Article ID: 1006-8775(2009) 01-0083-06

STUDY ON THE METEOROLOGICAL PREDICTION MODEL USING THE LEARNING ALGORITHM OF NEURAL ENSEMBLE BASED ON PSO ALGORITHMS

WU Jian-sheng (吴建生)[1], JIN Long (金 龙)[2]

(1. Department of Mathematic and Computer Sciences, Liuzhou Teachers College, Liuzhou 545004 China;
2. Guangxi Research Institute of Meteorological Disasters Mitigation, Naning 530022 China)

Abstract: Because of the difficulty in deciding on the structure of BP neural network in operational meteorological application and the tendency for the network to transform to an issue of local solution, a hybrid Particle Swarm Optimization Algorithm based on Artificial Neural Network (PSO-BP) model is proposed for monthly mean rainfall of the whole area of Guangxi. It combines Particle Swarm Optimization (PSO) with BP, that is, the number of hidden nodes and connection weights are optimized by the implementation of PSO operation. The method produces a better network architecture and initial connection weights, trains the traditional backward propagation again by training samples. The ensemble strategy is carried out for the linear programming to calculate the best weights based on the "east sum of the error absolute value" as the optimal rule. The weighted coefficient of each ensemble individual is obtained. The results show that the method can effectively improve learning and generalization ability of the neural network.

Key words: neural network ensemble; particle swarm optimization; optimal combination

CLC number: P456.7 **Document code:** A **doi:** 10.3969/j.issn.1006-8775.2009.01.014

1 INTRODUCTION

Rainfall forecasting is a challenging task, especially in a modern world where we are facing major environmental problems like global warming. Accurate forecasts of rainfall such as its spatial and temporal distributions are the best approach to help prevent casualties and damage caused by natural disasters [1-2]. Although the physically-based approach for rainfall forecasting has several advantages, given the short time scale, the small catchments area, and the massive costs associated with collecting the required meteorological data, it is not a feasible alternative in most cases because it involves many variables which are interconnected in a very complex way.

Artificial Neural Networks (ANNs) have recently become important alternative tools to conventional methods in modeling complex non-linear relationships [3-4]. As neural network approaches require rigorous theoretical support, effects of applications are strongly dependent upon the operator's experience, which will cause overfitting, heavily degrade the generalization ability of networks and limit applicability of neural networks in meteorological application [5-6]. Neural network ensemble is a learning paradigm where a number of neural networks are trained for the same task [7], which shows that the generalization ability of a neural network system can be significantly improved through training many neural networks and then combining their results [8]. This kind of technique has been widely used in different fields such as face recognition, optical character recognition, and medicine and so on [9-13].

Particle swarm optimization (PSO) is a population-based optimization technique for unconstrained minimization problems. Its development is based on the observations of social behaviors of animals such as bird flocks, fish schooling, and swarm theory [14 - 15]. In this paper, a novel optimization approach is proposed. The PSO algorithm is applied to

Received date: 2007-11-26; **revised date:** 2009-03-24

Foundation item: Natural Science Foundation of Guangxi (0832019Z); Natural Science Foundation of China (40675023)

Biography: WU Jian-sheng, associate professor, mainly undertaking in neural network applications and Intelligent Optimization Algorithm.

E-mail for correspondence author: jinlong01@163.com

optimize neural network architecture and connection weights. The evolved neural network architecture and connection weights are then input into a new neural network. The new neural network is trained using back propagation (BP) algorithm and ensemble individuals are acquired. The ensemble output is obtained by linear programming that takes the least sum of the absolute values of error as the optimal rule. The real data – monthly precipitation in Guangxi China is used to examine the proposed approach. Experimental results show that the proposed approach is easy to operate and can lead to a high prediction accuracy.

