Integrated neural network model with pre-RBF kernels

Hui Wen, Tao Yan, Zhiqiang Liu and Deli Chen
Institute of Electromechanical and Information Engineering, Putian University, Putian, Fujian, China

Abstract
To improve the network performance of radial basis function (RBF) and back-propagation (BP) networks on complex nonlinear problems, an integrated neural network model with pre-RBF kernels is proposed. The proposed method is based on the framework of a single optimized BP network and an RBF network. By integrating and connecting the RBF kernel mapping layer and BP neural network, the local features of a sample set can be effectively extracted to improve separability; subsequently, the connected BP network can be used to perform learning and classification in the kernel space. Experiments on an artificial dataset and three benchmark datasets show that the proposed model combines the advantages of RBF and BP networks, as well as improves the performances of the two networks. Finally, the effectiveness of the proposed method is verified.

Graphical abstract

Corresponding author:
Hui Wen, Institute of Electromechanical and Information Engineering, Putian University, 1133 Xueyuan Middle Street, Chengxiang District, Putian, Fujian 351100, China.
Email: wen_hui81@163.com

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
Keywords
Neural network, network integration, kernel mapping, radial basis function, back propagation

Highlights
An integrated neural network model with pre-RBF kernels is designed. The proposed network model can effectively combine the local nonlinear mapping ability of the hidden nodes of the RBF network with the global nonlinear classification ability of the BP network. The proposed network model improves the network performance of single RBF network and BP network obviously. The learning algorithm can effectively adapt to the proposed network model.

Introduction
In the field of machine learning, feedforward networks have been used extensively to solve various problems, including image classification, medical diagnosis, water quality inspection, where back-propagation (BP) neural networks and radial basis function (RBF) networks are typically used. The hidden nodes of BP neural networks generally use the unified sigmoid kernel to map the input samples, where the sigmoid kernel affords good generalization performance; however, the learning of BP networks involves repeated updating of weight parameters, which often results in slow convergence and converges to a local minimum. The optimization methods of BP neural networks primarily include the optimization of initialization weight, design of adaptive learning rate, addition of momentum term, correction of error cost function. Although these methods improve the shortcomings of BP neural networks, the effects of these methods are limited for more complex nonlinear problems.

In contrast to BP neural networks, the RBF network generally uses a Gaussian kernel with different parameters to map the input samples, where the Gaussian kernel exhibits good local response characteristics. The optimization process of the RBF network primarily includes the optimization of kernel parameters and linear output weights, which can be categorized into two stages: stage one improves the separability of samples by mapping the original samples through the hidden layer Gaussian kernels, and stage two completes the pattern classification by optimizing the linear hyperplane. Typical methods for optimizing the hidden nodes of RBF neural networks include fuzzy c-means clustering, sensitivity analysis, particle swarm optimization, and the dynamic adjustment of the number of hidden nodes; however, when these methods are used to optimize complex nonlinear problems, they often increase the burden of the subsequent linear output weight optimization.

The optimization BP and RBF networks are based on a single network model, from which relevant optimization algorithms are established. Because of the
characteristics of BP and RBF network models, the adaptability of these different optimization algorithms to different nonlinear problems is often limited. Hence, scholars have investigated the cascade neural network model,\(^ {19}\) which is composed of several independent sub-networks. However, the essence of a cascade neural network is to establish relevant learning algorithms for each independent sub-network. For a complex classification problem, the integration and effective connection of different sub-networks to form an entire network model is worth investigating.

In this context, convolutional neural networks (CNNs)\(^ {20}\) can be regarded as a network that integrates different sub-networks and establishes effective connections. In the CNN, convolution kernels are used to extract features, and then the connected multi-hidden layer BP neural network is used for classification. Tang et al.,\(^ {21}\) the BP neural network connected with convolution kernels is replaced by extreme learning machines, which improves the convergence speed of the CNN.

Inspired by the study above, an integrated neural network model with pre-RBF kernels is proposed herein. The main motivation of this research is to effectively combine the advantages of RBF network and BP network, and restrain the disadvantages of single RBF neural network and BP neural network. By integrating and connecting the RBF kernel mapping layer and BP neural network, a complementary network can be constructed. In the proposed integrated neural network model, the local features of the sample set can be effectively extracted to improve separability; subsequently, the connected BP network can be used to learning and classification in the kernel space. Therefore, the proposed network model effectively combines the local response ability of hidden nodes in the RBF network with the generalization ability of hidden nodes in the BP neural network, as well as effectively overcomes the shortcomings of the single RBF neural network and BP neural network in complex nonlinear problems.

