

A Multiple Evolutionary Neural Network Classifier Based on Niche Genetic Algorithm*

Jiang Wu^{1,2}, Changjie Tang¹, Taiyong Li^{1,2}, Yue Jiang¹, Shangyu Ye¹, Xun Lin^{1,2}, Jie Zuo¹

¹School of Computer Science,
Sichuan University,
Chengdu, China,
tangchangjie@cs.scu.edu.cn,

²School of Economic Information Engineering,
Southwestern University of Finance and Economics,
Chengdu, China,
wuj_t@swufe.edu.cn

Abstract

Classification is important in data mining. In this paper, a Multiple Evolutionary Neural Network Classifier Based on Niche Genetic Algorithm (MNC-NG) is presented, which establishes classifiers by a group of three-layer feed-forward neural networks with high accuracy and good diversity. The neural networks are trained by niche genetic algorithm based on clustering. The class label of the identifying data can first be evaluated by each neural network, and the final classification result is obtained according to the dynamic voting rule. Experimental results on 6 data sets show that MNC-NG increases the predictive accuracy by 5.6%, 5.5% and 8.5% respectively compared with BP, GA and LM training methods and by 6.0%, 6.1% and 4.0% compared with Naïve Bayesian classifier, C4.5 and SVM.

1. Introduction

Classification is very important in data mining. The popular classification algorithms include decision tree [1], Bayesian classifier [2], support vector machine [3], neural network [4], etc.

NN has been widely used in classification because of its noise toleration and modeling ability in complex problems. In order to improve the accuracy of classification, we propose a Multiple Evolutionary Neural Network Classifier Based on Niche Genetic Algorithm (MNC-NG). The algorithm establishes sub-classifiers by a group of neural networks. These neural networks are trained by niche genetic algorithm simultaneously. The class label of the identifying data

can first be evaluated by each neural network, and the final classification result is obtained according to the dynamic voting rule. Experimental results show that MNC-NG is more effective in improving the predictive accuracy than other traditional algorithms.

The rest of this paper is organized as follows. Section 2 introduces the framework of MNC-NG. Section 3 analyzes the principle of classifier ensemble. Section 4 describes the design of MNC-NG. Section 5 provides experimental evaluations. Finally Section 6 concludes this paper.

2. Framework of MNC-NG

The framework of MNC-NG is illustrated in Figure 1.

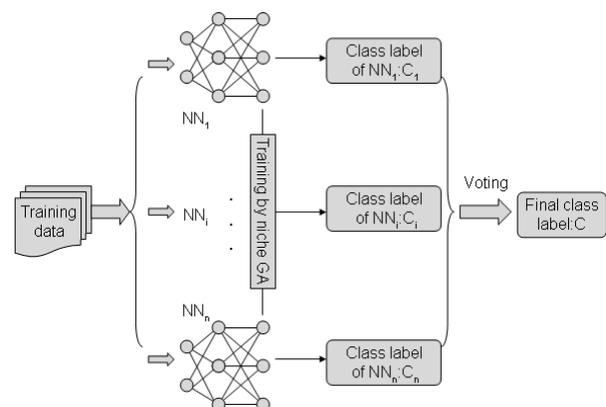


Figure 1. Framework of MNC-NG

3. Principle of classifier ensemble

Extensive experiments show that the unselective combination of all sub-classifiers does not always improve the ensemble predictive accuracy. Krogh and Vedelsby [5] put forward a formula of generalization

*This work was supported by the National Science Foundation of China under Grant No. 60773169 and National Key Project of Scientific and Technical Supporting Programs Funded by Ministry of Science & Technology of China under Grant No. 2006BAI05A01.

error in neural network ensembles. In this section, we analyze the relationship between ensemble performance and sub-classifiers, and discuss how to effectively combine sub-classifiers.

Consider a two-class classification problem. Assume that dataset consists of N instances (x_p, y_p) , where x_p is an attribute set, y_p is a class label, the value of y_p is either 0 or 1, $p = 1, 2, \dots, N$, S is the ensemble of M sub-classifiers (c_1, c_2, \dots, c_M) , each weight of sub-classifier is w_i ($i = 1, 2, \dots, M$), where $w_i \geq 0$ and $w_1 + w_2 + \dots + w_M = 1$ and $f_i(x_p)$ is the output (class label) of the i th sub-classifier for the input x_p , so the value of $f_i(x_p)$ is either 0 or 1.