2 LEARNING ALGORITHM OF NEURAL ENSEMBLE BASED ON PSO ALGORITHM

The position-speed relation model of PSO operates easily. The proposed method of using PSO to evolve neural networks includes three steps, i.e., (i) using the global searching ability of PSO to find an appropriate network architecture and connection weights, (ii) using BP algorithm to search peak value(s) in detail, and (iii) obtaining results using the ensemble strategy. Mathematically, optimization problems of PSO-neural networks can be described as follows:

$$\begin{cases} \min \ E(w,v,q,g) = \dfrac{1}{N_1}\sum_{k=1}^{N_1}\sum_{t=1}^{n}[y_k(t)-\hat{y}_k(t)]^2 < e_1 \\ \hat{y}_k(t) = \sum_{j=1}^{p} v_{jk}\cdot f[\sum_{i=1}^{m} x\cdot w_{ij}+q_j]+g_t \\ f(x) = \dfrac{1}{1+e^{-x}} \\ s.\ t \quad w\in R^{m\times p},\ v\in R^{p\times n},\ q\in R^p,\ g\in R^n \end{cases} \tag{1}$$

where x is the training sample and $\hat{y}_k(t)$ and $y_k(t)$ are the desired output and real data, respectively. The fitness function is defined as follows:

$$F(w,v,q,g)=\frac{1}{1+\min E(w,v,q,g)} \tag{2}$$

Here we introduce our scheme:

Step 1: Initialize positions and speeds of a number of particles. L particles are randomly generated and each of them includes two parts: position and speed. The position of each particle consists of network node link and connection weights. The hidden nodes are encoded as binary code string; 1 with connection and 0 without connection. The connection weights are encoded as float string, randomly generated within [-1, 1].

Step 2: Input training samples and calculate the fitness of each particle according to Eq.(2). Initialize individual best position $P_{best}(t)$ and the global best position $P_{gbest}(t)$.

Step 3: Compare individual current fitness with the fitness of its experienced best position. If the current fitness is better, it is set to be the current best position. Moreover, compare the fitness with the fitness of global best position. If the fitness is better, it is set to be the global best position.

Step 4: The equation of speed evolution for each particle can be written as follows.

$$v_{ij}(t+1)=w\cdot v_{ij}(t)+c_1r_1(P_{best}(t)-x_{ij}(t))+c_2r_2(P_{gbest}(t)-x_{ij}(t)) \tag{3}$$

$$w(t)=w_{\max}-\frac{w_{\max}-w_{\min}}{iter_{\max}}\cdot iter \tag{4}$$

where $w_{\max}$ and $w_{\min}$ denote the maximum and minimum of inertia weights, respectively, while $iter$ and $iter_{\max}$ denote the current iteration number and maximum iteration number, respectively.

Step 5: According to Riget et al. [16], the equation of network link can be written as follows:

$$x_{ij}(t+1)=\begin{cases} 0, & r\ge \dfrac{1}{1+\exp(-v_{ij}(t+1))} \\ 1, & r< \dfrac{1}{1+\exp(-v_{ij}(t+1))} \end{cases} \tag{5}$$

where r is the random number that distributes uniformly within [0, 1]. The equation of position evolution for each particle can be written as follows:

$$x_{ij}(t+1)=x_{ij}(t)+v_{ij}(t+1) \tag{6}$$

Step 6: Repeat Steps 2-5 until the stopping criteria are satisfied, e.g., the best fitness is satisfied or the maximum iteration number is reached.

Step 7: Decode each particle and obtain L groups of network architecture and connection weights. Thus, we can construct L different neural networks. Train these networks with training samples until the stopping criteria are satisfied.

Step 8: Input testing samples and output L results.

3 ENSEMBLE STRATEGIES

Different weight combination will lead to different prediction results. How do we find optimal weight combination to obtain optimal ensemble results? This issue has been little reported in the literature. In this paper, we use a quadratic programming method to study this issue. $\{y_t\}$ and $\{\hat{y}_t(i)\}$ respectively denote an observed data series and an output series which are obtained from L corresponding network models, where $t=1,2,\cdots,n$ are independent samples and $i=1,2,\cdots,L$ are ensemble individuals. The resulting value of ensemble can be described as follows:

$$\tilde{y}_t=\sum_{i=1}^{L}w_i\hat{y}_t(i),\quad w_i\ge 0 \tag{7}$$

where w_i is the weight of the ith model, which meets $\sum_{i=1}^{L}w_i=1$. In addition, let e_t be the estimated value, namely,

$$e_t = y_t - \hat{y}_t = y_t - \sum_{i=1}^{L} w_i \hat{y}_t(i) = \sum_{i=1}^{L} w_i (y_t - \hat{y}_t(i)) \quad (8)$$

