

유전자 알고리즘을 이용한 오차 역전파 신경망의 초기화

박형태*, 이형세

아주대학교 전자공학부

전화 : 031-219-2376 / 핸드폰 : 011-210-8154

An Initialization of Backpropagation Network Using Genetic Algorithm

Hyung Tae Park*, Hyeng Sea Lee

Dept. of Electronic Engineering, Ajou University

E-mail : edasom@ajou.ac.kr

Abstract

본 논문에서는 오차 역전파 알고리즘의 전역 최소값을 찾지 못하는 문제점에 대해서 설명하였고, 이 문제를 해결하기 위한 방법으로 유전자 알고리즘에 대해서 설명하였다.

오차 역전파 알고리즘은 기본적으로 경도 하강법을 따른다. 따라서 신경망의 각 가중값 행렬이 만드는 오차의 오차 평면이 대부분의 문제에서 다수의 국부 최소값들을 가지는게 일반적이는데, 가중값의 변화가 한 방향으로 진행하기 시작하여, 오차가 증가되어지는 언덕이 학습 계수보다 크다면 더 이상 학습은 진행되지 않고 거기에서 빠져 나가지 못한다. 따라서 초기의 위치가 중요한 역할을 하는데, 이 문제를 해결하기 위해서 유전자 알고리즘을 이용한 신경망 초기화 방법을 제안하였다.

끝으로, 간단한 실험으로 제안된 방법을 구현하고 결과에 대해서 논하였다.

분야에서 비선형 필터로서 많이 사용되고 있다. 이 중 비교적 간단하면서도 지도 학습(Supervised)에 효과가 뛰어난 오차 역전파 알고리즘(Error Back Propagation Algorithms)을 사용하는 다층구조 퍼셉트론이 자주 사용된다. 그러나 근본적으로 경도 하강법(Gradient Descent Algorithms)을 사용하는 오차 역전파 알고리즘은 해결하고자 하는 문제에 따라서 오차 표면(Error Surface)에서 전역 최소값(Global Minimum)을 찾지 못하고, 국부 최소값(Local Minimum)에 이르러 학습이 더 이상 진행되지 않는 경우가 종종 발생하는데, 이 경우 (Local Minima Problem)에 망의 초기화가 중요한 영향을 미치게 된다.

유전자 알고리즘(Genetic Algorithms)은 자연계에서 일어나는 표현형(Phenotype)의 자연선택(Natural Selection)에 의한 유전자(Genotype)의 진화(Biological Evolution)를 흉내 내어 정보처리에 이용하는, 역시 역사가 길지 않은 새로운 방법론이다. 여러 개체들 중에서 교배(Crossover)와 돌연변이(Mutation)을 통한 재생산(Reproduction), 그리고 선택(Selection)이 여러 세대

I. 서론

최근 인공신경망(Artificial Neural Network)이 그 학습 능력과 패턴 분류 능력으로 인해 각종 정보 통신

에 걸쳐 누적되면 집단(Population)중에 목표값(Goal)에 가까운 개체가 생겨나게 되어 근사 최적값을 찾을 수 있다. 따라서 이는 전략적 검색(Search with strategy)의 한 방법이라고 볼 수 있다.

본 논문에서는 오차 역전파 알고리즘과 유전자 알고리즘을 조합하여 신경망을 학습시키는 방법을 제안하는바, 유전자 알고리즘을 이용하여 전역 최소 값을 탐색하고, 탐색 되어진 값으로 신경망을 초기화시키고 나서 오차 역전파 알고리즘으로 신경망을 학습시킴으로써 신경망이 국부 최소값 근처에서 더 이상 학습되지 않는 문제를 해결 할 수 있음을 간단한 실험을 통해 보인다.

