Extending the Power and Capacity of Constraint Satisfaction Networks *

Xinchuan Zeng and Tony R. Martinez
Computer Science Department, Brigham Young University, Provo, Utah 84602
Email: zengx@axon.cs.byu.edu, martinez@cs.byu.edu

Abstract

This work focuses on improving the Hopfield network for solving optimization problems. Although much work has been done in this area, the performance of the Hopfield network is still not satisfactory in terms of valid convergence and quality of solutions. We address this issue in this work by combing a new activation function (EBA) and a new relaxation procedure (CR) in order to improve the performance of the Hopfield network. Each of EBA and CR has been individually demonstrated capable of substantially improving the performance. The combined approach has been evaluated through 20,000 simulations based on 200 randomly generated city distributions of the 10*-city traveling salesman problem. The result shows that combining the two methods is able to further improve the performance. Compared to CR without combining with EBA, the combined approach increases the percentage of valid tours by 21.0% and decreases the error rate by 46.4%. As compared to the original Hopfield method (using neither EBA nor CR), the combined approach increases the percentage of valid tours by 245.7% and decreases the error rate by 64.1%.*

Introduction

The work of Hopfield and Tank [1] showed that neural networks can be applied to solving combinatorial optimization problems. They suggested that a near-optimal solution of *traveling salesman problem* (*TSP*) can be obtained by finding a local minimum of an appropriate *energy function*, which is implemented by a neural network. For an N-city *TSP*, the network consists of $N \times N$ neurons and the links that connect these neurons. The weights are set to encode the information about the constraints and the cost function of a particular city distribution of *TSP*. Each neuron updates its input value based on the information received from all other neurons. They showed that the neural network can often find a near-optimal solution in a short time. This network is commonly referred to as the *Hopfield network*. The advantages of the Hopfield network over other heuristic methods for solving combinatorial optimization problems include massive parallelism and convenient hardware implementation. Another advantage is that the procedure of the Hopfield network is more general for applications. There is an *ad hoc* procedure for mapping the constraints and cost function into the weight settings of the network. This general procedure can be applied to solve many different types of combinatorial optimization problems.

Since Hopfield and Tank showed that neural computation can be effective for solving combinatorial optimization problems, some work has been done to improve the performance of the Hopfield network. The research focuses on analyzing and improving the original model in order to obtain a higher percentage of valid solutions and solutions with better quality. The work by Wilson and Pawley [2] showed that there was some difficulty getting the Hopfield model to yield valid tours. For randomly generated sets of the 10-city *TSP*, Wilson and Pawley reported that only 8% of their trials resulted in valid tours. After their report, Brandt et al. [3] and Aiyer et al. [4] showed that better performance can be achieved by modifying the energy function. Li combined the Hopfield network with the "augmented Lagrange multipliers" algorithm from optimization theory [5]. Catania et al. applied a fuzzy approach to tune the parameters in the Hopfield network [6].

Although the performance of the Hopfield network has been improved over the past decade, this model still has some basic problems [7, 8]. One of the problems is that the performance of the Hopfield network is inconsistent. The performance is good for some city distributions of *TSP*, but the performance is poor for other city distributions with the same size. The performance is usually better for city distributions with simple topologies but poor for those with complex topologies, where the solutions are often trapped in poor local minima

*To appear in *International Joint Conference on Neural Networks, 1999.*

or even invalid. Another problem is that the performance is sensitive to the choice of the parameters and initial input values of neurons. Different parameters in the energy function can lead to significant differences in the performance. For the same set of parameters, different settings of random noise in the initial input values (a small fraction of random noise is necessary to break the symmetry of the network) can yield solutions with varying quality or invalid solutions.

In this paper, we study an approach of combining a new activation function (EBA) and a new relaxation procedure (CR) and its effects on the performance of the Hopfield network. Previous work has shown that EBA can reduce the effects of noise [9] and CR can lead to a smoother relaxation process [10]. Both are capable of significantly improving the performance in terms of both valid tours and qualities of tours. This work shows that the performance can be further improved by combining the two methods. We have conducted a large number of simulations to evaluate the performance of the combined approach. Based on the results of 20,000 simulations on 200 randomly generated city distributions of the 10-city TSP, the combined approach can increase the percentage of valid tours by 21.0% and decrease the error rate by 46.4% as compared to CR without combining with EBA. As compared to the original Hopfield method (without using either CR and EBA), the combined approach can increase the percentage of valid tours by 245.7% and decrease the error rate by 64.1%.

