

論文98-35C-3-8

오류 역전파 학습에서 확률적 가중치 교란에 의한 전역적 최적해의 탐색

(Searching a Global Optimum by Stochastic Perturbation
in Error Back-Propagation Algorithm)

金 三 根 * , 閔 昌 宇 ** , 金 明 源 ***

(Sam-Keun Kim, Chang-Woo Min, and Myung-Won Kim)

요 약

다층 퍼셉트론(multi-layer perceptron)을 학습시키는 대표적인 학습 알고리즘인 EBP(Error Back-Propagation) 알고리즘은 패턴 인식, 적응 제어, 최적화 등의 복잡한 문제를 해결하는데 성공적으로 적용되어 왔다. 그러나 EBP 학습 알고리즘은 기본적으로 오차 표면의 기울기를 이용하여 학습을 수행하는 경사 하강법을 이용한 학습 알고리즘으로 국부적 최적해에 수렴할 수 있다는 문제점을 내포하고 있다. 더욱이 학습이 성공적으로 이루어지기 위해서는 주어진 문제의 적절한 신경망 구조를 결정해야 한다. 신경망 구조를 결정하는 것은 대개 은닉 노드의 개수를 결정하는 문제로서 이것은 주로 시행 착오에 의해서 결정된다. 이 논문에서는 위의 두 가지 문제를 해결하는 새로운 알고리즘을 제안한다. 제안한 학습 알고리즘은 가중치 공간에서 확률적 교란을 수행함으로써 다층 퍼셉트론 학습에서의 국부적 최적해에 수렴한 경우로부터 효과적으로 탈출한다. 제안한 확률적 교란 방법은 EBP 학습에 의하여 국부적 최적해에 수렴한 경우로부터 탈출하기 위하여 은닉 노드들과 관계된 가중치들을 확률적으로 재초기화한다. 또한 새로운 은닉 노드의 추가 방법 역시 확률적 가중치 교란의 특별한 경우로 볼 수 있다. 따라서 확률적 교란을 통하여 국부적 최적해 문제 및 신경망 구조 설계 문제를 일관된 방법으로 해결할 수 있다. Parity 문제, two-spirals 문제, 그리고 credit-screening 데이터에 대한 실험 결과는 제안한 알고리즘이 매우 효율적임을 보여준다.

Abstract

The Error Back-Propagation (EBP) algorithm is widely applied to train a multi-layer perceptron, which is a neural network model frequently used to solve complex problems such as pattern recognition, adaptive control, and global optimization. However, the EBP is basically a gradient descent method, which may get stuck in a local minimum, leading to failure in finding the globally optimal solution. Moreover, a multi-layer perceptron suffers from lacking a systematic determination of the network structure appropriate for a given problem. It is usually the case to determine the number of hidden nodes by trial and error. In this paper, we propose a new algorithm to efficiently train a multi-layer perceptron. Our algorithm uses stochastic perturbation in the weight space to effectively escape from local minima in multi-layer perceptron learning. Stochastic perturbation probabilistically re-initializes weights associated with hidden nodes to escape a local minimum if the EBP learning gets stuck to it. Addition of new hidden nodes also can be viewed as a special case of stochastic perturbation. Using stochastic perturbation we can solve the local minima problem and the network structure design in a unified way. The results of our experiments with several benchmark test problems including the parity problem, the two-spirals problem, and the credit-screening data show that our algorithm is very efficient.

* 正會員, 安城産業大學校 컴퓨터工學科
(Dept. of computer engineering, Anseong National
University)

** 正會員, 韓國 IBM 소프트웨어研究所

(Korea Software Development Institute, IBM, Korea)

*** 正會員, 崇實大學校 컴퓨터學部

(School of computing, Soongsil university)

接受日: 1997年12月3日, 수정완료일: 1998年3月2日

I. 서 론

EBP 학습 알고리즘은 패턴 인식, 적응제어 및 최적화 등의 많은 응용 분야에서 성공적으로 적용되어 온 다층 퍼셉트론을 학습시키는 대표적인 학습 알고리즘이다^[1,2]. 그러나 이러한 EBP는 국부적 최적해(local minima)에 수렴할 수 있다는 심각한 문제를 피할 수 없다. 이것은 EBP가 기본적으로 오차 표면(error surface)에 대한 기울기(gradient)만을 이용하여 오차가 감소하는 쪽으로만 학습하기 때문에 발생하는 문제라 할 수 있다. 또한 학습이 성공적으로 이루어지기 위해서는 주어진 문제의 적절한 신경망 구조를 결정해야 한다. 신경망 구조를 결정하는 것은 대개 은닉 노드의 개수를 결정하는 문제로서 이것은 주로 시행 착오에 의해서 결정된다. 일반적으로 은닉 노드의 개수는 문제의 복잡도에 의해서 결정되는데, 문제의 복잡도는 학습을 시키기 전에는 알 수가 없다. 더구나 많은 수의 훈련 데이터를 가지는 경우 학습 시간이 길어지므로 반복적 학습에 의하여 적절한 수의 은닉 노드를 찾기가 어렵게 된다.

이 논문에서는 위의 두 가지 문제를 동시에 해결할 수 있는 새로운 EBP 학습 방법으로 EBP/SP(Error Back-Propagation/Stochastic Perturbation) 학습 알고리즘을 제안한다. 일반적으로 신경망은 임의의 개수의 은닉층을 가질 수 있다. 그러나 은닉층의 활성화 함수(activation function)가 시그모이드 함수와 같이 적절하게 부드러운(smooth) 활성화함수를 가지며 하나의 은닉층을 갖는 다층 퍼셉트론은 임의의 모든 함수에 대하여 정확하게 근사시킬 수 있다^[3]. 따라서 하나의 은닉층을 가지는 다층 퍼셉트론을 EBP로 학습시킬 때 각 은닉 노드는 입력 패턴들의 특징을 구분해주는 하나의 기능적 단위의 역할을 한다^[4]. 이와 같이 은닉 노드가 다층 퍼셉트론의 기능적인 단위의 역할을 하기 때문에 EBP가 국부적 최적해에 수렴한 경우에도 모든 은닉 노드의 가중치를 재학습시킬 필요가 없다. 따라서 EBP가 국부적 최적해에 수렴했을 때에도 많은 은닉 노드들이 유효한 특징을 가질 수 있기 때문에 부적절하게 학습된 은닉 노드에 대해서만 재학습시켜서 국부적 최적해를 벗어나 전역적 최적해를 탐색할 수가 있다.

또한 국부적 최적해에 수렴한 상태에서 각 은닉 노드를 평가하였을 때 대부분의 은닉 노드가 전역적 최

적해에 가깝다면 은닉 노드의 개수가 불충분한 현상으로 판단하여 은닉 노드를 추가한다. 이때 현재 학습되고 있는 신경망 구조에 추가된 노드를 교란시킬 때의 방법과 마찬가지로 재초기화한 후 포함시킨다. 이것은 마치 추가된 노드가 예전부터 존재해온 것처럼 간주되어 기존의 노드들과 마찬가지로의 방법으로 평가되고, 교란시킬 노드의 선택에서도 동등한 기회를 부여받아서 결국 일관된 확률적 가중치 교란으로 국부적 최적해를 벗어난다.