The minimum-error method is an approach that minimizes the forecasting errors when individual forecasts are incorporated into a single forecast. The linear programming combination prediction method is used to get the minimum weight according to the minimum value of $Q = \sum_{t=1}^{N} |e_t|$, namely,

$$\begin{cases} Min \quad Q = \sum_{t=1}^{N} |e_t| \\ s.t. \quad \sum_{i=1}^{L} w_i = 1, \ w_i \geq 0 \end{cases} \quad (9)$$

That is, Q is the objective function of linear programming. In order to eliminate the absolute sign of the objective function, assume that

$$u_t = \frac{|e_t| + e_t}{2}, \quad v_t = \frac{|e_t| - e_t}{2} \quad (10)$$

Clearly, $|e_t| = u_t + v_t$ $e_t = u_t - v_t$, and $Q = \sum_{t=1}^{N} (u_t + v_t)$. Based on the above specification we can construct the linear programming model as follows:

$$\begin{cases} Min \ Q = \sum_{t=1}^{N} (u_t + v_t) \\ s.t. \ w_1 + w_2 + \cdots + w_m = 1 \\ \sum_{i=1}^{L} w_i (y_1 - y_1(i)) - u_1 + v_1 = 0 \\ \sum_{i=1}^{L} w_i (y_2 - y_2(i)) - u_2 + v_2 = 0 \\ \quad \cdots\cdots \\ \sum_{i=1}^{L} w_i (y_n - y_n(i)) - u_n + v_n = 0 \\ w_i \geq 0, u_t \geq 0, v_t \geq 0. (i = 1, \cdots, L; t = 1, \cdots, N) \end{cases} \quad (11)$$

The minimum value of $Q = \sum_{t=1}^{N} |e_t|$ can be obtained from Ma [17] by the linear programming method.

4 DATA PRE-PROCESSING

In order to increase the accuracy of the method, preprocessing should be implemented in advance. There are many methods to pre-process data, such as data cleaning, data integration, data normalization, data reduction, etc. This paper adopts data cleaning to pre-process meteorological data. Particularly, we use the singular spectrum analysis (SSA) method [18] to reconstruct the original precipitation series, and then use mean generating function (MGF) [19] to construct mean generating function matrix. This matrix is denoted as a self variable, while the original precipitation series as a function. We use the partial least-square regression [20] to extract input factors and let the original data be real output. Thus, we can build a neural network ensemble prediction method based on the PSO algorithm.

5 EMPIRICAL DATA AND RESULTS ANALYSI S

This paper investigates the application of the PSO and neural network ensemble to predicting average July precipitation in Guangxi. The length of the dataset is 49 for 1957-2005 and that of training sample is 39 for 1957-1995 and that of testing sample is 10 for 1996-2005. We use the pre-processing method to clean the meteorological data and get eight variables. The partial least-square regression model is established as follows.

$$\begin{aligned} Y = 0.026F_1 + 0.273F_2 + 0.362F_3 + 0.427F_4 + 0.610F_5 \\ + 0.512F_6 + 0.189F_7 + 0.167F_8 - 368.356 \end{aligned} \quad (12)$$

In order to measure the effectiveness of the proposed method, the Simple Average Ensemble (SAE) model and PSO-based BP (PSO-BP) model are established by the eight variables.

In order to measure the effectiveness of the proposed method, we compare the results of PSO-BP ensemble. Four types of errors, such as the root mean squares error (RMSE), the mean absolute percentage error (MAPE), the mean absolute error (MAE) and Pearson Relative Coefficient (RC), which have been documented in the literature, are also used here. Interested readers are referred to Sollich et al. [8] for more details.

