

# 은닉노드의 목표 값을 가진 2개 층 신경망의 분리학습

## 알고리즘\*

최범기, 이주홍, 박태수  
인하대학교 컴퓨터정보공학과

neural@inha.ac.kr, juhong@inha.ac.kr, taesu@datamining.inha.ac.kr

### Separate Learning Algorithm of Two-Layered Networks with Target Values of Hidden Nodes

Bum-Ghi Choi, Ju-Hong Lee, Tae-Su Park  
Dept. of Computer Science & Information Engineering, Inha University

#### 요약

BP알고리즘은 지역 최소점이나 고원 문제와 같은 수렴 실패문제와 학습 속도가 느리다고 알려져 있다. 이제까지 알려진 BP알고리즘의 대체 방법들은 수렴 속도와 인자에 따른 수렴의 안정성에 대한 불균형을 해소 하는데 치중했다. 기존의 전통적인 BP알고리즘에서 발생하는 위와 같은 문제를 해결하기 위하여, 본 논문에서는 적은 용량의 저장 공간만을 요구하며 수렴이 빠르고 상대적으로 안정성이 보장되는 알고리즘을 제안한다. 이 방법은 상위연결(upper connections), 은닉층-출력층(hidden to output), 하위연결(lower connections), 입력층-은닉층(input to hidden)에 대해 개별적으로 훈련을 시키는 분리 학습방법을 적용한다.

#### 1. 서론

인공신경망(Artificial neural network)은 망의 구성 방식에 따라 여러 가지로 분류되는데 가장 활발히 연구되고 유용한 구성은 층으로 이루어진 전방향 위상 또는 LFF로 표기되는 구성 방식이 있다. 본 논문에서 LFF 뉴런 모델은 이진 임계치 단위로써 뉴런의 단순한 모델에 대해 제안된 McCulloch 와 Pitts[1]의 방법을 기반으로 하였다. 1960년대에 Frank와 Rosenblatt[2] 연구팀은 퍼셉트론이라 불리는 네트워크에 대해 집중적으로 연구하였다. 여기서 퍼셉트론이란 한 층과 그 다음 층이 전방향으로 연결된 층으로 구성된 네트워크를 의미한다. 하지만 Minsky와 Papert[3]는 그들의 저서 'Perceptrons'에서 Rosenblatt의 단층 퍼셉트론은 선형 분리의 단순한 예인 XOR 문제도 해결할 수 없다고 지적하였다. 20년 정도 후에 Rumelhart, Hinton, Williams [4], 그리고 Parker[5]에 의해 BP알고리즘이라 알려진 역전파가 개발되었다. 비록 BP알고리즘이 네트워크에서 학습을 위한 임의적인 계산 작업에 아직 이용할 수 없었지만, 단순한 단층 퍼셉트론으로는 해결할 수 없는 XOR문제와 같은 많은 문제를 해결할 수 있었다. 현재에는 BP알고리즘과 그와 관련된 응용분야에 많은 연구가 집중되고 있다. 본 논문에서는 다층 네트워크보다 단지 2개 층 네트워크로 관심분야의 제한을 둔다. 그 이유는 본 논문에서 서술하는 모든 주제의 경우 2개 층이 더욱 효율적이기 때문이다.