이와 같이 제안한 EBP/SP 알고리즘은 국부적 최적해 문제와 신경망 구조의 자동 설계 문제를 기존의 여러 연구^[5-7]에서 적용한 방법들과는 다른 전혀 새로운 방법으로 해결한다. 즉 국부적 최적해 문제를 각 은닉 노드의 평가에 따른 확률적 교란에 의하여 해결한다는 점과, 또한 은닉 노드 추가 역시 이러한 확률적 교란의 개념을 일반화하여 교란 및 노드 추가를 일관된 방법으로 수행한다는 점이 크게 다르다.

본 논문은 다음과 같이 구성된다. 2장에서는 국부적 최적해 문제 및 신경망 구조 설계 문제를 해결하고자 한 여러 관련 연구에 대하여 기술한다. 3장에서는 다층 퍼셉트론에서 은닉 노드의 역할에 대하여 알아보고, 국부적 최적해 문제와 적절한 신경망 구조 설계를 해결하는 제안한 EBP/SP 학습 알고리즘을 구체적으로 설명한다. 4장에서는 제안한 방법의 일반화 성능에 대하여 설명하고, 5장에서는 여러 가지 벤치마크 문제들 및 실제 문제에 대한 실험 결과를 제시한다. 끝으로 6장에서는 결론 및 향후 연구 방향을 기술한다.

II. 관련 연구

신경망 학습에서의 국부적 최적해 문제와 신경망 구조 설계 문제는 매우 중요한 문제로 이를 해결하기 위한 많은 관련 연구가 있었다. 대표적인 연구로는 전역적 최적화 방법인 유전자 알고리즘(genetic algorithm) 기법을 적용한 연구를 들 수 있다. Ng^[5]는 신경망의 가중치 집합(weight set)을 교란시켜 다음 세대의 개체군을 형성하여 진화시킴으로써 EBP 학습의 국부적 최적해 문제를 해결하고자 하였다. 그러나 이러한 유전자 알고리즘 기법을 적용한 신경망 학습 방법은 학습이 오래 걸리기 때문에 많은 양의 데이터를 가지는 실용적인 문제에 적용하기 어렵다.

이외에 작은 개수의 은닉 노드를 갖는 구조로 출발

하여 만족할 만한 해에 도달할 때까지 은닉 노드 및 가중치를 추가하는 방식을 취하는 생성 알고리즘(constructive algorithm) 계열의 방법^[6,7]이 있다. Hirose^[6]는 작은 은닉 노드 개수로 학습을 시작하여 국부적 최적해에 수렴할 때마다 은닉 노드를 추가한 후, 전역적 최적해에 도달했을 경우 은닉 노드를 하나씩 삭제하는 방식으로 만일 다시 학습시켜 수렴하게 되면 또 다른 은닉 노드를 삭제하게 되고 그렇지 않고 수렴하지 못하면 바로 전에 얻은 상태를 최적의 해로 간주한다. 이와 유사하게 생성 알고리즘 계열의 접근 방법을 이용한 연구^[7]가 있다. 이 방법 역시 작은 개수의 은닉 노드를 가지고 학습을 시작하여 은닉 노드를 하나씩 추가해 가는 방식을 취한다. 한가지 특징은 많은 연구에서 흔히 사용되어온 평균 자승 오차를 최소화하도록 학습하는 것이 아니라 유사도(likelihood) 함수를 정의하여 이를 최대화시키는 학습 방법을 사용한다는 점이 특이하다. 새로운 은닉 노드가 추가될 때 유사도 함수는 항상 증가함을 보장해주고, 또 이러한 유사도 함수의 특성은 이 알고리즘의 종료를 보장해준다. 이러한 생성 알고리즘 계열의 방법들은 초기에 추가되는 은닉 노드가 부적절하게 학습된다면 결코 전역적 최적해를 얻지 못한다.

III. EBP/SP 학습 알고리즘

다층 퍼셉트론에서 은닉 노드를 보는 가장 단순한 관점은 선형 임계 함수(linear threshold function)에 대한 근사로 보고, 또 함수 근사에서 은닉 노드를 일종의 기저 함수(basis functions)로 본다^[4]. 이와 같이 패턴 분류나 함수 근사에서 각 은닉 노드가 자신의 영역에서 기능적 단위로서의 역할을 수행한다. 이 논문에서는 이와 같은 은닉 노드의 특성을 이용하여 EBP 학습의 문제점인 국부적 최적해 문제와 신경망 구조 설계 문제를 해결한다. 즉, 은닉 노드 단위의 확률적 가중치 교란 방법을 이용하여 일관된 방법으로서의 두 문제를 동시에 해결한다. 제안한 EBP/SP 알고리즘은 EBP가 국부적 최적해에 수렴했을 때 각 은닉 노드의 기여도(strength)를 평가하여 부적절하게 학습된 노드들만을 확률적으로 교란시킴으로써 국부적 최적해를 벗어나 전역적 최적해를 탐색하도록 한다. 동시에 작은 개수의 은닉 노드로 학습을 시작하여 은닉 노드가 부족한 현상을 확률적으로 결정하여 노드를

추가한다.

EBP/SP 알고리즘에서의 첫 번째 단계는 EBP 학습을 수행한다. EBP 학습은 기존의 표준 EBP 학습^[2]과 같다. 다음은 종료 조건으로 오차 E 가 오차 허용치 E_{tol} 보다 작으면 학습을 종료한다. 세 번째 단계는 EBP 학습에 의해 국부적 최적해에 수렴한 것인지를 조사한다. 다음은 각 은닉 노드의 유효한 정도를 평가하고, 각 은닉 노드를 교란(재초기화) 할 확률을 노드의 유효한 정도에 따라 계산한다. 마지막으로 앞에서 이미 계산된 각 은닉 노드의 평가된 값에 의한 확률을 이용하여 노드를 추가할 확률도 계산한다. 노드의 교란과 추가를 앞 단계에서 계산한 확률에 의하여 수행한 후 다시 EBP 학습을 진행한다.

1. 은닉 노드 평가

다층 퍼셉트론에서 은닉 노드는 특징 추출자(feature detectors)의 역할을 한다^[4]. 즉 은닉 노드는 학습 데이터 상에 존재하는 특징을 구분한다. 따라서 주어진 문제를 해결하는 적절한 특징을 찾고자 할 때 은닉 노드를 조정하는 것이 중요하다.

EBP/SP 알고리즘에서 확률적 교란의 개념은 간단하다. 즉 신경망 구조에서 중요하지 않게 평가된 은닉 노드들만의 가중치를 재초기화 한다. 재초기화 할 은닉 노드는 신경망에서 그 은닉 노드의 기여도(strength)에 따른 확률에 의해 선택된다. 여기서 은닉 노드 i 의 기여도는 현재의 신경망의 오차와 노드 i 가 제거된 신경망에 의해 평가된 오차의 차이로 정의된다. 평균 자승 오차(mean square error) E 는 다음과 같이 정의된다:

$$E = \frac{1}{P} \sum_p E_p = \frac{1}{PN_o} \sum_p E_p \sum_k^N (T_{pk} - O_{pk})^2. \quad (1)$$

여기서 F 는 훈련 패턴 개수이고, N_o 는 출력 노드 개수, 그리고 T_{pk} 와 O_{pk} 는 각각 패턴 p 에 대한 k 번째 출력 노드의 목표 출력값과 실제 출력값을 의미한다.