2. 유전자 알고리즘을 이용한 오차 역전파 신경망의 초기화

오차 역전파 알고리즘은 기본적으로 경도 하강법(Gradient Descent Algorithm)을 따른다. 따라서 신경망의 각 가중값 행렬(Matrices of Weights)이 만드는 고차의(High Dimensional)오차 표면(Error Surface)이 대부분의 문제에서 다수의 국부 최소값들(Local minima)을 가지는게 일반적인데, 일단 가중값들의 변화가 한 방향으로 진행하기 시작하여 다음 번 가중값들의 변화가 오히려 오차를 증가시키며 그 언덕이 학습 계수(Learning Rate)보다 크다면 더 이상 학습은 진행되지 않고 거기에서 빠져 나가지 못한다. 따라서 초기의 위치가 중요한 역할을 하는데, 임의로 초기값을 선택하는 것보다는 가중값들을 전역 최소값(Global Minimum) 근처에 위치시켜 놓는 것이 필요하다. 이 때 유전자 알고리즘 특성상 전역 최소값을 가질 수 있는 가능성이 높기 때문에, 유전자 알고리즘을 사용하여 가중값들을 초기화시켜 놓은 다음에 오차 역전파 알고리즘을 사용하면 국부 최소값에 빠져 학습이 되지 않을 확률이 낮아지게 된다. 유전자 알고리즘은 그 특성상 세대간의 표현형의 격차가 크므로 다소 큰 걸음의 검색을 통해 전역 최소값 근방을 찾을 수 있고, 그 다음엔 작은 계수를 가진 오차 역 전파 알고리즘을 사용하여 작은 걸음의 검색을 한다면 전체적으로 큰 걸음에서 작은 걸음으로의(Coarse-to-fine)검색을 행할 수 있게 된다. 신경망 각각의 가중값들을 모두 이어서 하나의 유전자(Genotype)를 만들고 이런 유전자들의 집단을 랜덤

하게 구성한 다음에, 이들을 대상으로 하여 유전자 알고리즘을 적용한다. 표현형(Phenotype)에 대한 평가(Evaluation)는 선택된 개체, 즉 신경망의 가중값들이 만들어내는 출력층 뉴런들의 출력이 목표값에 가까운 정도를 가지고 행한다. 초기 유전자 집단들중에서 신경망의 출력이 특히 목표값에 가까운 것들끼리 교배시켜서 새로 탄생한 유전자가 상대적으로 열등한 유전자들은 대체시켜 나가는 과정을 반복하며, 여러 세대에 걸쳐 별로 성능이 개선되지 않으면 유전자의 일부분에 돌연변이를 일으켜 새로운 세대는 그 부모 세대와는 약간 다른 개체가 되도록 유도한다.

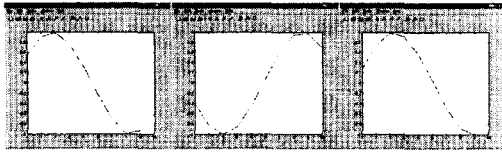
이런 과정이 반복되다가 오차가 특정 범위 내에 들면 신경망은 이미 어느 정도 학습이 된 상태와 다름없게 된다. 그러면 이렇게 초기화된 가중값을 가지고 다시 오차 역전파 알고리즘을 사용하여 좀더 세부적인(Fine) 튜닝(Tuning)을 행한다.

3. 실험 및 결과

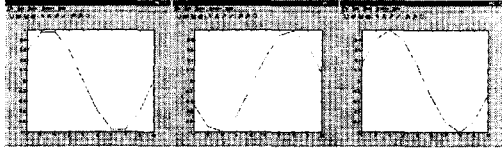
본 실험은 MatLab5.3R11을 사용하여 행하였다. 신경망은 MatLab Neural Network Toolbox를 사용하여 구현하였고, 유전자 알고리즘은 북캐롤라이나 주립 대학(North Carolina State University)의 산업공학에서 개발한 MatLab용의GAOT(Genetic Algorithms Optimization Toolbox)을 사용하여 구현하였다.

이 실험에 사용한 데이터는 각각 10개씩의 샘플을 갖는 실수 벡터로서, 주파수가 0.09Hz이고 위상이 0인 사인파(Sinusoidal wave)[그림1], 주파수가 0.09Hz이고 위상이 108인 사인파[그림2], 주파수가 0.09Hz이고 위상이 216인 사인파[그림3], 주파수가 0.10Hz이고 위상이 0인 사인파[그림4], 주파수가 0.10Hz이고 위상이 108인 사인파[그림5], 주파수가 0.10Hz이고 위상이 216인 사인파[그림6], 주파수가 0.11Hz이고 위상이 0인 사인파[그림7], 주파수가 0.11Hz이고 위상이 108인 사인파[그림8], 주파수가 0.11Hz이고 위상이 216인 사인파[그림9], 이렇게 아홉개의 데이터를 주파수별로 세 그룹으로 분류할 수 있다.

