A Simple Steepest Descent Method for Minimizing Hopfield Energy to Obtain Optimal Solution of the TSP with Reasonable Certainty

Y. Tachibana, G. H. Lee, H. Ichihashi & T. Miyoshi
College of Engineering, University of Osaka Prefecture
1-1 Gakuencho, Sakai, Osaka 593, Japan
ichi @ center. osakafu-u. ac. jp

Abstract

An effective algorithm for the Hopfield neural network model is proposed through its application to the traveling salesman problem. By the estimation of local minimum inside a hypercube of solution space, a threshold value for deciding integer valued solution can be properly chosen. An optimum solution of the TSP is obtained with reasonable certainty.

1. Introduction

Since Hopfield and Tank[3,4] showed an algorithm to solve the traveling salesman problem(TSP) by neural networks, many applications to combinatorial optimization problems have been reported. Unfortunately the Hopfield neural net does not seem likely to lead to better algorithms for finding acceptably low-cost minima for the TSP. The conventional algorithm already in use clearly outperforms many new approaches such as Hopfield neural net both in speed and the quality of solution found[5]. Therefore much effort has been made paying attention to achieving better solutions by modifying algorithms. Theoretical aspects on the stability of solution[1,12] in a multi-linear form of the energy function have been studied. It was shown that the Hopfield model is effective when it is endowed with an appropriate form and parameters of the energy function with the term restricting non-integer solution[6]. The conditions of asymptotic stability of solutions and the non-stability of non-solutions have been discussed recently [8]. All of these studies are based on the stability at a vertex of a hypercube of the solution space. Actually, in most cases, the local minimum of the energy function with general quadratic form exists inside the hypercube. Hence, in this study we propose a revised algorithm to obtain a better solution by selecting proper values of parameters of the energy function and by operating positions of local minimum inside the hypercube. By the estimation of local minimum inside the hypercube, a threshold value for deciding integer valued solution can be properly chosen, and an optimum solution of the TSP is obtained with reasonable certainty.

Another neural approach to the TSP is to apply the chaotic neural networks[2,9] which are obtained by introducing negative self-feedback connections to the Hopfield network. The ability to find optimal solution is drastically improved by the chaotic behavior of the neural networks. Quite recent study[11] by computer simulation insists that the solution ability can be obtained by a simple combination of steepest descent method and randomness instead of chaos. Our simulation results show that the quality of solution can be improved only by the revised algorithm of steepest descent method. The organization of this paper is as follows: In section 2, after briefly reviewing the Hopfield network and the TSP, we analyze the energy surface between a vertex corresponding to a solution and its adjoining vertex, and a revised algorithm is proposed. Section 3 presents a simulation study on performance of the algorithm to the TSP.

2. Solution Method to TSP by Minimizing Hopfield Energy

A typical neural network mathematical model proposed by Hopfield[3] is given by a system of ordinary differential equations as:

$$\frac{du_i}{dt} = \sum_j w_{ij} v_j + h_i \qquad \left(w_{ij} = w_{ji}\right)$$

with

$$v_i = s(u_i) = \frac{1}{2}\left(1 + \tanh\frac{u_i}{x_0}\right), \quad i = 1,\cdots,m$$

where, u_i is the intermediate state variable, $v_i \in (0,1)$ is the output of neuron i, h_i is the external input of neuron i, $w_{ij}\left(w_{ij} = w_{ji}\right)$ is the synaptic connection of neuron j to neuron i, $x_0(>0)$ is the gain scaling parameter of the sigmoid function s. The state of a network of N neurons is represented by a vector $v = (v_1,\cdots v_N) \in (0,1)^N$, namely by a point in a unit

0-7803-2768-3/95/$4.00 © 1995 IEEE

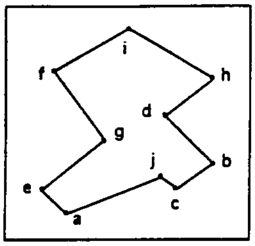
hypercube.