In the experiments, a simple 3-layer neural network is selected to handle the prediction problem. The number of neurons in the input layer is eight and the number of neurons in the output layer is one. PSO parameters are set as follows: the iteration times are 100, the population is 20, the minimum inertia weight is 0.1, and the maximum inertia weight is 0.9. BP parameters are set as follows: the learning rate is 0.9, the momentum factor is 0.7, the iteration times are 1000 and the global error is 0.001.

One can see that the maximum, average and minimum fitness and convergent speed are tending towards stability with the increase of iteration number in Fig.1. Therefore, network architecture and connection weights are in a near-optimal zone.

The architecture of the BP is 8-8-1 and illustrates a curve of error convergence for a PSO-BP network. Clearly, the iteration number in PSO-BP (Fig.3) is less than in BP (Fig.2). Therefore, the convergent speed is greatly improved as well as the convergence reliability. Table 1 shows a comparison of four types of errors in the 39 training samples. Table 2 lists a comparison of 10 testing samples.

Both BP and PSO-BP deal with data pre-processing. According to Table 1 and Fig.4, one can see that the learning results and generalization ability of PSO-BP are better than those of BP.

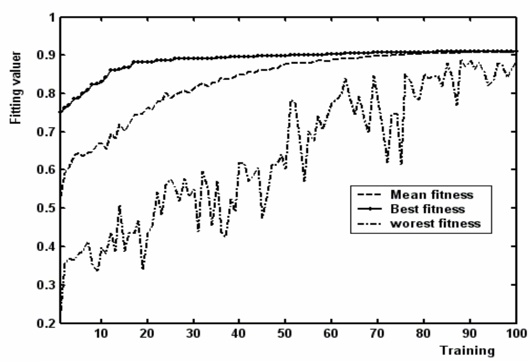

Fig.1 Curves of fitness in the training stage.

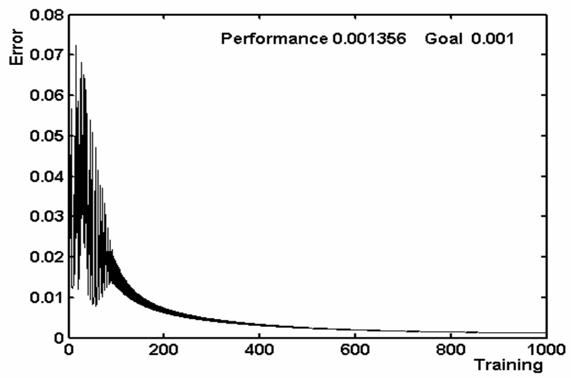

Fig.2 Curves of error convergence for a BP network.

Table 1 Comparisons of the four types of error.

Model	MAPE(/%)	RMSE	MAE	RC
PLS	17.49	41.80	32.44	88.46
SAE	16.80	36.17	28.96	89.23
PSO-BP	15.93	35.35	24.03	89.55

Fig.3 shows the fitting of training sample. PLS, SAE and PSO-BP deal with data pre-processing. From Table 1 and Fig.4, we can see that the learning results of PSO-BP are better than those of PLS and SAE. An important aspect to measure the performance of a method is to check its ability to generalize.

From Table 2, the differences between the different models are very significant. For example, for the rainfall testing samples in July, the MAE is 69.72 for the PLS model and 46.62 for the SAE model, while being 27.36 for the proposed PSO-BP model, which has obvious advantages over the other two models. The errors of the PSO-BP model are less than other models in forecasting June rainfall. The results imply that the

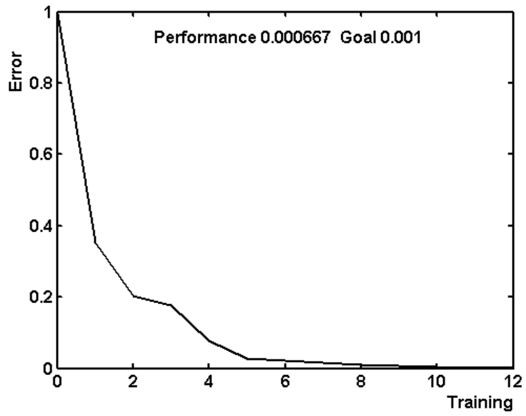

Fig.3 Curves of error converge nce for a PSO-BP network.

