Using the Hopfield Model with Mean Field Annealing to Solve the Routing Problem in Packet-Switched Communication Networks

Michael W. Dixon, Graeme R. Cole, and Matthew I. Bellgard
Computer Science Programme
School of Mathematical & Physical Sciences
Murdoch University
Murdoch, Western Australia 6150

ABSTRACT

The performance of the Hopfield neural network with mean field annealing for finding optimal or near-optimal solutions to the routing problem problem in a communication network is investigated. The proposed neural network uses mean field annealing to eliminate the constraint terms in the energy function. Unlike other systems which use penalty constraint terms there is no need to tune constraint parameters (this tuning has been found to be difficult and problem specific) and the neural network should avoid the problems of scaling. It also avoids the need to pre-determine the minimum number of hops corresponding to the optimal route. We have obtained very encouraging simulation results for the nine node grid network and fourteen node NFSNET-backbone network.

1. INTRODUCTION

The parallel nature of neural network models makes them suitable for simulation on parallel computer architectures. Neural networks are expected to be especially useful in the communications industry. Their promise as pattern recognizers and data correlators is naturally applicable to areas such as switching and queuing, transmission, error-correcting, data compression and routing of data communications traffic.

Neural networks have often been formulated to solve difficult (for the most part, NP-complete) optimization problems. The two main functions performed by a routing algorithm are the selection of the various origin-destination pairs and the delivery of the messages to their correct destination once the route has been selected. The second function is provided by using a variety of routing protocols and routing tables. This is where the proposed neural network will be used to quickly find good paths through a communication network. The proposed neural network algorithm will use a distributed routing approach where the computation of routes is shared among the network nodes with information exchanges between them as necessary. It will be able to adapt to changing conditions in the network (link failures and congestion) and change its routes to bypass the point of failure or congestion. Thus, the routing problem can be viewed as an optimization problem where the objective is to minimize total cost when selecting a route for a given origin-destination pair. This makes neural networks very good candidates for implementing the routing computations involved in the routing problem because of the potential of the neural network hardware approach for high computational speed, high degree of robustness and fault tolerance and low power consumption. As a result there has been widespread research into the application of neural networks to the routing problem with varying degrees of success [2, 6, 7, 10, 13]. The major limitations that have been encountered are: i) some systems have failed to consistently converge to a valid route, ii) valid routes are not typically optimal, iii) most of these systems do not scale to larger sized communication networks, and iv) pre-processing is often needed to simplify the problem (e.g. calculating the minimum number of hops from origin to destination).

In this paper, we present a version of the Hopfield model that employs mean field annealing [4, 9, 14, 15] in an attempt to solve the routing problem. The Hopfield model has achieved prominence because of the relative ease of building it into hardware using currently available VLSI technology. Current simulations on nine and fourteen node communication networks have shown that the system converges to a valid route for all origin-destination pairs. These routes are optimal or near optimal. Unlike other systems which use penalty constraint terms there is no need to tune parameters. This tuning has been found to be difficult and problem specific [1, 3]. Also, there is no need to pre-determine the minimum number of hops corresponding to the optimal route.

2. HOPFIELD MODEL

In 1985, Hopfield showed how the Hopfield model could be used to solve combinatorial optimization problems of the Travelling Salesman type [5]. The Hopfield model is a fully connected network of simple processing units, V_i, with numerically weighted symmetric connections, T_{ij}, between units V_i, V_j. Processing units have states (either discrete in $\{0,1\}$, or continuous in $[0,1]$ depending on whether the discrete or the continuous version of the network is being considered. Each processing unit performs simple and identical computations which generally involve summing weighted inputs to the unit, applying an internal transfer function, and changing state if necessary. The power of the Hopfield model lies in the connections between units and the weights of these

0-7803-2559-1/95 $4.00 © 1995 IEEE

connections. An energy function was defined by Hopfield on the states of the network (values of all the units). The energy function, E, in its simplest form is:

$$E = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} T_{ij}V_jV_i + \sum_{i=1}^{N} V_iI_i \qquad (1)$$

where V_i denotes the current state (value) of the i*th* neuron and I_i denotes its bias. Hopfield used the fact that the E in (1) is a Liapunov function (bounded from below) to show that, from any starting state, the network would always converge to some energy function minimum upon applying a sequence of asynchronous local state updates (that locally reduce energy).

