References

1 RIZZOLI, V., and LIPPARINI, A.: 'General stability analysis of periodic steady state regimes in nonlinear microwave circuits', *IEEE Trans. Microw. Theory Tech.*, 1985, **MTT-33**, (1), pp. 30–37

2 PLATZKER, A., STRUBLE, W., and HETZLER, K.T.: 'Instability diagnosis and the role of K in microwave circuits'. IEEE MTT-S Digest, 1993, pp. 1185–1588

3 PALAZUELOS, E., SUÁREZ, A., PORTILLA, J., and BARAHONA, J.: 'Hysteresis prediction in autonomous microwave circuits, using commercial software', *IEEE J. Solid-State Circuits*, 1998, **33**, (8), (to appear)

4 OGATA, K.: 'Modern control engineering' (Prentice-Hall, 1970)

5 BASU, S., MAAS, S.M., and ITOH, T.: 'Stability analysis for large signal design of a microwave frequency doubler', *IEEE Trans. Microw. Theory Tech.*, 1995, **MTT-43**, (12), pp. 2890–2898

6 HASLER, M.J.: 'Electrical circuits with chaotic behavior', *IEEE Special Issue on Chaotic Systems*, August 1987, pp. 1009–1021

Adaptive algorithm for training pRAM neural networks on unbalanced data sets

S. Ramanan, T.G. Clarkson and J.G. Taylor

A novel algorithm for training pyramidal pRAM neural networks on an unbalanced training set is proposed. The behaviour of the standard reinforcement learning algorithm is analysed and an adaptive learning rate algorithm that modifies the reinforcement learning algorithm based on readily available *a priori* class probability is developed.

Introduction: In an automatic target recognition (ATR) problem, we frequently face the problem of under-represented classes because of the presence of non-target objects in great numbers in real scenario images [1]. It is also difficult to obtain the required number of target objects. This poses severe problems in training neural networks as the learning process becomes biased towards the majority class, ignoring the minority classes and leaving them poorly trained at the end of the training stage. The learning process also becomes slower and it takes a longer time to converge to the expected solution.

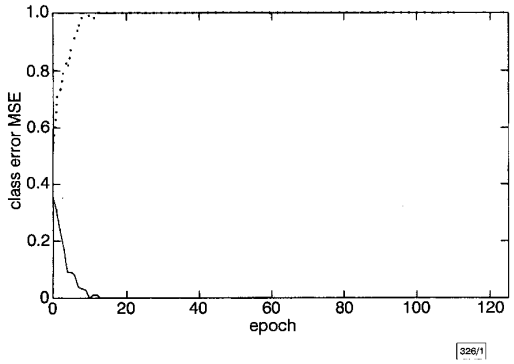

Fig. 1 *Error per exemplar for each class for standard algorithm*
——— class $\mathbf{C}_1$
– – – – class $\mathbf{C}_2$

Training on unbalanced training sets: To demonstrate the detrimental effect of an unbalanced training set on the learning process, two-class data sets *S*1, *S*2 and *S*3 with unbalanced ratio 1.36, 3.72 and 8.81, respectively, were used in the experiments. A pyramidal pRAM neural network [2] structure using the reinforcement learning rule was employed. Simulations were left running until there was no further decrease in training error and generalisation error. In the case of data set *S*1, the network settled on a solution which classified both classes correctly. However, this was not the case for data sets *S*2 and *S*3. The average error per pattern in each class is plotted for data set *S*2 in Fig. 1 as training progressed. It is apparent from these results that the network has converged onto a solution favouring the majority class [3]. The same experiment was repeated a number of times and the results resembled the previously obtained ones. Further investigation revealed that most of the memory contents (weights) of the network's output layer pRAM were close to zero. Similar behaviour was observed when the network was trained on training set *S*3.

