Classifier model based on neighbourhood rough set and genetic neural network

Ganzhou WU[1]*

[1] *School of Science, Guangdong University of petrochemical technology, Maoming Guangdong, China, 525000*

Received 23 September 2014, www.cmnt.lv

Abstract

In this study, the backpropagation (BP) neural network data classifier model optimized by genetic algorithm (GA) and neighborhood rough set is proposed. First, we integrate the attribute reduction technique with the neighborhood rough set, which is used to delete the redundant attributes of training samples. Second, we optimize the weights and thresholds of the BP neural network by using GA. As such, the training speed and generalization capability of the BP neural network are improved to obtain the optimal weights and thresholds. Finally, the experimental results show that the proposed algorithm performs well.

Keywords: BP neural network; genetic algorithm; neighborhood set; reduction

1 Introduction

High dimensionality, complex structure, and information-relatedness characterize nonlinear complex data. However, pattern recognition and data mining are based on the processing of these problems, which remains a significant challenge. The backpropagation (BP) neural network has large-scale computing capability and can easily perform the nonlinear mapping process, which is a unique advantage when dealing with large complex nonlinear systems. However, BP has some limitations. For example, BP easily falls into the local minimum and has a fixed learning rate. Genetic algorithm (GA), as a representative of evolutionary computation, shows superiority in solving robust, nonlinear, parallel, and complex problems and is not based on knowledge of the problem. Using GA to optimize BP network weights simplifies the structure of the network, thus improving training speed, convergence, and generalization capability. The use of GA to optimize neural networks has recently been the focus of considerable attention from many scholars. Zhang Lijun [1] proposed a GA to optimize Elman network initial weights and then applied this algorithm to predict stock prices. He then established an efficient GA–Elman dynamic recurrent neural network stock prediction model. Ding Shifei [2, 3] used a hybrid coding of GA to optimize RBF structure and weight and then adjusted the network using the pseudoinverse method or LMS, which can attain a better network structure, stronger classification capability, and shorter time to build the network. The GA is an adaptive and intelligent learning algorithm.

To our knowledge, granulation and approximation are two key issues in rough set methodology. Granulation refers to the segmentation of the universe into different subsets with a certain criterion. The generated subsets are also called elemental granules or elemental concepts. Meanwhile, approximation refers to the approximate description of the arbitrary subset of the universe with these elemental concepts. Pawlak's rough set model utilizes equivalence relations to partition the universe and generate mutually exclusive equivalence classes as elemental concepts. This model is only applicable to data with nominal attributes. In numerical spaces, the concept of neighborhood serves an important function [4, 5, 6]. Neighborhood relations can be used to generate a family of neighborhood granules from a universe that is characterized by numerical features. Neighborhood granules can then be used to approximate decision classes. Based on this observation, a neighborhood rough set model was constructed [7, 8, 9]. Although the basic definitions of neighborhood rough sets have been proposed several years ago, to the best of our knowledge, only a few studies have been conducted on the applications of this model as of this writing.

The purpose of classification is to construct a classification function or classification model (also known as classifier) based on the characteristics of the data set. This model can classify the unknown sample map to a given category.The BP neural network has been extensively used in the classification field. However, this approach is limited by slow convergence speed and low accuracy. A large amount of information with high-dimensional data characteristic, which often fail to meet the requirement of fast convergence and accuracy diagnosis, causes the application of the BP neural network to be restricted in the areas of classification. To solve these problems, this study proposes a new algorithm that integrates the neighborhood rough sets and genetic neural network.

The remainder of this paper is organized as follows: Section 2 describes the BP neural network and the notions and properties of the neighborhood rough set. Section 3 presents the classification based on the neighborhood rough set and genetic neural algorithm. Section 4 presents the numerical results of the algorithm. Section 5 concludes our study.

* ***Corresponding author's*** e-mail: wuganzhou@126.com .