### Integrated neural network model with pre-RBF kernels

#### Network structure

The construction principle of the proposed network structure is as follows: First, the original samples are input into the RBF kernel mapping layer, which extracts the local features of the original samples in different spatial regions and forms new feature vectors; subsequently, the BP network connected by the RBF hidden layer is used to complete the effective classification of samples in the feature space. Compared with the BP network, the proposed network structure improves the separability of input samples, which can accelerate the convergence speed of network weights and reduce the risk of falling into a local minimum; compared with the RBF network, the proposed network structure changes the linear weight optimization connecting the hidden and output layers to the nonlinear BP neural network, which exhibits better adaptability to complex nonlinear problems. Even if a certain deviation occurs in the original space mapping, the nonlinear BP network can compensate for the kernel mapping effect of the RBF hidden layer. Therefore, the proposed model can effectively combine the local nonlinear mapping ability of
the hidden nodes of the RBF network with the global nonlinear classification ability of the BP network and improve the shortcomings of single RBF and BP neural networks.

To effectively demonstrate the characteristics of the proposed network model, Figure 1 shows a schematic diagram of the integration of the Gaussian kernel with different parameters and the BP neural network. As shown, the pre-RBF kernels in the integrated network can improve the separability of the sample set; subsequently, the nonlinear BP network can be classified more effectively in the mapped kernel space.

Figure 2 shows the proposed neural network model constructed in this study. The proposed model comprises four parts: an input layer, an RBF kernel mapping layer, a BP hidden layer, and an output layer, where the number of BP hidden layers can be set to 1 or 2 depending on the actual problem. The RBF hidden layer is composed of a set of Gaussian kernel functions with different parameters. Let the number of Gaussian kernels in the RBF hidden layer be $K$, and the mapping of the input sample $x$ through the Gaussian kernels in the hidden layer be expressed as follows:

$$
\varphi_i(x) = \exp\left(-\frac{1}{2\sigma_i^2}||x - \mu_i||^2\right), \quad i = 1, 2, ..., K,
$$

(1)

where $\mu_i$ and $\sigma_i$ are the center and width of the $i$th Gaussian kernel, respectively. In this study, to simplify parameter $\sigma_i$, it is set to the same value $\sigma$, where $\sigma = \frac{d_{\text{max}}}{\sqrt{2K}}$, and $d_{\text{max}}$ is the farthest Euclidean distance between all the centers.

After using the Gaussian kernels of the RBF kernel mapping layer to map the original samples, to use the obtained kernel mapping value as the input of the
subsequent connected BP network, dual polarization processing is required to obtained the kernel mapping value. The formula for the dual-polarization transformation is

\[ \text{Figure 2. Structure diagram of integrated neural network with pre-RBF kernels: (a) with single BP hidden layer and (b) with two BP hidden layers.} \]
In the proposed integrated network structure, the BP hidden layer is composed of nodes from the RBF kernel mapping layer to the network output layer. Because the sigmoid function of the BP hidden layer is the hyperbolic tangent function, the output signal of node $j$ in the $c$th BP hidden layer can be expressed as

$$y_{j}^{(c)} = \phi_{j}(v_{j}) = a \tanh(bv_{j}),$$

where parameters $a$ and $b$ are constants.

The output $g_{j}(x)$ was used as the input of the BP hidden layer. If $c = 0$, $g_{j}(x)$ is the $j$th node of the BP network input layer, which can be expressed as

$$y_{j}^{(0)} = g_{j}(x)$$

The induced local region of the $j$th node of the $c$th BP hidden layer can be expressed as

$$v_{c}^{(c)}(j) = \sum_{i} \omega_{ji}^{(c)} y_{i}^{(c-1)}$$

where $y_{i}^{(c-1)}$ is the output of the node $i$ from layer $c - 1$, and $\omega_{ji}^{(c)}$ is the weight from layer $c - 1$ to layer $c$.

The output signal at node $k$ in the output layer can be expressed as

$$o_{k} = y_{j}^{(C)}$$

where $C$ is the sum of the input, hidden, and output layers of the BP network component.

