Research Article

Clustering Optimization Algorithm for Data Mining Based on Artificial Intelligence Neural Network

Shuyue Zhang and Chao Duan

Engineering College, Guangzhou College of Technology and Business, Guangzhou 510850, Guangdong, China

Correspondence should be addressed to Shuyue Zhang; zhangshuyue666@163.com

Received 10 November 2021; Revised 7 December 2021; Accepted 21 January 2022; Published 17 February 2022

Academic Editor: Mohamed Elhoseny

Copyright © 2022 Shuyue Zhang and Chao Duan. 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.

Social production and life have become increasingly prominent. Cluster analysis is the basis for further processing of the data. The concept of data mining and the application of neural networks in data mining are introduced. According to the related technology of data mining, this article introduces in detail the two-layer perceptron, backpropagation (BP) neural network, RBF radial basis function network for processing classification problems, and self-organizing map (SOM) self-organizing neural network for unsupervised clustering problems. According to the characteristics of self-adaptive and self-organizing capabilities of these algorithms, we learn and design and implement data mining clustering optimization algorithms. In this paper, the neural network-based data mining process consists of three stages: data preparation, rule extraction, and rule evaluation. This paper studies the teaching-type and decomposition-type rule extraction algorithms. After analyzing the BP decomposition-type algorithm, the correlation method is used to calculate the correlation of the input and output neurons. After sorting by the degree of correlation, the RBF neural network is used for node selection. This can greatly reduce the number of input nodes of the neural network, simplify the network structure, reduce the number of recursive splits of the subnet, and improve calculation efficiency. Taking the model as an example, the training error is calculated through data mining technology and clustering algorithm. Data mining clustering optimization algorithm mainly improves the popular neural network from two aspects: finer model design and model pruning, and simulates model complexity, computational complexity, and errors through simulation experiments. The rate is measured, and finally, the simulation experiment is performed. The results show that the proposed algorithm for differential distributed data mining has higher accuracy and stronger convergence ability and overcomes the shortcomings and shortcomings of several original genetic algorithm optimization neural network data mining models; it can effectively improve the searchability and search accuracy of the algorithm and improve the efficiency of data mining. Accuracy and accuracy have a wide range of applications.

1. Introduction

The amount of data on the Internet is exploding, and the impact of data on many areas of social production and life is becoming more and more prominent, using traditional data analysis methods. Based on this, data mining techniques and clustering optimization algorithms are generated. Firstly, we determine the mining task and then select the corresponding mining algorithm to implement the data mining operation. The mining process is a process of human-computer interaction and repeated many times. It mainly includes defining problems, establishing data mining libraries,

analyzing data, preparing data, establishing models, evaluating models, and implementing them. The whole process of data mining is inseparable from professional knowledge in the application field, database, data warehouse, or other information repositories.

Its research goal is to divide the collection of limited data objects in a database or data warehouse into a group of clusters. Theoretical analysis shows that the data mining clustering algorithm is very suitable for using neural computing. A cluster composed of clusters is a collection of a set of data objects, which are similar to objects in the same cluster and different from objects in other clusters. The

analysis results can not only reveal the internal connections and differences between the data but also provide an important basis for further data analysis and knowledge discovery. After the global analysis of the similarity between data objects, the similarity will be high. Data objects are grouped together in the same class while data objects with low similarity are grouped into different classes. Commonly used techniques include probability analysis and correlation analysis. The learning and training of sets result in the required patterns or parameters. Since various methods have their own functional characteristics and applicable fields, the choice of data mining technology will affect the quality and effect of the final results. In the actual application process, multiple technologies are usually combined to form complementary advantages.

Jorge observed that most of the calculations are essentially invalid. Jorge's observations inspired CN volatile (CNV), which eliminated most of these invalid operations, improved performance and energy, and had no precision loss compared with the most advanced accelerator [1]. Han proposed an energy-efficient inference engine (EIE) [2]. After learning and extracting microwave data, Wang uses a neural network model in the microwave design process through a process called training to provide instant answers for the learned tasks. Appropriate neural network structure and training algorithm are two main problems in developing neural network model of microwave application [3]. Tapas designed a classifier based on syntax bee colony and applied it to medical data mining [4, 5]. Tapas applied this method to 10 medical datasets and compared it with the Levenberg Marquardt algorithm trained multilayer perceptron classifier [6, 7]. Sadiq takes the students in the affiliated colleges of the University of Dibrugarh as the research object [8]. Chauhan studies different text mining techniques to extract relevant information as needed [9, 10]. Under the Map Reduce paradigm, Sudhakar implements three variants of Apriori algorithm [11], namely, trie and hash technology, by using three data structures: hash tree, trie, and hash table trie. Sudhakar focuses on the significance of these three data structures to Apriori algorithm on Hadoop cluster [12]. Xingyi proposed an evolutionary algorithm [13]. T ü Lin studies the spatial clustering problem without prior information [14]. The clustering method proposed by t ü Lin solves several challenges of clustering [15]. Pham proposed an automatic image in FCPFS [16].