To solve any particular problem, first a decision must be made on how to set the network parameters T and I, so that minimization of the Liapunov function (1) coincides with the minimization of the problem objective function and enforces satisfaction of the problem constraints; this process is termed 'mapping' the problem onto the network. Hopfield gives the motion equation of the i*th* neuron [5]:

$$\frac{dU_i}{dt} = \frac{U_i}{\tau} - \frac{\partial E}{\partial V_i} \qquad (2)$$

where the output is usually related to the state by a simple nondecreasing monatomic output function $g(U_i)$. Typically, a step function or a hyperbolic tangent function is used.

The local minimum to which the E function converges corresponds to the neural network's solution. One major difficulty with this approach is that the neural network may have local minima which do not correspond to correct and/or valid answers to the problem. Additional local minima beyond those intentionally placed in the state-space may be present and desired local minima may be absent.

3. MAPPING OF THE ROUTING PROBLEM

Solution of constrained optimization problems requires accounting for possible constraint violations. The constraint term(s) are explicitly stated in the energy function along with the cost term (objective). A neural network energy function for constrained optimization would be of the form:

$$E = \alpha[\text{"constraint(s)"}] + \beta[\text{"cost"}]$$

where α, β are scalar terms which are greater than zero and "cost" is an optimization cost function that is independent of the constraint term(s). Thus by minimizing the energy function E, we attempt to minimize the "cost" while at the same time maximize satisfaction of the constraints. The successful use of an energy function requires appropriate selection of the parameters α and β.

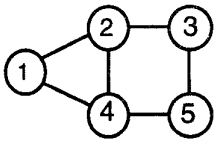

Figure 1: Five node network with bi-directional links.

For the routing problem NxN neurons will be used to represent a valid route from the source to destination node. The neurons are grouped into N groups of N neurons which are used to represent the path that the packet (message) will take. In a five node network communiction network (see Figure 1) twenty five neurons will be required to represent a route by an output matrix V.

	1	*2*	*3*	*4*	*5*
1	1	1	0	0	0
2	0	0	1	0	0
3	0	0	0	1	1
4	0	0	0	0	0
5	0	0	0	0	0

Table 1: A neural network representation of a route from node 1 to node 3.

The example in Table 1 illustrates a route the packet can take from node 1 to node 3 (e.g. where 1-1-2-3-3 is interpreted as 1-2-3). The same path can be represented in different ways, such as: 1-2-3-3-3, 1-2-2-3-3 and 1-1-1-2-3. The total length of the route for 1-2-3-3-3 would be

$$l_{route} = d_{11} + d_{12} + d_{23} + d_{33}.$$

Self looping is allowed and there is no cost associated with it. The actual cost would be the d_{12} and d_{23} terms in the route. A nine node grid network and the NFSNET-Backbone network have been chosen for simulating and testing our implementation of the Hopfield model (See Figures 2 & 3).

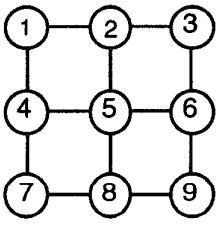

Figure 2: Nine node grid network with bi-directional links.

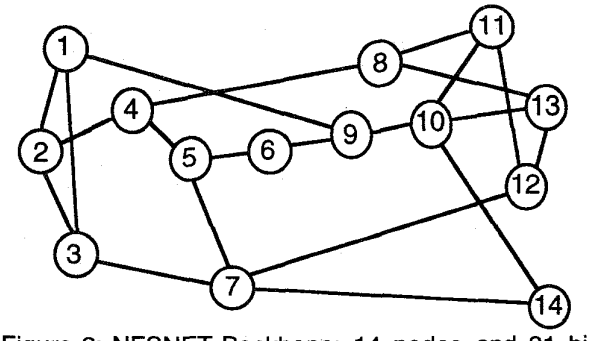

Figure 3: NFSNET-Backbone: 14 nodes and 21 bi-directional links.