The sum of $f_i(x_p)$ is called the ensemble output of M sub-classifiers for the input x_p , denoted as $f(x_p)$:

$$f(x_p) = \sum_{i=1}^M w_i f_i(x_p) \quad (1)$$

If the ensemble output $f(x_p) < 0.5$, then x_p belongs to class 0, otherwise it belongs to class 1:

$$f(x_p) = \begin{cases} 0 & f(x_p) < 0.5 \\ 1 & otherwise \end{cases} \quad (2)$$

E is the generalization error of ensemble classifier on the entire dataset:

$$E = \sum_{p=1}^N (y_p - f(x_p))^2 \quad (3)$$

\bar{E} is the weighted average of the generalization error of sub-classifiers on the entire dataset:

$$\bar{E} = \sum_{p=1}^N \sum_{i=1}^M w_i (y_p - f_i(x_p))^2 \quad (4)$$

\bar{D} is the difference between each sub-classifier and the ensemble output on the entire dataset:

$$\bar{D} = \sum_{p=1}^N \sum_{i=1}^M w_i (f_i(x_p) - f(x_p))^2 \quad (5)$$

Proposition 1 The generalization error of ensemble classifier satisfies: $E = \bar{E} - \bar{D}$ (6)

The analysis above shows that a good ensemble is one where the sub-classifiers in the ensemble are both accurate and tend to have diversity in their predictions. So, if we want to improve the generalization ability of ensemble classifier, we should improve the accuracy of sub-classifiers and at the same time boost their diversity.

4. Ensemble NN-classifier based on niche GA

The key steps of MNC-NG are as follows:

- (1) Construction. Niche GA is used to evolve c sub-classifiers satisfied with good diversity and predefined accuracy.
- (2) Combination. Sub-classifiers are combined with Dynamic Voting (DV).

4.1. Sub-classifier construction based on niche GA

Traditional GA always converges to a single best classifier and thus it is not appropriate for multimodal optimization. As improvement, we present a novel niche GA algorithm. In each GA iteration, the population is partitioned into several classes by clustering. Then the fitness of best individual in each cluster is unchanged and the other individuals in the population are penalized by a reduction in their fitness. This causes population diversity pressure that allows the population to maintain individuals at local optima.

4.1.1. Chromosome encoding. According to the structure of neural network, the chromosome encoding includes the weights of links and the biases of nodes. The length of a chromosome is computed as follows:

$$l = (n_i + n_o)n_h + n_h + n_o \quad (7)$$

Where n_i is the number of nodes of input layer; n_o is the number of nodes of output layer; n_h is the number of nodes of hidden layer. The encoding is as follows: the weights of links between input layer and hidden layer \rightarrow the weights of links between hidden layer and output layer \rightarrow the biases of hidden nodes \rightarrow the biases of output nodes.

4.1.2. Fitness function. Fitness function is used to measure the accuracy and generalization ability of the individual (neural network). The fitness function is defined as below:

$$f_i = \frac{1}{\left(\sum_{j=1}^n (C_{i,j} - T_j) / n\right) + 1} \quad (8)$$

Where $C_{i,j} - T_j = \begin{cases} 0 & C_{i,j} = T_j \\ 1 & otherwise \end{cases}$ (9)

$C_{i,j}$ is the predicted classification output (class label) of the i th individual on the j th case. T_j is the actual class label of the j th case. n is the number of cases. Obviously, when $C_{i,j} = T_j$ ($j = 1, 2, \dots, n$), the value of f_i reaches the maximum value 1.

4.1.3. Genetic operators.

(1) Crossover operator

Crossover operator is as follows: for two individuals $f_1 = [w_1, \dots, w_n, v_1, \dots, v_m]$, $f_2 = [w'_1, \dots, w'_n, v'_1, \dots, v'_m]$, MNC-NG decides if they should be recombined or not according to crossover rate P_c .

Sub-classifiers are combined with Dynamic Voting (DV).