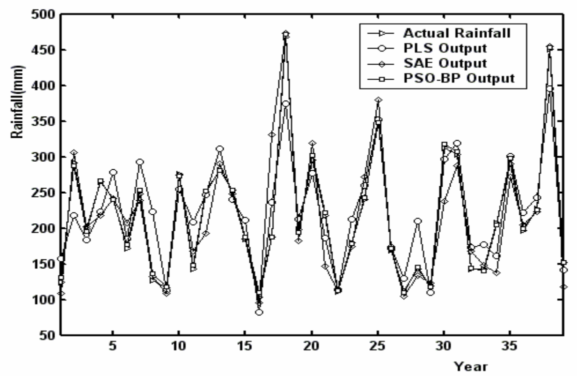

Fig.4 Fitting of training samples.

PSO-BP has significant forecasting ability with the same network input.

Furthermore, we use the same method to train the precipitation data for July of 1957 – 1995 and predict the precipitation of 1996 – 2005. The experimental results also show that the PSO-BP method converges more quickly and forecasts better than the other two methods.

6 CONCLUSIONS

The precipitation system is the most active in the climate system, and its interaction is also the most complex. Besides, because it is influenced by many changeable factors, accurate forecasts of rainfall are very difficult. In this paper, a novel evolutionary neural networks approach, based on the particle swarm optimization (PSO) algorithm, is presented for rainfall prediction. Data pre-processing is adopted for improving data quality. The results show that the method has the following characteristics:

(1) As the factors which involve the rainfall are complex, it is difficult to determine which factor is playing an important role in the fluctuations of the

climate system. Therefore, it is not an easy task to establish a forecasting model with high accuracy. This paper discusses the use of SSA-MGF method and PLS method to reduce the dimension for a number of technical indicators and to extract main information which affects the climate system so as to prevent disasters of dimensional importance. A direct look at the data and computer optimization is objectively better, and can reduce the dimension of the input matrix of neural networks. This will make the network structure smaller in scale and enhance the network in stability.

Table 2 Prediction results of testing samples.

Year	Real Data	PLS Model			SAE Model			PSO-BP Model		
		Prediction	MAE	MAPE%	Prediction	MAE	MAPE%	Prediction	MAE	MAPE%
1996	317.09	243.50	73.59	23.21	226.08	91.01	28.70	245.43	71.66	22.60
1997	287.4	330.15	42.75	14.88	301.74	14.34	4.99	323.76	36.36	12.65
1998	274.38	209.16	65.22	23.77	229.63	44.75	16.31	275.87	1.49	0.54
1999	296.43	259.01	37.42	12.62	277.17	19.26	6.50	299.07	2.64	0.89
2000	148.6	69.973	78.63	52.91	100.72	47.88	32.22	173.06	24.46	16.46
2001	339.69	256.35	83.34	24.53	291.43	48.26	14.21	377.08	37.39	11.00
2002	312.83	216.47	96.36	30.80	285.17	27.66	8.84	278.84	33.99	10.87
2003	121.68	95.745	25.94	21.31	91.20	30.48	25.05	149.31	27.63	22.71
2004	336.13	197.04	139.09	41.38	237.61	98.52	29.31	306.92	29.21	8.69
2005	130.55	80.19	50.36	38.58	86.467	44.08	33.77	139.28	8.73	6.69
Mean			69.27	28.40		46.62	19.99		27.36	11.31

(2) The training of neural network ensemble approach has been established by using particle swarm optimization (PSO), based on a neural network learning algorithm framework. In the learning process, PSO is used for solving the suitable network structure and connection weights, and the neural network algorithm is used again to train the learning samples. This method not only helps improve the performance of local search based on the guidance of gradient descent learning algorithm, but also is conducive to the global search characteristics of the PSO algorithm. The BP algorithm is further used to train these networks and determine the final prediction output by using the optimal combination of integration, which greatly improves the generalization ability of the system. PSO-BP is superior to PLS and SAE models in terms of forecasting accuracy under the same conditions of forecasting factor samples.