2 BP neural network structure and neighborhood rough set

2.1 BP NETWORK STRUCTURE

The BP network is a multilayer feedforward network, which uses nonlinear differential functions in weight training. With a simple structure and high plasticity, the BP network is currently the most extensively used neural network. As shown in Fig. 1, the BP network has three network structures, namely, the input, hidden, and output layers. The node number of the three layers is n, h, m. Neurons in different layers achieve full connectivity, whereas those in the same layer are disconnected. Neurons in the input layer receive the training samples and disseminate them through the hidden layer to the output layer. Neurons in the output layer receive the input response of the network. Then, to reduce the error between the target and the actual value, connection weight correction is conducted layer by layer, from the output layer to the input layer.

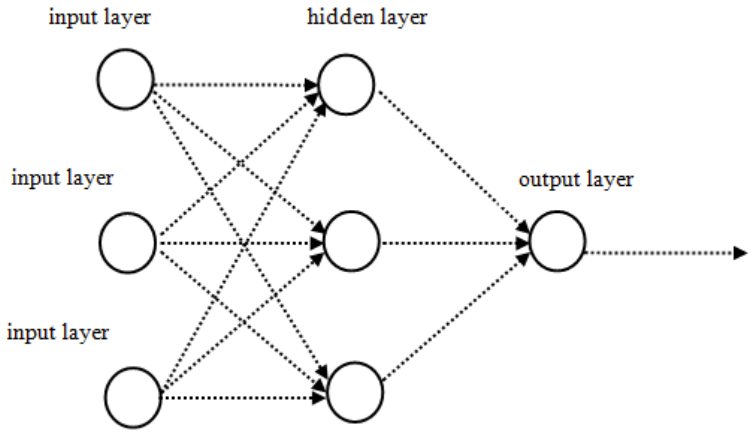

FIGURE 1 BP structure network

2.2 BASIC KNOWLEDGE OF NEIGHBORHOOD ROUGH SET

Attribute reduction is a core topic of rough set, and its main function is to delete the unimportant or irrelevant attributes. As such, the classification can become more effective. We discuss some concepts and definitions of neighborhood rough set.

Formally, the structural data for classification learning can be written as $IS=(U, A, V, f)$, where U is the nonempty set of samples (also called a universe or sample space) $\{x_1, x_2, \cdots, x_n\}$, A is the nonempty set of variables (also called features, inputs, or attributes) $\{a_1, a_2, \cdots, a_m\}$ to characterize the samples, V_a is the value domain of attribute *a*, and f is an information function, $f: U \times A \to V$. More specifically, (U, A, V, f) is also called a decision table if $A = C \cup D$, where *C* is the set of condition attributes and *D* is the output (also called decision).

Definition 1 [10]. Given the arbitrary $x_i \in U$ and $B \subseteq C$, the neighborhood $\delta_B(x_i)$ of x_i in subspace *B* is defined as follows:

$$\delta_B(x_i) = \{x_j \mid x_j \in U, \Delta_B(x_i, x_j) \le \delta\}, \tag{1}$$

where Δ is a metric function. $\forall x_1, x_2, x_3 \in U$ satisfies the following conditions:

1) $\Delta(x_1, x_2) \ge 0$;

2) $\Delta(x_1, x_2) = 0$, if and only if $x_1 = x_2$;

3) $\Delta(x_1, x_2) = \Delta(x_2, x_1)$;

4) $\Delta(x_1, x_3) \le \Delta(x_1, x_2) + \Delta(x_2, x_3)$.

$\delta_B(x_i)$ is the information granule centered with sample x_i. The size of the neighborhood depends on the threshold δ, which is a positive constant.

Definition 2 [10]. Given a set of samples U, N is a neighborhood relation on U and $\{\delta_B(x_i) \mid x_i \in U\}$ is the family of neighborhood granules. Then, we call (U, N) a neighborhood approximation space.