Suppose that the number of sub-classifiers in ensemble is c and the output of each sub-classifier is $f_j(x_p)$, where $j = 1, 2, \dots, c, p = 1, 2, \dots, n$. The output of the ensemble classifier is:

$$y_{ensemble} = \sum_{j=1}^c w_j f_j(x_p) \quad (17)$$

The algorithm MNC-NG is described as follows:

Algorithm 1. MNC-NG

Input: training dataset, ensemble size: c

Output: ensemble classifier

Begin

1. S = Initial Population;
 2. m = MaxGeneration; // iteration generations
 3. evaluate(S); // fitness evaluation
 4. preserve(*Best c NN-classifiers*);
 5. repeat
 6. genetic-operation(S);
 7. S = combine(S , *Best c NN-classifiers*);
// combine current population with the best c individuals in last generation
 8. SC(S); // modified Subtractive Clustering
 9. penalize(S); //fitness penalty
 10. evaluate(S);
 11. preserve(*Best c NN-classifiers*);
 12. $m = m - 1$;
 13. until ($m = 0$) || (All *Best c NN-classifiers* reach the predefined accuracy)
 14. *ensemble_classifier* = DV(*Best c NN-classifiers*);
 15. return *ensemble_classifier*
- End.

5. Experiments and performance evaluation

5.1. Experimental dataset and environment

We evaluate the performance of MNC-NG on 6 data sets that are from the UCI repository [10]. The main characteristics of the 6 data sets are shown in Table 1.

Table 1. Summary of 6 UCI benchmark data sets used in the experiments

Dataset	Records	Attributes	Classes	Hiddens
Credit	690	15	2	10
Diabetes	768	8	2	5
Glass	214	9	6	5
House-votes-84	435	16	2	10
Iris	150	4	3	5
Hepatitis	155	19	2	10

The parameters of MNC-NG are shown in Table 2.

Table 2. Parameters for the experiments

Parameter	Value
Subpopulation size (N)	400
Number of generation (M)	1000
Mutation rate (P_m)	0.044
Crossover rate (P_c)	0.7

All results of MNC-NG are averaged over 4-fold cross-validation experiments.

5.2. Compared with BP, GA and LM

In this section, we compare MNC-NG with the traditional training methods of single neural network: BP, GA, and LM in terms of classification accuracy.

Experimental results of various models are summarized in Table 3.

Table 3. Estimation of average predictive accuracy rate on the 6 data sets

Dataset	BP	GA	LM	MNC-NG
Credit	81.9	81.9	78.1	86.6
Diabetes	71.1	71	69.2	77.2
Glass	63.2	63	60.3	68.9
House-votes-84	89.1	90.3	86.5	95.6
Iris	90.9	90.1	88.9	96.7
Hepatitis	78.9	79.1	74.3	83.4

In terms of predictive accuracy, MNC-NG shows significantly superior performance to the traditional training methods of single neural network: BP, GA, and LM in all data sets.

Figure 3 compares the average predictive accuracies of various models on the 6 data sets. The result shows that compared with BP, GA, and LM, MNC-NG increases the predictive accuracy by 5.6%, 5.5%, and 8.5% respectively.

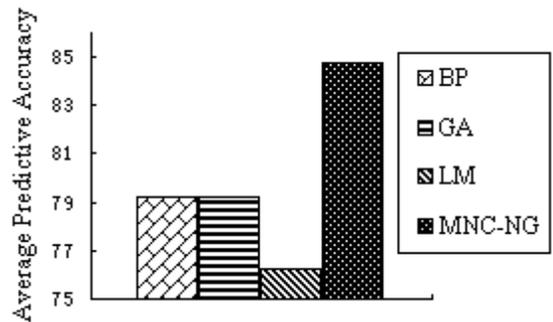


Figure 3. Comparison of average predictive accuracies of four models

5.3. Compared with NB, C4.5 and SVM

In this section, we compare MNC-NG with the traditional classification algorithm: NB, C4.5 and SVM.

Experimental results of various models are summarized in Table 4. In terms of predictive accuracy, MNC-NG also shows significantly superior performance to the traditional classification algorithms: NB, C4.5 and SVM in all data sets. MNC-NG increases the predictive accuracy by 6.0%, 6.1% and 4.0% compared with NB, C4.5 and SVM respectively.