The innovation of this paper is to optimize the data mining clustering algorithm based on artificial intelligence neural network and propose the K-means clustering algorithm to optimize the data mining algorithm, using the scalability and uniformly distributed data types of the K-means algorithm. Data mining can be better. And combined with the actual application, the algorithm is selected to ensure the real-time performance of the algorithm. Main research contents and some useful results are obtained. In this paper, theoretical research is the main part, supplemented by some simulation experiments, and the standard data and real data of UCI data warehouse, such as power load sequence, verify the simulation experiments.

2. Proposed Method

2.1. Neural Network. It is an algorithmic mathematical model that imitates the behavioral characteristics of animal neural networks and performs distributed and parallel information processing. This kind of network relies on the complexity of the system and achieves the purpose of processing information by adjusting the interconnection between large numbers of internal nodes. A machine learning method learns from samples. The purpose of neural characteristics (expressed in the form of conditional distribution $P(t|x)$) of the laws produces samples. It is mainly composed of three basic components:

- (1) Random sample generator, which is used to extract random sample x independently from a fixed but unknown distribution $P(x)$. Generally, sample x is a multidimensional vector, and distribution $P(x)$ is a multidimensional random distribution.
- (2) System internal mapping: according to the mapping, input a random vector x and return output y with a certain probability. Input vector x and output vector y obey a fixed but unknown conditional distribution $P(y|x)$. The definition of this mapping considers the influence of noise, and it is actually a kind of random mapping.
- (3) Learning machine (or algorithm), which can realize a certain kind of function $f(x, a)$ to approximate the internal mapping of the system.

A common BP neural network model is usually composed of an input layer, hidden layer, and output layer, as shown in Figure 1.

Each layer contains an unequal number of neuron nodes, where W_{ij} represents the connection weight between neurons. Each node neuron has multiple inputs and one output, which can be expressed as

$$T_i = f(ne t_i), \quad (1)$$

where T_i represents the output value of node i on a layer.

$$\begin{aligned} Net_i &= \sum W_{ij} \cdot x_i - \theta_j, \\ R &= \alpha Qy + \beta_1 B - W\beta_2 N + \gamma, \\ Z &= \left(1 - \frac{|M| - |P|}{|P|}\right) \Phi, \end{aligned} \quad (2)$$

$$L_P = B_0 + B_1 du \times dt + \sum_{i=1}^N B_j XG + \kappa_u.$$

Input a training sample P from the input layer. Through system analysis, the single sample training error of the BP neural network can be defined as

$$E_p = \frac{1}{2} \sum (d_{pj} - O_{pj})^2. \quad (3)$$

In the formula, the expected output value is

FIGURE 1: BP neural network model.

$$E = \frac{1}{p'} \sum_{p=1}^{p'} E_p. \quad (4)$$

With the deepening of research, more and more researchers tend to use evolutionary programming to study evolutionary neural networks and think that it is more appropriate to use evolutionary programming to study evolutionary neural networks. Therefore, the combination of evolutionary programming and the neural network model is a more effective model. The combination of evolutionary programming and neural network model can better imitate and evolve learning behavior. Based on this analysis, a typical EPNet evolutionary neural network model has been proposed, which has strong representativeness and pertinence. As a relatively mature neural network evolution system, EPNet has the following characteristics: first, EPNet emphasizes the plan of ANN behavior and uses some technologies, such as local training after each structural variation and node splitting, to maintain the behavior relationship between the parent and its offspring, while some previous EP systems have little emphasis on this behavior relationship. The usual way of structural variation is to randomly add or delete a hidden layer neuron or connection. Obviously, this method tends to destroy the behavior that the parents have learned and weaken the behavior relationship between the parents and the children. Secondly, EPNet uses different mutation operations according to a certain priority level and gives higher priority to mutation operations that can generate a simplified network structure. Take structural variation as an example. Before adding nodes, always try to delete nodes or connections first. If the deletion can increase the fitness of individuals, the subsequent mutation operation will not be used. Compared with the existing methods, which limit the network scale by adding the penalty term of network complexity in the fitness function, this method can avoid the laborious attempt to find the parameter of the penalty term. Finally, in order to eliminate the influence of permutation problems, the EP algorithm without a crossover operator is used in the EPNet system. The basic flow of the EPNet model is shown in Figure 2.

FIGURE 2: EPNet model structure diagram.