Background of Hopfield Network

The Hopfield network [1] includes a set of neurons and the links that connect the neurons. For an N-city *TSP*, there are $N \times N$ fully connected neurons in the network, in which the row index represents the city and the column index represents the order of the city in the tour. The weights of the connecting links are determined by the constraints and the cost function.

The above constraints and the cost can be represented by an energy function, which is used to determine the connecting weights between neurons. Hopfield's original energy function for an N-city *TSP* is given by [1]:

$$E = \frac{A}{2}\sum_{X=1}^{N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} V_{Xi}V_{Xj} + \frac{B}{2}\sum_{i=1}^{N}\sum_{X=1}^{N}\sum_{Y=1,Y\neq X}^{N} V_{Xi}V_{Yi} + \frac{C}{2}\left(\sum_{X=1}^{N}\sum_{i=1}^{N} V_{Xi} - N_0\right)^2 + \frac{D}{2}\sum_{X=1}^{N}\sum_{i=1}^{N}\sum_{Y=1,Y\neq X}^{N} d_{XY}V_{Xi}(V_{Y,i+1}+V_{Y,i-1}) \quad (1)$$

where X, Y are row indices, and i, j are column indices, V_{Xi} is the activation for neuron (X, i), and d_{XY} is the distance between cities X and Y. The first three terms enforce the constraints for a valid tour, and the last term represents the cost function for obtaining a short tour. The value of each parameter (A, B, C, and D) measures the importance of the corresponding term.

Each neuron (X, i) has an input value U_{Xi} and an activation (output) value V_{Xi}. The connecting weight between neuron (X, i) and (Y, j) is set according to:

$$W_{Xi,Yj} = -A\delta_{XY}(1-\delta_{ij}) - B\delta_{ij}(1-\delta_{XY}) - C - Dd_{XY}(\delta_{j,i+1}+\delta_{j,i-1}) \quad (2)$$

where δ_{ij} is equal to 1 if $i = j$, and equal to 0 otherwise.

Each neuron (X, i) is also connected to an external input current:

$$I_{Xi} = CN_0 \quad (3)$$

Before the relaxation of the network, the initial value of each U_{Xi} is set to be a constant value (determined by the condition: $\sum_{X=1}^{N}\sum_{i=1}^{N} V_{Xi} = N$) and is then perturbed with a small random noise value to break the symmetry of the network.

During relaxation, each neuron updates its input and activation value based on the weighted activations of other neurons and its own value. Specifically, the value $U_{Xi}^{(n+1)}$ at iteration step *(n+1)* is given by:

$$U_{Xi}^{(n+1)} = U_{Xi}^{(n)} + \Delta U_{Xi} \quad (4)$$

where $U_{Xi}^{(n)}$ is the input value at iteration step *(n)*. The value of ΔU_{Xi} is given by the following equation:

$$\Delta U_{Xi} = \left(-\frac{U_{Xi}}{\tau} + \sum_{Y=1}^{N}\sum_{j=1}^{N} W_{Xi,Yj}V_{Yj} + I_{Xi}\right)^{(n)}\Delta t \quad (5)$$

where $\tau(= RC)$ is the time constant of an RC circuit and was set to be 1.0 by Hopfield and Tank [1].

The activation $V_{Xi}^{(n+1)}$ at iteration step *(n+1)* is then determined by $U_{Xi}^{(n+1)}$ through an *activation (output) function*. In the Hopfield network, the activation function is the *sigmoid* function:

$$V_{Xi}^{(n+1)} = \frac{1}{2}\left(1 + tanh\left(\frac{U_{Xi}^{(n+1)}}{u_0}\right)\right) \quad (6)$$

where u_0 is the amplification parameter that reflects the steepness of the activation function.