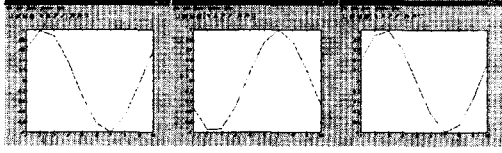
신경망의 구조는 입력층에 10개의 뉴런이 있고, 중간층(Hidden Layer)에 5개의 뉴런이 있으며, 출력층에 3개의 뉴런이 있다. 각 층마다 바이어스(Bias) 뉴런이 하나씩 주어졌다. 중간층 뉴런과 출력층 뉴런의 전달



[그림1]0.09Hz,0 [그림2]0.09Hz,3 [그림3]0.09Hz,6



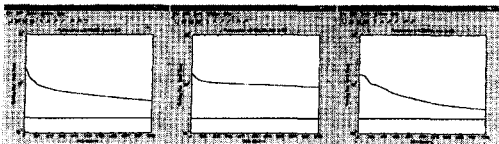
[그림4]0.10Hz,0 [그림5]0.10Hz,3 [그림6]0.10Hz,6



[그림7]0.11Hz,0 [그림8]0.11Hz,3 [그림9]0.11Hz,6

함수는 탄젠트 시그모이드(Tangent-Sigmoid)함수를 사용했으며, 학습시에 첫번째 유형(0.09Hz)의 데이터가 인가됐을 때엔 첫 번째 출력 뉴런은 1, 나머지 출력 뉴런은 -1을 가지도록 목표값을 주었고, 두 번째 유형(0.10Hz)의 데이터가 인가됐을 때엔 두 번째 출력 뉴런이 1, 나머지 출력 뉴런은 -1을 가지도록 목표값을 주었으며, 마지막 유형(0.11Hz)의 데이터가 인가됐을 때엔 마지막 출력 뉴런이 1, 나머지 출력 뉴런은 -1을 가지도록 목표값을 주었다. 학습 완료 평가값은 실제 출력과 목표값의 차의 제곱 평균값(Mean Square Error)이 0.2 이내에 들도록 했고, 최대 반복 횟수는 2000회로 제한하였다.

[그림10]부터 [그림15]까지는 오차 역전과 알고리즘만을 사용하여 실험한 결과다. [그림10]부터 [그림12]까지는 입력층과 중간층 사이의 가중값들에 대해서는 Nguyen-Widrow 초기화를 하고 중간층과 출력층 사이의 가중값들에 대해서는 랜덤 초기화를 한 후에 단순한 경도 하강법만을 사용하여 학습시킨 결과다. 학습 계수(Learning Rate)는 0.01이다. 세번의 실험에서 최대 반복 횟수인 2000회 이내에 학습이 완료된 것은 한 번도 없었다. 학습 시간은 Pentium III-750MHz, 64MB PC로 24초쯤 소요되었다.



[그림11, 12, 13] 학습 안됨

[그림13]부터 [그림15]까지는 위와 마찬가지로 입력층과 중간층 사이의 가중값들에 대해서는 Nguyen-Widrow 초기화를 하고 중간층과 출력층 사이의 가중값들에 대해서는 랜덤 초기화를 한 후에 적응 학습 계수(Adaptive Learning Rate)와 관성항(Momentum)을 적용하여 오차 역전과 알고리즘으로 신경망을 학습시킨 결과다. 세 경우 모두 100회와 150회 사이의 반복을 통해 학습이 완료되었다. 학습 시간은 3초 가량 소요되었다.

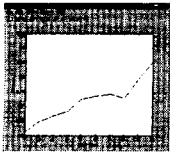


[그림14] 181 [그림15] 123 [그림16] 178

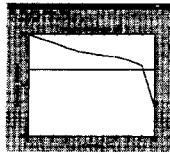
이제 그러면 유전자 알고리즘을 사용하여 실험을 할 차례다. 바이어스(Bias) 항을 제외하면 10개의 입력층 뉴런들과 5개의 중간층 뉴런들 사이의 가중값(IW)은 모두 50개이고, 5개의 중간층 뉴런과 3개의 출력층 사이의 가중값(LW)은 모두 15개이므로 신경망 전체는 모두 65개의 가중값들을 갖는다. 입력과 출력이 모두 -1에서 +1사이의 값을 가지므로 이 가중값들 또한 모두 그와 같은 범위의 실수값을 가질 수 있다. 행렬 IW와 LW를 이루는 각 원소(Element)들, 즉 각각의 가중값들을 순서대로 나열하여 이를 하나의 유전자로 본다. 이 때 유전자는 이진수 형태이어야 연산하기 편한데, 가중값들은 실수(Float)이므로 -1과 1사이를 65536등분하여 가중값들이 0부터 65535까지의 정수에 대응하는 -1과 1사이의 불연속한 실수만을 가질 수 있도록 제한한다. 전체의 유전자 집단은 20개의 개체를 가지고 시작된다.

평가 함수(Evaluation Function)는 목표값과 신경망의 출력의 차의 제곱 평균(Mean Square Error)이 0.5 이내가 될 때까지 재생산을 계속한다. 교배(Crossover)는 유전자 벡터의 어느 위치에서나 발생할 수 있고, 돌연변이(Mutation)는 3세대간의 차이가 0.05미만일 때 유전자 벡터내의 임의의 한 비트에 대해 발생된다.