BP 알고리즘은 지역 최소점이나 고원 문제와 같은 수렴 문제를 내포하고 있다고 알려져 있다. 고원은 매우 느리게 수렴

되는 문제에 대한 원인이 되고, 지역 최소점의 경우에는 모든 방향의 기울기가 0과 같아진다. 만약 훈련 프로세스가 지역 최소점에 빠지면, 프로세스의 업데이트 가중치 벡터는 정지하게 된다. 여기서 우리는 다음과 같은 문제를 제기한다. 'BP 알고리즘과 다른 방법들이 비용 함수의 전역 최소점에 도달함으로써 수렴을 보장할 수 있는가? 그리고, 고원에서 이른 포화상태로부터 수렴의 속도를 높일 수 있는가?'이다. 이 두 타입의 문제는 잘못된 가중치 전달 및 잘못된 인자의 선택으로 인한 것일 수 있다. 네트워크에 대한 초기 가중치에 대한 BP의 민감도는 이미 밝혀졌고 카오스[6]와 관련하여 과민성의 종류에 대하여 수학적으로 그 증거가 판명되었다. 특히 초기 가중치는 네트워크에서 훈련에 대해 수렴하기 위하여 매우 중요하다. 일반적으로 초기 가중치의 임의적인 집합은 네트워크에서 훈련에 대한 수렴을 보장하지 못한다. 지금까지 많은 방법들이 제안되어 왔고, 그 방법들은 두개의 클래스로 범주화시켰다. 그 중 한 연구 분야가 학습률과 모멘텀에 따른 동적 변동에 관한 것이고[7,8,9] 다른 방법들은 활동에 대해 좀더 향상된 함수를 선택하거나 새로운 가중치 갱신 규칙으로 오차를 평가하는 것이다. 이제까지 알려진 역전파의 대체 방법들은 매개변수의 넓은 영역에서 항정을 피하기 위한 수렴의 안정성과 수렴 속도, 또는 전체적인 성능과 저장 공간요구 사이의 불균형이 존재하는 것으로 알려져 있다.

본 논문에서는 임의의 초기 가중치에 대해 더욱 유연하고, 매개변수의 넓은 범위에 대해서도 수렴을 보장하며, 지역 최소점과 고원 문제를 극복할 수 있는 새로운 알고리즘을 제안한다. 새로운 알고리즘은 은닉 노드의 목표 값을 이용하여 분리 학습하는 알고리즘이다. 이 알고리즘은 출력 노드의 현재 오차

\*본 연구는 정보통신부 및 정보통신연구진흥원의 대학 IT연구센터 육성·지원사업의 연구결과로 수행되었음.



오차가 가장 큰 출력 하나만이 은닉 노드의 오차를 계산하는데 영향을 끼친다. 이 방법에서 가장 큰 오차를 가지는 오직 하나의 출력 노드만이 은닉 노드의 오차를 전파하는 과정에 영향을 끼친다.

이 알고리즘을 완성하기 위해서는 하위연결 층에서 훈련을 수행하기 위해 어느 정도의 은닉 노드를 선택해야 할지 결정해야 된다.

가장 많이 사용하는 방법은 하위연결에서 훈련을 하는 매 시간마다 하나의 은닉 노드로 하나만 선택하는 것이다. 훈련을 위해 노드들 중에 오직 하나만 선택하는 것은 불이익인 것처럼 보이지만, 계산 작업에 불이익을 주는 것이 아니라 수렴속도를 향상시키는데 도움을 준다. 우리는 다음의 4-2-4 인코더 디코더 문제의 짧은 실험을 통하여 위에서 언급한 것을 증명하였다. 알고리즘은 다음과 같이 요약된다.

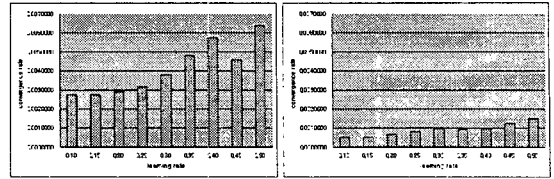
1. 상위연결을 프로세스가 수렴하거나 훈련이 느려지는 지점을 만날 때까지 훈련시킨다.
2. 만약 프로세스가 수렴하면 프로그램을 중지시키고, 그렇지 않다면 이미 선택되었던 것과 다른 은닉 노드와 가장 큰 오차를 가지는 출력 노드를 골라내고, 은닉 노드의 오차를 전파한다. 그리고 목표 값  $\gamma_j$ 와 은닉 노드의 오차  $k$ 를 생성한다.
3. 훈련이 느려지거나 수렴할 때 까지 은닉 노드의 이상치를 생성하면서, 선택된 은닉 노드에 대한 입력 노드들로부터 하위연결을 훈련시킨다.
4. 1로 되돌아간다.

