ESANN'2006 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2006, d-side publi., ISBN 2-930307-06-4.

Diversity creation in local search for the evolution of neural network ensembles

Pete Duell[1], Iris Fermin[2] and Xin Yao[1] *

1- School of Computer Science
The University of Birmingham, UK

2- Thales Research & Technology (UK) Ltd., Reading, UK

Abstract. The EENCL algorithm [1] automatically designs neural network ensembles for classification, combining global evolution with local search based on gradient descent. Two mechanisms encourage diversity: Negative Correlation Learning (NCL) and implicit fitness sharing. This paper analyses EENCL, finding that NCL is not an essential component of the algorithm, while implicit fitness sharing is. Furthermore, we find that a local search based on independent training is equally effective in both accuracy and diversity. We propose that NCL is unnecessary in EENCL for the tested datasets, and that complementary diversity in local search and global evolution may lead to better ensembles.

1 Introduction

One approach to the design of accurate and diverse ensembles is the EENCL algorithm (Evolutionary Ensembles with Negative Correlation Learning) [1]. Unlike many other ensemble methods, the individual networks are trained in parallel, rather than independently or sequentially. Individual networks learn by Negative Correlation Learning (NCL) [2] and evolutionary learning [3]. Diversity amongst the final population is encouraged by the negative correlation of the networks' outputs and through speciation by implicit fitness sharing [4][5]. Both accuracy and diversity are important for the creation of good ensembles.

EENCL [1] proved successful on some problems. However, little work has been carried out to analyse why EENCL is effective. Specifically we are interested in what contribution NCL makes to the performance of the algorithm, since it introduces an extra parameter and additional complexity. This paper uses additional datasets to analyse how the two learning mechanisms of global evolution and local search interact. Surprisingly, we find that NCL is not an essential component of EENCL for the datasets tested, and that a comparable performance can be achieved with a much simpler local search technique: Backpropagation. Our experiments show that by replacing NCL in EENCL with Backpropagation, we can achieve comparable classification accuracies, and also produce ensembles that are just as diverse in terms of the joint correct sets of the networks and also in terms of correlation of outputs. We suggest that the

*This work is partially funded by EPSRC and Thales Research & Technology (UK) Ltd.

diversity creation in local search provided by NCL based on correlation is unnecessary for these datasets. The fitness sharing stage of the algorithm enforces diversity in terms of a coverage of training patterns, which is not complementary to the diversity creation in the evolutionary stage. Most significantly, it may also be possible to create better ensembles if we design the local search and global evolution to work effectively together.

NCL was first proposed by Liu and Yao [6] as a means to generate an ensemble of neural networks, whose outputs would be negatively correlated. The networks are trained simultaneously and a penalty term is included during training to encourage the formation of decorrelated networks. Backpropagation is used to train the network but the error to be minimised is now [7]:

$$E_i(n) = \frac{1}{N}\sum_{n=1}^{N}\left[\frac{1}{2}\left(d(n) - F_i(n)\right)^2 + \lambda p_i(n)\right], \tag{1}$$

where N is number of training patterns, $E_i(n)$ is the error of network i on training pattern n, $d(n)$ is the target output, $F_i(n)$ is the output of network i, λ is the strength of penalty parameter and $p_i(n)$ is the penalty term, defined as:

$$p_i(n) = \left(F_i(n) - d(n)\right)\sum_{j\neq i}\left(F_j(n) - d(n)\right). \tag{2}$$

Evolutionary learning can benefit from combining a final population into an ensemble rather than selecting the 'best' individual [8]. Populations tend to converge towards similar solutions, as successful genetic material is re-used to form offspring. In contrast, ensembles are effective when their members are both accurate and diverse [1][2][7][9]. Speciation through fitness sharing creates a diverse set of solutions to exploit different niches in the fitness landscape [4][5]. Raw fitness scores are *shared* amongst *similar* individuals. The definition of *similarity* and the mechanism for *sharing* varies, here similar individuals are those which make the same correct classifications, and the reward for the correct classification is shared equally amongst all those individuals judged to be similar.

