A New Activation Function in the Hopfield Network for Solving Optimization Problems *

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 paper shows that the performance of the *Hopfield network* for solving optimization problems can be improved by using a new activation (output) function. The effects of the activation function on the performance of the Hopfield network are analyzed. It is shown that the *sigmoid* activation function in the Hopfield network is sensitive to noise of neurons. The reason is that the *sigmoid* function is most sensitive in the range where noise is most predominant. A new activation function that is more robust against noise is proposed. The new activation function has the capability of amplifying the signals between neurons while suppressing noise. The performance of the new activation function is evaluated through simulation. Compared with the *sigmoid* function, the new activation function reduces the error rate of tour length by 30.6% and increases the percentage of valid tours by 38.6% during simulation on 200 randomly generated city distributions of the 10-city *traveling salesman problem*.

1 Introduction

The subject of combinatorial optimization consists of a large set of important problems in computer science and engineering. A classical example of combinatorial optimization problems is the *traveling salesman problem* (*TSP*). *TSP* belongs to the class of *NP* problems. All exact methods known for determining an optimal tour require a computing effort that increases exponentially with N for an N-city *TSP* problem. Interest in solving combinatorial optimization problems by neural networks was motivated by the work of Hopfield and Tank [1]. They suggested that a suboptimal solution of *TSP* can be obtained by finding a local minimum of an appropriate *energy function*. The energy function can be implemented by a neural network. For a N-city *TSP* problem, 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]. Liang added adjusting neurons to the Hopfield network for solving the quadratic assignment problem [7].

Although the performance of the Hopfield net-

*In *Proceedings of the International Conference on Neural Networks and Genetic Algorithms, 1999.*

work has been improved over the past decade, this model still has some basic problems [8, 9]. 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 topology. However, for city distributions with complex topology, the solutions are often trapped in poor local minima, or the solutions are invalid. Another problem is that the performance is sensitive to the choice of the parameters and the initial input values of neurons. Different values of the 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 show that one important reason for the inconsistency of the Hopfield network is the non-robust *sigmoid* activation function which serves to map the current state input of each neuron to its activation (output). The *sigmoid* function is most sensitive in noise dominant range, and thus reduces the effects of signals between neurons. A new activation function is proposed and evaluated in this paper. The new activation function has different shape and has an adjustable parameter to control the threshold for the sensitive region. Thus it is more robust against noise and can amplify the signals between neurons while suppressing noise. In the simulation on 200 randomly generated city distributions of the 10-city *TSP*, the new activation function reduces the error rate of tour length by 30.6% and increases the percentage of valid tours by 38.6% as compared with the *sigmoid* function. It shows that the Hopfield network using the new activation function can yield solutions with better quality and a higher percentage of valid solutions.

2 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. The row index for a neuron represents the city and the column index represents the order of the city in the tour. The weights of the connecting links are determined according to the constraints and the cost function. *TSP* includes the following constraints: each city must be visited exactly once in a tour (each row has exactly one "on" neuron) and exactly one city is visited at any time in a tour (each column has exactly one "on" neuron). The cost function is constructed to reflect the objective of finding a valid tour with minimum tour length.

The constraints and the cost function for *TSP* can be represented by an energy function. The energy function is then used to determine the values of all weights in the network. Hopfield's original energy function for an N-city *TSP* is given by [1]:

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

where X and Y are row indices; i and j are column indices; V_{Xi} is the activation for each neuron; and d_{XY} is a measure of the distance between cities X and Y. The first two terms enforce the constraint that no city can be visited more than once. The third term reflects the constraint that each city should be visited. The last term represents the distance cost function. The value of each parameter (A, B, C, and D) is the measure of the importance of the corresponding term.

Each neuron has a current state input U_{Xi} and an activation (output) V_{Xi}. The initial values of U_{Xi} for all neurons are first set to be a small constant value which is determined by the condition: $\sum_{X=1}^{N}\sum_{i=1}^{N} V_{Xi} = N$. To break the symmetry of the network, a small fraction of random noise is added to the initial values of each U_{Xi}. The network is then relaxed until an equilibrium is reached.

After each iteration, U_{Xi} is updated based on the following equation:

$$\frac{dU_{Xi}}{dt} = -\frac{U_{Xi}}{\tau} - \frac{\partial E}{\partial V_{Xi}} \quad (2)$$

where $\tau = 1.0$ is the time constant.

V_{Xi} is also updated after each iteration and its value is determined by U_{Xi} through an *activation (output) function*. In the Hopfield network, the activation function is the *sigmoid* function given by:

$$V_{Xi} = \frac{1}{2}(1 + tanh(\frac{U_{Xi}}{u_0})) \quad (3)$$

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 connecting weights. They tested this model for a 10-city *TSP*. For a total of 20 runs, they claimed 16 valid tours, and half of the valid tours were optimal.