2.2. Data Mining. In order to realize the differential distribution, calculate the global kernel function and the hybrid kernel function, adopt the hybrid particle swarm optimization method, use the differential distributed data of limited samples for training, and pass a nonlinear mapping:

$$\Phi: x \in R^n \longrightarrow F,$$

$$F(S_1 \| S_2) = \frac{1}{2} F\left(S_1 \left\| \frac{S_1 + S_2}{2} \right.\right) + \frac{1}{2} F\left(P_2 \left\| \frac{S_1 + S_2}{2} \right.\right),$$

$$(\ln(S - \alpha R))M = (\ln - \alpha S)XM + \gamma,$$

$$M(X_i, w_j) = N(X_i)P(R_j X);$$

$$P(w_j | T_i) = \sum_{k=1}^K P(w_j | C_k) P(C_k | T_i). \quad (5)$$

Each particle in the particle swarm represents a possible solution to a problem. The intelligence of problem-solving is realized through the simple behavior of individual particles and the information interaction within the swarm. Due to its simple operation and fast convergence speed, PSO has been widely used in many fields such as function optimization, image processing, and geodesy.

The nonlinear time series of differential distributed data is projected to high-dimensional space f by the global call method of inertia weight, assuming that the training sample set of differential distributed data, $x_i \in R^n$, is the input vector for mining control of differentiated distributed data, and $y_i \in R^n$ is the target value of particle swarm optimization. Thus, the total standard value of load balance of differential distributed data output in big data information base is obtained as follows:

$$\begin{aligned} s(t) &= \sum_{k=1}^N p_k \sin(\omega_k n + \Phi_k) + \zeta(n), \\ H_G^*(x) &= \sum \frac{W_{\text{data}}(x)}{W_{\text{data}}(x) + W_g(x)}, \\ E &= \log \frac{T_R(X)}{(1/2)[T(X) + T_g(X)]}, \\ KL(V_1 \| V_2) &= E \log \frac{J_1}{J_2}. \end{aligned} \quad (6)$$

Among them, $\zeta(n)$ is the modulation error of data mining, Φ_k is the data fusion degree, and ω_k is the characteristic scale of distributed data. The data mining clustering in this study is shown in Figure 3.

The main data mining algorithms are as follows. (1) Association analysis: in nature, there are many related relationships among events, some of which are often known, and some of which are not easy to be found. For example, in the shopping basket, bread and milk are well-known collocations. When they are put together to promote sales, they can promote each other's sales. Without data analysis, people do not know the relationship between beer and diapers. Putting them together can also promote sales. In short, association analysis is the mining of association rules. For example, website designers can discover the relationship between visitors' habits and website pages according to visitors' logs; e-commerce can analyze customers'

preferences according to customers' browsing records and stay time, so as to make targeted recommendations. Association analysis is the basis of other data mining research and has achieved good results in practical application. (2) Sequential pattern mining: the core of sequential pattern association analysis is to find out the pre- and postrelevance of the development of things, so as to dig out the laws with certain causal properties. Association analysis generally only considers simple association relations, while sequential pattern mining should consider time, space, and other factors. For example, after buying a new mobile phone, we will generally consider buying accessories such as film, which is the sequence relationship. Generally, we will not buy accessories before buying a mobile phone! This kind of sequence relation is very obvious, which is a typical sequence relation. The main algorithms are the Apriori algorithm and pattern growth framework. (3) Classification algorithm: the classification algorithm is a very important mining algorithm. Its main idea is to establish a classification model, then input data, and then use the classification model to predict its category. Classification mining is usually represented by a predicate. The typical applications of classification algorithms include credit rating, curative effect diagnosis, and customer rating. The main algorithms are k-nearest neighbor classification, decision tree classification, Bayesian classification, and so on. (4) Clustering algorithm: the clustering algorithm is also called group analysis. In classification, the class of samples is predictable. There are much other text mining, web mining, and so on. The design of each module of data mining is shown in Figure 4.

Data mining process is as follows:

- (1) Data preprocessing: generally, the data is incomplete and polluted, and sometimes, there are inconsistencies in the data, such as code or name differences. The quality of data will determine the quality of mining. Generally speaking, low-quality data will produce poor quality data mining results. Therefore, before data mining, data preprocessing is generally needed to improve the quality of mining data.
- (2) Mining process: after data preprocessing, select the appropriate data mining algorithm according to the mining purpose and task. There are also many data mining algorithms. Different mining algorithms have different characteristics, and the scope of application is different. The mining results are different. No one mining algorithm is suitable for all types of mining, and different mining algorithms will produce different mining results.
- (3) Pattern evaluation: generally, the mining algorithm should remove useless output, provide the mining results to users in a way that users are interested in and easy to understand, and store the mining results effectively. That is, by setting a reasonable threshold of user interest to select the mode of user interest, it can effectively prevent the useful mode from drowning in the mode of many users not interested.

FIGURE 3: The data mining clustering in this study.

FIGURE 4: The design of each module of data mining.

Finally, data mining will output knowledge. Generally, this knowledge cannot be found by our intuition. Some knowledge even goes against our intuition, which is unexpected. But the more such knowledge, the more valuable it may be. The detailed process of data mining is shown in Figure 5.

