Feedforward Neural Network의 개선된 학습 알고리즘

윤여창
우석대학교 e-정보공학과
yoonyc@core.woosuk.ac.kr

A Modified Learning Algorithm for Feedforward Neural Network

YeoChang Yoon
Dept. of e-Information Engineering, Woosuk University

요 약

본 연구에서는 Feedforward Neural Network에 적용될 수 있는 개선된 학습 알고리즘을 개발하고자 한다. 제시된 알고리즘을 이용하여 학습을 할 때 학습 초기는 가장 단순한 경우로써 한 개의 학습 패턴과 은닉층으로부터 시작한다. 신경망 학습중에 지역 최소값에 수렴되면 weights scaling 기법을 이용하여 지역 최소값을 벗어나도록 한다. 지역 최소값의 탈출이 용이하지 않으면 은닉노드를 점차적으로 추가한다. 이러한 단계에서 새롭게 추가된 노드에 대한 초기값 선택은 선형계획법을 이용한 최적 처리절차를 이용한다. 최적 처리절차의 결과로써 은닉층의 노드가 추가된 후의 네트워크는 학습회수를 증가시키지 않아도 학습 허용오차를 만족시킬 수 있다. 본 연구에서 적용한 개선된 알고리즘을 이용하면 신경망 학습시의 수렴 정도를 높여주고 최소한의 단순 구조를 갖는 신경망으로 추정할 수 있게 하며, 이 결과를 모의실험을 통하여 살펴보고 기존의 연구 결과와 비교한다.

1. 서론

신경망학습과 관련된 알고리즘들이 현재 많이 연구되고 있고 이들 연구들이 성공적으로 다양한 응용분야에 적용되고 있지만 몇 가지 쟁점 부문에서 완전하게 문제를 해결하지는 못하고 있다. 이들 쟁점들에는 은닉노드의 개수 결정문제들과 학습의 수렴속도 등으로 요약될 수 있다. 본 연구에서는 학습과정이 문제의 해로 잘 수렴되도록 하고 해에 수렴되는 경우의 학습량에 대하여 논의한다. 기존의 많은 알고리즘을 이용할 때 신경망 학습은 지역 최소값으로 인하여 목적한 학습에 도달하지 못하고 종료되는 경우가 많다. 전역 최소값으로 수렴되는 경우라도 오차가 제로이거나 제로에 가까운 값이 아닐 수 있기 때문에 이 값이 문제의 해가 되지 않을 수 있다.

신경망학습 알고리즘에 관한 대부분의 연구 결과들은 수렴 속도를 어떻게 증진시키는가에 초점을 두고 있다[1]. 이들 연구의 가장 중요한 특성중의 하나는 전역 최소값을 찾지 못하고 지역 최소값으로 수렴되는데 있다. 즉 이들 알고리즘의 학습과정은 수렴 속도를 증가시키고는 있지만 지역 최소값에 수렴되므로써 나타나는 속도 증가가 대부분이다.

일반적으로 학습이 수렴된다는 것은 궁극적으로 알고리즘이 문제의 해를 찾을 수 있다는 것을 의미한다. 다시 말하여 학습 알고리즘은 신경망학습의 전 과정을 통하여 만날 수도 있는 지역 최소값을 벗어날 수 있게 해야 한다는 것을 의미한다. 이와 같은 문제들의 해결을 위하여 관련 연구들은 weight scaling과 mean field annealing을 이용하여 지역 최소값 문제를 해결하고 있다[2,3].