Definition 3 [10]. Given (U, N), for arbitrary $X \subseteq U$, two subsets of objects, called lower and upper approximations of X in terms of relation N, are defined as follows:

$$\underline{N}x = \{x_i \mid \delta(x_i) \subseteq X, x_i \in U\}, \tag{2}$$

$$\overline{N}x = \{x_i \mid \delta(x_i) \cap X \ne \Phi, x_i \in U\}. \tag{3}$$

The boundary region of X in the approximation space

is formulated as follows:

$$BNx = \overline{N}x - \underline{N}x. \tag{4}$$

The size of the boundary region reflects the degree of roughness of the set X in the approximation space.

2.3 CLASSIFIER MODEL OF NEIGHBORHOOD ROUGH SET

To begin the optimization process of GA, an initial population of candidate solutions is generated, defined as chromosomes. After the initial population is generated, the chromosomes are evaluated using an objective function. If the result is unsuitable, then the process proceeds to the next step and creates a mating pool using a selection operator. We utilize a roulette wheel selection method to choose an excellent chromosome in the mating
pool. Afterward, a new population is formed by utilizing crossover and mutation operators.

GA is used to train the neural network. The global optimization weights and threshold value are then obtained. Attribute reduction is used to reduce data inputted by the genetic neural network. Then, the representative decision attributes are selected to learn and predict the genetic neural network algorithm. Then, the trained neural network algorithm is integrated in the sensor on the base station to shorten the data processing time and improve the accuracy of classification. The classifier model of neighborhood rough set and genetic neural network is shown in FIGURE 2.

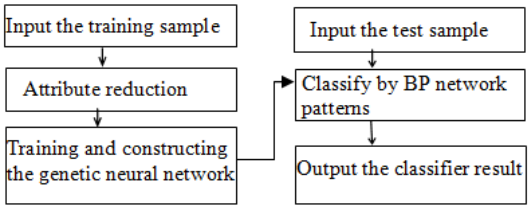

FIGURE 2. Classifier model of neighborhood rough set and genetic neural network

3 Classification based on neighborhood rough set and genetic neural algorithm

When the classifier model is implemented, then the classification based on neighborhood rough set and genetic neural algorithm is formed. The algorithm has two parts, namely, neighborhood rough set attribute reduction and GA to optimize the neural network.

3.1 ATTRIBUTE REDUCTION

Generally, we aim to recognize patterns in a relatively lower dimensional space to prevent the curse of dimensionality, reduce cost in measuring and processing information, and enhance the interpretability of learned models. Prior to neural network learning, the attribute reduction technique should be applied to high-dimensional data using neighborhood rough set to eliminate impractical information and simplify the neural network.

The construction of a neighborhood classifier has two steps. First, we search an optimal feature subspace, which has a similar discriminating power as the original data, but with a significantly reduced number of features. Then, we associate a neighborhood with each test sample in the selected subspace and assign the class with majority samples in the neighborhood to the test. The neighborhood rough set attribute reduction is as follows:

Step 1: Given (U, C, d) and the threshold δ, which is the threshold to control the size of the neighborhood, we specify the norm to be used.

Step 2: $\Phi \to red$, where *red* is the pool to contain the selected attributes.

Step 3: For each $a_i \in C - red$,

compute $\gamma_{red \cup a_i}(D) = \dfrac{POS_{red \cup a_i}(D)}{|U|}$,

compute $sig(a_i, red, D) = \gamma_{red \cup a_i}(D) - \gamma_{red}(D)$

Step 4: Select the attribute a_k, which satisfies

$$sig(a_k, red, D) = \max_i \left(sig(a_i, red, D) \right).$$

Step 5: If $sig(a_k, red, D) > 0$,

then $red \cup a_k \to red$.

go to Step 3

else

return *red*.

The most important problem in neighborhood-based classification is the threshold, which determines the size of the neighborhood. No sample will be included in the neighborhood if *d* is small. By contrast, the neighborhood cannot reflect the local information of the test if a large neighborhood is taken into consideration. After several tests, we finally select the optimal value as the threshold δ.

