

論文97-34C-12-8

최적화용 신경망의 성능개선을 위한 새로운 최적화 기법

(A New Optimization Method for Improving the Performance of Neural Networks for Optimization)

趙 鏞 鉉 *

(Yong Hyun Cho)

요 약

본 논문에서는 기울기하강법과 동적터널링 시스템을 조합 사용함으로써 최적화용 신경망의 성능을 개선할 수 있는 새로운 방법을 제안하였다. 제안된 방법에서는 빠른 수렴속도의 전역최적화가 가능하도록 하기 위하여 기울기하강법의 경신규칙을 이용하고, 국소최적해를 만났을 때 이를 벗어난 다른 뉴런 출력치의 설정을 위하여 동적터널링 시스템의 경신규칙을 이용함으로써 빠른 수렴속도로 전역최적해로의 수렴확률을 높였다. 제안된 경신규칙의 신경망을 외판원 경로문제와 최적태스크 분할에 각각 적용하여 그 타당성과 성능을 확인한 결과, 기울기하강법의 경신규칙을 이용한 홉필드 모델에 비해 최적해로의 수렴확률과 그 수렴속도가 우수함을 확인할 수 있었다.

Abstract

This paper proposes a new method for improving the performances of the neural network for optimization using a hybrid of gradient descent method and dynamic tunneling system. The update rule of gradient descent method, which has the fast convergence characteristic, is applied for high-speed optimization. The update rule of dynamic tunneling system, which is the deterministic method with a tunneling phenomenon, is applied for global optimization. Having converged to the local minima by using the gradient descent method, the proposed method estimates the initial point for escaping the local minima by applying the dynamic tunneling system. The proposed method has been applied to the travelling salesman problems and the optimal task partition problems to evaluate the performances. The simulation results show that the performance of proposed method is superior to that of Hopfield model using the update rule of gradient descent method.

I. 서 론

신경망은 단순한 처리능력을 가지는 뉴런(neuron)들이 대규모로 상호연결되어 분산지식표현(distributed

knowledge representation), 일반화(generalization), 그리고 학습(learning) 등의 속성을 가지고 있어 조합형 최적화와 같은 과제들을 해결하는 데에 많이 응용되고 있다^[1-8]. 그러나 복잡하고 계산집약적인 문제들을 풀기 위해서는 대규모의 뉴런들을 이용하여 신경망을 구성하게 되므로 해를 구하는 데에 있어서 아직까지는 비현실적인 시간이 요구되거나 국소최적해로의 수렴과 같은 문제가 발생되고 있다. 이를 해결하기 위한 연구로서는 문제의 속성에 따라서 신경망의 구성을 달리하거나 에너지함수의 가중치 파라미터들을 조정한다든지, 혹은 온도기획 방법의 변경 등을 시도하고 있

* 正會員, 大邱曉星가톨릭大學校 電子情報工學部
(School of Electronics and Information Engineering,
Catholic University of Taegu Hyosung)

※ “이 논문은 1996년도 한국학술진흥재단의 공모과제 연구비에 의하여 연구되었음”

接受日字:1997年3月25日, 수정완료일:1997年11月14日

다^[1-4].

최적화에 이용되고 있는 신경망에는 결정론적 방법과 확률적 방법이 있다. 홉필드 모델(Hopfield model)^[3]은 기존 최적화 알고리즘의 비용함수에 상응하는 에너지함수가 볼록함수(convex function)인 때에 결정론적 경신규칙(update rule)을 적용하여 에너지함수가 단조감소하도록 하며, 수렴속도가 빨라 최적화 신경망으로 널리 이용되고 있다. 그러나 홉필드 모델은 결정론적 방법으로 최적화 문제에 따라서는 국소최적해를 찾는 데에 그칠 수도 있다. 볼츠만 머신(Boltzmann machine)^[5]은 홉필드 모델의 병렬최적화 기법과 온도기회에 의한 에너지교란 개념을 적용하여 국소최적해로 부터의 탈출을 시도하고 있는 시뮬레이티드 아닐링(simulated annealing) 개념을 도입한 확률적 방법이다. 이 방법은 전역최적해로의 수렴확률은 높으나 초기 온도의 설정과 온도기회 방법에 따라서는 많은 수렴시간이 걸린다. 볼츠만 머신이 가지는 이러한 점을 개선하기 위하여 제안된 방법으로는 각 뉴런의 입력층에 가우스분포 잡음을 인가한 후 온도의 순차적 감소로 전역최적해를 찾는 가우스 머신(Gaussian machine)^[6], 그리고 빠른 온도기회 및 코오시분포 잡음을 이용한 코오시 머신(Cauchy machine)^[7] 등의 확률적 방법이 있다. 일반적으로 결정론적 방법들은 수렴속도가 빠르며 초기값을 잘 설정하면 전역최적해를 얻을 수 있으나 국소최적해를 만났을 때 이를 벗어날 수 없는 반면, 확률적 방법들은 전역최적해로의 수렴은 보장되나 수렴속도가 낮다.

한편, 최적화용 신경망에서 각 뉴런의 출력치로 표현되는 에너지함수는 여러개의 국소최적해를 가지는 다극점(multiextremal) 함수일 수 있다. 이러한 에너지함수에서 전역최적해를 찾을 수 있으면서도 수렴속도가 빠른 알고리즘의 개발이 요구된다. 이제까지의 연구들에서는 전역최적해로의 수렴확률을 개선하거나 또는 수렴속도를 개선하기 위한 연구 중에서 어느 하나에 주로 역점을 두었을 뿐, 그 두가지를 동시에 만족시키기 위한 연구는 찾아보기 힘들다. 또한 전역최적해로의 수렴확률이나 수렴속도의 개선을 위하여 취한 방법으로 온도기회 방법이나 에너지함수의 가중치 파라미터 조정 등을 시도하고 있다^[1-4]. 그러나 최적화 과정에서 국소최적해를 만났을 때에 이를 벗어남으로써 전역최적해로의 수렴확률과 수렴속도를 동시에 개선하려는 체계적인 시도는 거의 찾아보기 힘들다.