### Algorithm implementation

When the network model is established, the subsequent task is to establish a corresponding learning algorithm to optimize the network parameters. The algorithm implementation of the proposed network process includes these steps: (1) Initialize parameters and preprocess samples; (2) Optimize the parameters of RBF kernel mapping layer; (3) Double polarize each input sample; (4) Forward calculation of BP network; (5) Backward calculation and update the weights of each layer of BP network; (6) Decision output of sample label value. In this study, the parameters of the Gaussian kernel in the RBF kernel mapping layer are optimized using a fuzzy c-means clustering algorithm, and the weights of each layer of the BP network are optimized using the existing BP algorithm based on gradient descent.

Figure 3 shows the implementation algorithm of the proposed network, where the optimization of the Gaussian kernel parameters is provided in Table 1.

In Figure 3, the formula to calculate for the overall mean square error of the BP network is

$$g_{i}(x) = 2 \cdot \varphi_{i}(x) - 1$$ (2)
Table 1. Algorithm implementation of fuzzy c-means clustering.

For sample set \( \{ x_i \}_{i=1}^{N} \), set \( K \) is the number of clusters, \( \mu_i \) (\( i = 1, 2, \ldots, K \)) is the clustering center.
1. Selecting a random number between 0 and 1 to initialize membership matrix \( U \), \( u_{ij} \) is the element of matrix \( U \), where \( \sum_{i=1}^{K} u_{ij} = 1 \), \( \forall j = 1, \ldots, N \).
2. Calculate each clustering center \( \mu_i \), where \( \mu_i = \frac{\sum_{j=1}^{N} u_{ij}^m x_j}{\sum_{j=1}^{N} u_{ij}^m} \), \( m \) is a weighted index, \( m \in [1, \infty) \).
3. Optimize the objective function \( J(U, \mu_1, \ldots, \mu_K) = \sum_{i=1}^{K} J_i = \sum_{i=1}^{K} \sum_{j=1}^{N} u_{ij}^m d_{ij}^2 \). If \( J < \xi \), the algorithm stops. Here, \( d_{ij} = ||\mu_i - x_j|| \), \( \xi \) is the threshold.
4. Update the matrix \( U \), where \( \mu_i \) = \( \frac{1}{\sum_{k=1}^{K} (\frac{q_{ik}}{q_{ik}})^{2/(m-1)}} \). Return to step 2 for iteration.
\[ J(\omega) = \frac{1}{2} \sum_{j=1}^{L} e_j^2 = \frac{1}{2} \sum_{j=1}^{L} (d_j - o_j)^2, \tag{7} \]

where \( d_j \) is the output of the target output of the proposed network, \( o_j \) is the actual output of the network, and \( L \) is the number of output nodes.

The back calculation of the BP network is the updating process of the local gradient and can be expressed as

\[
\delta_j^{(n)} = \begin{cases} 
\epsilon_j^{(C)} & \text{for node } j \text{ in the output layer } C \\
\varphi_j^{(C)} \sum_k \delta_k^{(c+1)} \omega_{kj}^{(c+1)} & \text{for node } j \text{ in the sigmoid layer } c 
\end{cases} \tag{8}
\]

The weight updating process of layer \( c \) in BP network is expressed as follows:

\[
\omega_{ji}^{(c)}(m+1) = \omega_{ji}^{(c)}(m) + \eta \delta_j^{(c)}(m)v_j^{(c-1)}(m), \tag{9}
\]

where \( m \) is the iteration step, and \( \eta \) is the learning rate.

**Experimental comparison and analysis**

In this section, the performance of the proposed method is evaluated using an artificial dataset, namely the Concrete Circle and three benchmark datasets\(^{22}\) from University of California, Irvine (UCI): climate, heart disease, and blood transfusion. Table 2 provides a description of the classification datasets. The performance of the proposed method is compared with a BP algorithm based on stochastic gradient descent (SGBP),\(^ {11} \) a constrained optimization method based on a BP neural network (CO-BP),\(^ {23} \) an RBF network based on fuzzy c-means clustering (FCRBF),\(^ {14} \) and an optimized RBF network based on fractional order gradient descent with momentum (FOGDM-RBF).\(^ {24} \) In each dataset, all data samples in each dataset are scaled to \([-1, 1]\), the number of kernels in the RBF mapping layer is adjusted manually based on the distribution of the sample space, and the number of BP hidden layers is set to one and two layers; the number of BP hidden layer nodes was set between two and nine, the network learning rate was adjusted iteratively using the simulated annealing algorithm, and the sigmoid kernel parameters