3.2 GA FOR BP NETWORK OPTIMIZATION

We use the GA to determine the optimal BP neural network weights and threshold value. The steps to achieve the optimization of GA for the neural network are as follows:

Step 1: BP neural network weights, threshold value coding, and initial population.

All weights and threshold value for optimization of the neural network are considered a group of chromosomes of GA. We use floating point coding to code them. Each individual encoded is a chromosome, with the length $L = n \times h + h \times m + h + m$, where n is the input number, h is the hidden layer number, and m is the output number.

Step 2: Fitness function.

For calculating the fitness function, the sum of the squared errors (SSE) error returned by the BP neural network was used. The SSE is a metric of the error made by the neural network. Other evaluation systems could be simply implemented. The fitness function can be defined as follows:

$$f = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2, \tag{5}$$

where y_i is the actual output of the sample, $\hat{y}_i$ is the expected output of the sample, and n is the number of learning samples.

Step 3: Selection criterion.

Selection is the process of selecting individuals. Best performing gene signatures from each population are allowed to pass through the other generations to produce offspring. Selection operators are used by the classical "roulette" algorithm, which is shown as follows:

$$p_i = \frac{f_i}{\sum_{j=1}^{N} f_j}. \tag{6}$$

Step 4: Crossover operation.

In the crossover operation, some constraints have been implemented to maintain acceptability of the solution (e.g., to prevent the repetition of a gene in the same chromosome).

We let a_1, a_2 be the parent individuals. After crossover, offspring individuals b_1, b_2 are generated using the following equations.

$$b_1 = a_1(1-\beta) + a_2\alpha, \tag{7}$$

$$b_2 = a_2(1-\beta) + a_1\alpha, \tag{8}$$

where $\beta \in [0,1]$ is a parameter that changes with evolution generations.

Step 5: Mutation operation.

Variation is the operation change of the gene value of some individual strings in the chromosome group. The individual α is generated using the following equation:

$$\alpha = \begin{cases} \alpha + r_2(\alpha - \alpha_{max})(1 - g/g_{max}), \theta \geq 0.5 \\ \alpha + r_2(\alpha_{min} - \alpha)(1 - g/g_{max}), \theta < 0.5 \end{cases}, \tag{9}$$

where θ is a random number, with a value between 0 and 1, r_2 is also a random number, α_{max} and α_{min} are the upper and lower bounds of the gene value of the chromosome, g is the current generation, and g_{max} is the maximum number of generations.

Step 6: If the setting error is satisfied, then the individual gene value obtained using GA is coded as the initial weights and threshold of the BP neural network, or go back to Step 2.

Based on these steps, the chart of the BP neural network based on GA is shown in Figure 3.

