SEPA: Structure Evolution and Parameter Adaptation in Feed-Forward Neural Networks

Paulito P. Palmes, Taichi Hayasaka, and Shiro Usui

Department of Information and Computer Sciences
Toyohashi University of Technology, Toyohashi, 441-8580 Japan
{ppalmes,hayasaka,usui}@bpel.ics.tut.ac.jp
http://www.bpel.ics.tut.ac.jp/

Abstract. In developing algorithms that dynamically changes the structure and weights of ANN (Artificial Neural Networks), there must be a proper balance between network complexity and its generalization capability. SEPA addresses these issues using an encoding scheme where network weights and connections are encoded in matrices of real numbers. Network parameters are locally encoded and locally adapted with fitness evaluation consisting mainly of fast feed-forward operations. Experimental results in some well-known classification problems demonstrate SEPA's high consistency performance in classification, fast convergence, and good optimality of structure.

1 SEPA Strategy

Until now, ANN architecture design remains to be one of the most important areas of research due to the lack of general criteria in finding an optimal network topology for a particular class of problems. To address this issue, SEPA uses a GA-based model of ANN where weight adaptation is not gradient-based but stochastic (Fig. 1). To achieve a proper balance in ANN's network complexity and generalization capability, SEPA's fitness function considers three important parameters: training error; hidden nodes; and connections.

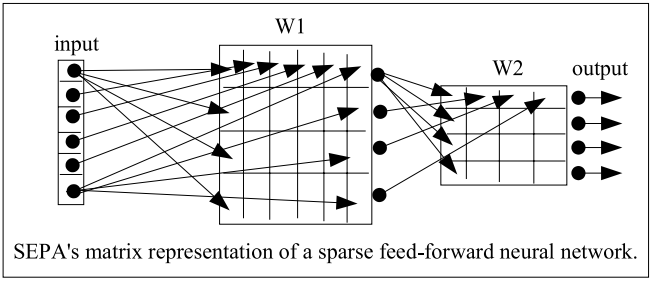

Fig. 1. Encoding scheme

E. Cantú-Paz et al. (Eds.): GECCO 2003, LNCS 2724, pp. 1600–1601, 2003.

Equation (1) is SEPA's fitness function where: α, β are complexity constants; n, N are no. of active hidden nodes and its total; and c, C are no. of active connections and its total. The degree of influence in network complexity is controlled by setting α and β parameters appropriately. SEPA uses $\alpha = 0.70$ and $\beta = 0.30$.

$$Q = \frac{1}{N}\sum_{i=1}^{N}(T_i - O_i)^2 + \alpha(\frac{n}{N}) + \beta(\frac{c}{C}) \tag{1}$$

SEPA's structure evolution is carried out by mutation using (2) and (3) with mutation probability set to 0.01. Structure exploration is carried out by adapting the mutation operator using Gaussian perturbation:

$$m' = m + \mathbf{N}_0(0, \sigma) \tag{2}$$

$$w'_{ij} = w_{ij} + \mathbf{N}_0(0, m') \tag{3}$$

where σ is the step size parameter (SSP) influencing the strength of mutation. SEPA uses step sizes of 5 (SSP5) and 100 (SSP100) in the experiment. SEPA's selection criterion uses an elitist roulette-wheel method where only the two best parents are retained.

2 Summary of Results

SEPA exhibits very low classification error variability (maximum=0.02) and small network size in all problems (Table 1). The experiments suggest that SSP can serve as a local optimizer in SEPA. For well-behaved classification problems, larger SSP produces smaller networks, faster convergence, and good classification performance. On the other hand, smaller SSP produces larger network structure with slower convergence rate but with good adaptability in noisy data.

Table 1. Experiments

HIDDEN NODES

a) cancer p=0.00

SSP	ave	var
5	5.84	2.71
100	3.84	1.65

b) iris p=0.00

SSP	ave	var
5	4.28	3.39
100	3.04	1.30

c) wine p=0.00

SSP	ave	var
5	4.78	3.60
100	3.24	1.45

d) glass p=0.00

SSP	ave	var
5	5	0.00
100	5.9	0.10

CLASSIFICATION ERROR

a) cancer p=0.64

SSP	ave	var
5	3.8%	0.00
100	3.6%	0.00

b) iris p=0.01

SSP	ave	var
5	10.4%	0.00
100	7.2%	0.00

c) wine p=0.35

SSP	ave	var
5	8.6%	0.00
100	7.8%	0.00

d) glass p=0.01

SSP	ave	var
5	40.0%	0.02
100	57.0%	0.00

ACTIVE CONNECTIONS

a) cancer p=0.00

SSP	ave	var
5	60.72	235.39
100	32.44	235.88

b) iris p=0.00

SSP	ave	var
5	23.68	41.86
100	19.20	24.33

c) wine p=0.00

SSP	ave	var
5	102.02	268.02
100	79.62	376.36

d) glass p=0.00

SSP	ave	var
5	58.20	14.40
100	74.90	0.10

MAXIMUM GENERATION

a) cancer p=0.00

SSP	ave	var
5	1093.0	2779005.10
100	77.2	4816.49

b) iris p=0.00

SSP	ave	var
5	452.00	148269.4
100	139.20	18346.3

c) wine p=0.00

SSP	ave	var
5	1311.4	2460293.92
100	233.8	74693.43

d) glass p=inf

SSP	ave	var
5	5000	0
100	5000	0