충분히 학습된 신경망의 경우 노드 i 의 기여도는 다음과 같이 정의될 수 있다:

$$S(i) = E'(i) - E. \quad (2)$$

여기서 $E'(i)$ 는 노드 i 가 제거된 신경망에 의해 평가된 오차를 의미한다. 따라서 $S(i)$ 는 은닉 노드 i 가 신경망에서 역할하는 정도를 의미하는 것으로 해석될 수 있다. 즉 만일 $S(i)$ 가 0에 가깝다면 은닉 노드 i 는

아주 미약한 역할을 한다는 사실을 의미하고, 반면에 $S(i)$ 가 1에 가깝다면 그 은닉 노드는 신경망에서 중요한 역할을 한다는 사실을 의미한다. $S(i)$ 는 추후 EBP/SP 알고리즘을 설명하는데 이용된다.

2. 은닉 노드 교란 방법

EBP/SP 알고리즘에서는 EBP 학습에 의해 국부적 최적해에 수렴한 것인 지를 다음과 같이 조사한다: 먼저 오차 변화율 ΔE 를 $|E(t) - E(t-\tau)|/\tau$ 에 의해 계산한다. 여기서 τ 는 ΔE 를 추정하기 위한 시간 지연(time lag) 즉 epoch 수를 의미하고, t 는 현재의 epoch 수이다. 만일 $\Delta E < \lambda$ 이고 $E > E_{\theta}$ ($\geq E_{\text{tol}}$)이면 국부적 최적해에 수렴한 것으로 판단한다. 여기서 λ 와 E_{θ} 는 각각 사용자에게 의해 주어지는 patience 파라미터와 오차 임계값을 의미한다. 이때 E_{θ} 는 E 가 E_{tol} 에 충분히 가까울 때 불필요한 교란을 막음으로써 성능을 개선하기 위해 사용된다.

만일 시간 t 에서 국부적 최적해에 수렴한 것으로 판단되면 각 은닉 노드를 평가한 값에 의한 확률에 따라 은닉 노드를 교란함으로써 국부적 최적해를 탈출한다. 각 은닉 노드의 교란 확률은 다음 식 (3)과 같이 결정된다:

$$P^{\text{trb}}(i) = e^{-\frac{\Delta E_i \cdot N_i}{E}} \quad (3)$$

여기서 $\Delta E_i = E_i - E$ 이고, E_i 는 노드 i 가 재초기화된 신경망에 의해 평가된 오차를 의미한다. E 는 신경망이 국부적 최적해에 수렴했을 때의 남아있는 오차를 의미하고, N_i 는 현재의 은닉 노드 개수를 나타낸다. 여기서 위의 식 (3)과 시뮬레이티드 어닐링(simulated annealing)^[18]에서의 Metropolis 기준과의 유사성을 비교해 볼 수 있다. 즉 ΔE_i 는 에너지 변화량 Δe 과 대응되며 E 는 온도 T 에 대응된다고 볼 수 있다. 따라서 식 (3)은 오차 E 가 클 때는 가중치 공간상에서 넓은 탐색 범위를 가지게 되고(전역적 탐색), 오차가 작을 때는 좁은 탐색 범위를 가지게 된다. 여기서 노드 i 와 연결된 가중치가 충분히 작은 값으로 재초기화된다고 가정하면 노드 i 의 기여도 $S(i)$ 는 ΔE_i 를 근사시킨다는 사실을 알 수 있다. 또한 식 (3)에서 오차가 크다면 많은 개수의 은닉 노드가 교란될 것이고, 반면에 오차가 작다면 적은 개수의 은닉 노드가 교란될 것이다.

3. 은닉 노드 추가

주어진 문제에 적절한 은닉 노드 개수를 결정하는데 있어서 기본적인 방향은 작은 개수의 은닉 노드로 학습을 시작하여 은닉 노드가 부족한 현상을 확률적으로 결정하여 추가하는 것이다. 이때 노드 추가에 사용되는 확률을 현재의 은닉 노드 중에서 가장 부적절하게 학습된 노드 즉 교란 확률이 가장 큰 노드 j 가 교란되지 않을 확률을 이용하여 계산한다:

$$1 - \{P^{\text{trb}}(j)\}, \quad j = \arg \min \Delta E_i \quad (4)$$

이와 같이 계산한 확률을 기본적인 노드 추가 확률로 정함과 동시에 현재의 점에서 계속 가중치 교란을 반복해도 더 낮은 최적해의 끝개 범주를 찾는데 실패하는 경우, 즉 국부적 최적해 탈출에 실패하는 횟수를 이용하여 노드 추가 확률을 계산한다. 즉, 모든 은닉 노드가 제 역할을 다하고 있으나 아직 줄여야 할 오차가 크다면 계속하여 국부적 최적해 탈출에 실패하는 현상이 발생할 수 있다. 다시 말해 반복적으로 국부적 최적해 탈출에 실패하는 상황을 극복하기 위해 국부적 최적해 탈출에 실패하는 횟수 N_j 를 노드 추가 확률에 반영하고자 한다. 즉, N_j 가 증가함에 따라 노드 추가 확률도 증가되도록 하여 결국 노드 추가에 성공하도록 한다. 이와 같이 하여 위에서 기술한 은닉 노드 부족 현상으로 발생하는 문제를 해결할 수 있다. 이때 노드 추가 확률을 N_j 회의 실패 후에 최초의 노드 추가에 성공하는 사건이 나타날 때까지의 확률들을 누적시킨 값으로 한다. 이것은 실패 횟수 N_j 가 기하분포(geometric distribution)^[9]를 따름을 의미한다. 이와 같이 N_j 가 기하분포를 따른다는 사실에 근거하여 다음과 같이 노드 추가 확률을 결정할 수 있다. 즉, 가중치 교란을 통하여 국부적 최적해를 탈출하는데 N_j 회 실패한 경우 은닉 노드 추가 확률은 아래의 식 (5)와 같다.

$$P(K < N_j) = \sum_{k=0}^{N_j-1} \{P^{\text{trb}}(j)\}^k \{1 - P^{\text{trb}}(j)\} \\ = 1 - \{P^{\text{trb}}(j)\}^{N_j} \quad (5)$$

위의 식 (5)는 N_j 가 증가하면 노드 추가 확률도 점차 증가하는 특성을 가진다. 한가지 문제는 N_j 가 1일 때의 확률 값이 매우 크게 되는 상황이 발생할 수 있다는 점이다. 즉, 국부적 최적해 탈출에 단 한번 실패

패했다고 해서 쉽게 노드를 추가하는 경우를 의미한다.