It has been shown by Hopfield[1] that the network converges to a set of stable states and these stable states correspond to a local minimum of the following energy function as:

$$E=-\frac{1}{2}\sum_i\sum_j w_{ij}v_iv_j-\sum_i h_iv_i$$

E certainly decreases as the time changes, and converges the stable equilibrium point. Strictly speaking, in case that the network operates with the initial value in a hypercube, it converges an asymptotic stable vertex or a local minimum inside the hypercube.

The TSP is a classic combinatorial optimization problem. The TSP is defined as the task of the salesman visiting all the N cities on his list once and once only, and returning to his starting point after traveling the minimum possible distance. We use the same expression for the solution of the TSP as Hopfield[1]. The neurons are arranged in $N\times N$ grid patterns, the visiting order is indicated by lateral position and the city names are indicated by longitudinal position, so each neuron is expressed by the suffix of city name $i,j=1,\cdots,N$ and the suffix of visiting order $k=1,\cdots,N$. Fig. 1 shows a route: c→j→a→e→g→f→i→h→d→b→c.



$$\begin{array}{cccccccccc|l}
 & & & & & & & & & & \text{city}\\
0&0&1&0&0&0&0&0&0&0&\text{a}\\
0&0&0&0&0&0&0&0&0&1&\text{b}\\
1&0&0&0&0&0&0&0&0&0&\text{c}\\
0&0&0&0&0&0&0&0&1&0&\text{d}\\
0&0&0&1&0&0&0&0&0&0&\text{e}\\
0&0&0&0&0&1&0&0&0&0&\text{f}\\
0&0&0&0&1&0&0&0&0&0&\text{g}\\
0&0&0&0&0&0&0&1&0&0&\text{h}\\
0&0&0&0&0&0&1&0&0&0&\text{i}\\
0&1&0&0&0&0&0&0&0&0&\text{j}\\
1&2&3&4&5&6&7&8&9&10&
\end{array}$$

visiting order

(a) Example of a city allocation (b) Allocation of variables

Fig. 1 Representation of a route in the TSP for N =10

The energy function $E(v)$ is given as:

$$E(v)=\frac{1}{2}\sum_i\left(\sum_k v_{ik}-1\right)^2+\frac{1}{2}\sum_k\left(\sum_i v_{ik}-1\right)^2 \qquad (1)$$
$$+\frac{A}{2}\sum_i\sum_k v_{ik}(1-v_{ik})+\frac{B}{2}\sum_i\sum_j\sum_k d_{ij}v_{ik}\left(v_{jk-1}+v_{jk+1}\right)$$

Where v_{ik} is the output of neuron ik at the discrete time, d_{ij} is the constant value of distance from the city j to i, A and B are positive constants, $v_{i0}=v_{iN}$ and $v_{iN+1}=v_{i1}$. The first term represents the constraint visiting all cities once for all. The second term represents the constraint visiting only one city at the same time. The third term is to restrict non-integer solution[6]. And the fourth term represents the total tour length.

In this paper the energy function of the TSP in Eq. (1) is regarded as an objective function of mathematical programming problem with continuous variables. The learning rule based on the steepest descent method is obtained as:

$$u_{ik}^{NEW}=u_{ik}^{OLD}-\tau\frac{\partial E(v)}{\partial u_{ik}}$$

$$\frac{\partial E(v)}{\partial u_{ik}}=\frac{\partial E(v)}{\partial v_{ik}}\cdot\frac{\partial v_{ik}}{\partial u_{ik}}$$

$$\frac{\partial E(v)}{\partial v_{ik}}=\left(\sum_m v_{im}-1\right)+\left(\sum_n v_{nk}-1\right)$$
$$+A\left(\frac{1}{2}-v_{ik}\right)+B\sum_i d_{ij}\left(v_{jk-1}+v_{jk+1}\right)$$

