COMBINING NEUROEVOLUTIONARY AND GRADIENT LEARNING FOR SOLVING CLASSIFICATION PROBLEMS*

Y. R. Tsoy
Tomsk Polytechnic University

Tomsk State University of Control Systems and Radioelectronics

The problem of input features transform is one of the central problems in training a classifier. There's a lot of different methods have been proposed. Some of them consider linear and non-linear transform of the features, including decomposition of the object description into some basis [1-4], while other try to remove insignificant features [5, 6]. Large review on feature space dimensionality reduction and feature selection can be found in [7]. But due to large variety of real-world problems no approach can be considered a universal. This fact can be used to encourage application of methods, which can adapt "on-the-fly" to the problem's properties, for such a transform of input features, which could ease a subsequent gradient learning.

To address this idea a novel combined method for training of artificial neural networks (ANN) is proposed in this paper. In this method ANN is divided in two parts (fig. 1): the first part (ANN-1) is trained using neuroevolutionary approach, while the second (ANN-2) is trained with use of traditional gradient method. Note that the ANN should not obligatory be multilayered since neuroevolution assumes in general evolutionary emergence of ANNs with irregular structure as well.

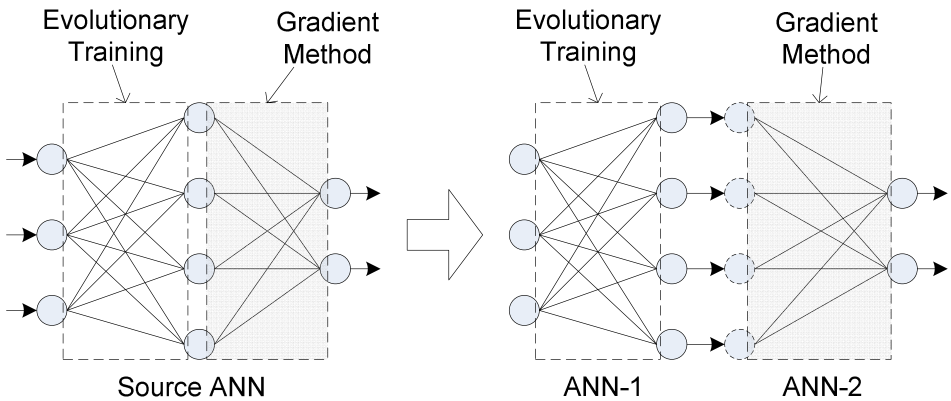

Fig. 1. General scheme for dividing of ANN for combined training. Input nodes for ANN-2 have activation $y(x) = x$.

In this paper ANN-1 and ANN-2 have no hidden layers. Since ANN-1 is not connected with the ANN output the objective function for ANN-1 training should guide the evolutionary search in some indirect manner. We'll test combined training with 3 objective functions for ANN-1.

1. The first objective function f_1 considers is maximization of

$$f = \frac{2\sum_{i=1}^{N}\sum_{j>i}^{N} R_{Y^i,Y^j}}{N(N-1)},, \qquad (1)$$

where $Y^i = \{y_k^i\}, i = 1,...,\alpha n_I, k = 1,...,N$ – vector of *i*-th output signal of ANN-1, obtained during application of all training sets.; $\overline{Y}^i$ and σ_{Y^i} – are mean and deviation of Y^i respectively; R_{Y^i,Y^j} – Pearson's correlation coefficient for vectors $\overline{Y}^i$ and $\overline{Y}^j$.

2. The second objective function f_2 considers minimization of (1).
3. The third objective function

* The research is supported by the Russian Foundation for Basic Researches grant (project no. 09-08-00309-a)

$$f_3 = \max\{e_j, j = 1,2,3\}, \qquad f_3 \to \min,$$

where e_j – training error of ANN-2 trained using RPROP algorithm for 50 epochs.

RPROP implementation is taken from the Encog[1] library, while the combined training is made using Mental Alchemy[2] library.

Overview of the test classification accuracy on some problems from the Proben1 [8] test set for combined training and comparison with results from [8] is given in the table 1. The best statistically significant results are given in bold.

Table 1. Mean classification accuracy on the test set for different algorithms. Deviation is given inside round brackets.

Problem	f_1	f_2	f_3	RPROP [8]
cancer1	**1,03 (0,53)**	1,26 (0,45)	2,07 (0,73)	1,38 (0,49)
card1	11,28 (0,30)	12,21 (0,39)	**10,17 (0,63)**	14,05 (1,03)
diabetes1	22,55 (0,95)	22,24 (0,65)	**21,51 (0,43)**	24,10 (1,91)
glass1	**25,66 (0,97)**	27,17 (0,97)	26,41 (0)	32,70 (5,34)
heart1	**17,04 (0,34)**	18,43 (0,42)	17,74 (1,02)	19,72 (0,96)
horse1	**24,51 (0,53)**	28,35 (2,25)	34,40 (3,84)	29,19 (2,62)

In all test problems the proposed combined training method allowed obtaining better results that traditional ANN training.

Note that it's often when 'indirect' training with objective functions f_1 and f_2, which is not pushed towards local best result for ANN-2, allows obtaining better generalization results than more 'greedy' training of ANN-1 using f_3.

Future research is aimed towards revealing of possibility to use combination of f_1, f_2 and f_3 for ANN-1 training and committee of ANNs trained with different objective functions via bagging scheme. Also interesting results can be obtained using evolutionary algorithm for simultaneous search of structure and connection weights of ANN-1.

References

1. Jaeger H., Haas H. Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication // Science. – 2004. – Vol. 304 (5667). – P. 78 – 80
2. LeCun Y., Bottou L., Bengio Y., Haffner P. Gradient-based learning applied to document recognition // Proceedings of the IEEE. – 1998. – P. 2278-2324.
3. Kecman V. Learning and Soft Computing – Support Vector Machines, Neural Networks, Fuzzy Logic Systems. – The MIT Press, Cambridge, MA, 2001.
4. Principal Manifolds for Data Visualisation and Dimension Reduction / A. Gorban, B. Kegl, D. Wunsch, A. Zinovyev (Eds.). – LNCSE 58. – Springer, Berlin – Heidelberg – New York, 2007.
5. Vafaie H., De Jong K. Genetic Algorithms as a Tool for Feature Selection in Machine Learning // Proceedings of the Fourth International Conference on Tools with Artificial Intelligence, 1992. TAI '92. – IEEE Press, 1992. – P. 200-203.
6. Tsoy Y.R. On Adaptive Increase of the Features Space Dimensionality // 12th International Conference. with International Participation on Artificial Intelligence (CAI-2010), Vol. 4. – Moscow: Physmathlit, 2010. – P. 134-140. (In Russian). http://qai.narod.ru/Workshop/Tsoy_cai2010.pdf
7. Zhao Z., Sharma S., Anand A., Morstatter F., Alelyani S., Liu H. Advancing Feature Selection Research. – 2010. http://featureselection.asu.edu/featureselection_techreport.pdf
8. Prechelt L. PROBEN1 – a set of neural network benchmark problems and benchmarking rules. Technical Report 21/94. – Fakultat fur Informatik, Universitat Karlsruhe, Karlsruhe, Germany, 1994.

[1] http://www.heatonresearch.com/encog

[2] http://code.google.com/p/mentalalchemy/