2.3. Cluster Optimization. We call the process of segmentation a clustering process and the method of segmentation a clustering algorithm. In clustering analysis, data are divided according to certain rules. The result of its function is to divide the data into classes so that the similarity between classes is small and the similarity within classes is large. At present, although there are many kinds of clustering algorithms, they all have their own characteristics and applicability. Taking the definition forms of five basic clustering

FIGURE 5: Basic process of data mining.

algorithms as examples, this paper makes the following explanation:

- (1) Partition method: this method can find the spherical mutually exclusive clusters, and the center of clusters is represented by mean value or center point. This algorithm is suitable for those clustering problems with a fixed number of clusters and a small data set. Among these methods, K-means and K-median are the most classical ones.
- (2) Hierarchical method: this method is based on the idea of hierarchical decomposition clustering. The disadvantage of this method is that it cannot be corrected, such as the error of merging or splitting, and this kind of method can carry out multilevel clustering at different granularity. That is to say, if you want to process very complex data, you have to know how to summarize and count these data in a systematic and purposeful way.
- (3) Density-based method: the cluster density described in this clustering method refers to the minimum number of samples in a single sample space. This algorithm can find clusters with different regular shapes without forcing the shape of clustering. It is suitable for clustering with an irregular number and random shape and has the advantage of reducing or even eliminating noise.
- (4) Which can analyze the validity of the data model, such as data fitting? It is suitable for data distribution that has been classified.
- (5) Grid-based method: the algorithm clusters the quantitative grid space, speed, and a strong computing advantage.

The boundary between different types of clustering algorithms is usually not very clear. Taking mean shift algorithm as an example, its basic idea is to move sample points from areas with low density to areas with high density: from the perspective of density estimation and density gradient estimation, it can be regarded as a density clustering algorithm; however, some K-means algorithms can be regarded as mean shift using special kernel function. At the same time, the maximum entropy clustering algorithm based on physical model can also be regarded as mean shift algorithm with a special kernel function. Five clustering methods are introduced as follows.

2.3.1. Segmentation and Clustering. K-means and K-medoids are two typical segmentation and clustering methods, which usually need the number of clusters input by users. Through continuous iterative optimization, the minimum distance within the cluster is the maximum distance between the clusters. Randomly select k clustering centers: $\mu_1, \mu_2, \dots, \mu_k$; repeat the following steps to convergence; for each sample, calculate its class:

$$c_i = \arg \min_j \|x^{(i)} - \mu_j\|^2,$$

$$V(M, N) = E_{x \sim p_{\text{data}}(x)} [\log M(x)]$$

$$+ R_{z \sim p_Z(z)} [\log (L - N(G(K)))],$$

$$Y(B, D) = R[\log C(X|Y)] + M[\log (V - G(M(L|K)))],$$

$$D_H^*(x) = \frac{Q_D(x)}{Q_D(x) + Q_D(x)}.$$

(7)

The K-means algorithm is more efficient for small-scale data sets, while the K-Medoids algorithm has better performance for large-scale data, but it has poor scalability. In addition, the K-means algorithm has two main defects: first, the clustering results are affected by the initial cluster center. There are many researches on the improvement of K-means in the scientific community. K-means++ is a representative one. And the difference lies in the selection strategy of the clustering center: the clustering center of K-means is located at the average of coordinates of all points in the cluster, while the clustering center of K-medoids must be a sample point in the cluster, and the cluster distance of all points in the cluster with it as the center is the smallest. Compared with K-means, the complexity of the segmentation clustering algorithm is low, and it is suitable to deal with large-scale data. CLARANS algorithm is a representative algorithm. It makes large-scale data clustering have high efficiency and good scalability through random search strategy. The segmentation clustering algorithm is usually easy to parallelize, and it is active in big data processing platforms in recent years.

2.3.2. Hierarchical Clustering Algorithm. According to the similarity between data points, the hierarchical clustering algorithm decomposes hierarchically and creates a nested clustering tree with a hierarchical structure. The bottom-up hierarchical decomposition corresponds to the agglomerative method and the top-down hierarchical decomposition corresponds to the splitting method. The basic flow of aggregation method and splitting method is shown in Figure 6. For example, in the first step, {a, b, c, d, e} is a cluster, in the first step, B and C elements which are both quadrilateral are clustered into a cluster, in the second step, D and e elements which are both circular are clustered into a cluster, in the third step, polygon elements are clustered into a cluster, and in the fourth step, all elements are clustered into a cluster. According to the different distance measurement methods between clusters, SL hierarchical clustering has a single linkage), CL hierarchical clustering has complete linkage), and AI hierarchical clustering has an average linkage). The typical hierarchical clustering algorithms are birch, cure, and chameleon. In recent years, some improved algorithms improve their efficiency and robustness.

FIGURE 6: The basic flow of cohesion method and splitting method.