The rest of this paper is organised as follows: section 2 describes the EENCL algorithm; section 3 describes the experiments with local search techniques and our results; section 4 concludes and indicates possible further work.

2 EENCL Algorithm

EENCL uses partial training with the Negative Correlation Learning (NCL) algorithm alongside an evolutionary process to form a population of neural networks suitable for combination into an ensemble [1]. EENCL exploits two mechanisms to ensure that the final networks are both accurate and diverse. Firstly, NCL encourages the negative correlation of the outputs of the networks in the population. Secondly the fitness of individuals in the population is evaluated with implicit fitness sharing [10] based on the coverage of patterns in the training set. Either the entire final population is used to form the ensemble or some subset. In all of the experiments in this paper the entire final population is used. The outputs of the individual networks are combined by either a simple average, majority vote or a winner-takes-all procedure. The algorithm proceeds according to the following steps [1]:

1. An initial population (M) is trained for a small number of epochs (n_e) according to the NCL algorithm.
2. n_b parents are randomly selected from M, according to a uniform distribution.
3. n_b parents' weights are mutated, by Gaussian mutation N(0,1) to form n_b offspring.
4. n_b offspring are added to population and trained for e epochs, holding the weights of M fixed.
5. Fitness of all $M + n_b$ individuals evaluated using a fitness sharing scheme based on the coverage of training patterns.
6. Fittest M from current population $M + n_b$ selected for next generation.
7. If total number of generations reached, go to step 8, otherwise go to 2.
8. Combine the population into an ensemble.

3 Experiments with local search technique

Liu and Yao [1] showed that the combination of evolution with fitness-sharing and NCL could produce competitive results in comparison to a number of other classification techniques. It is less clear to what extent these different learning mechanisms are responsible for this performance. We sought to establish whether NCL was a necessary component of the EENCL algorithm, or if similar results could be obtained by means of an alternative (and less complex) local search.

We applied the EENCL algorithm to four datasets (Australian Credit Card, Pima Indian Diabetes, Heart Disease and Wisconsin Breast Cancer) over 30 independent runs. All of these datasets are available by anonymous ftp from the UCI Machine Learning Repository at ics.uci.edu (128.195.1.1) in /pub/machine-learning-databases. Each set was equally divided into a training, validation and testing set. The validation set is not used in these experiments. Each network learns the same training set. The results of 30 runs on the test set are averaged to approximate the generalisation error of the resulting ensembles. Three different combination schemes are tested: a simple average, majority vote and winner-takes-all. In each case the classification accuracy is shown in table 1.

Here we also define a new algorithm, EE-Backprop. This algorithm is identical to EENCL, except that NCL is no longer used as the local search technique in steps 1 and 4 of the algorithm. In EE-Backprop, the penalty strength, λ, is set to 0. As can be seen in equation 1, the right-hand-side now disappears and we are left with a conventional mean-square-error. Hence, EENCL with a penalty of 0, is equivalent to using conventional Backpropagation for the local search. In all other ways, EE-Backprop is identical to EENCL and is also used with the same set of parameters, as detailed in section 3.1.

3.1 Experimental Setup

The EENCL algorithm was implemented as described by Liu and Yao [1]. The initial population is a set of randomly initialised MLP's with full connection and

Dataset	Algorithm	Average	Majority	Winner
Australian	EENCL	0.876	0.873	0.867
	EE-Backprop	0.873	0.870	0.868
Breast Cancer	EENCL	0.972	0.972	0.973
	EE-Backprop	0.972	0.972	0.972
Diabetes	EENCL	0.766	0.764	0.762
	EE-Backprop	0.763	0.764	0.757
Heart Disease	EENCL	0.789	0.785	0.772
	EE-Backprop	0.793	0.793	0.772