일반적으로 신경망의 구조는 적절하면서도(전역적 최적해를 찾을 수 있을 만큼 충분한 은닉 노드의 개수를 갖는 구조) 동시에 단순한 구조를 가져야 한다. 이것은 단순한 구조일수록 좋은 일반화 성능(generalization capability)을 갖는다는 Minimum Description Length(MDL)^[10]의 원리를 따른다. 그러나 위의 식 (5)에서와 같이 쉽게 노드를 추가한다면 결코 단순한 신경망 구조를 얻을 수 없을 것이다. 따라서 이러한 점을 보완하기 위하여 노드 추가 확률을 다음과 같이 변형한다.

$$P^{add} = \{1 - P^{trb}(j)\}^{\alpha} [1 - \{P^{trb}(j)\}^N], \quad \alpha = 0, 1, 2, \dots \quad (6)$$

즉, 식 (5)에 $\{1 - P^{trb}(j)\}^{\alpha}$ 의 항을 추가한다. 이것은 적어도 처음 몇 번(즉, α)은 항상 국부적 최적해 탈출에 실패했다고 간주하는 효과를 주게 된다. 이 α 값의 적절한 설정을 통하여 더 단순한 신경망 구조를 얻을 수 있다. 이와 같이 새로운 은닉 노드 추가를 위의 식 (6)과 같은 확률로 결정한다. 만일 식 (6)에 의해 노드를 추가하는 경우로 결정되면 새로운 은닉 노드를 추가하게 된다. 이때 새로운 노드를 추가한다는 의미는 마치 그 은닉 노드에 연결된 모든 가중치(입력 가중치와 출력 가중치)가 0인 노드가 신경망에 존재하고 있어서 그 노드에 연결된 모든 가중치를 0에 가까운 임의의 작은 값으로 재초기화하는 것과 같다. 따라서 새로운 은닉 노드 추가를 가중치 교란(식 (6) 참조)의 특수한 형태로 볼 수 있다.

제안한 EBP/SP 알고리즘의 가장 큰 장점은 일관된 방법에 의하여 가중치 교란 및 노드 추가를 수행한다는 점이다. 즉 확률적 가중치 교란에 의해 국부적 최적해를 벗어나 전역적 최적해를 탐색하고, 또 동시에 노드 추가를 수행함으로써 EBP에서 중요한 두 가지 문제를 해결하고 있다. EBP/SP의 구체적인 알고리즘은 그림 1과 같다.

IV. 일반화 성능

신경망의 일반화 성능은 학습해야 할 자료의 양과 해결해야 할 문제의 복잡도, 그리고 신경망 자체의 자유도(예: 가중치, 노드 수) 등에 크게 영향을 받는다. 일반적으로 신경망의 자유도를 제한시키면 학습 데이

터상에 존재하는 큰 경향만을 추출하고 작은 경향은 무시한다는 사실이 알려져 있다^[10]. 이와 같이 신경망의 능력을 제한하여 일반화 성능을 향상시키는 방법은 크게 두 가지로 분류된다: 민감도를 이용하는 방법(sensitivity methods)^[11,12]과 벌칙 항을 이용하는 방법(penalty-term methods)^[13-15].

단계 1. MLP의 가중치를 임의의 작은 수로 초기화한다.
 단계 2. 종료 조건을 조사한다.
 2.1. 만일 시간 t 에서 $E(t) < E_{thr}$ 이면 학습을 종료한다.
 2.2. 그렇지 않으면 1 epoch 학습한다.
 단계 3. 오차 변화를 $\delta E(t) = |E(t) - E(t-1)|$ 를 계산한다.
 단계 4. 만일 $\delta E(t) < \lambda$ 이고 $E(t) > E_0$ 이면 다음 단계들을 수행한다.
 4.1. $E = E(t)$, $N_j = 0$
 4.2. 각 은닉 노드를 평가한다:
 $\Delta E_j = E_i - E$
 4.3. 각 노드를 교란할 확률을 다음과 같이 계산한다:
 $P^{trb}(j) = e^{-\frac{\Delta E_j \cdot N_j}{E}}$
 4.4. 은닉 노드를 추가할 확률을 다음과 같이 계산한다:
 $P^{add} = \{1 - P^{trb}(j)\}^{\alpha} [1 - \{P^{trb}(j)\}^N]$, $\alpha = 0, 1, 2, \dots$
 여기서 $j = \arg \min \Delta E_j$ 이다.
 4.5. $P^{trb}(j)$ 와 P^{add} 의 확률로 교란 및 노드 추가를 수행한다. 이때 은닉 노드가 추가되면 $N_j = 0$ 으로 한다.
 4.6. $E(t) < E - \lambda$ 이면 단계 2로 간다. 그렇지 않으면 1 epoch 학습한다.
 4.7. 오차 변화를 $\delta E(t)$ 를 계산한다.
 4.8. $\delta E(t)$ 가 λ 보다 작으면
 $N_j = N_j + 1$, 가중치를 복원하고 단계 4-2로 간다.
 4-9. 단계 4-6으로 간다.
 단계 5. 단계 2로 간다.

그림 1. 제안한 EBP/SP 학습 알고리즘
 Fig. 1. The proposed EBP/SP algorithm.

일반적으로 민감도를 이용하는 방법들은 이미 학습된 신경망에서 민감도를 추정하여 가중치 혹은 노드를 제거하는 방법을 의미한다. 한편 벌칙 항을 이용하는 방법은 기존의 학습에 사용되는 오차 함수에 벌칙항을 추가하여 불필요한 가중치를 0 근처의 값이 되도록 함으로써 결국 그들을 학습 중에 제거하는 효과를 얻는 방법이다. 여기서 가중치를 실제로 제거하지 않더라도 신경망은 더 작은 크기의 신경망으로서 동작하게 된다. 벌칙 항을 이용하여 일반화 성능을 향상시키기 위한 방법 중에 가장 단순한 형태로 weight decay^[14]가 있다. Weight decay는 아래의 식 (7)과 같이 기존의 학습을 위한 오차 함수에 학습 데이터에 의해 부적절하게 결정된 가중치를 0을 향해 끌어들이는 효과를 주는 항(식 (7)의 오른쪽 두 번째 항)을 추가한 새로운

오차 함수를 이용하여 학습시킨다. 새로운 오차 함수는 다음과 같이 정의된다:

$$E = \frac{1}{2} \sum_k (O_k - T_k)^2 + \gamma \sum_i \frac{w_i^2}{1 + w_i^2}. \quad (7)$$