2.3.3. Density Clustering. The earliest idea of clustering based on density may come from the DBSCAN algorithm: the algorithm divides the regions with sufficient density into clusters and finds clusters of arbitrary shapes in the noisy spatial database. It defines the clusters as the largest collection of points with connected density, according to the local density of sample points; it can be divided into core points. DBSCAN is very sensitive to parameters, and slight changes of parameters may lead to abrupt changes in clustering results, which seems to be caused by the sensitivity of core points and boundary points to local density thresholds. Nonspherical clusters, GDBSCAN, ENDBSCAN, options, cancel, clusters, and other density clustering algorithms can identify clusters of any shape, besides DBSCAN like clustering algorithm, mean shift clustering algorithm, and density peak.

2.3.4. Grid Clustering. The grid clustering method divides the space into several grids and analyzes the data on the grid. The complexity of the clustering process is usually related to the number of grids and the number of sample points, so it is more efficient in processing some data. Common grid clustering methods include sting, wave cluster, clique, optgrid, and enclus. Sting algorithm is a multiresolution clustering method, the data space is divided into several rectangular cells, and each high-level rectangular cell is nested with many fine-grained rectangular cells. Generally speaking, the grid clustering method should consider how to divide the cells, how to choose the appropriate cell size, and how to store the updated cell information. If the grid cells are not refined enough, the accuracy will be lost, and if the grid

cells are too refined, the computation cost will be increased. The scalability of the grid clustering method depends on the strategy of storing and updating grid cells to a great extent. Because the boundary of grid cells is horizontal or vertical, the grid clustering method can only find grid-like clusters, not inclined boundaries.

2.3.5. Model Clustering. The model clustering algorithm assumes that the data is mixed according to a specific probability distribution, which is dedicated to finding the best fit between the data and the given model: statistical learning/machine learning methods (such as cobweb) and artificial neural network methods (such as SOM). Taking the SOM algorithm as an example, it is a neural network composed of the input layer and competition layer, as shown in Figure 7.

In the process of clustering, each node first initializes its own parameters randomly, then matches the best node for each input data, then updates the adjacent nodes according to the activated nodes, and finally updates the node parameters according to the gradient descent method, and then repeats the iterative updating until convergence. In the process of the development of model clustering, many researchers put forward the improved algorithm, which is applied to text data.

3. Experiments

3.1. Subjects. Three classic data sets Iris, wine, and zoo are for experimental verification. The accuracy and convergence are analyzed and verified. The characteristics of the test data set are described in Table 1.

FIGURE 7: Kohonen network structure.

The main parameters are set, learning factor $c_{max} = 2.5$, and the minimum value $c_{min} = 0.5$. Other parameters can be set flexibly according to the experimental situation.

3.2. Experimental Setup. At present, most clustering effect analysis often uses F-measure, which includes recall and precision. Recall and precision, respectively, examine the completeness and accuracy of experimental analysis. The definitions are as follows.

$$\text{Recall}(i, j) = \frac{N_{ij}}{N_j}, \quad (8)$$

$$\text{Precision}(i, j) = \frac{N_{ij}}{N_i},$$

where i is the known category; see the following formula:

$$F(i) = \frac{2 * \text{Precision}(i, j) * \text{Recall}(i, j)}{\text{Precision}(i, j) + \text{Recall}(i, j)}. \quad (9)$$

The commonly used measurement method of cluster analysis is the weighted average value of category i :

$$F - \text{measure} = \frac{\sum_{i \in n} |i| * F(i)}{\sum_{i \in n} |i|}. \quad (10)$$

The average value is the final F -measure measurement value, as shown in Table 2.

It can be seen from Table 2 that the ADPSO-k-mean algorithm has higher accuracy than the traditional k-mean algorithm and the PSO-k-mean algorithm and has relatively large optimization effect; especially in the Iris data set, the experimental accuracy has increased by 19.5% and 7%, respectively, with the most obvious effect.

4. Discussion

4.1. Test of Each Fitness Value When Each Algorithm Converges Stably. Also, Iris, wine, and zoo are used to test the

TABLE 1: Test data set.

Test data set	Number of data	Cluster number
Iris	151	4
Wine	177	2
Zoo	100	8

stability of the algorithm. The fitness values (f_{min} , f_{max} , and f_{ave}) of each algorithm are recorded when the algorithm converges stably, and the $f(x)$ on three kinds of data sets are, respectively, in the following formula:

$$f(x) = \frac{\zeta}{\sum_{j=1}^k \sum_{X_i \in C_j} d(X_i, C_j)}, \quad (11)$$

$$M = \frac{|S_1 \cap S_2|}{KL}.$$

From 103, 105, and 102, ζ is the constant $d(X_i, C_j)$, which represents the Euclidean metric distance from sample X_i to the corresponding cluster center C_j . Tested many times, the average value of all similar test data is taken as the final value (for example, by all tests, it is taken as the maximum fitness value). The test records are shown in Table 3.

