Neural Network Normalization for Genetic Search*

Jung-Hwan Kim, Sung-Soon Choi, and Byung-Ro Moon

School of Computer Science and Engineering
Seoul National University
Shilim-dong, Kwanak-gu, Seoul, 151-742 Korea
{aram,irranum,moon}@soar.snu.ac.kr

An arbitrary neural network has a number of functionally equivalent other networks. This causes redundancy in genetic representation of neural networks, which considerably undermines the merit of crossover in GAs [1]. This problem has received considerable attention in the past and has also been called the "competing conventions" problem [2].

We transform each neural network to an isomorphic neural network to maximize the genotypic consistency of two parents. We aim to develop a better genetic algorithm for neural network optimization by helping crossover preserve common functional characteristics of the two parents. This is achieved by protecting "phenotypic" consistency and, consequently, preserving building blocks with promising schemata.

Given a neural network, let N_i, N_h, and N_o be the numbers of input neurons, hidden neurons, and output neurons in the network, respectively. We denote a neural network $\mathfrak{N}$ by $\mathfrak{N} = \{\mathbf{h}_1, \ldots, \mathbf{h}_J\}$, where $\mathbf{h}_j = [w^i_{j1}, \ldots, w^i_{jN_i}, w^h_{1j}, \ldots, w^h_{N_h j}, w^o_{1j}, \ldots, w^o_{N_o j}]^T$ and w^i_{jk}, w^h_{kj}, and w^o_{kj} are the synaptic weights from input neuron k to hidden neuron j, from hidden neuron j to hidden neuron k, and from hidden neuron j to output neuron k, respectively. Let S_J be the set of all the permutations of the set $\{1, \ldots, J\}$ where J is the number of hidden neurons. We define the neural network isomorphism as follows:

Definition 1. *For two neural networks* $\mathfrak{N} = \{\mathbf{h}_1, \ldots, \mathbf{h}_J\}$ *and* $\mathfrak{N}' = \{\mathbf{h}'_1, \ldots, \mathbf{h}'_J\}$, $\mathfrak{N}$ *is* isomorphic *to* $\mathfrak{N}'$ ($\mathfrak{N} \simeq \mathfrak{N}'$) *if and only if there exists a permutation* $p \in S_J$ *such that* $\mathfrak{h}_{p(j)} = \mathfrak{h}'_j \ \forall j = 1, \ldots, J$.

From the definition, two isomorphic neural networks are constructed essentially in the same way. In other words, a neural network can be transformed into another isomorphic neural network by appropriate permutation of the hidden neurons.

We transform one of the parents in relation to the other so that high-quality schemata are well preserved and combined. We call such a transformation *normalization*. More formally, let $\mathcal{N}$ be the set of the neural networks and $\mathcal{N}_{\mathfrak{N}}$ be the set of the networks that are isomorphic to a network $\mathfrak{N} \in \mathcal{N}$. Suppose a distance measure $\mathfrak{d} : \mathcal{N} \times \mathcal{N} \to \mathbb{R}$ defined on a pair of networks that measures the genotypic distance of the two networks. Given parents $\mathfrak{N}, \mathfrak{M} \in \mathcal{N}$, the normalization

* This work was supported by Brain Korea 21 Project. The ICT at Seoul National University provided research facilities for this study.

K. Deb et al. (Eds.): GECCO 2004, LNCS 3103, pp. 398–399, 2004.

Table 1. Comparison with Several Normalization

	WBCD	CHDD	TDD
Normal	96.47 (96.91)	80.61 (82.74)	94.54 (94.72)
$\mathfrak{d}_1$_2Opt	96.78 (97.42)	82.03 (85.79)	95.19 (96.44)
$\mathfrak{d}_1$_KL	96.72 (97.42)	82.16 (84.26)	95.16 (96.47)
$\mathfrak{d}_2$_HM	96.89 (97.42)	82.07 (83.82)	94.83 (94.97)
$\mathfrak{d}_3$_2Opt	97.03 (97.43)	83.36 (85.15)	95.65 (96.53)
$\mathfrak{d}_3$_KL	97.21 (97.60)	84.26 (85.83)	95.77 (96.85)

operator transforms $\mathfrak{M}$ to $\mathfrak{M}' \in \mathcal{N}_{\mathfrak{M}}$ such that $\mathfrak{d}(\mathfrak{N}, \mathfrak{M}')$ is minimal among all the networks in $\mathcal{N}_{\mathfrak{M}}$.

For two neural networks $\mathfrak{N}$ and $\mathfrak{N}'$, we propose three distance measures. Among these, the first measure ($\mathfrak{d}_1$) uses Euclidean distance based on the weights of the connections. The other two measures are based on the degrees of learning of the networks. These measures are related to our assumption that sufficiently learned neurons constitute high-quality schemata. $\mathfrak{d}_2$ uses the degrees of learning of the individual hidden neurons and $\mathfrak{d}_3$ uses the degrees of learning of the individual connections.

We devised a Kernighan-Lin-style heuristic for the normalization which is fast but suboptimal to some degree. The heuristic iteratively improves the permutation in terms of the distance measure by exchanging pairs of the hidden neuron indices in the manner of sequential 2-Opt.

We selected three well known datasets from UCI repository: Thyroid Disease Database (TDD), Wisconsin Breast Cancer Database (WBCD), and Cleveland Heart Disease Database (CHDD). The effects of the proposed normalization approaches are examined with the above databases.

Table 1 shows the classification results. In the table, "Normal" represents the results of the neuro-genetic hybrid without normalization. $\mathfrak{d}_i$_B denotes the neuro-genetic hybrid with normalization by method B on distance measure $\mathfrak{d}_i$. "2Opt" and "KL" denotes the heuristics of 2-Opt and Kernighan-Lin style, respectively. "HM" indicates the Hungarian method and uses the learning degree of the hidden neurons.

The two values in each experiment show the mean and the best classification results, respectively, from 50 trials. The normalization methods for the neural network overall showed improvement over the one without normalization. Among the five methods of normalization, the method $\mathfrak{d}_3$_KL showed the best results. This tendency was consistent in all of the three test problems.

References

1. J. D. Schaffer, D. Whitley, and L. J. Eshelman. Combinations of genetic algorithms and neural networks: A survey of the state of the art. In *International Workshop on Combinations of Genetic Algorithms and Neural Networks*, pages 1–37, 1992.
2. D. Thierens. Non-redundant genetic coding of neural networks. In *IEEE International Conference on Evolutionary Computation*, pages 571–575, 1996.