for all i and k, $i=1,\cdots,N,k=1,\cdots,N$, where $v_{ik}=s(u_{ik})$ and τ is the positive learning rate. For simplicity, we assume :

$$\frac{\partial v_{ik}}{\partial u_{ik}}=1.$$

A solution of the TSP(i.e., a feasible solution which forms a route) is obtained such that all v_{ik} equal 0 or 1, namely it is on the vertex $V=(V_{11},\cdots,V_{NN})\in\{0,1\}^{N\times N}$ of a hypercube where the first and the second terms of the energy function equal 0. Eq. (1) is a multi-linear function when A=2, and V converges to a certain vertex. But it is known that by an energy function of the multi-linear form, to obtain a better solution is relatively hard compared with the general energy function of quadratic form. Eq. (1) is not a multi-linear form when $A<2$, and V converges to the local minimum in a hypercube of the solution space. In order to obtain a solution corresponding to a vertex of the hypercube, we introduce two threshold values θ_U and θ_L for selecting values of v_{ik} from 1.0 or 0.0. When repeating the learning rule, if $v_{ik}\geq\theta_U$ then v_{ik} is set to 1.0 and if $v_{ik}\leq\theta_L$ then v_{ik} is set to 0.0. In short, we put v_{ik} on a side face, an edge, or a vertex of the hypercube if v_{ik} is outside the interval$[\theta_L,\theta_U]$.

For choosing a proper threshold value, we focus on the energy surface between a vertex $V\in\{0,1\}^{N\times N}$ of a hypercube corresponding to a solution and its adjoining vertex.

① If $V_{am}=1$ at a vertex V corresponding to a solution, there exist b and c such that $V_{bm-1}=V_{cm+1}=1$. Since the following equation:

$$\frac{\partial E}{\partial v_{am}}=2(v_{am}-1)+A\left(\frac{1}{2}-v_{am}\right)+B(d_{ab}+d_{ac}),$$

holds for $v_{am}\in[0,1]$,when

$$A<4-2B(d_{ab}+d_{ac})$$

we have

$$\left.\frac{\partial E}{\partial v_{am}}\right|_{v_{am}=0}=-2+\frac{1}{2}A+B(d_{ab}+d_{ac})<0.$$

Moreover, if

$$A<2B(d_{ab}+d_{ac})$$

is satisfied then we have

$$\frac{\partial E}{\partial v_{am}}\Big|_{v_{am}=1} = -\frac{1}{2}A + B\left(d_{ab}+d_{ac}\right) > 0.$$

Consequently, if A is a small positive number, v_{am} at a local minimum of the energy E exists inside the unit interval (0,1). v_{am} at the local minimum can be written as:

$$v_{am} = \frac{1}{2-A}\left\{2 - \frac{1}{2}A - B\left(d_{ab}+d_{ac}\right)\right\}.$$

θ_U should be smaller than this value. And, at the local minimum, we have

$$\frac{dv_{am}}{dA} = \frac{1 - B\left(d_{ab}+d_{ac}\right)}{(2-A)^2}.$$

Therefore, as A increases, v_{am} at the local minimum of E increases if $1 > B\left(d_{ab}+d_{ac}\right)$. Similarly as B decreases, v_{am} at the local minimum of E increases if $A < 2$. In case that $A > 2B\left(d_{ab}+d_{ac}\right)$, v_{am} at the local minimum of E is greater than 1, but v_{am} cannot attain the local minimum because $v_{am} = s\left(u_{am}\right)$. The strong nonlinear transformation of the sigmoid function seems to affect the solution. If B is too small, the distance of a traveling route is not minimized in Eq. (1).