본 연구에서는 기울기하강(gradient descent)법과 동적터널링 시스템(dynamic tunneling system)을 조합 사용함으로써 최적화용 신경망의 성능을 개선할 수 있는 새로운 방법을 제안하였다. 제안된 방법에서는 빠른 수렴속도의 전역최적화가 가능하도록 하기 위하여 기울기하강법의 경신규칙을 이용하고, 국소최적해를 만났을 때 이를 벗어난 다른 뉴런 출력치의 설정을 위하여 동적터널링 시스템의 경신규칙을 이용하였다. 제안된 방법을 외판원 경로문제(travelling salesman problem : TSP)와 최적태스크 분할(optimal task partition) 문제에 각각 적용하여 기존의 홉필드 모델에 의한 결과와 비교 고찰하였다.

II. 기울기하강법과 동적터널링 시스템을 조합한 효율적 최적화

동적 흐름(flow)과 터널링 기능을 갖는 동적터널링 알고리즘은 다극점함수의 전역최소점을 찾는 전역최적화 알고리즘이다^[10]. 최적화는 다극점함수 $f(\mathbf{x})$ 에 동적최적화 시스템(dynamic optimization system)를 수행하여 국소최소점 \mathbf{x}^* 를 구하고, 구해진 \mathbf{x}^* 에 교란 ϵ 을 추가한 다음 동적터널링 시스템을 수행하여 $f(\mathbf{x}^*) \geq f(\mathbf{x}^0)$ 인 새로운 시작점 \mathbf{x}^0 를 구하는 두 시스템을 반복 적용함으로써 이루어진다. 이때 교란 ϵ 의 값은 무작위로나 혹은 결정론적으로 설정된다. 여기서 다극점함수 $f(\mathbf{x})$ 의 국소최소점을 구하는 동적최적화 시스템은

$$\frac{dx_i}{dt} = -\frac{\partial f}{\partial x_i} \quad (1)$$

와 같다. 이 과정은 임의의 초기값으로부터 국소최소점 \mathbf{x}^* 를 찾는 기울기하강법이다. 또한 구해진 국소최소점 \mathbf{x}^* 에 교란 ϵ 을 추가한 새로운 시작점 $\mathbf{x} = \mathbf{x}^* + \epsilon$ 으로 시작되는 동적흐름과 터널링 기능의 동적터널링 시스템은

$$\frac{dx_i}{dt} = -\frac{\frac{\partial f}{\partial x_i}}{[(\mathbf{x} - \mathbf{x}^*)'(\mathbf{x} - \mathbf{x}^*)]^{\bar{\lambda}}} - k f^*(f^{\wedge}(\mathbf{x})) \frac{\partial f}{\partial x_i} \quad (2)$$

과 같이 정의된다. 여기서 '는 전치를 의미하고 $f^{\wedge}(\mathbf{x})$ 는 $f^{\wedge}(\mathbf{x}) = f(\mathbf{x}) - f(\mathbf{x}^*)$ 로 정의된다. k 는 터널링 패널티(tunneling penalty)로 $\|\frac{\partial f}{\partial \mathbf{x}}\| / \epsilon$ 보다 같거나 큰 값을 가지나 시뮬레이션에서는 편의적으로 적당한

상수로 설정된다^[10]. 또한 λ 는 $\partial f/\partial \mathbf{x}$ 의 영(zero)점 \mathbf{x}^* 의 차수보다 작거나 같은 값으로 $\frac{\partial f}{\partial \mathbf{x}}$ 에서 영점 \mathbf{x}^* 를 제거하기 위한 충분한 큰 수로 설정된다^[10]. $f^*(z)$ 는 piecewise-linear 함수인 heaviside 함수로 정의한다. 즉,

$$f^*(z) = z, \text{ if } z > 0$$

$$0, \text{ if } z < 0 \quad (3)$$

이다. 식 (2)에서, 우측 첫 번째 항의 분모 $[(\mathbf{x} - \mathbf{x}^*)'(\mathbf{x} - \mathbf{x}^*)]^\lambda$ 항은 동적최적화 시스템으로 부터 얻어지는 국소최소점인 평형상태 \mathbf{x}^* 를 제거하기 위한 항이다. 또한 두 번째 항의 목적은 하위유역에 존재하는 즉, $f(\mathbf{x}^*) \geq f(\mathbf{x}^0)$ 인 새로운 시작점 \mathbf{x}^0 를 찾기 위한 것이다.

한편, 그림 1은 연속적이고 미분 가능한 6 개의 최소점을 가지는 다극점함수 $f(\mathbf{x})$ 에 동적최적화와 동적터널링 시스템을 반복 적용함으로써 전역최소점을 찾는 동적터널링 알고리즘의 기하학적 해석을 나타낸 것이다.

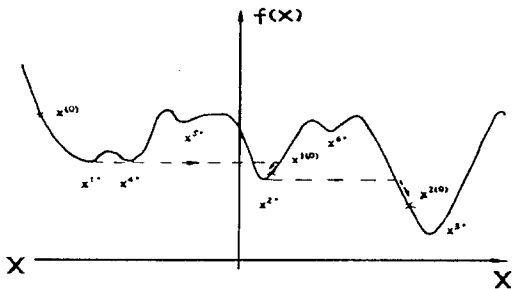


그림 1. 동적터널링 알고리즘의 기하학적 해석
Fig. 1. The geometric concept of the dynamic tunneling algorithm.