REFERENCES:

[1] HU Jian-lin, TU Song-bo, FENG Guang-liu. An exploration of heavy rain forecasting technique based on artificial neural networks [J]. J. Trop. Meteor., 2003, 19(4): 422-428.

[2] HSIEH W W. Nonlinear canonical correlation analysis of the tropical Pacific climate variability using Neural Network Approach [J]. J. Climate, 2001, 14(12): 2528-2539.

[3] GIORGIO C, GIOGIO G. Coupling Fuzzy Modeling and Neural Networks for River Flood Prediction [J]. IEEE Transactions on Systems, Man, and Cybernetic-Part C: Applications and Reviews, 2005, 25(3): 382-388.

[4] WU Jian-sheng, JIN Long, WANG Ling-zhi. The back propagation neural network meteorological forecast model research evolved and designed by genetic algorithms [J]. J. Trop. Meteor., 2006, 22(4): 411-416.

[5] HE Hui, JIN Long, QING Zhi-nian, et al. Downscaling forecast for the monthly precipitation over Guangxi based on the BP neural network model [J]. J. Trop. Meteor., 2007, 23(1): 72-77.

[6] JIN Long, KUANG Xue-yuan, et al. Study on the over-fitting of the artificial neural network forecasting model [J]. Acta Meteor. Sinica, 2004, 62(1): 62-69.

[7] HANSEN L K, SALAMON P. Neural network ensembles [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1990, 12(10): 993-1001.

[8] SOLLICH P, KROGH A. Learning with Ensembles: How Over-fitting can be useful [C]// Advances in Neural Information Processing Systems 8, Cambridge: MIT Press, 1996: 190-196.

[9] ZHOU Zhi-hua, CHEN Shi-fu. Neural network ensemble [J]. Chin. J. Comput., 2002, 25(1): 1-8.

[10] MAO J. A case study on bagging boosting and basic ensembles of neural networks for OCR [C]// Processing International Joint Conference on Neural Networks 1998. Anchorage: International Joint Conference on Neural Networks, 1998: 1828-1833.

[11] GUTTA S, WECHSLER H. Face recognition using hybrid classifier systems [C]// Proceeding International Joint Conference on Neural Networks 1996. Washington DC: 1996: 1017-1022.

[12] SOLLICH P, INTRATOR N. Classification of seismic signals by integrating ensembles of neural networks [J]. IEEE Trans. Signal Process., 1998, 46(5): 1194-1021.

[13] LI NING, ZHOU HUA-JIE, LING JIN-JIANG, et al.

Speculated lesion detection in digital mammogram based on artificial neural network ensemble [J]. Adv. Neural Networks ISNN, Springer Press, 2005, 3: 790-795.
[14] BONABEAU E, DORIGO M, THERAULAZ G. Inspiration for optimization from social insect behavior [J]. Nature, 2000, 406(6): 39-42.
[15] XIAOHUI H, EBERHART R. Multi-objective optimization using dynamic neighborhood particle swarm optimization [C]// Proceeding of Congress on Evolutionary Computation. Hawaii: Congress on Evolutionary Computation, 2002: 1677-1681.
[16] RIGET J, VESTERSTROM J S. A diversity-guided particle swarm optimizer-the ARPSO [R]. Technical Report 2002-02, Department of Computer Science, University of Aarhus, 2002.
[17] MA Zhen-hua. Operations Research and Optimizing Theory [M]. Beijing: Tsinghua Press. 1998, 235-425.
[18] VAUTARD. SSA: a toolkit for noisy chaotic signals [J]. Physical D, 1992, 58: 95-126.
[19] WEI Feng-ying, CHAO Hong-xing. The Mathematics Forecast Model and Application of Long Period Time [M]. Beijing: Meteorological Press, 1990.
[20] WANG Hui-weng. The Model and Application of Partial Least-Squares Regression [M]. Beijing: National Defenses Science and Technology University Press, 1999, 258-365.

Citation: WU Jian-sheng and JIN Long. Study on the meteorological prediction model using the learning algorithm of neural ensemble based on PSO algorithms. *J. Trop. Meteor.*, 2008, 15(1): 83-88.