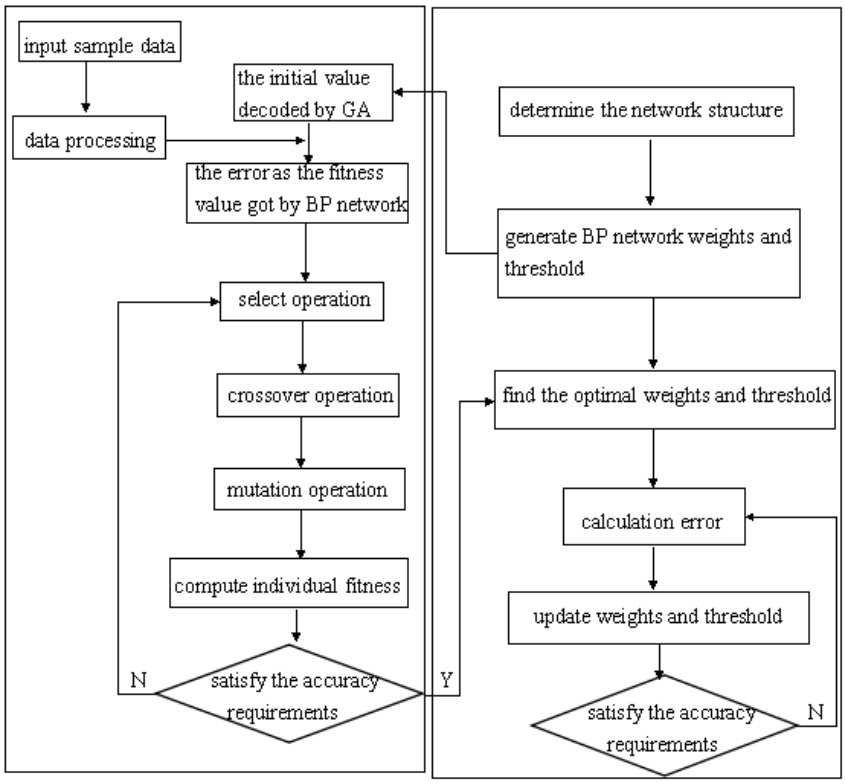

FIGURE 3. Chart of BP neural network based on GA

4 Simulation experiment

To test the proposed classification model, some data sets are downloaded from the standard UCI datasets [http://www.ics.uci.edu/mlearn/databases/ionospere]. The data set has 351 samples. Each sample has 34 features, which are used to predict the quality of the radar. We tested and selected 300 samples as training samples The remaining 51 samples were used as the simulation samples. First, given that each sample has 34 features, prior to BP neural network training, attribute reduction of the high-dimensional data using the neighborhood rough set is necessary. Attribute reduction can not only simplify the neural network input but also improve the real-time and diction efficiencies. We conduct attribute reduction on these data sets using the neighborhood rough set. We use $\delta = 0.15$ and $\delta = 0.25$

in the neighborhood rough set. Moreover, all of the three norms were used. The numbers of selected features based on different thresholds and norms are shown in Table 1. We observed that most of the candidate attributes are deleted. With the same
threshold, one-norm-based attribute reduction obtains the least features, whereas infinite-norm-based algorithms obtain the most features. We also observed that, even with the same threshold, the sizes of neighborhoods with different norms are different. Therefore, we selected nine features as the BP input layers, with $\delta = 0.15$ and two-norm-based attribute reduction.

TABLE 1 Numbers of selected features.

Norm	threshold radius	numbers of selected features
1-Norm	0.15	7
1-Norm	0.25	9
2-Norm	0.15	9
2-Norm	0.25	16
infinite-Norm	0.15	13
infinite-Norm	0.25	26

Second, after attribute reduction, we use MATLAB to establish a BP network with nine input layer neurons and one output layer neuron. The structure of the BP network is 9–13–1. The activation functions are "tansig" for the first two layers and "pure linear" for the last layer. The stop criterion for the training phase is 50,000 epochs or an SSE of less than 0.004. The initial learning rate is set to 0.3 and modified using the descent gradient momentum algorithm. These options provide a solution that is capable of reaching good equilibrium between the learningand generalization capabilities of the system. Finally, we use GA to optimize the network weights and threshold, with the initial population size of 100, genetic algebra of 500, and study accuracy of 10^{-6}. The crossover probability is 0.7, whereas the mutation probability is 0.01. After the preprocessing stage, the experimentation phase was conducted. Good convergence capability was achieved using the previously described parameters, as shown in Figure.4.

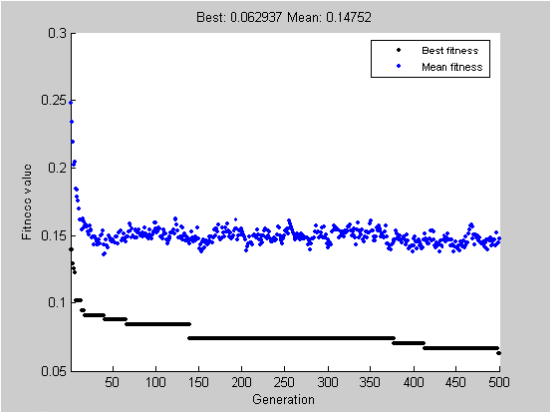

FIGURE 4 Best fitness and mean fitness values of generations.