Adaptive learning rate: Reinforcement learning is used to search for optimal parameter settings in feedforward pRAM networks for pattern recognition applications. The version of the reinforcement learning rule used for training the pRAM net is given below [4].

$$\Delta\alpha_u(t) = \rho[(a-\alpha_u)r + \lambda(\bar{a}-\alpha_u)p](t) \times \delta_{u,i} \quad (1)$$

where ρ and λ are learning rates, r and p are reward and penalty factors, α_u is the memory content and a is the pRAM output. It is proved in [5] that in the case of an unbalanced training set, the length of the weight change vector for class $\mathbf{C}_k$, $\Delta\underline{\alpha}(\mathbf{C}_k)$, is proportional to the size of the training set n_k, i.e.

$$\frac{\mathbf{E}(\|\Delta\underline{\alpha}(\mathbf{C}_1)\|^2)}{\mathbf{E}(\|\Delta\underline{\alpha}(\mathbf{C}_2)\|^2)} \simeq \frac{n_1^2}{n_2^2} \quad (2)$$

where $\mathbf{E}(\cdot)$ denotes the expectation with respect to memory contents (weights) $\underline{\alpha}$ and n_1, n_2 are sizes of the training sets. In the case of $n_1 > n_2$, the length of the weight change vector of the dominant class $\mathbf{C}_1$ is very large. It is also proved [5] that the dot product of $\Delta\underline{\alpha}(\mathbf{C}_1)$ and $\Delta\underline{\alpha}(\mathbf{C}_2)$ is always negative in the first few iterations, i.e.

$$\Delta\underline{\alpha}(\mathbf{C}_1) \cdot \Delta\underline{\alpha}(\mathbf{C}_2) < 0 \quad (3)$$

Since the length $\Delta\underline{\alpha}(\mathbf{C}_1)$ is much larger than the length of $\Delta\underline{\alpha}(\mathbf{C}_2)$, the overall weight change vector $\Delta\underline{\alpha}$ will be in the direction of the majority class. Therefore, $\Delta\underline{\alpha}$ does not always point in the best direction to minimise the error of both classes in a two class problem in an unbalanced training set case. To overcome this bias, we choose a direction b which bisects the angle between $\Delta\underline{\alpha}(\mathbf{C}_1)$ and $\Delta\underline{\alpha}(\mathbf{C}_2)$ because it is guaranteed to be a downhill direction for both classes (see Fig. 2).

$$b = \frac{\|\Delta\underline{\alpha}(\mathbf{C}_1) + \Delta\underline{\alpha}(\mathbf{C}_2)\|}{\left\|\frac{\Delta\underline{\alpha}(\mathbf{C}_1)}{\|\Delta\underline{\alpha}(\mathbf{C}_1)\|} + \frac{\Delta\underline{\alpha}(\mathbf{C}_2)}{\|\Delta\underline{\alpha}(\mathbf{C}_2)\|}\right\|} \cdot \left\{\frac{\Delta\underline{\alpha}(\mathbf{C}_1)}{\|\Delta\underline{\alpha}(\mathbf{C}_1)\|} + \frac{\Delta\underline{\alpha}(\mathbf{C}_2)}{\|\Delta\underline{\alpha}(\mathbf{C}_2)\|}\right\} \quad (4)$$

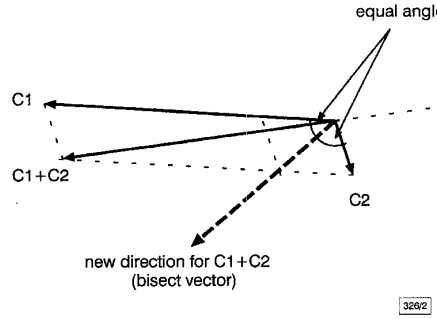

Fig. 2 *Weight change vectors* $\Delta\underline{\alpha}(\mathbf{C}_1)$, $\Delta\underline{\alpha}(\mathbf{C}_2)$ *and* $\Delta\underline{\alpha}^{new}$