| Datasets          | No. of classes | No. of features | No. of training samples | No. of testing samples |
|-------------------|----------------|-----------------|-------------------------|------------------------|
| Concrete Circle   | 2              | 2               | 1200                    | 3300                   |
| Climate           | 2              | 18              | 270                     | 270                    |
| Heart disease     | 2              | 13              | 151                     | 151                    |
| Blood transfusion | 2              | 4               | 374                     | 374                    |

Table 2. Information description of different classification datasets.
were set as $a = 1.1716$ and $b = 0.6667$. The operating environment of the experiment was an Intel (R) core i7-9700, 3.00 GHz CPU, 8GB RAM, and MATLAB 2013. Each experiment was repeated 10 times.

Concrete circle classification problem

Figure 4 shows a graphical representation of the Concrete Circle classification dataset, where the two classes of samples are mixed interactively and can be used to measure the characteristics of the proposed method. Figure 5 shows a comparison of the mean square error (MSE) learning curves of the proposed method involving SGBP and FCRBF. The MSE of SGBP in the training set is relatively large, which indicates that the learning effect of BP network on complex nonlinear problems is limited. Compared with SGBP, FCRBF has the advantages of good stability, the learning effect is improved to a certain extent. Compared with SGBP and FCRBF, the MSE of the proposed method was smaller, and the learning curve converged faster; therefore, the proposed method exhibited better learning performance for the training sample set.

Compared with the SGBP and FCRBF, Figure 6 and Table 3 show that the classification effect of the proposed method is better. From the classification accuracy of the test set, when the number of BP hidden layers is set to 1, the proposed network is approximately 6.7% and 31% higher than FCRBF and SGBP respectively; when the number of BP hidden layers is set to 2, the proposed network is approximately 7.5% and 21.3% higher than FCRBF and SGBP respectively. This shows that the proposed method can adapt well to the complex sample set, as evident by the better learning performance, and the classification effect on the test set was significantly higher.
UCI benchmark classification problems

Under the benchmark datasets, the performance comparison results of the proposed method and other methods are shown in Tables 4 to 6. The classification accuracy of the proposed method outperforms SGBP, FCRBF, CO-BP, and FOGDM-RBF in varying degrees on the benchmark datasets. For the Climate dataset, when the number of BP hidden layers is set to 1, the testing accuracy of the proposed method is approximately 1.3%–2% higher than those of SGBP, CO-BP, FCRBF, and FOGDM-RBF. When the number of BP hidden layers is set to 2, the testing accuracy of the proposed method is approximately 1.3%–2.3% higher than those of SGBP, CO-BP, FCRBF, and FOGDM-RBF. For the Heart Disease data set, the testing accuracy of the proposed method is higher than that of SGBP.

Figure 5. Comparison of MSE learning curves of different methods on Concrete Circle dataset: (a) SGBP, (b) FCRBF, and (c) proposed method.

**UCI benchmark classification problems**

Under the benchmark datasets, the performance comparison results of the proposed method and other methods are shown in Tables 4 to 6. The classification accuracy of the proposed method outperform SGBP, FCRBF, CO-BP, and FOGDM-RBF in varying degrees on the benchmark datasets. For the Climate dataset, when the number of BP hidden layers is set to 1, the testing accuracy of the proposed method is approximately 1.3%–2% higher than those of SGBP, CO-BP, FCRBF, and FOGDM-RBF. When the number of BP hidden layers is set to 2, the testing accuracy of the proposed method is approximately 1.3%–2.3% higher than those of SGBP, CO-BP, FCRBF, and FOGDM-RBF. For the Heart Disease data set, the testing accuracy of the proposed method is higher than that of SGBP.
Figure 6. Comparison of classification effects of different methods on Concrete Circle dataset: (a) SGBP, (b) FCRBF, and (c) proposed method.