Table 1: Testing classification accuracies for the EENCL and EE-Backprop algorithms over four datasets, with three combination schemes. No statistically significant difference between the algorithms is found, using a Student t-test with a confidence of 1%.

a single hidden layer. Output nodes are encoded using a 1-of-c scheme and all nodes are sigmoidal logistic. The node with the highest output is considered to be the classification of the network. The initial population M is set to 25, and the number of offspring per generation, n_b, is 2. The population is allowed to evolve for 200 generations. Both the initial population and offspring are trained for 5 epochs, n_e. The learning rate is set to 0.1 and the networks are trained using mean-square-error. For EENCL the NCL penalty term λ is set at 0.75. For EE-Backprop λ is 0. This is identical to the experimental setup described by Liu, arrived at in the original study after limited experimentation [1].

3.2 Classification accuracies for EENCL and EE-Backprop

The classification accuracies for EENCL and EE-Backprop are shown in table 1. Interestingly, no statistically significant difference between the techniques could be found, using a Student t-test with a 1% confidence interval. EENCL was found to be most effective with the winner-takes-all combination scheme over the Australian and Diabetes datasets in Liu's experiments [1]. Unlike Liu, we do not observe that EENCL is more suited to the winner-takes-all combination scheme for all problems (only in the Diabetes problem is the accuracy highest). This may be because winner-takes-all works well with specialised networks, and our ensembles have been trained on a smaller proportion of the datasets than Liu's, and therefore had less opportunity to specialise.

EE-Backprop is a simpler algorithm: each network learns independently and does not require the setting of a penalty strength parameter. However, it is just as effective as EENCL in terms of classification accuracy, so it is difficult to justify the added complexity of NCL. NCL alone is able to significantly improve on Backpropagation and many other algorithms [7] but it appears that when used in conjunction with fitness sharing in an evolved ensemble, NCL's effectiveness is no longer apparent.

Dataset	**Algorithm**	Ω_i	σ	$\Omega_{\forall i}$	σ
Australian	EENCL	195.16	5.90	138.80	15.52
	EE-Backprop	195.11	5.77	139.20	11.16
Breast Cancer	EENCL	225.98	1.40	219.07	3.69
	EE-Backprop	225.98	1.59	218.10	4.77
Diabetes	EENCL	183.07	6.91	86.33	14.49
	EE-Backprop	182.32	8.31	81.57	25.12
Heart Disease	EENCL	66.75	11.01	29.67	3.75
	EE-Backprop	66.73	4.40	26.10	4.36

Table 2: Correct response sets for EENCL and EE-Backprop across 4 datsets. Ω_i is the mean individual correct response set over 30 runs. $\Omega_{\forall i}$ is the mean joint correct repsonse. σ is the standard deviation. No significant difference was found using a Student t-tests with a 1% confidence interval, except for $\Omega_{\forall i}$ for the Heart Disease problem, where EE-Backprop is significantly lower.

3.3 Diversity analysis for EENCL and EE-Backprop

One method to analyse diversity in the final ensemble is to compare their correct response sets [7]. Here we define the correct response set of a network i, Ω_i, as the set of examples it correctly classifies. We also define the *joint* correct response set between networks i and j, $\Omega_{i,j}$, as the set of examples that both networks classify correctly. Table 2 shows the mean individual correct response sets, Ω_i, and the mean joint correct response sets for all networks in the final population, $\Omega_{\forall i}$, over 30 runs for both EENCL and EE-Backprop. Liu found that NCL produced significantly lower joint correct response sets to independent training with Backpropagation [7]. Our results show that this is not the case for EENCL and EE-Backprop, where no significant difference could be found in $\Omega_{\forall i}$ except for the Heart Disease set, where $\Omega_{\forall i}$ is significantly *lower* for EE-Backprop.