그림에서 보는 바와같이 임의의 값 $\mathbf{x}^{(0)}$ 를 초기값으로 하여 함수 $f(\mathbf{x})$ 에 기울기하강법의 동적최적화 시스템을 적용하면 국소최소점 \mathbf{x}^{1*} 에 도달하고, \mathbf{x}^{1*} 에 교란 ϵ 을 추가한 초기값으로 동적터널링 시스템을 적용하면 $\mathbf{x}^{(10)}$ 에 수렴하게 된다. 또한 $\mathbf{x}^{(10)}$ 를 초기값으로 동적최적화 시스템을 적용하면 \mathbf{x}^{2*} 에 도달하고, 동적터널링 시스템에 의하여 $\mathbf{x}^{(20)}$ 를 구하며, 다시 $\mathbf{x}^{(20)}$ 를 초기값으로 하여 동적최적화 시스템에 의해서 전역최소점인 \mathbf{x}^{3*} 를 얻을 수 있다. 여기서 \mathbf{x}^{4*} 와 \mathbf{x}^{5*} 그리고 \mathbf{x}^{6*} 는 터널링 현상에 의하여 제거된다.

따라서 기울기하강법은 그 동작 특성상 국소최적해

로 빠질 확률이 있지만 수렴속도가 매우 빠른 속성을 가지며, 동적터널링 시스템은 전역최적화를 위한 알고리즘이다. 그러므로 기울기하강법과 동적터널링 시스템의 장점을 조합하면 더욱 더 우수한 수렴특성을 가지는 최적화가 가능할 것이다. 즉, 전역최적해로의 수렴을 빠르게 하는 데는 기울기하강법의 경신규칙을 적용하고 전역최적해로의 수렴을 확실히 하는 데는 동적터널링 시스템의 경신규칙을 적용한 신경망을 이용하면 고속으로 최적화 문제의 전역최적해를 구할 수 있다. 최적화를 위한 신경망에서 뉴런 출력치 벡터 \mathbf{x} 와 에너지함수 $E(\mathbf{x})$ 의 최소값을 찾는 과정은 동적터널링 알고리즘으로 상태벡터 \mathbf{x} 에 대한 함수 $f(\mathbf{x})$ 의 최소값을 찾는 과정으로 대응시킬 수 있다. 신경망의 학습에 동적최적화 시스템으로 기울기하강법의 경신규칙을 적용하여 에너지함수의 값이 최소인 뉴런의 출력치를 구하고, 국소최적해를 만났을 때 이를 벗어난 다른 출력치의 설정을 위하여 구해진 출력치에 교란을 추가한 출력치를 초기값으로 하여 터널링 기능을 가지는 동적터널링 시스템의 경신규칙을 적용한다. 이렇게 하면 빠른 시간내에 전역최소의 에너지를 보장하는 최적화를 시킬 수 있음을 알 수 있다.

먼저, 신경망에서 에너지함수 $E(\mathbf{x})$ 를

$$E(\mathbf{x}) = - (1/2) \sum_i \sum_j T_{ij} x_i x_j - \sum_i I_i x_i \quad (4)$$

로 정의한다. 여기서 i 와 j 는 뉴런을 나타내는 번호이며, T_{ij} 와 I_i 는 각각 연결세기 행렬과 입력바이어스이다. 에너지함수 $E(\mathbf{x})$ 에 동적터널링 시스템을 적용할 때 뉴런 출력치 \mathbf{x} 의 변화는

$$\mathbf{x}(t+1) = \mathbf{x}(t) + \Delta \mathbf{x}(t)$$

$$\Delta \mathbf{x} = \frac{\partial E(\mathbf{x})}{\partial \mathbf{x}} \left[\frac{1}{[(\mathbf{x} - \mathbf{x}^*)'(\mathbf{x} - \mathbf{x}^*)]^\lambda} - kE^*(E^\wedge(\mathbf{x})) \right] \quad (5)$$

로 구할 수 있다. 이는 동적터널링 시스템의 경신규칙에 의한 뉴런 출력치의 변화량을 나타낸 것이다. 여기서 $E^\wedge(\mathbf{x}) = E(\mathbf{x}) - E(\mathbf{x}^*)$ 이다. 또한 $E^*(z)$ 는 $z > 0$ 이면 z 이고, $z < 0$ 이면 0의 값을 갖는다.

따라서 제안된 경신규칙에 따른 뉴런의 출력경신을 살펴보면 다음과 같다. 먼저, 기울기하강법에 의한 i 번째 뉴런의 입력 U_i 의 변화는

$$\frac{d U_i}{dt} = \sum_j T_{ij} x_j - \tau U_i + I_i \quad (6)$$

를 이용하여 구할 수 있다. 여기서 τ 는 시정수이다. 또한 식 (5)의 동적터널링 시스템에 의한 뉴런의 입력 U_i 의 변화 역시

$$\frac{dU_i}{dt} = [\sum_j T_{ij} x_j - \tau U_i + I_i] \left[\frac{1}{(U_i - U_i^*)^{2\alpha} - k} E^*(E^{\wedge}(\mathbf{x})) \right] \quad (7)$$

로 계산될 수 있다. 이때 식 (6)과 (7)에서 구해진 각각의 뉴런입력 U_i 를 구하여 뉴런 활성화함수를 통과시키면 실제의 뉴런출력 x_i 를 경신시킬 수 있다.

기울기하강법과 동적터널링 시스템을 이용한 경신규칙으로 최적화 성능을 개선하기 위해 제안한 최적화 알고리즘은 다음과 같다.