Table 3. Experimental comparison of different methods on Concrete Circle dataset.

| Methods                        | No. of kernels | Training MSE | Testing misclassification (%) |
|--------------------------------|----------------|--------------|-------------------------------|
| SGBP (one hidden layer)        | 9              | 0.824521     | 1080 (−32.73)                |
| SGBP (two hidden layers)       | 8, 5           | 0.627534     | 732 (−22.18)                 |
| FCRBF                          | 50             | 0.296667     | 276 (−8.36)                  |
| Proposed method (one BP hidden layer) | 22, 9          | 0.010063     | 55 (−1.67)                   |
| Proposed method (two BP hidden layers) | 16, 8, 4       | 0.010840     | 28 (−0.85)                   |
obviously. When the number of BP hidden layers is set to 1, the testing accuracy of the proposed method is approximately 1.6%–5% higher than those of CO-BP, FCRBF, and FOGDM-RBF. When the number of BP hidden layers is set to 2, the

Table 4. Performance comparisons of different methods on Climate dataset.

| Methods                  | Number of kernels | Training time (s) | Training accuracy (%) | Testing accuracy (%) |
|--------------------------|-------------------|-------------------|-----------------------|----------------------|
| SGBP (one hidden layer)  | 7                 | 0.57              | 92.52                 | 91.48                |
| SGBP (two hidden layers) | 7, 6              | 0.45              | 92.76                 | 91.86                |
| FCRBF                    | 16                | 0.49              | 92.39                 | 91.22                |
| CO-BP (one hidden layer) | 7                 | 0.46              | 92.78                 | 91.94                |
| CO-BP (two hidden layers)| 7, 5              | 0.38              | 93.26                 | 92.37                |
| FOGDM-RBF                | 16                | 0.82              | 92.85                 | 91.62                |
| Proposed (one BP hidden layer) | 16, 6         | 0.88              | 93.87                 | 93.24                |
| Proposed (two BP hidden layers) | 16, 6, 4    | 0.79              | 93.92                 | 93.63                |

Table 5. Performance comparisons of different methods on Heart Disease dataset.

| Methods                  | Number of kernels | Training time (s) | Training accuracy (%) | Testing accuracy (%) |
|--------------------------|-------------------|-------------------|-----------------------|----------------------|
| SGBP (one hidden layer)  | 6                 | 0.42              | 75.74                 | 63.12                |
| SGBP (two hidden layers) | 7, 4              | 0.35              | 78.85                 | 65.23                |
| FCRBF                    | 10                | 0.67              | 80.63                 | 76.42                |
| CO-BP (one hidden layer) | 7                 | 0.39              | 77.53                 | 74.14                |
| CO-BP (two hidden layers)| 7, 4              | 0.37              | 79.62                 | 75.57                |
| FOGDM-RBF                | 10                | 0.84              | 80.92                 | 77.56                |
| Proposed (one BP hidden layer) | 10, 4         | 0.73              | 81.68                 | 79.18                |
| Proposed (two BP hidden layers) | 10, 4, 2    | 0.68              | 82.20                 | 79.89                |

Table 6. Performance comparisons of different methods on Blood Transfusion dataset.

| Methods                  | Number of kernels | Training time (s) | Training accuracy (%) | Testing accuracy (%) |
|--------------------------|-------------------|-------------------|-----------------------|----------------------|
| SGBP (one hidden layer)  | 7                 | 1.84              | 72.58                 | 43.38                |
| SGBP (two hidden layers) | 7, 4              | 1.75              | 73.52                 | 46.19                |
| FCRBF                    | 30                | 2.82              | 78.63                 | 76.36                |
| CO-BP (one hidden layer) | 8                 | 1.53              | 77.53                 | 68.14                |
| CO-BP (two hidden layers)| 8, 5              | 1.25              | 78.84                 | 70.31                |
| FOGDM-RBF                | 30                | 3.23              | 79.31                 | 77.14                |
| Proposed (one BP hidden layer) | 30, 8         | 3.41              | 79.68                 | 77.76                |
| Proposed (two BP hidden layers) | 30, 7, 6   | 3.21              | 79.92                 | 78.13                |
testing accuracy of the proposed method is approximately 2.3%–4.3% higher than those of CO-BP, FCRBF, and FOGDM-RBF. For the Blood Transfusion data set, the testing accuracy of the proposed method is higher than those of SGBP obviously. When the number of BP hidden layers is set to 1, the testing accuracy of the proposed method is approximately 0.6%–9.6% higher than those of CO-BP, FCRBF, and FOGDM-RBF. When the number of BP hidden layers is set to 2, the testing accuracy of the proposed method is approximately 1%–10% higher than those of CO-BP, FCRBF, and FOGDM-RBF. As shown, compared with other methods, the proposed method adds new kernel parameters and training time in the learning process; however, its training accuracy and classification performance were significantly higher than those of other methods, indicating that the proposed method exhibits better adaptability to different classification problems, and the effectiveness of the proposed method was further verified.