In many combinatorial optimization problems the neurons are grouped into vectors or clusters (each column will be considered a group). For all of these vectors a 1-out-of-N constraint is used, where only one from a group of N neurons can be equal to 1 and all other neurons in the group are to be equal to 0. Sometimes the optimization process ends up with a violation of these constraints, i.e. there are some vectors which have all neurons equal to 0 or more than one neuron is equal to 1. This follows from the fact that a neural network may converge to a local minima of the energy function [1, 3, 5]. A way to prevent this from happening is to use mean field annealing (MFA) to replace the classical sigmoid activation function of the form

$$V_i = g(U_i) = \left[1 + \tanh(U_i / T)\right] / 2 \tag{3}$$

with a multi-dimensional activation function of the form shown in (4) below [4, 8, 14, 15].

$$V_{ia} = \frac{e^{U_{ia}}}{\sum_{b=1}^{N} e^{uib}} \tag{4}$$

(a =1,2,...,N)(i=1,2,...,N)

According to Cichocki and Unbehauen [4] the activation function generalizes the standard sigmoid function. By employing the multi-dimensional (generalized) activation function (4) the output variables V_{ia} of the neurons grouped in the vector V_i are not independent but always satisfy the constraint that the sum of the column must be equal to 1.

Our proposed energy equation is straightforward because it will only deal with the following distance term and an auxiliary term which adds local positive feedback around neurons which stabilize the neural network and eliminate chaotic behaviour (parasitic oscillations) during the optimization process [9]:

$$E = \frac{1}{2}\sum_{x=1}^{N}\sum_{y=1}^{N}\sum_{i=2}^{N-1} GD_{xy} V_{xi}(V_{y,i+1} + V_{y,i-1}) + \frac{P_1}{2}\sum_{x=1}^{N}\sum_{i=1}^{N} V_{xi}^2 \tag{5}$$

where the subscripts x and y refer to the nodes while the subscript i refers to the position in the route. The GD_{xy} term is a function of both capacity and distance across a link. Decreasing distance and increasing capacity cause the generalized distance to decrease. For the energy function in (5) the generalized MFA equation will take the following form:

$$V_{xi} = \frac{e^{U_{xi}}}{\sum_{J=1}^{N} e^{U_{xj}}}$$

where

$$U_{xi} = -\frac{1}{T}\frac{\partial E(V)}{\partial V_{xi}} = -\sum_{y=1}^{N} GD_{xy}(V_{y,i+1} + V_{y,i-1}) + P_1 V_{xi} \tag{6}$$

For each given Origin-Destination pair, the first and the last column in the neuron array are fixed (NxM matrix). The states of the neurons not in the first or last column are updated according to equation (6).

4. GENERALIZED DISTANCE MATRIX

The model was first tested using the nine node grid network shown in Figure 2. The links, nodes and distances of the network can be represented by a 9x9 distance matrix D with entries D_{ij} where each entry D_{ij} represents the distance between node i and node j. For non-existent links a large value L was used to prevent the neural network from using the links. Also, a 9x9 capacity matrix C with entries C_{ij} was used to represent the capacity between the nodes. All the links were assumed to be bi-directional and to have identical capacity of 100. The cost function used for our simulations is a function of the capacity C_{ij} and the distance D_{ij}. Using the distance and capacity matrices a (combined) generalized matrix was calculated as

$$GD_{ij} = (D_{ij}/C_{ij}) \tag{7}$$

which is represented by GD in equation (6).

5. PARAMETER SELECTION AND SENSITIVITY

The mapping of the routing problem to a neural network was straightforward. The only additional effort needed to complete an application of the MFA method to the routing problem was the selection the MFA equation parameters (temperature T and P_1) and the decay term.

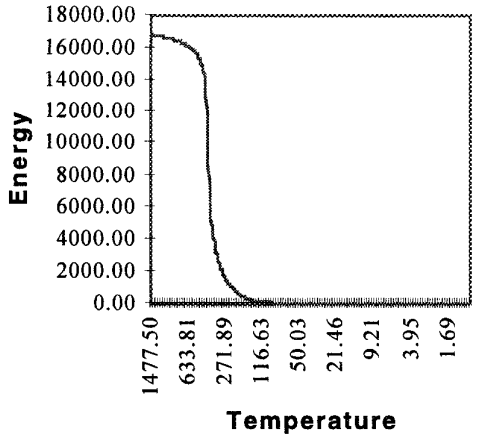

Figure 4: Plot of average energy of the network plotted against temperature.