Table 2 shows a comparison of the performance of these modes with the following steps: the optimal training step to convergence (after each model training, the step of the optimal model to convergence); the error sum of squares that covers the predicted and actual values of thedifferences of the square and is used to measure the proximity between the predicted and actual values, such that when the accuracy is the same as that in the simulation, a smaller SSE indicates higher model precision; and the accuracy of the simulation of samples.

TABLE 2 Comparison of experimental performance

The model of network	The steps of convergence	Error sum of squares	Accuracy rate (%)
BP neural network	837	4.3219	78.43
BP network based on GA and neighborhood rough set	365	1.8760	96.08

From Table 2, we observed that the BP network based on GA and neighborhood rough set improved the convergence speed. From the error sum of squares and the accuracy of simulation, the BP neural network based on GA and neighborhood rough set is significantly better than that of the traditional BP neural network.

5 Conclusion

We observed that the neighborhood rough set has significant power in attribute reduction experiments using UCI data sets. The classification is maintained or improved, although most features are deleted from the original data with neighborhood rough set-based attribute reduction. This finding indicates that the neighborhood rough set-based attribute reduction algorithm can select useful features and eliminate redundant and irrelevant information.

The BP neural network has some limitations. For example, a training network with random weight will cause a falling into local convergence. Based on these defects, this study proposes a new algorithm that optimizes the network weight using the GA based on the neighborhood rough set. On the basis of the experiment and Table 2, we can obtain the optimum model by optimizing the network weight, which verifies the effectiveness of the new algorithm. Using GA to optimize the BP network will take some time.

However, GA can improve convergence speed, shorten the training time, enhance operational efficiency, and obtain the optimal model for the simulation to yield the best result. The new algorithm will significantly improve performance, including such as factors as convergence speed, and will shorten run time, thus enhancing network efficiency.

References

[1] Zhang LJ, Yuan D 2008 Stock market forecasting research based on GA-Elman neural network East China Econ Manage 22(9):79–82.
[2] Ding SF, Su CY, Yu JZ 2011 An optimizing BP neural network algorithm based on genetic algorithm Artif Intell Rev 36(2):153–162.
[3] Ding SF, Jia WK, Su CY 2011 Research of neural network algorithm based on factor analysis and cluster analysis Neural Comput& Applic 20(2):297–302.
[4] W. Jin, Anthony K.H,Tung, J. Han, et al 2006 Ranking outliers using symmetric neighborhood relationship PAKDD 12(2):577–593.
[5] *Deleted by CMNT Editor*
[6] *Deleted by CMNT Editor*
[7] Q.H. Hu, J.F. Liu, D.R. Yu 2008 Mixed feature selection based on granulation and approximation Knowledge-Based Systems 21(2), 294–304.
[8] Qinghua Hu, Shuang An, Daren Yu 2010 Soft Fuzzy Dependency for Robust Feature Evaluation Information Sciences 180(22),4384-4400.
[9] Zhou,C,Yan, Y.Chen, Q 2006 Improving nearest neighbor classification with can weighted distance Pattern Recognition39(3), 635-645.
[10] *Deleted by CMNT Editor*

Author

Ganzhou Wu, 1980.8.10, Maoming, P.R. China

Current position, grades: Guangdong University of Petrochemical technology
University studies: received his mater degree from Yangtze University.
Scientific interest: His research interest fields include optimization and machine learning.
Publications: more than 13 papers published
Experience: Graduated from Yangtze University, Bachelor Degree of mathematics in 2003; Graduated from Yangtze University, Master Degree of mathematics; now working in Guangdong University of Petrochemical technology, with 9 years teaching experience.