여기서 O_k 와 T_k 는 각각 신경망의 실제 출력값과 목표 출력값을 의미하고, w_i 는 신경망의 각각의 가중치를, W 는 전체 가중치 집합을 의미하며, γ 는 실험에 의하여 결정되는 파라미터로 신경망의 학습 과정에서 각 가중치 크기의 출력 오차에 대한 상대적 중요도를 나타낸다. 이 γ 를 점차 증가시켜 가면 학습에 걸리는 시간이 많이 소요되고, 이와 달리 감소시켜 가면 순수한 EBP 학습에 접근하기 때문에 학습 시간과 향상시킬 성능을 고려하여 적절한 값을 정해줄 필요가 있다. 이 논문에서는 제안한 방법의 일반화 성능을 향상시키기 위한 방법으로 위의 weight decay를 이용한 실험을 통하여(credit-screening 데이터에 대한 실험 참조) 제안한 방법이 충분히 오래 동안 학습되더라도 과적응이 발생하지 않음은 물론 일반화 성능이 향상됨을 확인할 수 있었다. 이외에 일반화 성능을 향상시키기 위한 방법 중에 벌칙 항을 추가하는 여러 방법들, 즉 은닉 노드의 에너지 항을 정의하여 추가함으로써 은닉 노드의 활성 값을 제한하여 최적의 은닉 노드 개수를 정하고자 하는 방법^[13], 입력에 대한 출력값의 민감도를 줄이기 위해 은닉 노드의 활성 값을 제한하는 추가 항을 이용하는 방법^[15] 등을 EBP/SP 알고리즘에 적용해도 역시 비슷한 효과가 있을 것으로 기대된다. 이 논문에서 제안한 EBP/SP 알고리즘은 EBP 알고리즘의 근본적인 문제인 학습 데이터 상의 국부적 최적해 문제를 다루기 위한 방법에 주안점을 두는 것으로 이러한 일반화 성능을 향상시키는 문제 즉 테스트 데이터 상에서의 국부적 최적해 문제를 해결하는 문제에 대한 더 세부적인 연구는 추후 연구 과제이다.

V. 실험 결과

이 절에서는 제안한 EBP/SP 학습 알고리즘의 성능 평가를 위해 여러 가지 문제에 대한 실험에 의해 얻은 결과를 제시하고 분석한다. 또한 다른 관련 연구와의 성능 면에서의 직접적인 비교를 통하여 EBP/SP 학습 방법의 성능을 분석한다. 관련 연구와의 성능 비

교를 위해 사용된 방법들로는 순수한 EBP와 EBP/SP, 유전자 알고리즘에 근거한 Ng의 방법^[5] 등을 비교한다. 실험에 사용된 문제들은 UCI(University of California, Irvine)의 machine learning databases로부터 얻은 패턴 분류 문제들이다(parity, two-spirals, credit-screening). 이 연구에 사용된 모든 실험은 동일한 조건하에서 실행하였다. 즉 학습에 필요한 파라미터값을 모든 방법의 실행(run)에 동일하게 주었다. 모든 실험은 학습률 η 와 관성 항 β 를 고정시킨 채 가중치의 초기 값 설정만을 다르게 하여 수행한 것이다. 또 모든 방법에 사용된 학습 데이터에서 출력 노드의 시그모이드(sigmoid) 함수의 특성을 반영하기 위하여 입력 패턴에 대한 목표 출력값을 0.1과 0.9로 정하여 학습시켰다. 이 논문의 그림과 표에서 "EBP"는 순수한 EBP 학습 방법을, "EBP/SP"는 제안한 학습 방법을, "EBP#"는 EBP/SP에 의해 얻어진 평균 은닉 노드 개수를 갖는 신경망 구조에 순수한 EBP를 적용한 학습 방법을 의미하고 "Ng"는 유전자 알고리즘에 기반한 학습 방법인 Ng의 방법^[5]을 의미한다. 학습의 종료는 최대 반복 횟수를 정해주는 방법과 각 문제에 따라 100%의 정확도를 보이는 평균 지수 오차를 정해주는 방법을 병행 사용하여 둘 중의 어느 하나만을 만족하면 학습을 종료하도록 하였다. 또한 모든 표에서 "Ng"는 정해진 신경망 구조의 가중치 공간만을 탐색하는 방법으로 평균 은닉 노드 개수는 EBP#의 경우와 같다. 또 "Ng"의 모든 실험은 Ng의 방법^[5]에서 제시하는 파라미터값을 그대로 사용한 결과이다.

1. Parity 문제

N-비트 패리티 문제는 신경망의 학습 알고리즘의 성능 평가를 위해 자주 사용되는 문제로 입력은 N차원 공간상에서 2^N 의 패턴들로 구성되며, 각 패턴은 N-비트의 이진 벡터이다. 출력값 i 는 패턴에서 1의 개수가 홀수 개이면 1, 그렇지 않으면 0이다. 여기서는 4-비트, 5-비트, 6-비트 parity에 대하여 실험하였다. 표 1은 5개의 서로 다른 가중치 초기 값에 대하여 EBP, EBP#, EBP/SP 모두의 경우에 동일하게 학습률 η 는 0.1, 관성 항 β 는 0.8을 사용하고, 종료 조건으로 최대 epoch는 100000, 혹은 출력 오차가 0.002가 될 때까지 수행한 결과를 보여준다. 표에서 EBP, EBP#, EBP/SP의 평균 반복 횟수는 4-비트, 5-비트,

6-비트 각각의 경우에 대하여 최대 반복 횟수 100000 까지 혹은 평균 자승 오차가 0.002가 될 때까지의 평균 반복 횟수를 의미한다. "Ng"에 사용된 파라미터들로 초기 가중치는 $[-0.3, 0.3]$ 범위에서 임의로 추출된 값을 사용하고, 돌연변이 확률 P_m 은 0.8, 교란되는 가중치의 탐색 범위 D는 0.5, 그리고 개체군의 크기(POPSIZE)는 5개로 하였다. 표에서도 알 수 있는 바와 같이 제안한 EBP/SP 알고리즘의 실험 결과는 모든 경우에 100%의 평균 분류율을 보여준다. 표 1의 평균 분류율은 훈련 데이터에 대한 성능으로 문제의 특성상 테스트 데이터에 대한 성능 평가는 무의미한 것으로 판단되어 생략한다.

표 1. 4, 5, 6-비트 parity 문제에 대한 실험 결과

Table 1. The experimental results for the 4, 5, 6-bit parity problems.

Parity 문제	시행 횟수	평균 은닉 노드 개수	평균 반복 횟수	평균 분류율(%)	
				평균	표준편차
4-bit	EBP	5	3	81097.2	97.5, 3.423
	EBP#	5	4	74139	96.3, 3.423
	EBP/SP	5	4	22424	100.0, 0.0
	Ng	5	4	-	98.75, 2.795
5-bit	EBP	5	3	100000	96.88, 0.0
	EBP#	5	5	100000	96.88, 0.0
	EBP/SP	5	4.7	36370	100.0, 0.0
	Ng	5	5	-	99.38, 1.395
6-bit	EBP	5	4	100000	94.4, 3.046
	EBP#	5	6	58073.6	99.4, 0.854
	EBP/SP	5	6.2	40973	100.0, 0.0
	Ng	5	6	-	99.69, 0.698

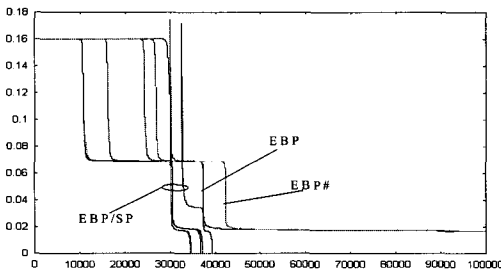


그림 2. 5-비트 parity 문제에 대한 EBP, EBP#, EBP/SP의 학습 곡선 비교

Fig. 2. Learning curves of EBP, EBP#, and EBP/SP algorithms for the 5-bit parity problem.

