OPTIMISATION OF TUPLE SIZE IN A HYBRID STOCHASTIC SEARCH

J. M. Bishop & R. J. Mitchell
Department of Cybernetics,
University of Reading, Whiteknights,
Reading, Berkshire, UK, RG6 2AY.

Abstract

A major difficulty when applying neural networks to pattern recognition problems is stimulus equivalence - the invariance of symbolic information independent of transformation within a search space. Adult humans can generally recognise patterns accurately despite such transformations and distortions. Classically it has been suggested that this ability is due to a normalisation process that occurs before the classification procedure begins.

One criticism of the early connectionist models, such as the perceptron, was that there is no clear mechanism by which such normalisation can occur [1]. However recent PDP techniques, such as knowledge replication [2], Hinton mapping [3] and stochastic networks [4] have helped overcome this problem.

Hybrid stochastic networks encompass a family of parallel algorithms that have been successfully applied to various search & optimisation problems [5]. They consist of a stochastic search engine which directs the mapping of n-tuple RAMs into a search space; the success or failure of each mapping being defined by the output of each RAM.

A stochastic search is most efficient when such testing is sharp and the response to genuine features is high compared to noise. However if the response is too sharp then the hybrid network is unable to generalise. This paper will describe the hybrid stochastic network and the factors influencing the choice of n-tuple size.

1 Introduction

Stochastic networks embody a search technique which uses a random diffusion process to find the best fit of a given data model within a specified search space. The model and search space are defined in terms of micro-features between which comparisons can be made. For example, when searching for the best fit of a string of numbers within a larger string, the features may be the digits which comprise the strings, or groups of such digits.

A stochastic network consists of a collection of mapping cells, where each cell maintains a pointer into the search space and a pointer to a micro-feature of the model. Unlike Hinton Mapping [3], where each mapping cell defines a fixed mapping, stochastic networks use a set of variable (*dynamic*) mapping cells to test possible locations of model micro-features in the search space.

Stochastic networks can operate asynchronously, such that each cell probes the search space independently of other cells; or synchronously whereby all cells update simultaneously. In either case each neuron performs the following two phase update cycle:

i. A *test phase* whereby the cell selects a micro-feature of the model to search for at the location defined by its current mapping.

   If the micro-feature is detected then the mapping defined by the cell is maintained and the cell becomes *active*.

ii. A *diffusion phase* which ensures that *inactive* cells adopt mappings from currently active cells.

NB. The selection of a different micro-feature of the model during each update cycle prevents a mapping cell becoming trapped in a local minima.

If the search conditions are favourable, a condition is rapidly reached whereby the distribution of mappings into the search space reaches statistical equilibrium. The most popular mapping then defines the position of the model within the search space.

2 Method

Each cycle of activity in a stochastic network consists of two phases: firstly current mappings are evaluated and secondly successful mappings are propagated to other cells.

eg. INITIALISE (*Mapping*);

REPEAT

TEST (*Mapping*);

DIFFUSE (*Mapping*);

UNTIL TERMINATION;

Mapping describes the cells current mapping of an object micro-feature in the search space. Each cell is in either an active or inactive state, defined by the accuracy of this mapping. Cell mappings are first initialised to specific values. These values are usually selected randomly unless extra information is available to cluster mappings into a specific region of the Search Space.

During the test phase each cell's activity changes to reflect the currently perceived accuracy of its present mapping. The accuracy of a mapping is usually defined by a local comparison between a specified model micro-feature and the corresponding feature in the search space defined by the cells mapping.

Diffusion spreads mappings from active cells to inactive cells by a stochastic selection process. In a fully connected network, each inactive cell selects another cell at random. If this cell is currently active, then its mapping is transferred to the inactive cell, otherwise the inactive cell is re-initialised with a new mapping into the search space. Figure one shows the first three iterations of a stochastic search for a three digit model *[123]* in a search space defined by eight digits *[765**123**89]*.

A stochastic network terminates when it is at statistical equilibrium; that is, when the distribution of mappings into the search space remains constant. However in some cases, for example when features in the search space are badly distorted by noise, (or when the model is not within the search space), this equilibrium condition may never be reached.