Parameter analysis and discussion

In this study, the RBF and BP networks were effectively connected and integrated to construct the proposed network model. The construction of the proposed network primarily includes three parameters: number of hidden nodes in the RBF kernel mapping layer, number of BP hidden layers, and number of BP hidden nodes. Table 7 shows the performance of the model when these parameters change under the Concrete Circle dataset. As shown, when the three parameters are combined arbitrarily, the proposed method can maintain a relatively stable and high classification performance. This further verifies that the proposed can effectively combine the local nonlinear mapping ability of the hidden nodes of the RBF network with the global nonlinear classification ability of the BP network. By selecting the Heart Disease dataset and adjusting the parameters of the BP hidden layer and the number of hidden nodes in the RBF kernel mapping layer, the performance of the proposed method is compared with other methods, as presented in Figure 7. As shown, compared with other methods, the proposed method can maintain a relatively higher classification performance, and the overall stability of the network is better. This indicates that although the proposed method increases the number of training parameters, it reduces the dependence on the selection of BP hidden layer parameters and the number of hidden nodes in the RBF kernel mapping layer. Therefore, the advantages of RBF network stability and BP network generalization performance can be effectively combined. The proposed method can effectively overcome the shortcomings of single BP neural networks and RBF neural networks in complex nonlinear problems.

Conclusion

In this study, an integrated neural network model with pre-RBF kernels is proposed. The proposed network effectively connects and integrates RBF and BP networks, and the established learning algorithm can optimize the network
Table 7. Performance comparison of different parameters of proposed method on Concrete Circle dataset.