According to the analysis of test results in Table 3: on the whole, the improved ADPSO-IKM algorithm has a relatively small $f(x)$ fluctuation range in these three types of data sets. The PSO-k algorithm and ADPSO-IKM algorithm improve 9.95%, 12.44%, and 20.85%, respectively, in three data sets. And the improvement of the center K-means algorithm itself ensures the effective search and better convergence performance of the algorithm. In order to further illustrate the convergence of ADPSO-IKM algorithms and the fitness value of each algorithm with the increase of iterations, the convergence graphs on three datasets are drawn. The format design of the training sample of the digit recognizer is shown in Table 4.

After training the network, we test the function approximation ability of the network with the test sample set of an unknown result. The output of the test sample set in the network model is shown in Figure 8.

It can be seen from the final test results that the network model can reach an accuracy rate of approximately 78.5%.

The drug is used as a classification attribute, and there are 5 categories in total; others are used as input attributes. All nominal attributes must be processed numerically; for example, BP is converted to 0: high; 1: low; and 2: normal. We design an RBF network so that it can correctly reflect the drug classification of the sample data after training. The training sample data is shown in Table 5.

We take the drug as the classification attribute, a total of 5 categories, and others as input attributes. All nominal attributes must be processed numerically; for example, BP is converted to 0: high, 1: low, and 2: normal. We design an RBF network so that it can correctly reflect the drug classification of the sample data after training. The training sample data is shown in Table 6.

The attributes of these animals are mapped to the two-dimensional output plane of SOM, and the self-organizing

TABLE 2: Experimental results of F -measure.

Algorithm	Data Set		
	Iris	Wine	Zoo
K- mean value	0.706	0.622	0.555
PSO-K-mean value	0.830	0.700	0.602
ADPSO-k-mean value	0.875	0.711	0.620
KHM-CPSO	0.884	0.720	0.673
ADPSO-IKM	0.900	0.728	0.696

TABLE 3: Results of algorithm fitness value.

Data set	$f(x)$	K- mean value	PSO-K-mean value	ADPSO-k-mean value	ADPSO-IKM
Iris	$f_{\min}$	8.2424	10.2494	10.2716	10.2980
	$f_{\max}$	10.2495	10.3833	10.4043	10.4540
	f_{ave}	9.5843	10.3019	10.3785	10.4011
Wine	$f_{\min}$	5.3290	5.8155	5.8791	5.8914
	$f_{\max}$	5.8966	5.9026	5.9130	5.9779
	f_{ave}	5.5981	5.8290	5.9011	5.9533
Zoo	$f_{\min}$	6.0471	6.4329	6.5706	6.5910
	$f_{\max}$	6.3126	6.5872	6.6015	6.6305
	f_{ave}	6.2156	6.4138	6.5881	6.6222

TABLE 4: The format design of the training sample of the digit recognizer.

Number	X0	X1	X2	X3
0	0	0	1	1
1	1	0	1	0
2	0	0	1	0
3	1	1	0	0
4	1	1	1	0
5	1	1	1	1
6	1	1	0	0
7	1	1	1	0
9	1	1	0	1

clustering process is used to test the rule that the attributes of samples between adjacent clusters are similar. There are 13 kinds of animals in the training set, and each animal is represented by a 9-dimensional vector. The training samples are shown in Table 7.

After 10,000 times of network training, the SOM network maps the pattern features of the high-dimensional space input data to the two-dimensional output plane in an orderly manner. The training results show that 15 neurons arranged in a $5 * 3$ rectangular (Gridtop) structure finally form 9 effective clusters. The clustering results are shown in Table 8.

The learning error change of the BP network combined with the genetic algorithm for the approximation of the sine function is shown in Figure 9. Because the genetic algorithm has a global search property, the optimized initial weight generated iteratively is compared with the real solution space in the direction of the optimal solution is closer; this makes the error of the BP network training process based on the gradient descent method tend to drop quickly at the beginning.

The change of the learning error of the pure BP algorithm is shown in Figure 10. The error convergence effect of BP network learning combined with the genetic algorithm is still better than the pure BP algorithm.

We reduced the network training times by 2 K/time, and the training results are shown in Table 9.

The attribute field includes the content of certain chemical substances in the DFM sample, the average daily alcohol consumption, etc. Take 180 of the data set samples as the training sample set and 50 as the test sample set. Part of the data is shown in Table 10.

Display the clustering results in the form of statistical histograms: this is mainly done by using the open-source chart drawing toolkit JfreeChart. The visual display of its clustering statistical histogram is shown in Figure 11.

In the system's BP neural network parameter settings, set the target error to 0.001, the number of training times to be 1,000, and the learning rate to be 0.01, respectively. The "weight" method reads the initial connection weight selected before and then starts the formal network training. After system operation analysis, the output error curves are shown in Figure 12.