- 단계 1 : 최적화 문제의 목적함수와 제약조건을 포함하는 비용함수 $f(\mathbf{x})$ 를 정의한다.
- 단계 2 : 정의된 비용함수 $f(\mathbf{x})$ 를 신경망의 에너지함수 $E(\mathbf{x})$ 와 사상시켜 연결세기 행렬 T_{ij} 및 입력바이어스 I_i 를 구한다.
- 단계 3 : 기울기하강법의 경신규칙을 이용한 신경망을 수행하여 종료조건이 만족되면 최적화 알고리즘을 종료하고, 그렇지 않고 계산된 에너지함수의 값이 허용치 이상으로 알고리즘의 반복에 따라 변하지 않으면 단계 3을 종료한다.
- 단계 4 : 단계 3에서 계산된 뉴런 출력치에 교란 ϵ 을 추가하여 동적터널링 시스템의 경신규칙을 수행한다. 이때 알고리즘의 종료조건이 만족되면 최적화를 종료하고, 그렇지 않고 계산된 에너지함수의 값이 단계 3에서 계산된 에너지함수의 값 이하이면 단계 3으로 간다.

여기서 최적화 알고리즘의 종료조건으로는 에너지함수의 값이 이미 설정된 허용치 이하이거나 반복횟수(number of iterations)가 미리 설정된 횟수일 때, 또는 뉴런의 출력변화가 일어나지 않는 때로 한다. 이와 같이 제안된 방법에서는 기울기하강법의 경신규칙을 수행함으로써 국소최적해로 수렴될 때에, 교란을 가한 뉴런의 출력치를 초기값으로 하여 터널링 기능을 가진 동적터널링 시스템의 경신규칙을 실행하게 된다. 따라서 기존의 홉필드 모델이 가중치 파라미터나 뉴런의 초기 출력치등과 같은 요인에 따라 국소최적해로 수렴

되는 문제를 해결할 수 있으며, 기울기하강법의 빠른 수렴속도도 그대로 살릴 수 있다. 만약 기울기하강법과 동적터널링 시스템의 경신규칙을 1회 적용에 의하여 국소최적해로 수렴될 경우에는 제안된 방법의 단계 3에서 단계 5까지의 과정을 반복함으로써 결국 전역최적해로 수렴될 수 있다.

기울기하강법과 동적터널링 시스템을 이용한 제안된 경신규칙의 흐름도를 도시하면 그림 2와 같다. 그림에서 $E_g(\mathbf{x})$ 와 $E_d(\mathbf{x})$ 는 각각 기울기하강법과 동적터널링 시스템의 경신규칙을 이용하여 계산된 에너지함수 값들이다.

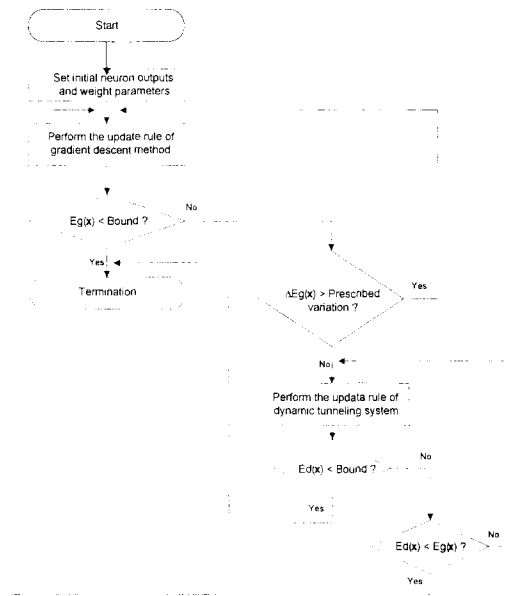


그림 2. 제안된 경신규칙의 흐름도

Fig. 2. The flowchart of the proposed update rule.

III. 응용 예 및 시뮬레이션 결과고찰

NP complete 최적화 문제들을 신경망으로 풀기 위해서는 먼저 문제의 정확한 정의, 목적함수, 제약조건, 그리고 목적함수와 제약조건을 포함하는 비용함수를 정의하여야 한다^{1,3,4}. 또한 정의된 비용함수를 신경망의 에너지함수와 비교하여 연결세기 행렬 및 입력바이어스를 각각 구하고, 가중치 파라미터의 값들을 적절히 설정하여 해당 신경망을 구동시킨다.

제안된 방법의 타당성과 그 성능을 확인하기 위하여 최적화 문제로 잘 알려진 외판원 경로문제와 최적태스

크 분할을 응용실험 대상으로 택하였다^[9]. 외판원경로 문제는 n개의 도시가 주어졌을 때에 임의의 한 도시에서 출발하여 모든 도시를 한 번씩만 방문하고 다시 출발했던 도시로 되돌아 오는 여러가지 경로 중에서 총 거리의 합이 최소가 되는 경로를 구하는 문제이다. 최적태스크 분할은 n개의 부태스크들로 구성된 태스크를 p개의 프로세서에 분할하는 것으로 이때 전산처리량과 통신량의 균형을 위한 문제이다. 여기서는 일반적으로 $n \geq p$ 의 관계를 가진다. 따라서 제안된 방법을 이들 문제에 각각 적용하여 그 타당성을 확인하였으며, 기존의 기율기하강법의 경신규칙을 이용하는 홉펠드 모델의 결과와 비교 고찰하였다.

실험에서 뉴런의 초기 출력치들의 범위는 -0.5와 +0.5 사이의 값으로 하였으며, 랜덤시드(random seed)의 변경으로 뉴런의 출력치들을 변화시킬 수 있도록 하였다. 알고리즘의 종료는 두 문제 모두 반복횟수가 10,000회 이상이거나 모든 뉴런의 출력변화가 일어나지 않을 때 또는 에너지함수의 값이 허용치 이하일 때로 하였다. 이때 k 와 λ 의 값은 두 문제 모두에 0.5와 5로 설정하였으며, 기율기하강법의 반복에 의한 에너지함수 값의 변화량이 10^{-6} 이하이면 동적터널링 시스템의 경신규칙으로 전환토록 하였다. 또한 여기서 이용되는 홉펠드 모델은 기율기하강법의 경신규칙을 이용하는 연속적인 모델이며 그 감쇠정수의 값은 0.001로 하였다.