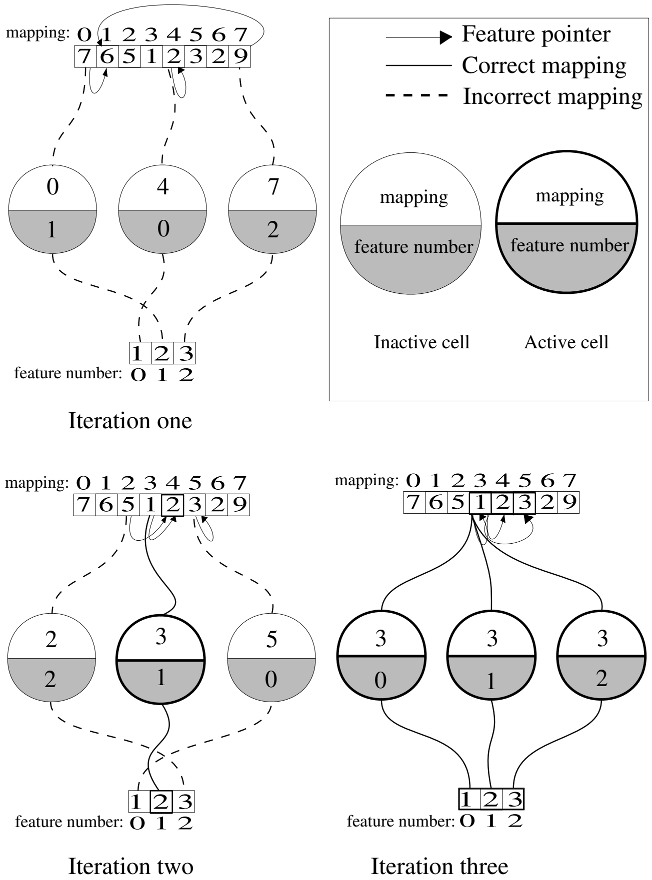

Figure 1: An example search

3. Theoretical

An elementary analysis of a simple stochastic network has been carried out and a Markov model model of the network derived [5]. The model accurately predicts the behaviour of a stochastic network under conditions of no noise, and no reselection.

If $P_{(x, n)}$ is the probability of x mapping cells being *active* at iteration n then it has been shown that:

$$P_{(x + q,\, n + 1)} = \binom{M - x}{q} \left(\frac{M - x}{M}\right)^{(M - x - q)} \left(\frac{x}{M}\right)^{q} \qquad (1)$$

The above expression defines the transition probabilities $\mathbf{D}_{ij}$, of a Markov chain $\mathbf{D}$, giving the probability of moving from state [i] to state [j] due to stochastic diffusion. The matrix of transition probabilities at the nth iteration is thus defined by:

$$P_n = \mathbf{D}^n \qquad (2)$$

Where P_n is the probability of moving from state [i] to state [j] at iteration n.

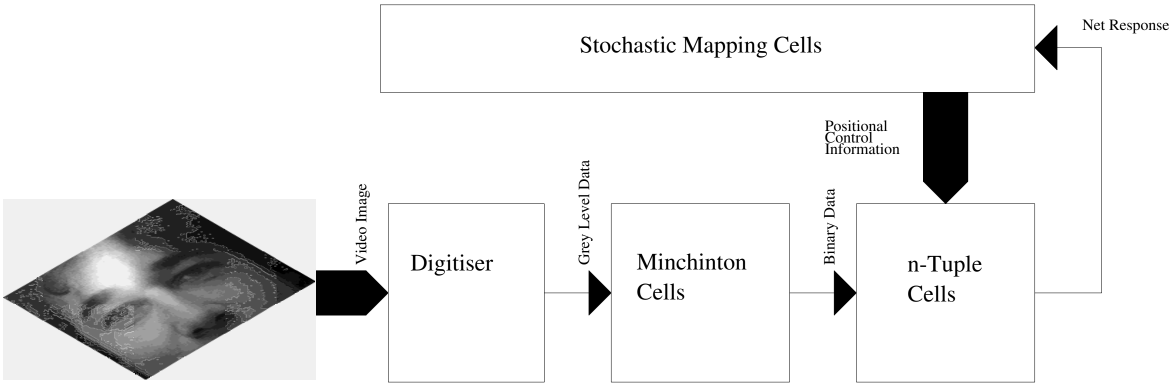

Figure 2: The Stochastic Search Network

4. Hybrid Systems

The stochastic search network describes a method of random search well suited for implementation in hardware on a parallel architecture. A key element of the search is the *test phase* where a micro-feature from the search space defined by a cell mapping is compared with a micro-feature of the object model. Various methods of performing this test have been investigated including: direct comparison, the use of a distance metric, and neural techniques. It has been established that neural methods are well suited to this application as they are capable of *generalising* when recognising distorted model micro-features.