Hopfield and Tank [1] showed that the network is guaranteed to converge to a local minimum in the case of symmetric ($W_{Xi,Yj} = W_{Yj,Xi}$) connecting weights.

Evidence Based Activation Function

We have analized the effects of the activation function on the performance of the Hopfield network and found that the activation function has an important impact on the performance [9]. The original activation function, i.e., the *sigmoid* function given in Eq. (6), is sensitive to random noise in the initial input values of neurons in the network. We proposed a new activation function – *evidence based activation* (EBA) function – which changes its activation significantly only when substantial positive or negative evidence is summed into a neuron. EBA has the following form [9]:

$$V_{Xi} = \frac{0.5(1 + tanh(\frac{U_{Xi}+x_0}{u_0}))}{1 + tanh(\frac{x_0}{u_0})} \qquad (U_{Xi} < 0)$$

$$V_{Xi} = \frac{tanh(\frac{x_0}{u_0}) + 0.5(1 + tanh(\frac{U_{Xi}-x_0}{u_0}))}{1 + tanh(\frac{x_0}{u_0})} \qquad (U_{Xi} \geq 0) \quad (7)$$

Fig. 1 shows the shapes of EBA with different threshold x_0 compared to the *sigmoid* function. The steep part of *sigmoid* function (curve "A") is around the region where U_{Xi} is close to zero, and thus it puts heavy emphasis on minor noise perturbation instead of the useful signals related to the constraints and cost encoded in the network.

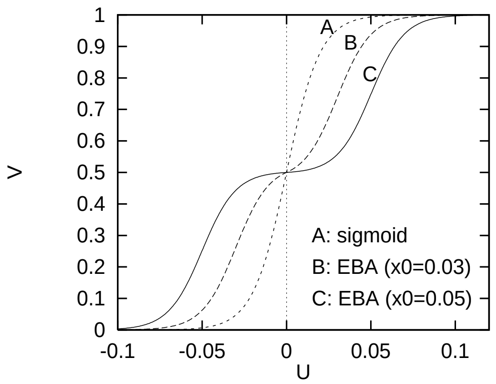

Figure 1: Comparison of the *sigmoid* function to the evidence based activation (EBA) with different thresholds (x_0).

In contrast, EBA (curve "B" and "C") has an adjustable threshold x_0 and becomes steep only when the absolute value of V_{Xi} is larger than the threshold. Thus random noise smaller than the threshold will not be amplified, while useful signals between neurons, which usually have larger magnitudes than noise, will be amplified. This mechanism has the capability of reducing the effects of noise in the formation of tours.

Simulation on 200 randomly generated city distributions of the 10-city *TSP* has shown that EBA increases the percentage of valid tours by 38.6% and reduces the error rate of tour length by 30.6% as compared to the *sigmoid* function.

Controlled Relaxation

We have proposed an approach [10] that uses a new relaxation procedure – *controlled relaxation* (CR) – to replace the original one. Unlike the original procedure that first updates the input and then the activation for each neuron (using Eq. (2) - (6)), CR updates the activation directly without using the update of the input as an intermediate step.

For convenience of describing CR, we can reformulate Eq. (5) into the form:

$$\Delta U_{Xi} = (-\frac{U_{Xi}}{\tau} + Net_{Xi})^{(n)} \Delta t \qquad (8)$$

where Net_{Xi} is the *net input value* received by the neuron (X, i) from other neurons, and can be written as:

$$Net_{Xi} = \sum_{Y=1}^{N} \sum_{j=1}^{N} W_{Xi,Yj} V_{Yj} + I_{Xi} \qquad (9)$$

In CR, the activation value $V_{Xi}^{(n+1)}$ for neuron (X, i) at iteration $(n + 1)$ is updated according to:

$$V_{Xi}^{(n+1)} = V_{Xi}^{(n)} + R(T_{Xi} - V_{Xi}^{(n)}) \qquad (10)$$

where T_{Xi} is the *target value* for the activation V_{Xi}, and $R (0.0 < R < 1.0)$ is a *relaxation rate*. T_{Xi} is derived from Net_{Xi} by using the *sigmoid* function:

$$T_{Xi} = \frac{1}{2}(1 + tanh(\frac{Net_{Xi}}{u_0'})) \qquad (11)$$

where u_0' is the steepness parameter (its role is similar to that of u_0 in Eq. (6)).