3 New activation function

The quality of the solutions obtained by the Hopfield network depends on the random noise in the initial input values of neurons in the network. In the simulation reported by Wilson and Pawley [2], 10 sets of randomly produced 10-city *TSP* coordinates are tested. There were a total of 50 runs with different initial input values for each city distribution. They reported that 8% of the runs converged to a valid tour, 48% of the runs froze into invalid tours, and 44% of the runs did not converge in 1000 iterations. They showed that the quality of the solutions is sensitive to the initial input values of neurons. Different sets of initial values for a fixed set of parameters can yield solutions with different quality or even invalid solutions.

We propose that one important factor for the above problem is the activation function. The activation function in the Hopfield network is the *sigmoid* function as expressed in Eq. (4). The steep part of this function is around the region where U_{Xi} is close to zero, as shown by the curve labeled A in the Fig. 1. Thus this activation function puts too much emphasis on minor noise perturbation instead of the signals related to the cost and the constraints encoded in the network.

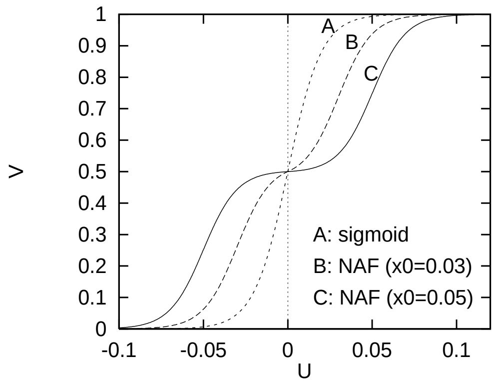

Figure 1: Comparison of the *sigmoid* and the new activation function (NAF) with different thresholds (x_0).

Based on the above analysis, we propose the following new activation function:

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

$$V_{Xi} = \frac{tanh(\frac{x_0}{u_0}) + 0.5(1 + tanh(\frac{U_{Xi}-x_0}{u_0}))}{1 + tanh(\frac{x_0}{u_0})} \qquad (U_{Xi} \geq 0) \tag{4}$$

where x_0 represents the threshold for V_{Xi} to become steep, and u_0 measures the steepness of the activation function. The shape with differing values of x_0 is shown in Fig. 1. The new activation function becomes steep only when the absolute value of V_{Xi} is larger than x_0. Thus random noise smaller than the threshold will not be amplified, while the signal between neurons, which usually has a larger magnitude than noise, will be amplified. Only when substantial positive or negative evidence is summed into a neuron, will the neuron make a significant change in its activation function. This mechanism has the capability of reducing the effects of noise in the formation of tours.

4 Simulation results

The performance of the new activation function is evaluated by simulation and is compared with that of the *sigmoid* function. Evaluation is accomplished with 200 randomly generated 10-city *TSP* city distributions, including wide varieties of distributions. The performance of different algorithms usually depends on the topology of the city locations in a city distribution, and different algorithms may favor different types of distributions. Using a large number of city distributions can reduce this effect and obtain a better evaluation of the algorithms. For each of the 200 city distributions, there are 100 runs with different settings of random noise added to the initial input values of the neurons. Each evaluated parameter is first averaged over 100 runs for each city distribution and then averaged over 200 city distributions.

The original energy function of the Hopfield network is used for both the case using the *sigmoid* function and the case using the new activation function. To evaluate specific effects of the different activation functions, the same set of parameters in the energy function are used in both cases. The values of the parameters are those by Hopfield and Tank [1]:

$$A = B = D = 500, \quad C = 200, \quad N_0 = 15 \tag{5}$$

The value of dt in Eq. (2) is set to be 10^{-5}. The value of u_0 is fixed at 0.02 for both Eq. (3) and Eq. (4). The fraction of random noise in the initial values of neurons is set to be 0.001.

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.

Fig. 2 shows the errors of the *sigmoid* function and the new activation function with different thresholds. For the city distribution i, the error of a valid tour j is defined by:

$$Err_{i,j} = \frac{d_{i,j} - d_{i,opt}}{d_{i,opt}} \tag{6}$$

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

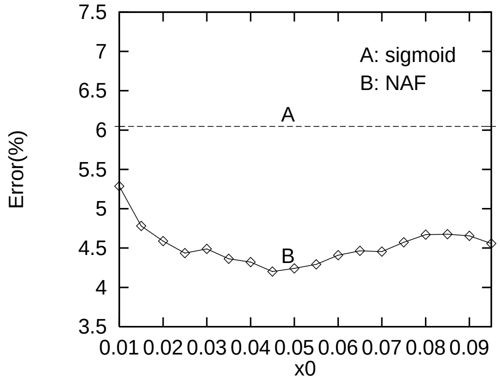

Figure 2: The errors (%) of the new activation function (with different threshold x_0) compared to that of the *sigmoid* function.