그림 2는 EBP/SP 알고리즘의 효과를 잘 표현해주는 대표적인 그래프로 EBP와 EBP#의 경우 충분히 오래 동안 학습을 진행시켜도 결코 국부적 최적해를 벗어나지 못하는 반면에 EBP/SP는 가중치 교란 및 새로운 은닉 노드 추가에 의하여 학습률과 관성 항을 전혀 조정하지 않고서도 매우 효과적으로 국부적 최적해를 벗어나 전역적 탐색을 수행함으로써 수렴 시간을 크게 단축시킨다.

2. Two-spirals 문제

Two-Spirals 문제의 경우 신경망은 두 개의 실수 값을 갖는 입력과 하나의 출력값을 가진다. 학습 데이터로는 총 194개의 X-Y 좌표 값으로 구성되는데, xy 평면상에 2개의 나선을 그리는 점들의 집합이다. 각 나선은 각각 1과 0을 나타내는 총 97개의 점들로 이루어진다. 입력 패턴은 나선형의 두 선이 원점을 기준으로 세 번 꼬여있는 모양이다. 입력 패턴의 반이 1의 출력값을 가지는 한 나선이고, 나머지 반은 다른 나선으로 0의 출력값을 가진다(그림 3 (a) 참조). 일반적으로 two-spirals 문제는 EBP 계열의 알고리즘이 해결하기 어려운 문제로 잘 알려져 있다^[7]. 그림 3 (b)는 EBP/SP에 의하여 학습된 신경망의 대표적인 출력 패턴으로 은닉 노드의 개수가 28개이며 100%의 분류율을 보이는 경우의 출력 패턴을 보여준다.

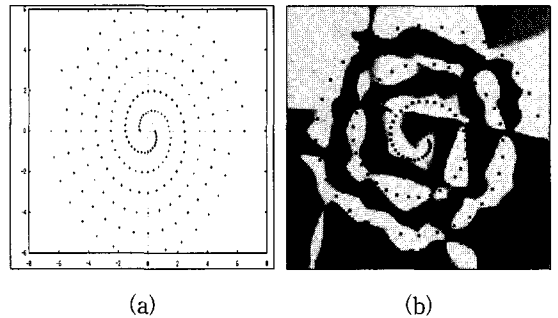


그림 3. Two-spirals 문제의 학습 데이터(a)와 EBP/SP에 의하여 학습된 신경망의 출력 패턴(b)
Fig. 3. Training points for the two-spirals problem (a) and output pattern for one network trained with EBP/SP (b).

표준 EBP와 제안한 EBP/SP를 비교하기 위해 5개의 서로 다른 초기 가중치 집합을 사용하여 실험하였다. 표 2는 5개의 서로 다른 가중치 초기 값에 대하여 EBP, EBP#, EBP/SP 모두의 경우에 동일하게 학습률 η 는 0.1로 하고, 관성 항은 사용하지 않았으며(관

성 항 β 를 0.0), 종료 조건으로 최대 epoch는 200000, 혹은 출력 오차가 0.004가 될 때까지 수행한 결과를 보여준다. 표에서 EBP, EBP#, EBP/SP의 평균 반복 횟수는 각각의 경우에 대하여 최대 반복 횟수 200000까지 혹은 평균 지승 오차가 0.004가 될 때까지의 평균 반복 횟수를 의미한다. 표 2에서 보여준 평균 분류율은 표 1에서와 마찬가지로 훈련 데이터에 대한 성능으로 문제의 특성상 테스트 데이터에 대한 성능 평가는 무의미한 것으로 판단되어 생략한다. “Ng”에 사용된 파라미터들로는 5.1절의 실험에서와 같다. 표에서도 알 수 있는 바와 같이 제안한 EBP/SP 알고리즘의 실험 결과는 모든 경우에 100%의 학습 성능을 보여준다. 한편 표 2에서 EBP는 20개의 은닉 노드를 가진 경우의 실험 결과이고, EBP#의 경우는 EBP/SP의 실험 결과에 의하여 얻어진 평균 은닉 노드 개수인 26개의 은닉 노드를 사용한 경우이다. EBP/SP의 경우는 처음에 20개의 은닉 노드로 시작하여 EBP 학습이 국부적 최적해에 수렴하게 되면 가중치 교란 및 노드 추가를 확률적으로 수행하여 전역적 최적해에 수렴하게 된 실험 결과이다. 표에서 알 수 있듯이 제안한 EBP/SP 알고리즘의 경우 EBP나 EBP#에 비해 학습 시간이 크게 단축되었음을 알 수 있다.

표 2. Two-spirals 문제에 대한 실험 결과
Table 2. Experimental results for the two-spirals problem.

Two-spirals 문제	시행 횟수	평균 은닉 노드 개수	평균 반복 횟수	평균 분류율(%)	
				평균	표준편차
EBP	5	20	200000	87.11	2.389
EBP#	5	26	200000	98.04	1.173
EBP/SP	5	26	105987.8	100.0	0.0
Ng	5	26	-	93.71	4.757

그림 4는 EBP, EBP#, EBP/SP의 학습 곡선을 비교한 그래프로 EBP와 EBP#의 경우 충분히 오래 동안 학습을 진행시켜도 결코 국부적 최적해를 벗어나지 못하는 반면에 EBP/SP는 가중치 교란 및 새로운 은닉 노드 추가에 의하여 학습률과 관성 항을 전혀 조정하지 않고서도 효과적으로 국부적 최적해를 벗어남을 보여준다. 한편, 식 (6)에서의 α 의 크기는 얻어지는 해에 대한 신경망 구조의 복잡도와 수렴 속도에 영향을 준다. 즉, α 가 클수록 더 단순한 신경망 구조를 얻

게 되지만 수렴 속도는 느리게 된다. 그림 5는 이와 같은 α 의 크기에 따른 신경망의 복잡도 및 수렴 속도 간의 관계를 보여준다.

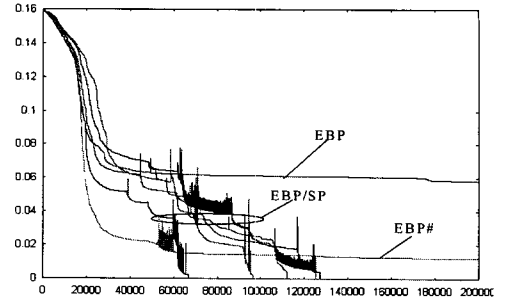


그림 4. Two-spirals 문제에 대한 EBP, EBP#, EBP/SP의 학습 곡선 비교
Fig. 4. Learning curves of EBP, EBP#, and EBP/SP algorithms for the two-spirals problem.