One major difference between the original and CR is the following. The original procedure first updates the input value by Eq. (4), and then updates the activation value through the *sigmoid* function given by Eq. (6). CR however directly updates the activation value using Eq. (10) without using the update of the input as an intermediate step.

Another major difference is that CR has a relaxation rate to gain a better control over the relaxation process, for which there is no direct analog in the original procedure.

Simulation on 200 randomly generated city distributions of the 10-city *TSP* has shown that the new relaxation procedure increases the percentage of valid tours by 185.7% and reduces the error rate of tour length by 33.1% as compared to the original procedure.

Combining EBA and CR

In this work, we combine EBA and the CR, and examine how combining can further improve performance.

The combined approach works as follows. We use exactly the same steps as those in CR except that we use EBA, instead of the *sigmoid* function, as the activation function. Specifically, the combined approach uses Eq. (7) as the activation function, rather than Eq. (6). That is, the combined approach method replaces Eq. (11) (the *sigmoid* function) by the following function (using EBA in Eq. (7)):

$$T_{Xi} = \frac{0.5(1 + tanh(\frac{Net_{Xi} + x_0'}{u_0'}))}{1 + tanh(\frac{x_0'}{u_0'})} \qquad (U_{Xi} < 0)$$

$$T_{Xi} = \frac{tanh(\frac{x_0'}{u_0'}) + 0.5(1 + tanh(\frac{Net_{Xi} - x_0'}{u_0'}))}{1 + tanh(\frac{x_0'}{u_0'})} \qquad (U_{Xi} \geq 0) \tag{12}$$

All the other steps in the combined approach are the same as those using CR.

Simulation Results

We have evaluated the performance of combining EBA and CR through simulations. The evaluation in our experiment was based on 200 randomly generated 10-city *TSP* city distributions, including wide varieties of topologies.

Many previous studies used only one city distribution (the one used by Hopfield and Tank in their original paper [1]) or a small number of city distributions (for example, 10 city distributions in [2]) in their simulations. This may lead to unreliable conclusions when comparing two algorithms. The reason is that the performance of an algorithm often depends on the topology in a city distribution, and different algorithms may favor different types of topologies. Using a large number of city distributions can reduce this effect and allow a better evaluation of the algorithms.

In the simulation, 100 runs are conducted for each of the 200 city distributions. For each of 100 runs, different random noise is added to the initial input values of the neurons.

For a fixed set of parameters (dt, u_0, etc), the quantities to be evaluated (error rate and percentage of valid tours) are first averaged over 100 runs for each city distribution, and then averaged over the entire 200 city distributions. Thus, 20,000 runs are needed to obtain each data point that is shown in the figures of the simulation results.

We experimented with different sets of 20,000 runs for a fixed set of parameters. The results show that the estimated quantities are fairly stable. Their values vary within a range of about 1% among different sets of 20,000 runs. This demonstrates that the number of runs in our simulation is large enough to make a reasonable estimation of the evaluated quantities.

The original energy function of the Hopfield network is used in the simulation, and the parameters in the energy function are those used by Hopfield and Tank [1]:

$$A = B = D = 500, \quad C = 200, \quad N_0 = 15 \tag{13}$$

The value of dt in Eq. (5) is set to be 10^{-5}, and the value of u_0 is fixed at 0.02 for both Eq. (6) and Eq. (7).

The fraction of random noise in the initial values of neurons is set to be 0.001 in the simulation. We tried several different values for the fraction in the range from 0.0001 to 0.01. The performance of the network is only slightly sensitive to this parameter, and the choice of 0.001 works slightly better than others.

Fig 2 shows the percentage of valid tours obtained by combining CR and EBA, compared to that using only CR without EBA, and also compared to that using neither CR nor EBA. Each data point in the figure is the weighted average percentage of valid tours over $N_{CityDist}$ (= 200) city distributions:

$$Valid = \frac{\sum_{i=1}^{N_{CityDist}} Valid_i}{N_{CityDist}} \tag{14}$$

where $Valid_i$ is the percentage of valid tours for city distribution i and is defined in the following.