4.2. Performance Comparison of Different Algorithms

4.2.1. Convergence of Each Algorithm on Iris. In the Iris data set, from the iterative trend line within 20 times of PSO-K-means algorithm, it can be seen that the ADPSO algorithm can significantly expand the overall optimization space of the algorithm between 20 and 30 times of iteration and avoid falling into the local extreme situation, and the iteration to 30 times has become stable convergence state. The convergence on Iris is shown in Figure 13.

FIGURE 8: The output of the test sample set in the network model.

TABLE 5: The training sample data.

Number	X0	X1	X2	X3
0	0	0	1	1
1	1	0	1	0
2	0	0	1	0
3	1	1	0	0
4	1	1	1	0
5	1	1	1	1
6	1	1	0	0
7	1	1	1	0
9	1	1	0	1

TABLE 7: The training samples.

Animal	Body type	Number of limbs	Mane
A	Small	2	0
B	Small	2	0
C	Small	2	0
D	Small	4	0
E	Small	4	0
F	Middle	2	0
G	Middle	2	0
H	Middle	2	1
I	Small	2	0

TABLE 6: The training sample data.

Age	Sex	BP	Na/K
23	F	High	25.35
47	M	Low	13.09
47	F	Low	10.11
28	F	Normal	7.80
61	F	Low	18.04
22	F	Normal	8.61
49	M	Normal	16.28
41	F	Low	11.04
60	M	Normal	15.17

TABLE 8: The clustering results.

Animal	Body type	Water	Body type	Number of limbs
Clustering0	Small	0	Small	2
	Small	0	Small	2
	Small	0	Small	2
Clustering2	Small	0	Small	4
	Small	1	Small	4
Clustering4	Middle	0	Middle	2
	Middle	0	Middle	2
	Middle	0	Middle	2
Clustering6	Small	0	Small	2

4.2.2. *Convergence of Each Algorithm on Wine.* The high latitude wind data set, from the overall trend chart of the algorithm, has the same convergence trend, and compared with the k-mean algorithm, the convergence effect is obvious. ADPSO-IKM algorithm iterates 40 times to reach a stable convergence state. The convergence on wine is shown in Figure 14.

4.2.3. *Convergence of Each Algorithm on Zoo.* In the data set zoo, the ADPSO-IKM algorithm approximately iterates 30 times and converges to the best situation. The introduction of the IKM algorithm can obviously accelerate the clustering speed of iterations between 25 and 35 times and also can quickly find excellent solutions. The convergence on wine is shown in Figure 15.

FIGURE 9: The learning error change of the BP network combined with the genetic algorithm for the approximation of the sine function.

FIGURE 10: The change of the learning error of the pure BP algorithm.

TABLE 9: The training results.

Training times	Water	Error of BP network learning combined with genetic algorithm	Pure BP algorithm learning error
8000	0.001	Error of BP network learning combined with genetic algorithm	0.002
6000	0.001	Error of BP network learning combined with genetic algorithm	0.007
4000	0.002	Error of BP network learning combined with genetic algorithm	0.029

TABLE 10: Part of the data.

A1	A2	A	A4	A5
37	78	20	24	86
98	70	21	28	10
88	55	25	11	7
88	62	13	20	1

FIGURE 11: The visual display of its clustering statistical histogram.

FIGURE 12: The output error curves.

FIGURE 13: Convergence of each algorithm on Iris.

FIGURE 14: Convergence of each algorithm on wine.

On the whole, the algorithm iterates on the dataset and tends to be stable and convergent. Compared with other algorithms, the ADPSO-IKM algorithm has better convergence performance and can reduce the influence of initial

cluster center selection on the volatility of the K-means algorithm. Moreover, the improved neighborhood fusion idea can make the algorithm extend to the effective search area. In a small iteration range, it can get a good clustering

FIGURE 15: Convergence of each algorithm on the zoo.

FIGURE 16: The new data classification result of the drug classification problem.

convergence effect. According to the overall trend chart of the three data sets, the ADPSO-IKM algorithm has good optimization performance and can quickly converge to the clustering effect of a fixed value. In the ADPSO-k average algorithm and the introduction of the IKM algorithm, the overall clustering optimization effect of the algorithm has

not been significantly improved, and the research in this area needs to be improved. The new data classification result of the drug classification problem is shown in Figure 16.

The prediction result based on the improved algorithm is shown in Figure 17.

FIGURE 17: The prediction result based on the improved algorithm.

5. Conclusions

In this era of massive data, data mining is extremely important, its application is more and more extensive, and its importance is more and more obvious. As long as an enterprise has a data warehouse or database with analysis value and demand, it carries out purposeful data mining. Data mining will mean a new wave of productivity growth and the arrival of the consumer surplus wave. In the data mining project, if a reasonable network model can not be determined in advance when dealing with some complex problems with the BP model, used for global optimal search, this algorithm is very effective for improving the accuracy and accuracy of data mining in CRM and obtaining a lot of valuable data.