분리 학습 방법은 가중치를 갱신하는 규칙과 변하지 않는 네트워크 구조, 목표 값과 은닉 노드의 오차와 학습과정에 대하여 다른 상태를 줌으로써 수렴문제로부터 탈출하기 위한 새로운 아이디어를 제공한다. 계산적인 이점은 분리 학습의 반복 당 계산적인 시간이 BP 알고리즘보다 적게 걸린다는 사실을 통하여 쉽게 알 수 있다. 다른 장점은 위에서 언급되었던 알고리즘들과의 결합으로 인해 좀더 향상된 성능을 보인다는 것이다.

### 3. 실험 결과

4-2-4 인코더 디코더 문제를 사용한 실험으로 분리 학습 알고리즘과 BP 알고리즘을 비교하였다. 여기서 우리는 실험한 시도로서 2000번 이상의 반복을 수행하였다. 탄젠트 쌍곡선의 함수를 활성화 함수로 사용하였다. 학습율과 이득에 대한 매개변수의 범위를 설정해야 한다. 학습율 매개변수는 0.05의 간격으로, 범위는 0.1에서 0.85이다.

실험은 각 매개변수의 집합에 대하여 -1에서 1사이의 임의의 초기 가중치를 이용하여 5번 반복 수행한다. 그리고 결과를 평균하여, 각 매개변수의 집합의 성능을 대조한다. 수렴율은 주어진 초기 가중치로부터 수렴되는 시간의 평균 값의 역으로 정의 된다. (그림 2)는 분리 학습 알고리즘과 BP 알고리즘의 수렴율에 대한 비교를 보여준다. 전체적으로 넓은 범위의 매개변수에서 수렴과 수렴율은 분리학습이 더 좋은 성능을 보인다.



(그림 2) 분리 학습과 BP알고리즘의 수렴율비교

### 4. 결론

지금까지 LFF에서의 분리 학습 알고리즘에 대하여 기술하였고, 4-2-4 인코더 디코더 문제의 실험을 통하여 전통적인 BP알고리즘과 성능을 비교하였다. 논리적으로 정당하고 좋은 실험결과에도 불구하고, 더 증진시킬 수 있는 두 가지 측면이 존재한다. 알고리즘 측면에서 더 관심을 기울여야 할 점은 상위나 하위의 한 연결의 집합에서 다른 연결의 집합으로 언제 전환하는가를 결정하는 것이다. 이 전환점은 현재의 비용 값과 이전 비용 값의 차이가 이전 비용 값의 60분의 1보다 더 작아질 때 일어날 수 있다. 그러나 전환 기준에 관한 보다 심도 깊은 연구를 통하여 성능을 더 개선시킬 수 있다. 또한 기준에 BP 알고리즘의 대안으로 소개되었던 알고리즘과 병합된 방법들도 고려될 수 있다.

### 참고문헌

- [1] McCulloch, W.S., Pitts, W., "A logical Calculus of Ideas Immanent in Nervous Activity" Bulletin of Mathematical Biophysics 5, 115-133, 1962.
- [2] Rosenblatt, F., "Principle of Neurodynamics", New York: Spartan, 1962.
- [3] Minsky, M.L and Papert, S.A., "Perceptrons", Cambridge: MIT Press, 1969.
- [4] Rumelhart, D.E., G.E. Hinton, and Williams, R.J., "Learning Internal Representations by Error propagation". In Parallel Distributed Processing, vol. 1, chap8, 1986.
- [5] Parker, D.B., "Learning Logic", Technical Report TR-47, Center for Computational Research in Economics and Management Science, Massachusetts Institute of Technology, Cambridge, MA, 1985.
- [6] Kolen, J. F. and Pollack, J. B., "Back Propagation is Sensitive to Initial Conditions", Complex System 4, 269-280, 1990.
- [7] Jacobs, R. A., "Increased Rates of Convergence Through Learning Rate Adaptation", Neural Networks 1, 293-280, 1988.
- [8] Vogl, T. P., J.K. Magis, A.K. Rigler, W.T. Zink, and D.L. Alkon., "Accelerating the Convergence of the Back-Propagation Method", Biological Cybernetics 59, 257-263, 1988.
- [9] Allred, L. G., Kelly, G. E., "Supervised learning techniques for backpropagation networks", In Proc. of IJCNN, vol. 1, 702-709, 1990.