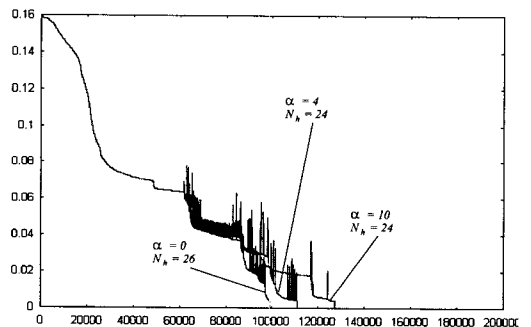


그림 5. α 의 크기에 따른 신경망 구조의 복잡도와 수렴 속도간의 관계
Fig. 5. The relationship between the convergence rate and the complexity of resulting network, depending on the magnitude of α .

3. Credit-screening 문제

이 데이터는 신용카드 승인 데이터로서 307개 (44.5%)의 승인 데이터와 383개(55.5%)의 비 승인 데이터로 구성되어 있으며, 15개의 특징을 가지고 있는데, 신경망에의 적용을 위해 인코딩(encoding)에 의하여 51개의 입력 개수를 얻는다. 345개의 훈련 데이터와 175개의 테스트 데이터를 이용하여 승인과 비승인의 두 클래스로 분류하는 것이 문제의 해이다. 표 3은 credit-screening 데이터에 대하여 EBP, EBP#, EBP/SP 알고리즘을 적용한 결과로 실험에서 학습에 사용된 파라미터값들로는 5개의 서로 다른 가중치 초

기 값에 대하여 EBP, EBP#, EBP/SP 모두의 경우에 동일하게 학습률 η 는 0.1, 관성 항 β 는 0.8을 사용하였으며, 종료 조건으로 최대 epoch는 30000, 혹은 출력 오차가 0.001이 될 때까지 수행한 결과이다.

표 3. Credit-screening 문제에 대한 실험 결과(1)

Table 3. Experimental results for the credit-screening problem(1).

학습 알고리즘	평균 은닉 노드 개수	시행 회수	평균 반복 횟수	평균 분류율(%)			
				훈련 데이터		테스트 데이터	
				평균	표준편차	평균	표준편차
EBP	2	5	30000	95.65	0.0	83.37	0.318
EBP#	6	5	10135	99.77	0.243	79.09	1.792
EBP/SP	6	5	21354	100.0	0.0	76.86	0.958

표에서 알 수 있듯이 훈련 데이터에 대해서는 EBP/SP 알고리즘이 단연 우수하지만 테스트 데이터에 대한 결과를 보면 EBP/SP 알고리즘의 경우 과적응이 발생했음을 알 수 있다. 따라서 이러한 과적응의 발생을 제거하여 일반화 성능을 향상시키기 위해 4장에서 기술한 방법을 적용하여 실험하였다.

표 4는 EBP/SP와 EBP#의 일반화 성능을 향상시키기 위하여 weight decay 항을 이용한 실험 결과를 보여 준다. 실험에서 학습에 사용된 파라미터값들로는 5개의 서로 다른 가중치 초기 값에 대하여 EBP#과 EBP/SP 모두의 경우에 동일하게 학습률 η 는 0.1, 관성 항 β 는 0.8을 사용하였으며, 가중치 크기의 출력 오차 E에 대한 상대적 중요도를 나타내는 γ 는 0.00005~0.0001 범위의 값을 사용하였다. 종료 조건으로 최대 epoch는 10000, 혹은 출력 오차가 0.001이 될 때까지 수행하였다.

표 4. Credit-screening 문제에 대한 실험 결과(2)

Table 4. Experimental results for the credit-screening problem(2).

학습 알고리즘	평균 은닉 노드 개수	시행 회수	평균 반복 횟수	평균 분류율(%)			
				훈련 데이터		테스트 데이터	
				평균	표준편차	평균	표준편차
EBP# with WD	8	5	10000	95.65	1.178	82.33	0.778
EBP/SP with WD	7.6	5	10000	96.81	1.728	84.88	1.165

WD: weight decay

위와 같이 일반화 성능을 향상시키기 위한 방법들

적용한 결과로 표 4를 보면 EBP/SP 알고리즘의 훈련 데이터에 대한 실험 결과는 물론 테스트 데이터에 대한 실험 결과 역시 EBP# 보다 우수함을 알 수 있다. 이와 같은 사실로 미루어 EBP/SP 알고리즘은 기존의 오차 함수에 벌칙 항을 추가한 새로운 오차 함수를 이용하여 학습시킴으로써 일반화 성능을 향상시키는 문제도 쉽게 해결할 수 있음을 알 수 있다. 또한 EBP/SP 알고리즘에서 일반화 성능 향상을 위해 사용한 방법은 추가적인 계산 량이 별로 증가하지 않는다는 점을 장점으로 들 수 있으며, 일반화 성능을 향상시키기 위한 방법 역시 가중치 교란 방법에 자연스럽게 포함됨을 알 수 있다. 한편 새로운 파라미터 γ 에 의해 추가된 항의 중요도를 바꾸어 가면서 학습시킬 수 있다. 학습의 초기 상태에서는 매우 작은 값으로 가중치를 초기화시키므로 추가된 항은 크게 영향을 미치지 못하고 식 (6)의 첫 번째 항인 오차 역전파 항에 의해 학습이 진행된다. 학습이 점차 진행됨에 따라 가중치의 크기가 커지게 되고 추가된 항인 weight decay항과 경쟁하게 된다. 실험에서 γ 의 값을 적절하게 설정해주는 것이 쉽지 않았는데 γ 를 체계적으로 결정해주는 방법에 대한 연구가 있어야 하겠고, γ 의 값을 바꾸어 가면서 적응적으로 신경망을 학습시키는 방법도 연구되어야 할 것이다.

4. 계산 시간 비교

EBP/SP 알고리즘의 계산 시간은 기존의 EBP 알고리즘의 계산 시간과 근사적으로 같다고 할 수 있다. 즉 EBP/SP가 각 은닉 노드에 대하여 ΔE_i 를 계산하는 시간을 제외하고는 EBP의 계산 시간과 근사적으로 같다고 할 수 있다. 이때 한 노드에 대하여 ΔE_i 를 계산하는 시간이 약 1/2 epoch의 계산 시간과 근사적으로 같다고 하면 EBP/SP는 $n + \frac{rN_h}{2}$ (epoch 수)의 epoch가 걸린다는 개략적인 값을 얻을 수 있다. 여기서 n 과 r 은 각각 전체 epoch 수와 전체 교란 횟수를 의미한다. 그러나 이것은 r 이 n 에 비해 충분히 작은 경우이며, 이 경우에 EBP/SP는 기존의 EBP에 비하여 아주 작은 양의 계산을 추가로 하게 된다.