다층 FNN(Feedforward Neural Network)을 이용한 문제 해결에 가장 중요한 고려사항 중 하나는 은닉층 노드의 개수 문제이다. 은닉노드의 개수가 불충분하면 특별한 문제를 해결하는데 있어서 네트워크가 능력을 다하지 못하게 하며 반대로 너무 많은 은닉노드는 네트워크를 일반화시키지 못하게 할 수 있다. 여기서 적절한 은닉노드 개수는 입력값 공간의 차원에 따라 다를 수 있다[4,5]. 은닉노드의 개수가 불충분하면 정확한 모형에 필요한 모수 개수가 충분하지 않아 과소평가된 의사결정을 하게 된다[4] 그러므로 이러한 문제를 해결하기 위한 방법으로써 학습에 필요한 은닉노드의 개수를 충분히 크게 하거나 또는 prune방법을 이용할 수 있다[6,7]. 그렇지만 이러한 학습은 지역 최소값을 벗어나게는 해 주지만 학습에 필요한 모수 개수보다 더 복잡한 네트워크를 학습시키기 때문에 수렴속도가 늦어지고 또한 과다적합으로 인하여 잘못된 결론을 초래하게 된다. 따라서 일부 알고리즘에서는 은닉층이 하나인 단순한 모형에서부터 신경망을 학습하기 시작하여 점차적으로 복잡한 네트워크로 확장시키면서 학습하기도 한다[6,7,8,9,10]. 즉 이 경우에 만약 신경망이 수렴되지 않으면 은닉층의 노드를 하나씩 추가하고 네트워크는 더 높은 차원의 모수공간에서 학습하게 된다. 따라서 학습속도 뿐만 아니라 수렴 가능성도 높아지게 된다 그러나 여기서 발생될 수 있는 또 다른 문제점은 확장된 네트워크가 가장 최적의 초기 가중값을 어떻게 다시 결정하는가에 있으며 또한 가장 적은 학습량을 이용하여

요구한 허용오차로 어떻게 수렴될 수 있게 하느냐가 중요한 문제가 될 수 있다[6,8,9].

본 연구에서는 은닉노드의 개수와 학습패턴을 각각 하나씩 증가시키면서 학습주기마다 확인되는 허용오차의 변화를 이용하여 최적의 모형을 선택할 수 있도록 하는 개선된 알고리즘을 이용하여 기존의 FNN의 학습결과와 비교해 보고자 한다. 이때 학습 패턴들은 허용오차가 만족되지 않을 때 하나씩 차례로 증가시키며 하나의 패턴은 모든 패턴이 학습될 때까지 먼저 학습된 패턴들과 결합하여 학습되도록 선택한다. 학습 초기에는 하나의 패턴으로 시작하고 여기서의 허용오차를 만족할 수 있도록 가중값을 선택한다. 학습주기마다 결정되는 모형은 학습자료로부터 추가적인 패턴으로 결정된다. 일반적으로 새롭게 추가된 패턴을 이용하면 허용오차보다 더 큰 오차를 발생시킬 수 있다. 그러므로 만약 오차가 주어진 허용오차를 만족하면서 감소하고 있다면 또 다른 패턴이 같은 방법으로 결합되면서 학습할 수 있다. 만약 그렇지 않다면 네트워크의 모수를 증가 시킨다. 이러한 방법으로 증가된 은닉층 노드와 이와 연결된 새로운 초기 가중값들을 결정할 수 있는 방법을 개발할 필요가 있다. 여기서 시스템 오차를 제로에 가깝게 감소시킬 수 있는 가중값들의 선택은 선형계획과 같은 최적화 기법들을 이용한다. 이와 같은 방법을 이용한 개선된 학습 알고리즘을 수식화하여 그 학습 결과를 일반적인 학습 알고리즘[11]의 결과와 비교해 본다.

2. 개선된 학습 알고리즘

네트워크는 입력값 N_i, 은닉층 노드 N_h, 출력층 노드 N_o 개로 구성된 2층 FNN를 고려하자. 먼저 w_{jk} 는 k번째 입력노드와 j번째 은닉층 노드간의 연결 가중값이고, v_{ij} 는 i번째 출력층 노드와 j번째 은닉층 노드간의 연결 가중값이라 하자. 그리고 은닉층과 출력층의 절편항은 $W=[w_{jk}]\in R^{N_h\times(N_i+1)}$ 과 $V=[v_{ij}]\in R^{N_o\times(N_h+1)}$ 이다. 학습자료인 입력패턴은 $x=[x_0,x_1,\cdots,x_{N_i}]^T$ 이다. 여기서 $x_0=1$ 은 가중값 w_{j0} 이고 $j=1,2,\cdots,N_h$ 이다. 출력패턴은 $d=[d_0,d_1,\cdots,d_{N_o}]^T$ 이다. 따라서 네트워크의 입출력 관계는 다음과 같다.