Table 4. Estimation of average predictive accuracy rate on the 6 data sets

Dataset	NB	C4.5	SVM	MNC-NG
Credit	81.9	80.2	82.1	86.6
Diabetes	69.2	71.3	72.8	77.2
Glass	63.2	63.9	65	68.9
House-otes-84	89.3	91.3	92.1	95.6
Iris	90.1	90.1	92.9	96.7
Hepatitis	78.9	75.3	79.5	83.4

5.4. Ensemble size

In this section, we use MNC-NG to examine ensemble characteristics and provide data analysis on how to construct an optimal ensemble size.

Figure 4 shows the relationship between the average predictive accuracy and the ensemble size on the 6 data sets in MNC-NG, where the number of classifiers ranges from 1 to 20.

Form Figure 4, we can find that the predictive accuracy steadily improves up to an ensemble size of 9. The performance improvement is unobvious when the number of sub-classifiers is more than 9.

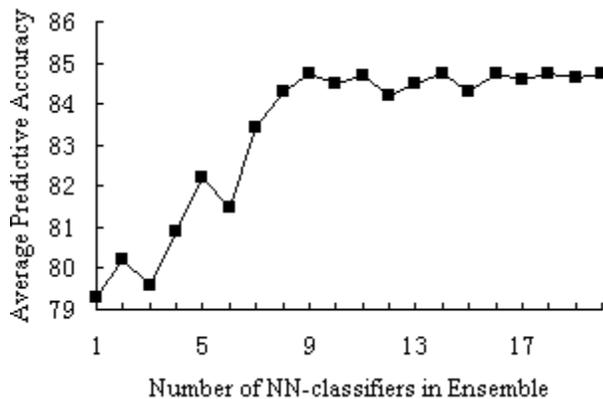


Figure 4 Relationship between ensemble size and average predictive accuracy

6. Conclusion

In this paper, we propose a Multiple Evolutionary Neural Network Classifier Based on Niche Genetic Algorithm (MNC-NG). The neural networks are trained by niche genetic algorithm based on clustering. The class label of the identifying data can first be evaluated by each neural network, and the final classification result is obtained according to the dynamic voting rule. Experimental results show that MNC-NG is more effective in improving the predictive accuracy than traditional algorithms.

7. References

- [1] Quinlan JR. C4.5: Programs for Machine Learning. *San Francisco: Morgan Kaufmann Publishers*, pp.1-25, 1993.
- [2] Friedman N, Geiger D, Goldszmidt M. Bayesian network classifiers. *Machine Learning*, 29(2-3),pp.131-163, 1997.
- [3] M.Hearst.Trends & Controversies: Support Vector Machines, *IEEE Intelligent Systems*, 13(4), pp.18-28,1998.
- [4] Arulampalam G, Bouzerdoum A. A generalized feedforward neural network architecture for classification and regression. *Neural Networks*, 16(5-6), pp561-568, 2003.
- [5] Krogh A , Vedelsby J . Neural Network Ensembles, Cross Validation, and Active Learning. *Advance in Neural Information Processing Systems 7*, pp. 231-238, 1995.
- [6] J. Cohen, A Coefficient of Agreement for Nominal Scales, *Educational and Psychological Measurement* 20, pp. 37-46, 1960.
- [7] S.L. Chiu. Fuzzy Model Identification Based on Cluster Estimation. *Journal of Intelligent and Fuzzy System* 2(3), pp. 267-278, 1994.
- [8] S.L. Chiu. A Cluster Estimation Method with Extension to Fuzzy Model Identification. In *Proceedings of the 3rd IEEE Conference on Fuzzy Systems*, pp. 1240-1245, 1994.
- [9] Liu Yu, Qin Zheng, Lu Jiang, and Shi Zhewen, Multimodal Particle Swarm Optimization for Neural Network Ensemble, *Journal of Computer Research and Development* 42(9), pp. 1519-1526, 2005.
- [10] Blake, C.L., Merz, C.J. (1998) UCI repository of machine learning databases [http://www.ics.uci.edu/~mllearn/MLRepository.html], university of California, Irvine, Department of Information and Computer Sciences.