1. 외판원경로 문제

외판원경로 문제는 어느 한 도시를 반드시 한 번은 방문해야 하고 매 방문마다 반드시 한 도시만을 방문하여야 한다는 제약조건 하에서, 각 방문 도시간의 거리가 최소되도록 모든 도시를 방문하는 목적의 문제이다. 이때 목적함수와 제약조건을 각각 에너지함수 $E_{obj}(\mathbf{x})$ 와 $E_{con}(\mathbf{x})$ 로 표현한 전체 에너지함수 $E_{TSP}(\mathbf{x})$ 는 다음과 같다. 즉,

$$E_{obj}(\mathbf{x}) = (A/2) \sum_i \sum_{j \neq i} \sum_k d_{ij} x_{ik} (x_{j, k+1} + x_{j, k-1})$$

$$E_{con}(\mathbf{x}) = (B/2) \sum_i \sum_k \sum_{j \neq k} x_{ik} x_{ji} + (C/2) \sum_i \sum_k \sum_{j \neq k} x_{ik} x_{ji} + (D/2) \left(\sum_i \sum_j x_{ij} - n \right)^2$$

$$E_{TSP}(\mathbf{x}) = E_{obj}(\mathbf{x}) + E_{con}(\mathbf{x}) \quad (8)$$

이다. 여기서 A, B, C, 그리고 D는 가중치 파라미터

이고 i와 j 및 k와 l은 각각 도시번호 및 방문순서이다. 또한 d_{ij} 는 도시 i와 j 사이의 거리이며, x_{ik} 는 i번째 도시가 k번째에 방문되면 1, 그렇지 않으면 0인 할당행렬 요소값이다.

실험은 전체 도시의 수가 각각 7과 10인 두 문제를 대상으로 하였으며, 가중치 파라미터의 값으로는 $A=B=200$, $C=50$, 그리고 $D=100$ 으로 하였다^[9]. 그리고 이들 두 규모의 문제에 대한 종료조건으로 에너지함수의 허용치는 각각 -3650과 -7500으로 하였다. 그림 3은 외판원경로 문제 각각의 도시좌표를 나타낸 것이다.

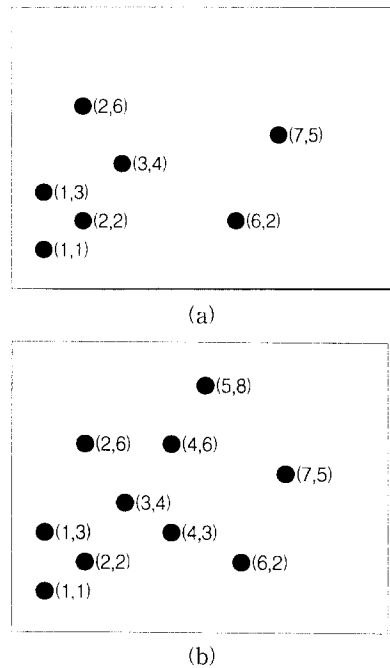


그림 3. (a) 7 도시 및 (b) 10 도시 외판원경로 문제의 예

Fig. 3. Examples of (a) 7-city and (b) 10-city travelling salesman problem.

그림에서 (x,y)는 도시의 2차원 좌표를 나타낸다.

그림 4는 그림 3(a)에서와 같은 7 개 도시 TSP에서 랜덤시드를 10과 90으로 하여 홉펠드 모델과 제안된 방법을 각각 적용한 결과로서 반복횟수 N에 따른 정규화된 에너지함수 E를 나타낸 것이다. 그림에서 보는 바와같이 랜덤시드의 값이 이와같은 두 경우에는 뉴런의 초기 출력치가 최적해 근처에 설정되어 제안된 방법과 홉펠드 모델 둘 다 최적해로 수렴되었다. 특히 랜덤시드가 10인 경우 제안된 방법과 홉펠드 모델은

들 다 최적해로의 수렴속도가 동일하다. 이는 랜덤시드에 따른 뉴런의 초기 출력치가 최적해 근처에 설정되어 오직 기울기하강법의 경신규칙만 적용된 경우이다. 랜덤시드가 90인 경우에는 제안된 방법의 수렴속도가 홉필드 모델에 비하여 약 3.6배 정도 빠르다. 이는 홉필드 모델의 경우에 그 수렴특성이 뉴런의 초기 출력치에 의존됨을 확인할 수 있다.

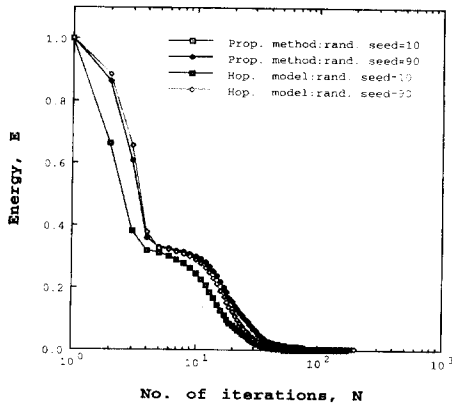


그림 4. 7 도시 외판원경로 문제의 정규화된 에너지 함수 E

Fig. 4. Normalized energy function E of 7-city travelling salesman problem.