The initial starting temperature was identified by the phase transition which is a critical point where the energy decreases significantly. This very sharp phase transition represents when the neural network has stabilized into a local minima. This was done by running a number of simulations and plotting the energy with respect to temperature. See Figure 4 for an example of the average of all the simulations which shows the energy drops significantly around the 800 mark. Thus, we can set the starting temperature at 800 and avoid unnecessary computations.

The annealing process was stopped when all the neuron values were within the range [0.0, 0.1] or within the range [0.9, 1.0] or when the temperature reached 1. The stopping temperature of 1 was chosen because the neural network never changed a route after cooling to this point. Not all of the neurons will go to 0 or 1. This is a consequence of the fact that the MFA variables V_i represent mean values of the neuron variables U_i at temperature T. Hence the values should be interpreted probabilistically.

The reinforcement term P_1 was chosen by running the simulation hundred of times to find the value that gave the best results. We found that tying it to the temperature produced the best results. If the P_1 term is set too high it will cause oscillations and the neurons will not settle to any valid route.

The decay term was chosen for its ability to find a path quickly and provide good results. Simulations were done varying the decay term from 0.95 throught to 0.995. The value of 0.95 was able to find valid paths quickly but produced inferior results to the higher decay terms. The value of 0.985 was found to be the optimal decay term for speed and for producing high quality results.

6. SIMULATION RESULTS

For our simulations the following parameters were set: Starting Temperature of 800, Ending Temperature of 1, the P_1 parameter was set equal to Temperature*1.75 and Decay term to 0.985. Since the neural network should have no prior favour for a particular path, all the U_{xi} were set to a random number between 0 and -1. This random noise helped to break up the symmetry caused by the symmetric network topology. All neuron output voltages V_{xi} were set to 0, except for the origin and destination nodes. Each simulation was stopped when the system reached the final temperature of 1 or if all the columns met the stopping condition explained in the previous section.

In order to be valid a route need not be optimum but must not contain any non-existent links. Between 50 and 400 iterations were required to reach the final solutions. Simulations were run for all origin-destination pairs to find valid routes for the nine node grid network and the fourteen node NFSNET network and the summary of results are shown in the following tables.

Hops	*Number of optimal routes*	*Number of near optimal routes*	*Total number of routes*
1	24	0	24
2	28	0	28
3	16	0	16
4	4	0	4

Table 2: Summary of simulation results for the nine node grid network.

Hops	*Number of optimal routes*	*Number of near optimal routes*	*Total number of routes*
1	42	0	42
2	68	2	70
3	58	12	70

Table 3: Summary of simulation results for the fourteen node NFSNET-Backbone network.

Sourc	*Dest*	*Sim Routes*	*Dist*	*Optimal Routes*	*Dist*
3	5	3-7-5	26	3-2-4-5	23
3	6	3-7-5-6	33	3-2-4-5-6	30
3	10	3-1-9-10	40	3-7-13-10	32
5	3	5-7-3	26	5-4-2-3	23
5	11	5-4-8-11	30	5-6-9-10-11	26
5	14	5-4-8-14	31	5-6-9-10-14	26
6	3	6-9-1-3	40	6-5-4-2-3	30
6	12	6-5-7-12	32	6-9-10-14-12	23
6	8	6-5-4-8	30	6-9-10-14-8	27
8	6	8-4-5-6	30	8-14-10-9-6	27
10	3	10-9-1-3	40	10-13-7-3	32
11	5	11-8-4-5	30	11-10-9-6-5	26
12	6	12-7-5-6	32	12-14-10-9-6	23
14	5	14-8-4-5	31	14-10-9-6-5	26

Table 4: Comparison of all non-optimal simulation routes with actual optimal route.

7. DISCUSSION

All the source-destinations paths in the nine node grid network were valid and optimal (see Table 2). For the

fourteen node NFSNET-network all of the source-destination paths selected were valid and 92.3% contained the optimum route (see Table 3). This rate of optimum route selection is higher than any currently published rate. The neural network was able to find optimal routes where the path was fairly short. The 7.7% of nonoptimal routes for the NFSNET network were mainly longer routes which contained many hops. As the data shows in Tables 3 & 4, the neural network had greater difficulty converging to a global minimum when the optimum route took more hops to get from the source to destination than did another valid route. Usually, the neural network will favour a route which requires a smaller number of hops, even though the optimum route takes more hops.