One such neural comparator that has been investigated is the use of individual *n-tuple RAM neurons* to perform micro-feature comparisons; ie. each mapping cell (**M**) is associated with a RAM neuron, which is mapped into the search space at the location defined by **M**'s current mapping; the output of the RAM defining the result of the cell's test phase. Such a hybrid system has been successfully used as part of the Telecom CONNEX project, achieving over 95% success in locating eyes in 2D facial images [5]. However, when using a RAM neuron to perform micro-feature comparisons on grey level images an extra pre-processing layer must be introduced to convert the multi-value data into binary data suitable for processing by the RAM neuron (see Figure 2). For the CONNEX project this pre-processing was performed by Minchinton cells [6].

5. Optimisation of Tuple Size

When using weightless RAMs to perform micro-feature comparisons, care must be exercised in selecting a suitable tuple size. A stochastic search is most efficient when the response of the micro-feature comparison is sharp, ie. the response to genuine micro-features is high compared to noise (implying a relatively high tuple size). However if this response is too sharp then the network is unable to generalise. The use of Minchinton cell pre-processing further complicates this choice, since due to its differential operation, it increases the RAM cell's sensitivity to high frequency noise.

3a. Unseen eye

3b. Non eye feature

Figure 3. Test data for evaluation of tuple size

To estimate the optimal tuple size two experiments were carried out: the first to investigate the effect of increasing tuple size on the generalisation of a RAM neuron with Minchinton cell pre-processing; the second to investigate the effect of tuple size on the RAM cell's response to non-features.

A weightless network of 600 RAM cells was trained on a selection of nine known eye images. The networks tuple size was varied between four and eight and the response of the network to a new eye image (figure 3a) and to a prominent non-eye feature (figure 3b) noted. The results are given in Table 1.

Defining P_e as the probability of response to a new eye feature and P_n as the probability of response to a non-eye feature; to maximise the effectiveness of the Hybrid network we need to maximise P_e & minimise P_n ie. *if P_e & P_n are of equal importance we should maximise* $[P_e * (1 - P_n)]$.

Tuple size	% Response to new eye image	% Response to non eye image	$P_e * (1 - P_n)$
4	72	40	0.43
5	65	32	0.44
6	51	24	0.39
7	47	17	0.39
8	40	11	0.36

Table 1. Results of Tests

6. Conclusion

The use of a weightless RAM cell with Minchinton cell pre-processing to perform the test phase in a stochastic network has been evaluated. Experimental results indicate that there is an optimal tuple size for such a cell to maximise it's response to new genuine model micro-features and minimise it's response to non micro-features in the image. If these two factors are of equal weight, it has been demonstrated that a tuple size of five will offer the highest performance.

7. References

1. Rumelhart, D.E. & McClelland, J.L., 'PDP Models & General Issues in Cognitive Science', In "Parallel Distributed Processing", Vol. 1, pp. 114-115. The MIT Press, 1986.

2. Fukushima, K., 'Neocognitron: A hierarchical Neural Network Capable of Visual Pattern Recognition', Neural Networks, Vol. 1, pp. 119-130. Pergamon Press, 1988.

3. Hinton, G., 'A parallel computation that assigns canonical object-based frames of reference', Proc. 7th Int. Jnt. Conf. AI, 1981.

4. Bishop, J.M., 'Stochastic Searching Networks', Proc. 1st IEE Conf. Artificial Neural Networks, London 1989.

5. Bishop, J.M. & Torr, P., 'The Stochastic Search Network', in 'Neural Networks for Images, Speech and Natural Language', (eds. Lingard, R. & Nightingale, C). Chapman Hall, 1992.

6. Bishop, J.M., Minchinton, P.R. & Mitchell, R.J., 'Real Time Invariant Grey Level Image Processing using Digital Neural Networks', Proc IMechE Eurodirect '91, pp. 187-188. London, 1991.