②If $V_{am} = 0$ at a vertex V corresponding to a solution, we have

$$\frac{\partial E}{\partial v_{am}} = 2v_{am} + A\left(\frac{1}{2} - v_{am}\right) + B\left(d_{ab}+d_{ac}\right),$$

where $v_{am} \in [0,1]$. And,

$$\frac{\partial E}{\partial v_{am}}\Big|_{v_{am}=0} = \frac{1}{2}A + B\left(d_{ab}+d_{ac}\right) > 0.$$

If the following condition holds

$$A < 4 + 2B\left(d_{ab}+d_{ac}\right)$$

then we have

$$\frac{\partial E}{\partial v_{am}}\Big|_{v_{am}=1} = 2 - \frac{1}{2}A + B\left(d_{ab}+d_{ac}\right) > 0.$$

v_{am} at the local minimum is given as:

$$v_{am} = -\frac{1}{2-A}\left\{\frac{1}{2}A + B\left(d_{ab}+d_{ac}\right)\right\}.$$

Hence, a local minimum of E exists in the left side of the origin when A<2. θ_L can be chosen as $0 < \theta_L << 1$. When A or B increase, the local minimum moves to left, but v_{am} cannot attain the local minimum because $v_{am} = s\left(u_{am}\right)$ and the strong nonlinearity of the sigmoid function seems to affect the solution.

We see that the positive numbers A and B should be small and satisfy the conditions in ① and ② so that the local minimum exists near the vertex of the hypercube as close as possible. In this way the values of parameters A and B are decided, and if we choose the threshold value satisfying the following condition in case of ① :

$$\theta_U < v_{am}\text{(at the local minimum)}$$

we can expect obtaining a solution of the TSP even if v_{ik} converges to a local minimum point inside the hypercube.

3. Numerical experiment of the TSP

In this section we review the results of numerical experiment of the TSP for $N = 10$ and 24. The city coordinate values are normalized within the unit square $[0,1]^2$.

The city allocations used in the experiment is shown in Fig. 2[10]. The points from a to j represent the place of cities. The polygonal lines tying these cities represent the shortest paths, i.e., the optimal solutions.

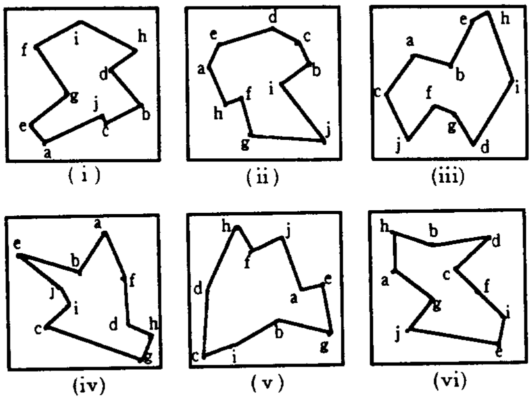


Fig. 2 City allocations used in the experiment

By the preparatory experiment the coefficients are decided as: $B = 0.5, x_0 = 1.0, \tau = 0.1$ and $\theta = (\theta_L, \theta_U) =$ (0.01,0.70). And the initial values of u_{ik} are chosen from interval$\left[-1.0\times10^{-2}, 1.0\times10^{-2}\right]$. The upper limit of the iteration number is set to 10,000. Namely we consider that we cannot get a solution when the number of iteration exceed the limit.

Fig. 3 shows the probability of obtaining an optimal solution corresponding to the values of coefficient A on (i)-(vi). Fig. 4 for city allocatishows the average error rate of the solution with respect to the value of coefficient A for city allocation (i)-(vi).

$$(\text{error rate}) = \frac{(\text{total route length}) - (\text{best route length})}{(\text{best route length})}$$

From these experimental results, it is shown that we are able to obtain an optimal solution with reasonable certainty only by the simple steepest descent method, if the coefficients A,B and the threshold values are chosen properly.