VI. 결론 및 향후 연구 방향

이 논문에서는 국부적 최적해 문제와 주어진 문제에 적절한 신경망 구조를 체계적으로 결정하는 새로운 학습 알고리즘을 제안한다. 제안한 EBP/SP 알고리즘은

가중치 공간상에서 확률적 교란을 수행하여 다층 퍼셉트론 학습에서 문제점인 국부적 최적해에서 벗어난다. 특히, EBP/SP 알고리즘은 이와 같은 확률적 교란 방법을 일관된 방법으로 새로운 노드의 추가도 가능하도록 확장한다. EBP/SP는 EBP가 국부적 최적해에 수렴했을 때 국부적 최적해로부터 벗어나기 위해 은닉 노드와 관련된 가중치를 재초기화한다. 새로운 은닉 노드 추가 또한 이와 같은 확률적 교란의 특수한 경우로 볼 수 있다. 이 논문에서는 확률적 교란이라는 일관적이고 효율적인 방법으로 국부적 최적해 문제 및 신경망 구조 설계 문제를 해결할 수 있음을 보였다. 그 결과로 제안한 EBP/SP 알고리즘은 신경망의 학습 시간을 크게 개선시킨다. 또한 EBP/SP 알고리즘은 고정된 크기의 신경망을 학습시키기 위해서는 새로운 노드의 추가 부분을 생략하면 된다. 즉 새로운 노드의 추가 없이 확률적 교란만을 통하여 국부적 최적해를 벗어날 수 있다. 이 경우 역시 제안한 EBP/SP는 기존의 EBP보다 훨씬 더 좋은 성능을 보였다.

신경망을 학습시켜 최소의 오차를 얻었을 경우에 과적응이 발생한 경우일 수 있다. 이것은 바람직하지 못한 경우로 본 논문에서는 오차 함수에 일반화 능력을 향상시키는 항을 추가하여 학습시킴으로써 EBP/SP를 사용한 다층 퍼셉트론의 일반화 능력을 향상시킬 수 있음을 보였다.

향후 연구로는 제안한 EBP/SP 알고리즘을 다른 신경망 모델들의 학습에 적용할 수 있도록 확장하는 것과 이론적으로 그 수렴성을 증명하는 것이다.

참 고 문 헌

- [1] Simon Haykin, *Neural Networks - A Comprehensive Foundation*, Macmillan College Publishing Co., Inc., pp.138-235, 1994.
- [2] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning Internal Representation by Error Propagation," *Parallel Distributed Processing*, vol. 1, pp.318-362, 1986.
- [3] Kurt Hornik, "Multilayer Feedforward Networks are Universal Approximators," *Neural Networks*, vol. 2, pp. 359-366, 1989.
- [4] D. E. Rumelhart, R. Durbin, R. Golden, & Y. Chauvin, "Backpropagation: The basic theory," in *Backpropagation - Theory, Architectures, and Applications*, Y. Chauvin & D. E. Rumelhart(Eds.), pp. 1-34, Lawrence Erlbaum Associates, Publishers, 1995.
- [5] S. C. Ng, S. H. Leung and A. Luk, "Evolution of Connection Weights Combined with Local Search for Multi-layered Neural Network," *IEEE*, pp. 726-731, 1996.
- [6] Yoshio Hirose, Koichi Yamashita and Shimpei Hijiya, "Back-Propagation Algorithm Which Varies the Number of Hidden Units," *Neural Networks*, vol. 4, pp. 61-66, 1991.
- [7] Rudy Setiono, "A Neural Network Construction Algorithm Which Maximizes the Likelihood Function," *Connection Science*, vol. 7, no. 2, pp. 147-166, 1995.
- [8] Timothy Masters, *Advanced Algorithms for Neural Networks - A C++ Sourcebook*, John Wiley & Sons, Inc., pp. 74-102, 1993.
- [9] Robert V. Hogg and Elliot A. Tanis, *Probability and Statistical Inference*, Macmillan Publishing Co., Inc., pp.73-75, 1977.
- [10] G. D. Kendall and T. J. Hall, "Ockham's nets: Self-adaptive Minimal Neural Networks," *Artificial Neural Networks 2*, pp. 183-186, 1992.
- [11] Y. Le Cun, J. S. Denker, and S. A. Solla, "Optimal brain damage," in *Advances in Neural Information Processing Systems*, vol. 2, pp. 598-605, 1990.
- [12] A. S. Weigend, D. E. Rumelhart, and B. A. Huberman, "Generalization by Weight-Elimination with Application to Forecasting," in *Advances in Neural Information Processing Systems*, vol. 3, pp. 875-882, 1991.
- [13] Yves Chauvin, "A Back-Propagation Algorithm with Optimal Use of Hidden Units," in *Advances in Neural Infor-*

mation Processing Systems, vol. 1, pp. 519-526, 1989.

[14] *Stuttgart Neural Network Simulator (SNNS), User Manual, Version 4.0*, Institute for Parallel and Distributed High Performance Systems (IPVR), University of Stuttgart, 1995.

[15] So-Young Jeong, Minho Lee, and Soo-Young Lee, "Hybrid Learning Algorithm with Low Input-to-Output Mapping Sensitivity for Iterated Time-Series Prediction," Proc. of International Conference on Neural Networks, vol. 2, pp. 1168-1171, 1997.

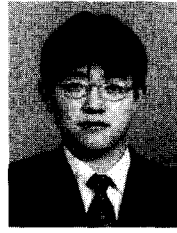
저 자 소 개



金 三 根(正會員)

1985년 부산대학교 계산통계학과 졸업(이학사). 1988년 숭실대학교 전자계산학과 석사학위 취득(공학석사). 1998년 숭실대학교 전자계산학과 박사학위 취득(공학박사). 현재 안성산업대학교 컴퓨터공학과 조교수. 관심

분야는 신경회로망, 최적화 이론, 진화알고리즘, 패턴인식 등



閔 昌 宇(正會員)

1995년 숭실대학교 컴퓨터학부 졸업(학사). 1998년 숭실대학교 컴퓨터학부 석사학위 취득(공학석사). 현재 한국 IBM 소프트웨어 연구소, 소프트웨어 엔지니어. 관심분야는 유연추론, 신경회로망, 진화알고리즘, 퍼지시스

템, 인공생명 등



金 明 源(正會員)

1972년 서울대학교 공과대학 응용수학과 졸업(학사). 1981년 University of Massachusetts (Amherst) Computer Science 석사학위 취득. 1986년 University of Texas(Austin) Computer Science

박사학위 취득. 1975년~1978년 한국과학기술연구소 연구원. 1982년~1985년 Institute for Computing Science & Computer Applications(Univ. of Texas 연구원. 1985년~1987년 AT&T Bell Labs.(Naperville) Member of Technical Staff. 1987년~1994년 한국전자통신연구소, 책임연구원. 1991년~1993년 충남대학교 전자계산학과 겸임 부교수. 1992년~1993년 한국신경회로망연구회 회장. 1994년~현재 숭실대학교 컴퓨터학부 부교수. 1993년~1995년 정보과학회 뉴로컴퓨팅연구회 위원장. 1993년~현재 IEEE Neural Networks Council 한국지부장. 1997년~현재 뇌정보처리 협의회 부회장. 1997년~현재 뇌연구촉진법 제정 준비위원. 1998년 한국 뇌학회 부회장. 관심분야는 유연추론, 신경회로망, 퍼지 시스템, 진화알고리즘, 패턴인식, 자동 추론, 기계학습, creativity engineering 등