| Proposed network | Number of Gaussian kernels | Number of sigmoid kernels | First layer MSE | Second layer MSE | Training MSE | Testing misclassification (%) |
|------------------|-----------------------------|---------------------------|----------------|----------------|--------------|----------------------------------|
| With a RBF kernel mapping layer and a single BP hidden layer | 8 | 4 | — | 0.082817 | 100 (−3.03) |
| | 8 | 5 | — | 0.102739 | 199 (−6.03) |
| | 8 | 6 | — | 0.119589 | 187 (−5.67) |
| | 10 | 5 | — | 0.021503 | 114 (−3.45) |
| | 12 | 4 | — | 0.104048 | 200 (−6.06) |
| | 12 | 6 | — | 0.091376 | 151 (−4.58) |
| | 12 | 8 | — | 0.029211 | 103 (−3.12) |
| | 14 | 5 | — | 0.092418 | 144 (−4.36) |
| | 14 | 7 | — | 0.050239 | 89 (−2.70) |
| | 16 | 9 | — | 0.013122 | 85 (−2.58) |
| | 20 | 5 | — | 0.037044 | 82 (−2.48) |
| | 22 | 9 | — | 0.010063 | 55 (−1.67) |
| | 22 | 5 | — | 0.035133 | 95 (−2.88) |
| | 24 | 5 | — | 0.044146 | 115 (−3.48) |
| | 24 | 8 | — | 0.013729 | 49 (−1.48) |
| | 26 | 6 | — | 0.023165 | 72 (−2.18) |
| | 26 | 9 | — | 0.018021 | 68 (−2.06) |
| | 28 | 4 | — | 0.045763 | 111 (−3.36) |
| | 28 | 6 | — | 0.017106 | 109 (−3.30) |
| | 30 | 8 | — | 0.015366 | 75 (−2.27) |
| With a RBF kernel mapping layer and two BP hidden layers | 8 | 8 | 4 | 0.047380 | 71 (−2.15) |
| | 8 | 4 | 5 | 0.077025 | 77 (−2.33) |
| | 10 | 7 | 5 | 0.017515 | 54 (−1.64) |
| | 10 | 6 | 3 | 0.068280 | 87 (−2.64) |
| | 10 | 4 | 9 | 0.021314 | 52 (−1.58) |
| | 10 | 3 | 5 | 0.059005 | 69 (−2.09) |
| | 12 | 5 | 5 | 0.027460 | 44 (−1.33) |
| | 12 | 5 | 8 | 0.029915 | 78 (−2.36) |
| | 14 | 6 | 7 | 0.023192 | 47 (−1.42) |
| | 16 | 9 | 4 | 0.024536 | 60 (−1.82) |
| | 16 | 7 | 5 | 0.014804 | 48 (−1.45) |
| | 16 | 8 | 4 | 0.010840 | 28 (−0.85) |
| | 18 | 6 | 5 | 0.009685 | 47 (−1.42) |
| | 18 | 6 | 7 | 0.005082 | 37 (−1.12) |
| | 18 | 5 | 4 | 0.036158 | 56 (−1.70) |
| | 18 | 8 | 5 | 0.013817 | 43 (−1.30) |
| | 20 | 7 | 5 | 0.008889 | 40 (−1.21) |
| | 20 | 4 | 6 | 0.035327 | 72 (−2.18) |
| | 24 | 8 | 6 | 0.010363 | 46 (−1.39) |
| | 26 | 5 | 8 | 0.018966 | 45 (−1.36) |
| | 30 | 6 | 4 | 0.017765 | 61 (−1.85) |
parameters. Experiments on an artificial dataset and several benchmark datasets show that the proposed model can effectively combine the local nonlinear mapping ability of the hidden nodes of the RBF network with the global nonlinear classification ability of the BP network, as well as improve the network performance of a single BP network and an RBF network. Overall, the testing accuracy of the proposed method is higher than that of SGBP obviously. For the Concrete Circle dataset, the testing accuracy of the proposed method is approximately 1.3%–2.3% higher than those of FCRBF and SGBP under different training parameters. For the Climate dataset, the testing accuracy of the proposed method is approximately 6.7%–31% higher than those of SGBP, CO-BP, FCRBF, and FOGDM-RBF under different training parameters. For the Heart Disease data set, the testing accuracy of the proposed method is approximately 1.6%–5% higher than those of CO-BP, FCRBF, and FOGDM-RBF under different training parameters. For the Blood Transfusion data set, the testing accuracy of the proposed method is approximately 0.6%–10% higher than those of CO-BP, FCRBF, and FOGDM-RBF under different training parameters. The advantages of the proposed method can be effectively verified. However, the training samples are presented via batch learning, which fails to consider the online learning method of using sequence samples. Hence, we will investigate the learning method involving sequence samples in our future study.

Figure 7. Classification performance comparison of proposed method with other methods based on different kernel parameters: (a) BP hidden nodes and (b) RBF kernel mapping.
Acknowledgements

We thank the Putian Science and Technology Bureau (2018RP4004). We thank Dr. Hang Xu for his guidance and help in the revised draft.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by the Natural Science Foundation of Fujian Province (Nos. 2019J01815, 2019J01816, 2020J05213, and 2020H0047), New Century Excellent Talents in Fujian Province University (2018, Yantao), and the Department of Education of Fujian Province (JT180486, FJJKCG20-101).

ORCID iD

Hui Wen https://orcid.org/0000-0003-4971-9836

Supplemental material

Supplemental material for this article is available online.

References

1. Rashno A, Nazari B, Sadri S, et al. Effective pixel classification of Mars images based on colony optimization feature selection and extreme learning machine. Neurocomputing 2017; 226: 66–79.
2. Zhang YD, Satapathy SC, Guttery DS, et al. Improved breast cancer classification through combining graph convolutional network and convolutional neural network. Inf Process Manage 2021; 58(2): 102439.
3. Wang SH, Rao RV, Chen P, et al. Abnormal breast detection in mammogram images by feed-forward neural network trained by Jaya algorithm. Fundam Inform 2017; 151(1–4): 191–211.
4. Lin HJ, Dai QY, Zheng LL, et al. Radial basis function artificial neural network able to accurately predict disinfection by-product levels in tap water: taking haloacetic acids as a case study. Chemosphere 2020; 248: 125999.
5. Deng Y, Zhou XL, Shen J, et al. New methods based on back propagation (BP) and radial basis function (RBF) artificial neural networks (ANNs) for predicting the occurrence of haloketones in tap water. Sci Total Environ 2021; 772: 145534.
6. Hong HC, Zhang ZY, Guo AD, et al. Radial basis function artificial neural network (RBF ANN) as well as the hybrid method of RBF ANN and grey relational analysis able to well predict trihalomethanes levels in tap water. J Hydrol 2020; 591: 125574.
7. Ding SF, Su CY and Yu JZ. An optimizing BP neural network algorithm based on genetic algorithm. Artif Intell Rev 2011; 36(2): 153–162.
8. Yang SS, Siu S and Ho CL. Analysis of the initial values in split-complex back propagation algorithm. *IEEE Trans Neural Netw* 2008; 19(9): 1564–1573.

