data mining: an overview," *ACM Transactions on Intelligent Systems and Technology*, vol. 6, no. 3, pp. 1–41, 2015.
- [2] A. M. Wilson, L. Thabane, and A. Holbrook, "Application of data mining techniques in pharmacovigilance." *British Journal of Clinical Pharmacology*, vol. 57, no. 2, pp. 127–134, 2015.
- [3] V. Krishnaiah, G. Narsimha, and N. Subhash, "Heart disease prediction system using data mining techniques and intelligent fuzzy approach: a review," *International Journal of Computer Applications*, vol. 136, no. 2, pp. 43–51, 2016.
- [4] A. Keramati, H. Ghaneei, and S. M. Mirmohammadi, "Developing a prediction model for customer churn from electronic banking services using data mining," *Financial Innovation*, vol. 2, no. 1, p. 10, 2016.
- [5] S. Moro, P. Rita, and B. Vala, "Predicting social media performance metrics and evaluation of the impact on brand building: a data mining approach," *Journal of Business Research*, vol. 69, no. 9, pp. 3341–3351, 2016.
- [6] F. Castellani, D. Astolfi, and P. Sdringola, "Analyzing wind turbine directional behavior: SCADA data mining techniques for efficiency and power assessment," *Applied Energy*, vol. 185, pp. 1076–1086, 2015.
- [7] P. Ahmad, S. Qamar, and S. Qasim Afser Rizvi, "Techniques of data mining in healthcare: a review," *International Journal of Computer Applications*, vol. 120, no. 15, pp. 38–50, 2015.
- [8] S. Asian, G. Ertek, C. Haksoz, S. Pakter, and S. Ulun, "Wind turbine accidents: a data mining study," *IEEE Systems Journal*, vol. 11, no. 3, pp. 1567–1578, 2017.
- [9] T. Lorberbaum, K. J. Sampson, J. B. Chang et al., "Coupling data mining and laboratory experiments to discover drug interactions causing QT prolongation," *Journal of the American College of Cardiology*, vol. 68, no. 16, pp. 1756–1764, 2016.
- [10] D. P. Mishra, S. R. Samantaray, and G. Joos, "A combined wavelet and data-mining based intelligent protection scheme for microgrid," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 1–10, 2015.
- [11] T. Guo and J. V. Milanovic, "Online identification of power system dynamic signature using PMU measurements and data mining," *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 1–9, 2015.
- [12] L. Chen, X. Li, and Y. Yang, "Personal health indexing based on medical examinations: a data mining approach," *Decision Support Systems*, vol. 81, no. C, pp. 54–65, 2015.
- [13] H. Karami, M. Ehteram, S.-F. Mousavi, S. Farzin, O. Kisi, and A. El-Shafie, "Optimization of energy management and conversion in the water systems based on evolutionary

- algorithms,” *Neural Computing & Applications*, vol. 31, no. 10, pp. 5951–5964, 2019.
- [14] J. Wang, Z. Pan, and Q. Zhang, “Intra-continental basalt data mining: the diversity of their constituents and the performance in basalt discrimination diagrams,” *Acta Petrologica Sinica*, vol. 32, no. 7, pp. 1919–1933, 2016.
 - [15] I. Medeiros, N. Neves, and M. Correia, “Detecting and removing web application vulnerabilities with static analysis and data mining,” *IEEE Transactions on Reliability*, vol. 65, no. 1, pp. 1–16, 2015.
 - [16] M. Kumar Jena and S. Ranjan Samantaray, “Data-mining-based intelligent differential relaying for transmission lines including UPFC and wind farms,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 1, pp. 8–17, 2015.
 - [17] Y. Ganin, E. Ustinova, and H. Ajakan, “Domain-adversarial training of neural networks,” *Journal of Machine Learning Research*, vol. 17, no. 1, pp. 2096–2030, 2015.
 - [18] M. Lv, L. Chen, and Z. Xu, “The discovery of personally semantic places based on trajectory data mining,” *Neurocomputing*, vol. 173, pp. 1142–1153, 2015.
 - [19] L. R. Margolies, G. Pandey, and E. R. Horowitz, “Breast imaging in the era of big data: structured reporting and data mining,” *American Journal of Roentgenology*, vol. 206, no. 2, pp. 1–6, 2015.
 - [20] Q. Sun, L. Shi, Y. Ni, D. Si, and J. Zhu, “An enhanced cascading failure model integrating data mining technique,” *Protection and Control of Modern Power Systems*, vol. 2, no. 1, p. 5, 2017.
 - [21] W. Li, L. Wang, X. Cai, J. Hu, and W. Guo, “Species co-evolutionary algorithm: a novel evolutionary algorithm based on the ecology and environments for optimization,” *Neural Computing & Applications*, vol. 31, no. 7, pp. 2015–2024, 2019.
 - [22] F. Mili and M. Hamdi, “A comparative study of expansion functions for evolutionary hybrid functional link artificial neural networks for data mining and classification,” *International Journal on Human Machine Interaction*, vol. 1, no. 1, pp. 47–59, 2016.
 - [23] J. D. . Kirkpatrick, “2MASS data mining and the M, L, and T dwarf archives,” *Proceedings of the International Astronomical Union*, vol. 38, no. 6, pp. 337–395, 2016.