We also measured the average pair-wise correlation between the networks in each ensemble. For the Australian Credit Card and Breast Cancer datasets, EENCL reduced correlation to a greater degree than EE-Backprop as expected. This was expected because EENCL explicitly seeks to minimise correlation, whereas EE-Backprop seeks only to minimise error during local search. The other datasets however provided counter-intuitive results, with EENCL creating more highly correlated ensembles. One possible explanation for these unexpected results is that they are a consequence of using two different ways of encouraging diversity. EENCL can only effectively de-correlate the networks if the offspring that are trained are fit enough to survive, otherwise the offspring will be discarded. In EENCL fitness is awarded according to accuracy and also coverage of training patterns, which is not necessarily the same as correlation of outputs. NCL warps the mean-square-error landscape, and then descends this new landscape [9]. Fitness sharing however, operates on a different landscape: a warping of the classification accuracy landscape, according to the coverage of training patterns amongst the ensemble.

ESANN'2006 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2006, d-side publi., ISBN 2-930307-06-4.

4 Conclusions and Future Work

Our analysis shows that NCL is not integral to the success of the EENCL algorithm for these datasets, whereas implicit fitness sharing is. We have demonstrated that EE-Backprop produces comparable classification accuracies. Likewise we find that both techniques produce similar joint response sets, showing that EENCL is no more effective in producing specialisation within the ensemble. We obtained surprising results which show that on some problems, EE-Backprop was able to produce lower correlation amongst the ensemble than EENCL, but that this did not necessarily translate into improved classification accuracy. We hypothesise that the explanation for how a method which explicitly seeks to reduce correlation such as EENCL can produce higher correlated networks than EE-Backprop, which only implicitly reduces correlation, is to be found in the different and not necessarily complementary representations of diversity in EENCL. EENCL also requires the setting of a penalty strength parameter, which does not significantly improve performance over EE-Backprop. We propose that better results could be achieved if both local search and global evolution had complementary implementations of diversity, (e.g. both based on the coverage of training patterns or on correlation of outputs). Further experiments with complementary diversity are necessary to determine if such an approach will lead to better ensembles. Effective local search becomes increasingly important as the dataset grows, so larger problems may also aid the comparison.

References

[1] Y. Liu, X. Yao, and T. Higuchi. Evolutionary ensembles with negative correlation learning. *IEEE Transactions on Evolutionary Computation*, 4(4):380–387, Nov 2000.

[2] Y. Liu. *Negative Correlation Learning and Evolutionary Neural Network Ensembles.* PhD thesis, University College, The University of New South Wales, Australian Defence Force Academy, Canberra, Australia, 1998.

[3] X. Yao. Evolving artificial neural networks. In *Proceedings of IEEE*, volume 87:9, pages 1423–1447, Sep 1999.

[4] J. H. Holland. *Adaptation in Natural and Artificial Systems.* Ann Arbor, MI:University of Michigan Press, 1975.

[5] D. E. Goldberg and J. Richardson. Genetic algorithms with sharing for multimodal function optimisation. In J. J. Grefensette, editor, *Proceedings of 2nd International Conference on Genetic Algorithms*, pages 41–49. Hillsdale, NJ: Lawrence Earlbaum, 1987.

[6] Y. Liu and X. Yao. Negatively correlated neural networks can produce best ensembles. *Australian Journal of Intelligent Information Processing Systems*, 4(3/4):176–185, 1997.

[7] Y. Liu and X. Yao. Simultaneous training of negatively correlated neural networks in an ensemble. *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics*, 29(6):716–725, Dec 1999.

[8] X. Yao and Y. Liu. Making use of population information in evolutionary artificial neural networks. *IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics*, 28(3):417–425, Jun 1998.

[9] G. Brown. *Diversity in Neural Network Ensembles.* PhD thesis, School of Computer Science, University of Birmingham, 2003.

[10] D. E. Goldberg J. Horn and K. Deb. Implicit niching in a learning classifier system: Nature's way. *Evolutionary Computation*, 2(1):37–66, 1994.