9. Yam JYF, Chow TWS and Leung CT. A new method in determining initial weights of feedforward neural networks for training enhancement. *Neurocomputing* 1997; 16(1): 23–32.

10. Lee HM, Chen CM and Huang TC. Learning efficiency improvement of back-propagation algorithm by error saturation prevention method. *Neurocomputing* 2001; 41: 125–143.

11. Istook E and Martinez T. Improved back propagation learning in neural networks with windowed momentum. *Int J Neural Syst* 2002; 12(3–4): 303–318.

12. Vetela JE and Reifman J. Premature saturation in back-propagation networks: mechanism and necessary conditions. *Neural Netw* 1997; 10(4): 721–735.

13. Rimer M and Martinez T. CB3: an adaptive error function for back propagation training. *Neural Process Lett* 2006; 24(1): 81–92.

14. Niros AD and Tsekouras GE. A novel training algorithm for RBF neural network using a hybrid fuzzy clustering approach. *Fuzzy Set Syst* 2012; 193: 62–84.

15. Staiano A, Tagliaferri R and Pedrycz W. Improving RBF networks performance in regression tasks by means of a supervised fuzzy clustering. *Neurocomputing* 2006; 69(13–15): 1570–1581.

16. Wang H, Feng R, Han Z, et al. ADMM-based algorithm for training fault tolerant RBF networks and selecting centers. *IEEE Trans Neural Netw Learn Syst* 2018; 29(8): 3870–3878.

17. Feng HM. Self-generation RBFNs using evolutionary PSO learning. *Neurocomputing* 2006; 70: 41–251.

18. Yu H, Xie T, Bartczak T, et al. An incremental design of radial basis function networks. *IEEE Trans Neural Netw Learn Syst* 2014; 2: 1793–1803.

19. Ao SI. A hybrid neural network cybernetic system for quantifying cross-market dynamics and business forecasting. *Soft Computing* 2011; 15(6): 1041–1053.

20. Pradhan R, Aygun RS and Maskey M. Tropical cyclone intensity estimation using a deep convolutional neural network. *IEEE Trans Image Process* 2018; 27(2): 692–702.

21. Tang JX, Deng CW and Huang GB. Extreme learning machine for multilayer perceptron. *IEEE Trans Neural Netw Learn Syst* 2016; 27(4): 809–821.

22. Blake C and Merz C. UCI repository of machine learning databases. University of California, Irvine, Department of Information and Computer Sciences. http://archive.ics.uci.edu/ml

23. Li Z, Wang F, Sun T, et al. A constrained optimization method based on BP neural network. *Neural Comput Appl* 2018; 29(2): 413–421.

24. Xue H, Shao Z-P and Sun H-B. Data classification based on fractional order gradient descent with momentum for RBF neural network. *Network* 2020; 31(1–4): 166–185.

**Author biographies**

Hui Wen was born in 1981, he received the Ph.D. degree in intelligent information processing from Shenzhen University, China. He is currently an associate professor in institute of electromechanical and Information Engineering, Putian University. His current research interests include machine learning and neural networks.
Tao Yan was born in 1981, he received the Ph.D. degree in Shanghai University, China. He is currently a professor in institute of electromechanical and Information Engineering, Putian University. His current research interests include 3D video analysis, video coding and pattern recognition.

Zhiqiang Liu was born in 1983, he received the Ph.D. degree in Communication University of China, China. He is currently a lecturer in institute of electromechanical and Information Engineering, Putian University. His current research interests include machine learning, remote sensing image processing.

Deli Chen was born in 1970. He graduated from Tianjin Institute of light industry with a bachelor’s degree in 1995. He is currently an associate professor in institute of electromechanical and Information Engineering, Putian University. His main research interests include computer network, data communication and pattern recognition.