표 1은 7 개의 도시에 대한 TSP에서 홉필드 모델과 제안된 방법을 각각 적용하여 랜덤시드를 다르게 100 번씩 시도한 결과, 최적해로 수렴된 경우들에 대한 반복횟수 및 CPU 시간의 평균 \bar{x} 와 표준편차 σ 를 각각 나타낸 것이다. 표에서 N_h , N_d , 그리고 N_p 는 각각 홉필드 모델, 동적터널링 시스템, 그리고 제안된 방법의 반복횟수이며, t_h 와 t_p 는 각각 홉필드 모델과 제안된 방법에 소요된 CPU 시간이다. 특히 N_p 와 t_p 는 홉필드 모델과 동적터널링 시스템에 각각 소요된 반복횟수와 CPU 시간의 합이다. 표에서 보는 바와같이 제안된 방법은 홉필드 모델에 비하여 최적해로의 수렴확률과 수렴속도에서 각각 약 1.1배와 약 5.0배 정도 개선되었다. 이 규모의 문제에서는 홉필드 모델과 제안된 방법 둘 다 모두 최적해로의 수렴률이 비슷하였다. 이는 랜덤시드에 의해 설정되는 뉴런의 초기 출력치가 우연히도 최적해로의 수렴가능한 영역에 설정된 경우로 보이며, 실제로 홉필드 모델에서 이와같이 적절하

게 뉴런의 초기 출력치의 설정을 보장하기 위한 체계적인 방안은 별로 없다. 또한 홉필드 모델의 표준편차가 큰 값으로 이는 그 수렴특성이 뉴런의 초기 출력치에 크게 의존함을 확인할 수 있다.

표 1. 7 도시 외판원경로 문제의 100 번 시도에 대한 실험결과

Table 1. Results of the 7-city travelling salesman problem for 100 trials.

	Hopfield model			Proposed method		
	N_h	t_h	E	N_p (N_h, N_d)	t_p	E
\bar{x}	514.55	7.5	-3650	102.79 (99.14,3.65)	1.51	-3650
σ	128.05	1.93	20.82	25.21 (22.55,7.58)	0.56	27.82
P_r	94 %			100 %		

\bar{x} : mean, σ : standard deviation, P_r : convergence ratio

표 2. 10 도시 외판원경로 문제의 100 번 시도에 대한 실험결과

Table 2. Results of the 10-city travelling salesman problem for 100 trials.

	Hopfield model			Proposed method		
	N_h	t_h	E	N_p (N_h, N_d)	t_p	E
\bar{x}	549.27	11.55	-7500	151.99 (147.91,4.08)	8.47	-7500
σ	472.82	9.89	4.45	26.78 (26.27,6.00)	1.52	41.02
P_r	44 %			100 %		

\bar{x} : mean, σ : standard deviation, P_r : convergence ratio

표 2는 그림 3(b)에서와 같은 10 개 도시 TSP에서 100 개의 서로 다른 랜덤시드로 홉필드 모델과 제안된 방법을 각각 적용하여 최적해로 수렴된 경우만을 나타낸 것이다. 표에서도 제안된 방법은 홉필드 모델에 비해 최적해로의 수렴확률과 수렴속도에서 각각 약 2.3배와 약 3.6배 정도 개선되었다. 또한 표에서 보면 홉필드 모델의 경우, 최적해로의 수렴확률이 낮은 값을 나타내어 기울기하강법의 경신규칙을 이용할 때 뉴런의 초기 출력치에 따라서는 국소최적해로 수렴될 수 있음을 보여준다. 그러나 제안된 방법에서는 기울기하강법의 경신규칙에 의해 국소최적해로 수렴될 경우 터널링 현상을 가진 동적터널링 시스템의 경신규칙을 이용함으로써 이를 벗어나 전역최적해로 수렴될 수

있음을 알 수 있다. 여기서도 제안된 방법은 홉펠드 모델에 비하여 반복횟수와 CPU 시간에 대한 표준편차의 값이 각각 더욱 작음을 확인할 수 있다.

한편 표 1과 2에서 보는 바와같이 일반적으로 문제의 규모가 클수록 전역최적해로의 수렴확률 및 그 수렴속도 면에서 더욱 더 큰 성능의 차이를 확인할 수 있다. 이는 문제가 대형화 됨에 따라 뉴런의 초기 출력치의 설정이 더욱 어려워지며, 그 성능 또한 초기 출력치에 더욱 크게 의존함을 알 수 있다. 그러나 제안된 방법은 이런 경우에도 뉴런의 초기 출력치에 많은 영향을 받지 않는 우수한 성능이 있음을 알 수 있다.

2. 최적태스크 분할

최적태스크 분할은 각 프로세서에는 한 개이상의 부태스크가 할당되어야 하고 각 부태스크는 반드시 어느 하나의 프로세서에라도 할당되어야 한다는 제약조건 하에서, 전산처리량과 통신량의 균형 있는 할당이 되도록 하는 문제이다. 이때 목적함수와 제약조건을 각각 에너지함수 $E_{obj}(\mathbf{x})$ 와 $E_{con}(\mathbf{x})$ 로 표현한 전체 에너지함수 $E_{OTP}(\mathbf{x})$ 는 다음과 같다. 즉,

$$E_{obj}(\mathbf{x}) = (A/2) \sum_i \sum_k \sum_j W_i W_j x_{ik} x_{jk} + (B/2) \sum_i \sum_k \sum_j \sum_l e_{ij}(x_{ik} x_{jl} - 2x_{ik} x_{jl})$$

$$E_{con}(\mathbf{x}) = (C/2) \left(\sum_i \sum_k x_{ik} - n \right)^2 + (D/2) \sum_i \sum_k \sum_j \sum_l x_{ik} x_{jl}$$

$$E_{OTP}(\mathbf{x}) = E_{obj}(\mathbf{x}) + E_{con}(\mathbf{x}) \tag{9}$$

이다. 여기서도 A, B, C, 그리고 D는 가중치 파라미터이고 i와 j는 부태스크 번호이며, k와 l은 프로세서 번호이다. e_{ij} 는 i와 j 번째 부태스크 사이의 통신량이며, W_i 는 i 번째 부태스크의 전산처리량이다. x_{ik} 는 i 번째 부태스크가 k 번째 프로세서에 할당되면 1 그렇지 않으면 0의 값을 갖는 할당행렬 요소값이다.