Double circle problems with 24 cities as in Fig. 5 are tested, where 12 cities are allocated on the inner circle and the rest on the outer circle. Optimal solutions of these problems are shown in Fig. 5. In Fig. 5, (a) is called " C" type and the radius of the inner circle is

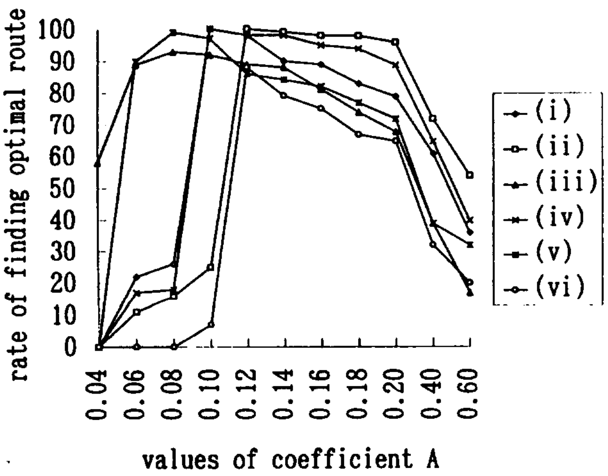


Fig. 3 Values of Coefficient A vs. probability of obtaining an optimal solution for city allocations(i)-(vi)

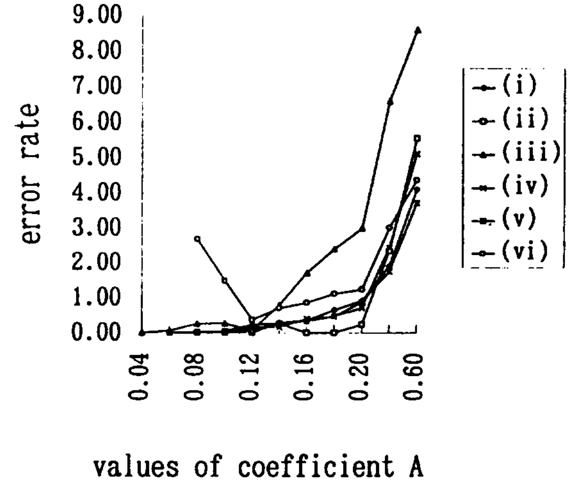


Fig. 4 Coefficient A vs. error rate for city allocations (i)-(vi)

small, (b) is called " O" type and the radius of the inner circle is big. These problems are known to be very hard to obtain the optimal solution by the original Hopfield neural networks and even by its revised version[12].

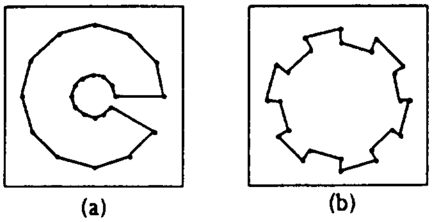


Fig. 5 Examples of the TSP for N=24
(a) External radius is 0.50 and internal radius is 0.15.
(b) External radius is 0.50 and internal radius is 0.45.

By the preparatory experiment, the coefficients are decided as $B=0.5, x_0=1.0, \tau=0.2$ and $\theta=(0.01, 0.70)$. And the initial value of u_{ik} is chosen randomly from interval $[-1.0\times10^{-2}, 1.0\times10^{-2}]$. The upper limit of the iteration numbers is set to 5,000.

Table 2. shows the experimental results of the 100 trials. From the experimental result for the "C" type problem of (a) in Fig. 5, it is shown that we are able to obtain the optimal solution with certainty only by the simple steepest descent method, if the coefficients A,B and the threshold values are chosen properly. And, from the experimental result for the "O" type problem of (b) in Fig. 5, it is shown that the error rate is very small if the coefficients A, B and the threshold values are chosen properly. Though the optimal solution could not be obtained, many routes which are similar to "O" type as in Fig. 6 were obtained.