Therefore, by using the results obtained in eqns. 2 – 4, we can show that

$$\Delta\underline{\alpha}^{new}(t) = K \cdot \left\{\Delta\underline{\alpha}(\mathbf{C}_1) + \frac{n_1}{n_2} \cdot \Delta\underline{\alpha}(\mathbf{C}_2)\right\} \quad (5)$$

where K is the magnitude factor. As the direction of steps taken in the weight space contributes to bypass the convergence onto poor local minima, we can omit the magnitude factor in eqn. 5 and simplify the modified algorithm. Further, multiplication of $\Delta\underline{\alpha}(\mathbf{C}_2)$ by n_1/n_2 can be associated with the learning algorithm. This can be accomplished by having an adaptive learning rate (ALR), i.e. a learning rate that varies according to the class of the training pattern presented to the network. If the selected base learning rate is ρ, in the adaptive training strategy, ρc_1 and ρc_2 should be as follows

$$\rho c_1 = \rho \quad \text{and} \quad \rho c_2 = \frac{n_1}{n_2} \cdot \rho \quad (6)$$

We rewrite the standard reinforcement algorithm for the pRAM neuron as follows:

$$\Delta\alpha_u(t) = \rho_t[(a - \alpha_u)r + \lambda_t(\bar{a} - \alpha_u)p](t) \times \delta_{u,i} \quad (7)$$

where ρ_t and λ_t are expressed by

$$\rho_t = \begin{cases} \rho c_1 & \text{if } \mathbf{C} = \mathbf{C}_1 \\ \rho c_2 & \text{if } \mathbf{C} = \mathbf{C}_2 \end{cases} \quad (8)$$

$$\lambda_t = \lambda_0 \quad \forall\, \mathbf{C} \in \{\mathbf{C}_1, \mathbf{C}_2\} \quad (9)$$

Simulations and results: The same experiments were repeated for the three data sets, but this time the learning rates were adapted according to the class of the input pattern presented. The results showed that the adaptive algorithm required fewer iterations compared with the standard algorithm, for convergence onto to an acceptable solution, particularly in the case of S_1. It is clear from this experiment that the training set with an unbalanced ratio close to 1 does not suffer from convergence onto poor local minima but the adaptive algorithm speeds up the convergence. Fig. 3 shows the error per example for each class for the pRAM net trained on data set S_2 using the ALR algorithm. Comparing the plots in Figs. 1 and 3, we show that by use of the adaptive algorithm the error for each class is reduced simultaneously during the training phase. Furthermore, the ALR algorithm guides the learning process towards a solution that classifies most of the patterns from both classes.

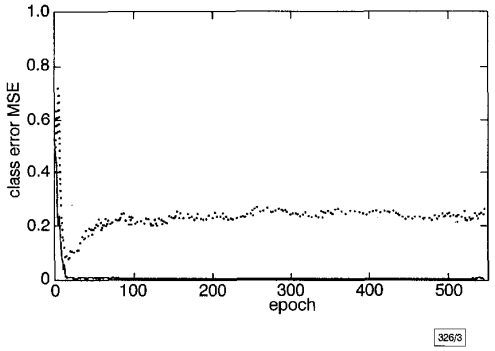

Fig. 3 *Error per exemplar for each class for ALR algorithm*

——— class $\mathbf{C}_1$
– – – – class $\mathbf{C}_2$

Conclusion: We have proposed a novel algorithm for training pyramidal pRAM neural networks on an unbalanced training set. The detrimental effect of an unbalanced training set on the learning process is shown experimentally. Based on the analysis of the standard reinforcement learning algorithm on an unbalanced training set, an adaptive learning rate (ALR) algorithm has been developed. The ALR algorithm uses the readily available *a priori* class probability to guide the learning process and to avoid being trapped into poor local minima.

7 May 1998
Electronics Letters Online No: 19980921

S. Ramanan (*Department of Electronic Systems Engineering, University of Essex, Colchester, Essex, CO4 3SQ, United Kingdom*)

T.G. Clarkson (*Department of Electronic and Electrical Engineering, King's College London, London WC2R 2LS, United Kingdom*)

J.G. Taylor (*Department of Mathematics, King's College London, London WC2R 2LS, United Kingdom*)

References

1 RAMANAN, S., PETERSEN, R.S., CLARKSON, T.G., and TAYLOR, J.G.: 'pRAM nets for detection of small targets in sequences of IR images', *Neural Netw.*, 1995, **8**, (7/8), pp. 1227–1237