$$y_i = f\left(\sum_{k=0}^{N_i} w_{jk}x_k\right),\ j=1,2,\cdots,N_h,$$

$$z_i = f\left(\sum_{j=0}^{N_i} v_{ij}y_j\right),\ i=1,2,\cdots,N_o.$$

여기서 y_j 는 j번째 은닉노드의 출력값이고 z_i 는 i번째 출력노드의 출력값이다. 그리고 $f(\cdot)$ 는 시그모이드형 변환함수로서 $f(u)=1/(1+\exp(-u))$ 다. 또한 $f^{-1}(\cdot)$ 은 존재하고, 모든 $u\in R$ 에 대하여 $f(0)=0$, $f(u)=-f(-u)$ 이다. $i=1,2,\cdots,N_o$ 에서 $y_0=1$ 과 v_{i0} 는 출력층의 절편항이다. 그리고 주어진 패턴에 대한 오차함수 E_p 는 다음과 같다.

$$E_p=\frac{1}{2}\sum_{i=1}^{N_o}(z_i-d_i)^2.$$

신경망학습에서 N_p 패턴은 $p=1,2,\cdots,N_p$ 일 때 $x^p=[x_0^p,x_1^p,\cdots,x_{N_i}^p]^T$ 와 $d^p=[d_0^p,d_1^p,\cdots,d_{N_o}^p]^T$ 이다. 또한 $x_0^p=1$ 이다. 따라서 신경망은 전체 오차 $E(N_p)$ 가 아주 작은 허용오차 $\varepsilon>0$ 를 만족시키는 네트워크를 찾는 것이다. 여기서 $E(N_p)$는 다음과 같다.

$$E(N_p)=\frac{1}{N_p}\sum_{p=1}^{N_p}E_p.$$

$L-1<N_p$ 패턴에 대하여 학습된 신경망이 허용오차 ε 를 만족하고 있다고 가정하자. 그리고 학습할 문제로 새로 결합될 패턴번호 L 이라는 패턴이 있다고 하자. 즉 오차 $E(L)<\varepsilon$ 를 만족하는 신경망을 찾는 것이 새로운 학습목표가 된다. 시스템에서 발생시킨 가중값으로 학습을 시작하면서 은닉노드가 N_h 개인 신경망을 두가지 방법으로 연속적으로 학습한다. 첫번째는 새로운 문제에서 L 패턴에 대한 특정 오차를 구할 수 있도록 학습하는 경우이고, 두번째는 지역 최소값에 빠지는 경우로서 큰 오차로 인한 시스템의 학습 종료가 발생되는 경우이다. 첫번째 경우에서는, 같은 과정을 순차적으로 새로운 패턴에 반복 적용함으로써 신경망이 목적한 바를 구할 수 있도록 학습할 수 있다. 두번째 경우에서는, 단지 은닉노드의 개수가 N_h 인 신경망 구조로는 오차를 ε 보다 작게 줄일 수 없다는 것을 의미한다. 다시말하여 신경망은 오차를 더욱 감소시키기 위하여 최소한 은닉노드의 개수가 N_h+1 로 확장할 필요가 있다. 따라서 이와 같은 경우의 학습문제는 다시 N_h+1 개의 은닉노드와 L 개의 패턴을 갖는 새로운 신경망이 된다.