For each city distribution i, there are a total of $N_{total,i}$ (=100) runs with different initial input values. The maximum number of iterations allowed for each run is set to be 1000. If a valid tour can not be reached within

1000 iterations, the network will stop and the tour is counted as invalid.

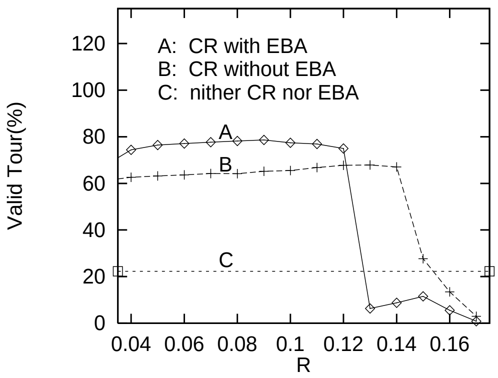

Figure 2: The percentage of valid tours obtained by combining the controlled relaxation (CR) and the evidence based activation (EBA) function (curve "A"), compared to that using only CR without EBA (curve "B"), and to that using neither CR nor EBA (curve "C").

If $N_{valid,i}$ is the number of valid tours among the total of $N_{total,i}$ runs, then $Valid_i$ is defined by:

$$Valid_i = \frac{N_{valid,i}}{N_{total,i}} \tag{15}$$

The first observation is that the percentage of valid tours achieved by the combined approach (curve "A") is higher than that by only using CR without combining EBA (curve "B") when $R \leq 0.12$, and their overall trends in performance are rather similar. Both are significantly better than that of the original network which uses neither CR nor EBA (curve "C").

A second observation is that although the combined approach performs better than the non-combined approach when $R \leq 0.12$, it performs worse when R becomes larger. This shows that there is a certain degree of difference in their response to the relaxation rate R, in addition to a similar overall trend. Despite this difference, the combined approach gives substantial improvement over the non-combined approach when $R \leq 0.12$.

Fig. 3 shows the error rate obtained by combining CR and EBA, compared to that using only CR without EBA, and also compared to that using neither CR nor EBA. Each data point is averaged over 100 different runs for each city distribution and then averaged over the 200 different city distributions as defined in the following.

For city distribution i, the error of a valid tour j is defined by:

$$Err_{i,j} = \frac{d_{i,j} - d_{i,opt}}{d_{i,opt}} \tag{16}$$

where $d_{i,j}$ is the tour length of a valid tour j and $d_{i,opt}$ is the optimal (shortest) tour length of city distribution i.

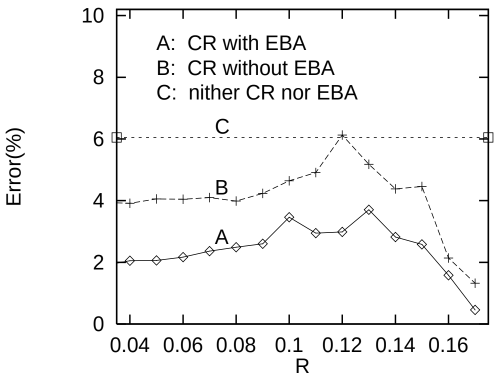

Figure 3: The error rate obtained by combining the controlled relaxation (CR) and the evidence based activation (EBA) function (curve "A"), compared to that using only CR without EBA (curve "B"), and to that using neither CR nor EBA (curve "C").

The error for city distribution i is the averaged error over all valid tours:

$$Err_i = \frac{\sum_{j=1}^{N_{valid,i}} Err_{i,j}}{N_{valid,i}} \tag{17}$$

The error shown in Fig. 3 is the average error of valid tours in all city distributions and is weighted by the percentage of valid tours for each city distribution:

$$Err = \frac{\sum_{i=1}^{N_{CityDist}} (Valid_i Err_i)}{\sum_{i=1}^{N_{CityDist}} Valid_i} \tag{18}$$

The results in Fig. 3 shows that all error rates obtained by the combined approach (curve "A") are smaller than those by the non-combined method (curve "B"). Both have much smaller error rates in overall than those using neither CR nor EBA (curve "C").