Frustrated systems like Spin-glass and the Travelling Salesman Problem are known to possess a hierarchy structure in the state space [10]. Thus, it is easy to see that the routing problem would suffer from the same problem. Such an ultrametric state space is characterized by "families" of solutions, in which all the states are close in some measure, e.g. Hamming distance. These different families are hierarchically connected. The distance between nodes in a tree is defined by the number of branches that must be traversed in order to move from one node to the other. Thus, the distance between states in different families is larger than the distance between states within families.

Currently, we are investigating ways of ensuring that all routes are optimal routes by using a hill climbing term such as that presented in [12] to prevent the neural network from getting trapped in local minima. Also, more tests on larger networks are being done to ensure that the neural network will always converge to valid routes no matter how large the communication network.

8. CONCLUSION

We have investigated the performance of a mean field theory neural network algorithm which does not use constraint terms for solving the routing problem in a communication network. The two advantages of using mean field annealing are the ability to find "good" routes quickly in a communication network, and the elimination of the constraint terms in the energy function which should help overcome the scaling problem. So far the results are very encouraging and more testing will have to be done on larger communication networks.

9. REFERENCES

[1] V.B. Aiyer, M. Niranjan & F. Fallside, "A Theoretical Investigation into the Performance of the Hopfield Model," *IEEE Transactions on Neural Networks*, Vol. 1, No. 2, pp. 204-215, 1990.

[2] J. Alspector, R. Goodman & T.X. Brown, *Proceedings of the International Workshop on Applications of Neural Networks to Telecommunications*, Lawrance Erlbaum Associates, Inc., Hillsdale, New Jersey, 1994.

[3] A.R. Bizzarrie, "Convergence Properties of a Modified Hopfield-Tank Model," *Biological Cybernetics*, Vol. 64, pp. 293-300, 1991.

[4] A. Cichocki & R. Unbehauen, *Neural Networks for Optimization and Signal Processing*, John Wiley & Sons, 1993.

[5] J.J. Hopfield & D.W. Tank, "Neural Computations of Decision in Optimization Problems," *Biological Cybernetics*, Vol. 52, pp. 141-152, 1986.

[6] F. Kamoun & M.K. Ali, "A Neural Network Shortest Path Algorithm for Optimum Routing in Packet-Switched Communications Networks," *in Proceedings of the Conference Globecomm*, Phoenix, Arizona, pp. 0120-0124, 1991.

[7] Y.C. Ouyang, *Development of a Neural Network Based Routing Algorithm for Communication Networks*, PhD Thesis, Memphis State University, USA, 1992.

[8] Peterson & B. Soderber, "A New Method for Mapping Optimization Problems onto Neural Networks," *International Journal of Neural Systems*, Vol. 1, pp. 3-22, 1989.

[9] C. Peterson, "Parallel Distributed Approached to Combinatorial Optimization: Benchmark on Travelling Salesman Problems," *Neural Computation*, Vol. 2, pp. 261-269, 1990.

[10] R. Rammal, G. Toulouse and M.A. Virasoro, "Ultrametricy for Physicists", *Reviews of Modern Physics*, Vol. 58, No. 3, 1986.

[11] H.E. Rauch & T. Winarske, "Neural Networks for Routing Communication Traffic", *IEEE Control Systems Magazine*, Vol. 8, No. 2, pp. 26-31, April 1988.

[12] Y. Takefuji & K. Lee, "An Artificial Hysteresis Binary Neuron: A Model Suppressing the Oscillatory Behavior of Neural Dynamics," *Biological Cybernetics*, Vol. 64, pp. 353-356, 1991.

[13] S.C.A. Thomopoulos, L. Zhang & C.D. Wann, "Neural Network Implementation of the Shortest Path Algorithm for Traffic Routing in Communication Networks," *in Proceedings of the IEEE International Joint Conference on Neural Networks*, Singapore, pp.2693-2702, 1991.

[14] D.E. Van den Bout & T.K. Miller, "Improving the Performance of the Hopfield-Tank Neural Network Through Normalization and Annealing", *Biological Cybernetics*, Vol. 62, pp. 129-139, 1989.

[15] D.E. Van den Bout & T.K. Miller, "Graph Partitioning Using Annealed Neural Networks", *Biological Cybernetics*, Vol. 62, pp. 129-139, 1990.