실험은 부태스크의 수 n이 4와 6이고 프로세서 수 p가 각각 3인 두 문제를 대상으로 하였으며, 가중치 파라미터의 값으로는 A=40, B=5, C=20, 그리고 D=30으로 하였다¹⁹⁾. 또한 이들 문제에 대해서도 종료조건으로 에너지함수의 허용치는 각각 -98과 -175로 하였다. 그림 5는 최적태스크 분할 각각의 전산처리 및 통신량을 나타낸 것이다.

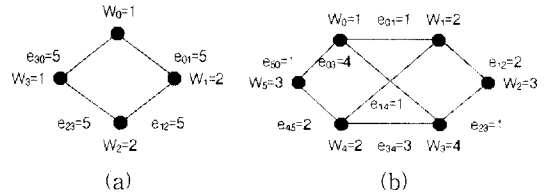


그림 5. (a) 4 개 부태스크와 3 개 프로세서 및 (b) 6 개 부태스크와 3 개 프로세서의 최적 태스크 분할 예
Fig. 5. Examples of (a) 4-subtask and 3-processor, (b) 6-subtask and 3-processor optimal task partition.

그림 6은 그림 5(a)에서와 같은 4 개 부태스크와 3 개 프로세서 최적태스크 분할에서 랜덤시드를 10과 90으로 하여 홉펠드 모델과 제안된 방법을 각각 적용한 결과로서 반복횟수 N에 따른 정규화된 에너지함수 E를 나타낸 것이다. 그림 6에서도 그림 3의 TSP 경우에서 처럼 랜덤시드의 값이 이와같은 두 경우에는 뉴런의 초기 출력치가 최적해 근처에 설정되어 제안된 방법과 홉펠드 모델 둘 다 최적해로 수렴되었다. 여기서도 제안된 방법이 홉펠드 모델에 비하여 최적해로의 수렴속도가 약 2.5배 정도 빠르며, 홉펠드 모델의 경우에는 그 수렴특성이 뉴런의 초기 출력치에 의존됨을 확인할 수 있다.

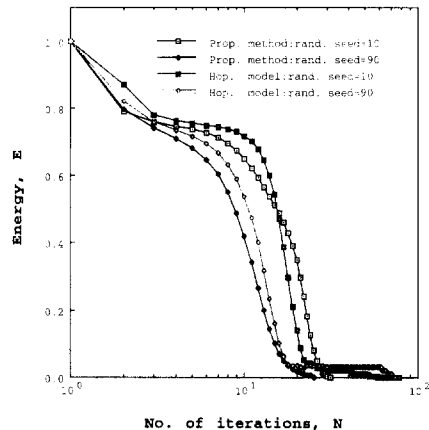


그림 6. 4 개 부태스크와 3 개 프로세서 최적태스크 분할의 정규화된 에너지함수 E
Fig. 6. Normalized energy function E of 4-subtask and 3-processor optimal task partition.

표 3은 4 개 부태스크와 3 개 프로세서 최적태스크

분할에서 홉필드 모델과 제안된 방법을 각각 적용하여 랜덤시드를 다르게 100 번씩 시도한 결과, 최적해로 수렴된 경우들에 대한 반복횟수 및 CPU 시간의 평균 \bar{x} 와 표준편차 σ 를 각각 나타낸 것이다. 표에서도 TSP의 결과에서 처럼 제안된 방법은 홉필드 모델에 비해 최적해로의 수렴확률과 수렴속도에서 각각 약 1.1배와 약 2.4배 정도 개선되었다. 또한 두 방법 다 비슷한 최적해로의 수렴률을 보여준다. 이는 기울기하강법의 경신규칙을 이용하는 홉필드 모델의 경우에도 뉴런의 초기 출력치만 잘 설정되면 최적해로의 수렴확률과 수렴속도에서 우수한 성능이 있음을 알 수 있다. 여기서도 제안된 방법은 동적터널링 시스템의 경신규칙에 의해서 국소최적해로부터의 탈출을 확인할 수 있다. 반복횟수 및 CPU 시간의 각각에 대한 표준편차의 경우 제안된 방법은 홉필드 모델에 비해 더 작은 값을 가짐을 알 수 있다.

표 3. 4 개 부태스크와 3 개 프로세서 최적 태스크 분할의 100 번 시도에 대한 실험결과

Table 3. Results of the 4-subtask and 3-processor optimal task partition for 100 trials.

	Hopfield model			Proposed method		
	N_h	t_h	E	$N_o (N_h, N_d)$	t_o	E
\bar{x}	88.09	0.12	-98	36.48 (36.42,0.06)	0.03	-98
σ	34.75	0.33	0	24.62 (24.41,0.34)	0.17	0.05
P_r	88 %			100 %		

\bar{x} : mean, σ : standard deviation, P_r : convergence ratio

표 4는 그림 5(b)에서와 같은 6 개 부태스크와 3 개 프로세서의 최적태스크 분할에서 100 개의 서로 다른 랜덤시드로 홉필드 모델과 제안된 방법을 각각 적용하여 최적해로 수렴된 경우만을 나타낸 것이다. 표에서 보는 바와같이 제안된 방법은 홉필드 모델에 비해 최적해로의 수렴확률은 약 3.0배 정도 개선되었으나 수렴속도에서는 홉필드 모델의 경우가 제안된 방법에 비하여 오히려 약 1.5배 정도 빠름을 알 수 있다. 이는 기울기하강법의 경신규칙을 이용하는 홉필드 모델은 국소최적해로의 수렴에 대한 탈출은 불가능하나 빠른 수렴특성이 있음을 확인할 수 있다. 여기서도 여

전히 제안된 방법은 홉필드 모델에 비하여 우수한 수렴성능이 있음을 확인할 수 있다. 또한 제안된 방법은 홉필드 모델에 비하여 반복횟수와 CPU 시간에 대한 표준편차가 각각 더 큰 값을 가진다. 이는 홉필드 모델이 제안된 방법에 비해 상대적으로 낮은 최적해로의 수렴확률을 가질 때에만 일어나는 것으로 추측된다.