The traditional clustering algorithm is difficult to deal with data with multidimensional and uncorrelated characteristics. The selection of clustering methods directly determines the quality of data mining. In order to improve the quality of clustering, people continue to explore and explore better clustering analysis methods. The group intelligence, self-adaptability, and robustness are shown by the group intelligence optimization algorithm; combined with the group intelligence optimization, cluster analysis develops rapidly.

The generalization ability of particle swarm learning is used to calculate the clustering center of data mining, to realize data mining optimization. The clustering algorithm of neural network data mining proposed in this paper can realize the clustering completed by the K-means method. At the same time, the improved neural network algorithm can automatically merge the clustering results with smaller granularity according to the preset warning value, thus effectively preventing the occurrence of unreasonable

clustering results caused by too many specified clustering numbers. Because the artificial neural network has the characteristics of highly nonlinear to noisy data, BP neural network is more popular. Because the artificial neural network has the characteristics of highly nonlinear learning ability and fault tolerance to noisy data, the artificial neural network and BP neural network are more popular. The ability of artificial neural networks to extract rule knowledge needs to be further strengthened. The artificial neural network uses the “black box” model to process data and mine knowledge. In some data mining applications, people often expect the system to express deep-level knowledge of laws in an intuitive way similar to “if-then.” Therefore, we need to know how to break this black box and explicitly present the useful knowledge hidden.

Data Availability

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 Young Innovative Talents Project in Guangdong Province in 2020 (no. 2020KQNCX109).

References

- [1] J. Albericio, P. Judd, T. Hetherington, T. Aamodt, N. E. Jerger, and A. Moshovos, “Cnvlutin,” *ACM SIGARCH - Computer Architecture News*, vol. 44, no. 3, pp. 1–13, 2016.

- [2] S. Han, X. Liu, H. Mao et al., "Eie," *ACM SIGARCH - Computer Architecture News*, vol. 44, no. 3, pp. 243–254, 2016.
- [3] F. Wang, V. K. Devabhaktuni, and C. Xi, "Neural network structures and training algorithms for RF and microwave applications," *International Journal of RF and Microwave Computer-Aided Engineering*, vol. 9, no. 3, pp. 216–240, 2015.
- [4] L. Qiao, Y. Li, D. Chen, S. Serikawa, M. Guizani, and Z. Lv, "A survey on 5G/6G, AI, and robotics," *Computers & Electrical Engineering*, vol. 95, Article ID 107372, 2021.
- [5] T. Si and S. Sujaudhin, "A comparison of grammatical bee colony and neural networks in medical data mining," *International Journal of Computer Application*, vol. 134, no. 6, pp. 1–4, 2016.
- [6] S. Hussain, J. Hazarika, and P. Buragohain, "Educational data mining on performance of under graduate students of dibugarh university using R," *International Journal of Computer Application*, vol. 114, no. 11, pp. 975–8887, 2015.
- [7] X. Zhou, X. Liang, X. Du, and J. Zhao, "Structure based user identification across social networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 6, pp. 1178–1191, 2018.
- [8] S. R. Chauhan and A. Desai, "A review on knowledge discovery using text classification techniques in text mining," *International Journal of Computer Application*, vol. 111, no. 6, pp. 12–15, 2015.
- [9] H. Chen, Y. Shang, and K. Sun, "Multiple fault condition recognition of gearbox with sequential hypothesis test," *Mechanical Systems and Signal Processing*, vol. 40, no. 2, pp. 469–482, 2013.
- [10] S. Singh, R. Garg, and P. K. Mishra, "Performance analysis of Apriori algorithm with different data structures on Hadoop cluster," *International Journal of Computer Application*, vol. 128, no. 9, pp. 975–8887, 2015.
- [11] X. Zhang, Y. Tian, and R. Cheng, "A decision variable clustering-based evolutionary algorithm for large-scale many-objective optimization," *IEEE Transactions on Evolutionary Computation*, vol. 99, 2016.
- [12] Z. Lv, Y. Han, A. K. Singh et al., "Trustworthiness in industrial IoT systems based on artificial intelligence," *IEEE Transactions on Industrial Informatics*, vol. 99, p. 1, 2020.
- [13] A. Zl, A. Dc, A. Rl, and B. Aa, "Artificial intelligence for securing industrial-based cyber-physical systems," *Future Generation Computer Systems*, vol. 117, pp. 291–298, 2021.
- [14] T. . İnkaya, S. Kayaligil, and N. E. Özdemirel, "Ant Colony Optimization based clustering methodology," *Applied Soft Computing*, vol. 28, pp. 301–311, 2015.
- [15] X. Xu, D. Cao, Y. Zhou, and J. Gao, "Application of neural network algorithm in fault diagnosis of mechanical intelligence," *Mechanical Systems and Signal Processing*, vol. 141, Article ID 106625, 2020.
- [16] P. H. Thong and L. H. Son, "A novel automatic picture fuzzy clustering method based on particle swarm optimization and picture composite cardinality," *Knowledge-Based Systems*, vol. 109, pp. 48–60, 2016.