2 CLARKSON, T.G., NG, C.K., and GUAN, Y.: 'The pRAM : An adaptive VLSI chip', *IEEE Trans. Neural Netw.*, 1993, **4**, (3), pp. 408–412

3 RAMANAN, S., PETERSEN, R.S., CLARKSON, T.G., and TAYLOR, J.G.: 'Adaptive learning rate for training pyramidal pRAM nets'. Proc. ICANN'94 Conf., Sorrento, Italy, 1994, pp. 1360–1363

4 GORSE, D., and TAYLOR, J.G.: 'Review of the theory of pRAMs'. Proc. Weightless Neural Network Workshop '93, York, UK, 1993, pp. 29–34

5 RAMANAN, S.: 'pRAM neural networks for automatic object recognition'. PhD Thesis, University of London, 1997

325nm bandwidth supercontinuum generation at 10Gbit/s using dispersion-flattened and non-decreasing normal dispersion fibre with pulse compression technique

H. Sotobayashi and K. Kitayama

325nm bandwidth (at 20dB) supercontinuum generation at 10Gbit/s using dispersion-flattened uniform normal dispersion fibre along with a pulse compression technique is experimentally demonstrated with a pumped 3.8ps, 1552nm hybrid modelocked semiconductor laser. Application to a 50nm wavelength tunable source is also demonstrated.

Introduction: Short optical pulse sources of wide wavelength tunability and high repetition rate are important for future optical communication applications. A supercontinuum (SC) pulse source is promising because pico-second pulses at several tens of Gbit/s can be generated over an extremely broad spectral range. It is some time since 200nm bandwidth SC generation in optical fibres was first reported [1]. Details of the fibre used in the experiments have been unknown until recently [2, 3]. Dispersion-flattened and decreasing fibre whose dispersion decreases from an anomalous dispersion regime to a normal dispersion regime is the key to ultra-broadband SC generation. However, dispersion-flattened and decreasing fibres require a prohibitively difficult fabrication process. In this Letter, we describe a more practical and simplified approach to SC generation.

Design theory: Techniques to generate SC in optical fibres are divided into two categories. The first is spectrum broadening by pulse compression using soliton effects in the anomalous dispersion regime. The second is spectrum broadening by the accumulation of frequency chirping caused by optical Kerr effects in the normal dispersion regime. There is an important advantage of the latter technique over the former one. Nonlinear pulse propagation in the anomalous dispersion regime tends to generate multiple pulses with up- and down-chirping. In contrast, the output pulse after propagation in the normal dispersion regime is always single pulse and frequency chirping is almost linear up-chirping. Consequently, the output pulse by filtering out from SC is always single pulse. Normal dispersion is also useful to generate a reduced intensity fluctuation spectrum [4]. Increasing the input peak power can generate a wider spectrum. These characteristics are attractive for the application of a wavelength tunable source.

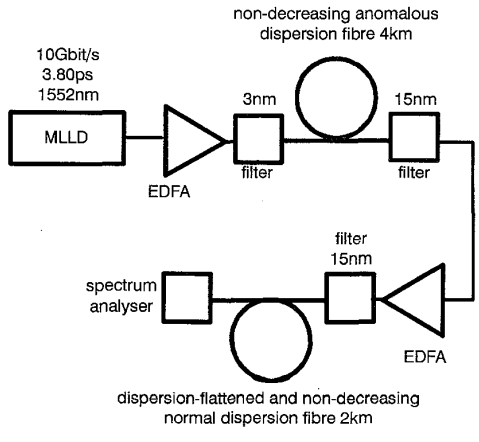

Fig. 1 *Experimental setup*

Experiments: Fig. 1 shows the experimental setup, which is divided into two parts. In the first part, pulses are compressed in a non-decreasing anomalous dispersion fibre using higher-order soliton effects to increase the peak power. In the second part, these narrow pulses with high peak power are launched into a dispersion-flattened and normal dispersion fibre to generate SC.

The input pulses, with pulsewidth 3.8ps and centre wavelength 1552nm from a 10GHz hybrid modelocked semiconductor laser,