표 4. 6 개 부태스크와 3 개 프로세서의 최적태스크 분할의 100 번 시도에 대한 실험결과

Table 4. Results of the 6-subtask and 3-processor optimal task partition for 100 trials.

	Hopfield model			Proposed method		
	N_h	t_h	E	$N_o (N_h, N_d)$	t_o	E
\bar{x}	146.94	0.39	175	217.47 (209,87,7.60)	0.53	-175
σ	19.34	0.50	1.27	337.31 (323.46,13.90)	0.92	2.35
P_r	33 %			100 %		

\bar{x} : mean, σ : standard deviation, P_r : convergence ratio

한편 표 2와 4에서 보는 바와같이 홉필드 모델은 문제의 규모가 클수록 국소최적해로의 수렴확률이 높아 뉴런의 초기 출력치의 설정은 더욱 어려우며, 그 성능 또한 초기 출력치에 더욱 크게 의존함을 알 수 있다. 그러나 제안된 방법은 이상의 선정된 문제 규모에 있어서는 랜덤시드에 무관하게 최적해로 수렴되었으며, 기존의 홉필드 모델에 비하여 최적화의 수렴특성도 우수함을 확인할 수 있었다.

IV. 결론

본 연구에서는 기울기하강법과 동적터널링 시스템을 조합 사용함으로써 최적화용 신경망의 성능을 개선할 수 있는 새로운 방법을 제안하였다. 제안된 방법에서는 빠른 수렴속도의 전역최적화가 가능하도록 하기 위하여 기울기하강법의 경신규칙을 이용하고, 국소최적해를 만났을 때 이를 벗어난 다른 뉴런 출력치의 설정을 위하여 동적터널링 시스템의 경신규칙을 이용함으로써 빠른 수렴속도로 전역최적해를 구할 수 있도록 하였다. 제안된 방법의 타당성과 성능을 확인하기 위하여 도시 수가 각각 7개 및 10개인 외판원경로 문제와 부태

스크 수가 각각 4와 6이고 프로세서 수가 각각 3개인 최적태스크 분할에 적용하여 시뮬레이션하였다. 10개 도시 외판원경로 문제의 경우, 제안된 방법은 기존의 기율기하강법의 경신규칙을 이용하는 홉필드 모델에 비해 최적해로의 수렴확률이 약 2배 정도 개선되었고, 수렴속도는 약 4배 정도 개선되었다. 또한 문제의 규모가 클수록 일반적으로 뉴런의 초기 출력치는 신경망의 성능에 많은 영향을 미치나 제안된 방법은 이 경우에도 우수한 최적화 성능이 있다. 한편 7개 도시보다는 10개의 도시 외판원경로 문제에서 그 성능 개선의 정도가 큼을 확인할 수 있다. 최적태스크 분할의 경우에도 외판원경로 문제에서와 같이 제안된 방법은 우수한 수렴특성을 가짐을 알 있다.

향후 제안된 경신규칙의 신경망을 좀 더 큰 규모의 다양한 최적화 문제에 적용하기 위한 연구가 보완적으로 진행되어야 하겠다.

참 고 문 헌

- [1] J. McClelland and D. Rumelhart, *Parallel distributed processing*, MIT Press, Cambridge, MA, 1986.
- [2] A.J. Maren, C.T. Harston, and R.M. Pap, *Handbook of neural computing application*, Academic Press Inc. London, 1990.
- [3] J. J. Hopfield and D. Tank, "Neural computation of decision in optimization problems," *Biol. Cybern.*, vol. 52, pp. 141-152, 1985.
- [4] J. Cervantes and R. Hilderbrant, "Comparison of three neuron-based computation schemes," *IEEE Conf. on Neural Networks*, San Diego, vol. 3, pp. 657-671, June 1987.
- [5] P. K. Mazaika, "A mathematical model of the Boltzmann machine," *IEEE Conf. on Neural Networks*, San Diego, vol. 3, pp. 157-163, June 1987.
- [6] Y. Akiyama, A. Tamashita, M. Kajimura, and H. Asio, "Combinatorial optimization with Gaussssian machines," *International Conf. on Neural Networks*, Washington D.C., vol. 1, pp. 533-540, June 1989.
- [7] Y. Takefuji and H. Szu, "Design of parallel distributed Cauchy machine," *International Conf. on Neural Networks*, Washington D.C., vol. 1, pp. 529-532, June 1989.
- [8] J. A. Freeman and D.M. Skapura, *Neural networks : algorithms, applications, and programming techniques*, Addison Wesley Co. Inc., 1991.
- [9] 조용현, 최홍문, "초기값의 최적설정에 의한 최적화용 신경회로망의 성능개선," *대한전자공학 회 논문집*, 제 30권 B편 8호, pp. 691-700, 1993년 8월
- [10] Y. Yao, "Dynamic tunneling algorithm for global optimization," *IEEE Trans. on Systems, Man, and Cybern.*, vol. 19, no. 5, pp. 1222-1230, Sept/Oct. 1989.

저 자 소 개



趙鏞鉉(正會員)

1979년 2월 경북대학교 전자공학과 졸업(공학사). 1981년 2월 경북대학교 대학원 전자공학과 졸업(공학석사). 1993년 2월 경북대학교 대학원 전자공학과 졸업(공학박사). 1983년 9월 ~ 1984년 2월 삼성전자(주) 근무. 1984년 3월 ~ 1987년 2월 한국전자통신연구소 연구원.

1987년 3월 ~ 1997년 2월 영남전문대학 전자과 부교수. 1997년 3월 ~ 현재 대구효성가톨릭대학교 전자정보공학부 조교수. 주관 심분야는 신경망 및 병렬처리, 회로 및 시스템, 통화로계 스위칭 등임