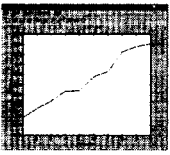
[그림16]은 유전자 알고리즘으로 신경망을 초기화한 결과인데, 10세대만에 제곱평균오차가 0.5 이내에 들었다. [그림17]은 이 초기화 결과를 바탕으로 오차 역전과 알고리즘을 사용하여 신경망을 학습시킨 결과다. [그림18]와 [그림19], 그리고 [그림20]과 [그림21] 역시 마찬가지 결과를 보여준다.



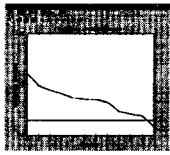
[그림17] GA평가(10세대)



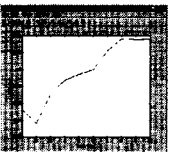
[그림18] 10 epochs



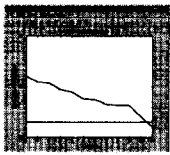
[그림19] GA 평가(10세대)



[그림20] 11 epochs



[그림21] GA평가(10세대)



[그림22] 11 epochs

세 경우 모두 유전자 알고리즘을 통해 10세대까지 망의 가중값들을 재생산하여 그중 가장 평가가 좋게 나온 조합을 신경망의 초기값으로 사용하여 적응학습 계수와 관성항 없는 오차 역전파 알고리즘을 써서 신경망을 학습시킨 결과는 [그림10]에서 [그림11]사이의 경우와 달리 세 경우 모두 각각 10회 정도의 반복 학습을 통해 수렴되었다. 이는 망의 초기화 과정에서 유전자 알고리즘을 사용하여 이미 최소 오차 영역 근처에 망의 가중값들이 도달해 있기 때문이다. 학습 시간은 같은 컴퓨터를 써서 4분 가량 소요되었다

4. 결론

오차 역전파 알고리즘을 사용한 신경망의 학습은 이미 많이 연구되었고 또 많은 분야에서 응용되고 있다. 본 실험에서 비교해 본 바와 같이 오차 역전파 알고리즘에도 단지 경도 하강법만을 단순히 사용하는 방법에서 경도 하강법에 적응 학습계수와 관성항을 함께 사용하는 방법이나 Levenberg-Marquardt 알고리즘을 사용하는 방법 등 학습 수렴시간을 줄이고 국부 최소값에서 벗어나게 하기 위한 많은 방법이 연구되어 왔다. 본 연구는 순수한 경도 하강 오차 역전파 알고리즘만을 사용하는 경우에 생기는 국부 최소 문제를 해결하기 위한 방법으로서, 유전자 알고리즘을 사용하여

신경망의 가중값들을 초기화시켜놓은 다음에 경도 하강 오차 역전파 알고리즘을 사용하여 신경망을 학습시키는 방법을 제안하였다.

순수한 경도 하강 오차 역전파 알고리즘만을 사용하여 학습되지 않는 문제를 적응 학습계수와 관성항을 사용한 오차 역전파 알고리즘만을 사용하여 학습되지 않는 문제를 적응 학습계수와 관성항을 사용한 오차역전파 알고리즘의 경우와 유전자 알고리즘을 사용한 경우의 학습 성능을 비교해 보았다. 결과로 적응 학습계수와 관성항을 사용한 오차 역전파 알고리즘의 경우 120회에서 180회 사이의 반복 학습을 통해 학습이 되고 학습에 소요된 시간은 3초 가량이다. 유전자 알고리즘으로 초기화한 다음에 오차역전파 알고리즘으로 학습하는 경우에는 학습에 소요된 시간이 4분 가량이었다. 그리고 두 경우 모두 국부 최소문제를 피할 수 있었다.

국부 최소 문제를 피하고자 제안한 알고리즘이 국부 최소문제를 피하는 데에는 성공했지만 적응 학습계수와 관성항을 사용한 오차 역전파 알고리즘에 비해 그 학습 속도가 비교가 되지 않을 정도로 많이 소요되기 때문에 별 실용성은 없어 보인다. 따라서 좀더 복잡한 문제에 대해서 실험해 보고 두 방법이 일반적으로 어떤 특징을 갖는지 좀더 연구해 볼 필요가 있을 것으로 보인다.

참고문헌(또는 Reference)

- [1] Wasserman, Philip D. Neural Computing : Theory and Practice, 1989, Van Nostrand Reinhold
- [2] Goldberg, David E., Genetic Algorithms in Search, Optimization & Machine Learning, 1989, Addison-Wesley
- [3] K S Tang, C Y Chan, K F Man and S Kwong "Genetic Structure for NN Topology and Weights Optimization"