Thus, the results in Fig. 2 and Fig. 3 demonstrate that the combined approach performs better than the non-combined approach in terms of both the percentage of valid tours and the error rate. We can quantitatively calculate the improvement at $R = 0.06$ (approximately optimal value for both methods) and obtain the following result: the combined approach increases the percentage of valid tours by 21.0% (77.1% vs 63.7%) and reduces the error rate by 46.4% (2.17% vs 4.05%) as compared to the non-combined approach which uses only CR without EBA.

Similarly, if it is compared to the original Hopfield method (without neither CR nor EBA), the combined approach increases the percentage of valid tours by 245.7% (77.1% vs 22.3%) and reduces the error rate by 64.1% (2.17% vs 6.05%).

Summary

In this paper, we have studied the effects of combining EBA and CR on the performance of the Hopfield network for solving TSP. EBA has the capability of reducing the effects of noise, and CR enables a smoother relaxation process. Both EBA and CR have been shown previously capable of significantly improving the performance of the Hopfield network. This work demonstrates that combining the two methods can further improve performance.

The evaluation of the combined approach is based on a large number of simulations on 200 randomly generated city distributions of the 10-city TSP. The combined approach increases the percentage of valid tours by 21.0% and decreases the error rate by 46.4%, compared to those using only CR without combining EBA. As compared to the original Hopfield method (using neither CR nor EBA), the combined approach increases the percentage of valid tours by 245.7% and decreases the error rate by 64.1% (Fig. 1).

In future we plan to extend this research in the following directions: (i) evaluate its performance on other optimization problems in order to better evaluate its generality; (ii) use adaptive relaxation rates to replace constant relaxation rates in order to achieve a more optimal relaxation process; (iii) introduce a learning mechanism to determine the parameters in the network; (iv) experiment with the bipolar version of the EBA function and evaluate its performance; (v) apply it into other real-world application domains, in particular continuous speech recognition.

Acknowledgments

This research is funded in part by a grant from *fonix* Corp.

References

[1] Hopfield, J. J. and Tank, D. W.: Neural Computations of Decisions in Optimization Problems. Biological Cybernetics, vol. 52, pp. 141-152, 1985.

[2] Wilson, G. V. and Pawley, G. S.: On the Stability of the Traveling Salesman Problem Algorithm of Hopfield and Tank. Biological Cybernetics, vol. 58, pp. 63-70, 1988.

[3] Brandt, R. D., Wang, Y., Laub, A. J. and Mitra, S. K.: Alternative Networks for Solving the Traveling Salesman Problem and the List-Matching Problem. Proceedings of IEEE International Conference on Neural Networks, San Diego, CA. II: 333-340, 1988.

[4] Aiyer, S. V. B., Niranjan, M. and Fallside, F.: A Theoretical Investigation into the Performance of the Hopfield Model. IEEE Transactions on Neural Networks, vol. 1, no. 2, pp. 204-215, 1990.

[5] Li, S. Z.: Improving Convergence and Solution Quality of Hopfield-Type Neural Networks with Augmented Lagrange Multipliers. IEEE Transactions On Neural Networks, vol. 7, no. 6, pp. 1507-1516, 1996.

[6] Catania, V., Cavalieri, S. and Russo, M.: Tuning Hopfield Neural Network by a Fuzzy Approach. Proceedings of IEEE International Conference on Neural Networks, pp. 1067-1072, 1996.

[7] Cooper, B. S.: Higher Order Neural Networks-Can they help us Optimise?. Proceedings of the Sixth Australian Conference on Neural Networks (ACNN'95), pp. 29-32, 1995.

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

[9] Zeng, X. and Martinez, T. R.: A New Activation Function in the Hopfield Network for Solving Optimization Problems. Fourth International Conference on Artificial Neural Networks and Genetic Algorithms (*ICANNGA'99*), 1999.

[10] Zeng, X. and Martinez, T. R.: A New Relaxation Procedure in the Hopfield Network for Solving Optimization Problems. To appear in Neural Processing Letters, 1999.