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

$$Err_i = \frac{\sum_{j=1}^{N_{valid,i}} Err_{i,j}}{N_{valid,i}} \tag{7}$$

where $N_{valid,i}$ is the number of valid tours among the total of $N_{total,i}$ runs with different initial input values for the city distribution i. In the simulation $N_{total,i} = 100$ for all city distributions.

The percentage of valid tours for city distribution i is defined by:

$$Valid_i = \frac{N_{valid,i}}{N_{total,i}} \tag{8}$$

The error shown in Fig. 2 is the average error of valid tours in all city distributions. It is weighted by the percentage of valid tours for each city distribution and is given by:

$$Err = \frac{\sum_{i=1}^{N_{CityDist}} (Valid_i Err_i)}{\sum_{i=1}^{N_{CityDist}} Valid_i} \tag{9}$$

where $N_{CityDist}$ is the total number of city distributions and is equal to 200.

The result in Fig. 2 shows that the error rate of the new activation function is lower than that of the *sigmoid* function, i.e. the tours obtained by using the new activation function have higher quality (shorter tour length). The improvement depends on the value of the threshold x_0. When the threshold increases from 0.01 to 0.045, the error rate decrease from 5.29% to 4.20%. This supports our conjecture that a higher threshold can suppress the effects of noise while still amplifying signals. The signal that represents the quality of the tour is from the last term in the energy function (1), which reflects the distance between the cities. When the threshold increases further, the error rate starts to grow. It implies that the threshold is so high that some signals as well as noise is suppressed by the activation function. The threshold should be properly chosen in order to reduce noise and keep the signals. The value of the threshold x_0 with lowest error rate is about 0.045, and the corresponding error rate is 4.20%. Compared with the error rate of 6.05% when using the *sigmoid* function, the error rate is reduced by 30.6%.

Fig. 3 shows the percentages of valid tours of the *sigmoid* function and the new activation function with different thresholds.

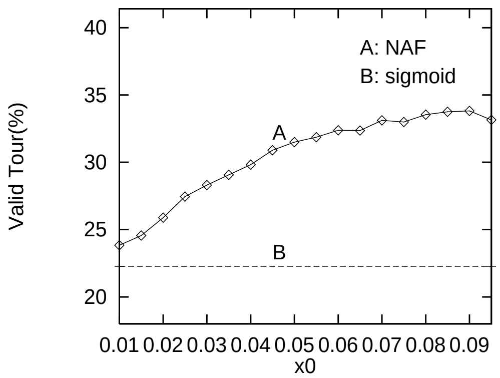

Figure 3: The percentages of valid tours of the new activation function (with different threshold x_0) compared to that of the *sigmoid* function.

The percentage of valid tours is the weighted average over 200 city distributions:

$$Valid = \frac{\sum_{i=1}^{N_{CityDist}} Valid_i}{N_{CityDist}} \tag{10}$$

The percentage of valid tours obtained by the new activation function is higher than that by the *sigmoid* function. When x_0 changes from 0.01 to

0.06, the percentages of valid tours increases from 23.8% to 32.4%. This suggests that the new activation function amplifies the signals that represent the constraints for a valid tour while reducing noise. The signals that represent the constraints are from the first three terms in the energy function (1). When the threshold increases further, the percentage of valid tours no longer increases. If the threshold is too high, the error rate starts to increase as shown in Fig. 2. The threshold should be properly chosen considering the trade off between the error rate and the percentage of valid tours. For the parameters in the simulation, the optimal value of x_0 is about 0.045. The corresponding percentage of valid tours is 30.9%. Compared with that of the *sigmoid* function (22.3%), the percentage of valid tours is increased by 38.6%.

5 Summary

In this paper, we have analized the effects of the activation function on the performance of the Hopfield network for solving *TSP* and proposed that the inconsistency of the performance of the Hopfield network is related to the activation function. The analysis reveals that the *sigmoid* activation function in the Hopfield network is sensitive in the region where noise is predominant and thus has negative effects on the performance of the network. We proposed a new activation function that is more robust against noise. The new activation function has an adjustable parameter that controls the threshold of input to be amplified, and it has the capability of amplifying the signals while suppressing noise. The proposed new activation function has been shown to outperform the *sigmoid* function based on the simulation results. In simulation on 200 randomly generated city distributions of the 10-city *TSP*, the new activation function reduces the error rate of tour length by 30.6% and increases the percentage of valid tours by 38.6% as compared to the *sigmoid* function, showing the capability of the new activation function to improve the performance of the Hopfield network.

The new activation function has also been applied in a study of speech recognition. Preliminary results show that it is also capable of improving the performance in this domain. In future we plan to evaluate its performance on other optimization problems in order to better evaluate the generality of this approach. Another extension which we are currently pursuing is to apply a dynamic activation function with varying parameters in the Hopfield network.

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] Liang, Y.:Combinatorial Optimization by Hopfield Networks Using Adjusting Neurons. Information Sciences, vol. 94, pp. 261-276, 1996.

[8] 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.

[9] 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.