Table 2. Results of the experiment for the TSPs of (a) and (b) in Fig. 5

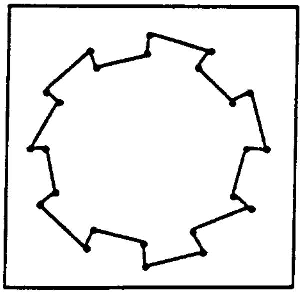
(a)

A	B	optimal solution	feasible solution	error rate
0.00	0.6	100	100	0.00
0.02	0.6	100	100	0.00
0.04	0.6	85	88	0.06
0.06	0.6	41	41	0.00
0.08	0.6	32	32	0.00
0.10	0.6	8	8	0.00
0.12	0.6	16	16	0.00
0.14	0.6	0	0	-
0.16	0.6	0	0	-
0.18	0.6	0	0	-
1.00	0.6	0	0	-

(b)

A	B	optimal solution	feasible solution	error rate
0.00	0.6	0	0	-
0.02	0.6	0	0	-
0.04	0.6	0	0	-
0.06	0.6	0	44	0.55
0.08	0.6	0	98	0.72
0.10	0.6	0	100	0.88
0.12	0.6	0	100	0.99
0.14	0.6	0	100	1.04
0.16	0.6	0	100	1.13
0.18	0.6	0	100	1.15
0.20	0.6	0	100	1.18

Fig. 6 Best route for "O" type problem



4. Conclusion

In this paper it is confirmed by numerical experiment that the optimal solution of the combinatorial optimization problem can be obtained with reasonable certainty only by the method of steepest descent. Though the proposed method is effective when the size of the combinatorial optimization problem is relatively small, this neural approach seems to have more advantages for the mixed integer programming problems such as clustering. Future reseach is directed to these problems, and the influence of nonlinear transformation by the sigmoid function is left for further study.

5. References

[1] S.Abe,"Determination of Weights of the Hopfield Neural networks and Its Evaluation",Transactions of Information Processing Society of Japan, Vol. 34, No. 1, pp. 21-28 (1993)

[2] K.Aihara, T. Takabe, M. Toyoda,"Chaotic Neural Networks", Physics Letter A, Vol. 144, No.6,7, pp 333-340 (1990)

[3] J.J.Hopfield,"Neurons with Graded Response Have Collective Computational Properties Like Those of Two-state Neurons", Proceedings of National Academy of Science of the United States of America, Vol. 81, pp. 3088-3092(1984).

[4] J.J.Hopfield and D.W.Tank,"Neural Computation of Decisions of Optimization Problems", Biological Cybernetics,Vol.52, pp. 141-152 (1985).

[5] D. Johnson, "More Approaches to the Traveling Salesman Guide", Nature, Vol. 330, No. 10, pp. 525-525 (1987)

[6]H. Kita, H. Odani, Y. Nishikawa, "Optimization by Means of a Neural Network Model -An Application to the Placement Problem," Transactions of the Institute of Systems, Control and Information Engineers, Vol. 4. No. 1, pp. 48-56 (1991)

[7] H. Kita," The Hopfield Neural Network and Simulated Annealing", Journal of Japan Society for Artificial Intelligence, Vol.7, No.6, pp. 970-979 (1992)

[8] S. Matsuda, "The Stability of the Solution in Hopfield Neural Network," Transactions of the Institute of Electronics,Information and Communication Engineers, Vol.J77-D-II, No. 7, pp. 1366-1344 (1994)

[9] H. Nozawa,"A Neural Network Model as a Globally Coupled Map andApplications Based on Chaos", Chaos, Vol. 2, No.3 pp.377-386 (1992)

[10] H. Nozawa,"Chaos and Optimization", Chaos Seminar, K. Aihara ed., Kaibundo, pp. 195- 221 (1994)

[11] Y. Obuchi,"On the Optimizing Ability of Chaos Neural Networks",Transactions of the Institute of Electronics, Information and Communication Engineers, Vol.J77-D-II,No. 11, pp. 2318-2321 (1994)

[12]Y.Uesaka,"Mathematical Aspects of Neuro-dynamics for Combinatorial Optimization", Transactions of the Institute of Electronics, Information and Communication Engineers, Vol. E74, No. 6, pp. 1368-1372(1991)