$L-1$ 패턴에 대한 원래의 신경망 구조가 최소 개수의 은닉노드를 갖는다면 L 패턴에 대한 신경망 학습 결과도 거의 유사하게 최소 개수의 노드를 갖는다. 어떤 하나의 패턴에 대하여, 한 개의 은닉노드 즉 $N_h=1$ 인 신경망에 대하여 오차가 제로가 되는 가중값 $W\in R^{1\times(N_i+1)}$ 와 $V\in R^{N_o\times 2}$ 는 항상 존재함을 증명할 수 있다[12]. 그러면 시스템의 허용오차를 만족하기 위하여 두번째 패턴을 첫번째 패턴과 함께 학습할 수 있다. 학습과 모형 확장이라는 두가지 과정은 모든 패턴이 함께 학습될 때까지 반복한다. 이와 같이 최소 개수의 은닉노드를 갖으며, 해에 수렴될 수 있는 신경망을 구할 수 있도록 학습과정을 개선한 학습 알고리즘은 다음과 같다.

알고리즘 2.1: 개선된 학습:

입력 : 출력 패턴은 각각 입력 패턴 N_p 와 연결된다.

출력 : 주어진 허용오차 ε 을 만족하고 가능한 최소 개수의 은닉노드로 학습된 네트워크.

단계 1: 먼저 $l=1$ 로 설정한다. 학습 자료로 부터 한 개의 패턴을 선택한다. 허용 오차 ε 를 만족하기 위해 선택된 패턴을 이용하여 은닉노드 하나로 신경망을 학습한다.

단계 2: 만약 $l < N_p$가 만족되면 특정 범주에 맞는 다음 번 패턴을 선택한다. 그리고 $l = l + 1$로 설정하고 단계 3으로 간다. 그렇지 않으면 끝.

단계 3: 만약 학습 알고리즘이 ε 을 만족하는 범위 안에서 $E(l)$을 감소시킬 수 있으면 단계 2로 간다. 그렇지 않으면 단계 4로 간다.

단계 4: 마지막으로 변화된 신경망의 가중값을 저장한다. 은닉노드의 개수를 하나 증가시키고, 새로운 모형의 초기 가중값을 할당한다. 단계 3으로 간다.

단계 4에서 초기 가중값 결정은 이차계획법(quadratic programming formulation)을 이용한다[12]

3. 실증분석

적은 개수의 학습자료를 이용하여 많은 개수의 모수들로 이루어진 MLP네트워크를 의도적으로 학습하기 위하여 다음과 같은 함수를 추정한다고 하자.

$$f(x) = e^{-(x-1)^2} + e^{-(x+1)^2}, \ x \in [-2.6, 2.6].$$

실험을 단순화하기 위하여 12개의 모의 학습자료 x를 구간 [-2.6, 2.6]의 균등분포에서 발생시키며, 발생된 표본은 평균 0이고 분산 0.005인 정규분포를 따르는 오차를 포함한다고 하자. 적용된 학습 알고리즘의 단계 3은 LM(Levenberg-Marquardt) 알고리즘을 이용한다. 학습의 중단시점은 최종 학습 허용오차 ε = 0.01이고, 초기 가중값의 범위는 $|c| < w$ 이다. 여기서 $w = 0.001 \sim 10$ 이다. 또한 오차 판단기준은 MSE이다. 일반적인 학습 알고리즘[11]과 본 연구의 개선된 알고리즘을 이용한 학습 결과는 표 1과 같다. 여기서 초기 가중값의 범위는 0.01 주변에서 두 방법간의 최선의 결과를 보여주고 있으며, 5와 10인 경우에 발생되고 있는 학습의 발산이 개선된 알고리즘에서는 수렴으로 진행되고 있음을 알 수 있다.

표 1. 초기값의 범위와 학습결과

w	일반 학습알고리즘		개선된 알고리즘	
	학습량	MSE	학습량	MSE
0.001	13172	0.013661	1227	0.012571
0.01	4644	0.010080	452	0.010102
0.1	18726	0.011074	684	0.010874
1	5863	0.011755	576	0.011545
2	6963	0.01227	635	0.012221
5	∞		701	0.012567
10	∞		817	0.012983

4. 결 론

본 연구에서는 FNN을 위한 개선된 학습 알고리즘과 기존의 학습 알고리즘을 이용하여 학습량과 수렴정도를 살펴보았다. 개선된 알고리즘은 한 개의 학습 패턴과 은닉노드로부터 시작하며, 패턴들은 하나씩 차례로 학습한다. 네트워크의 학습은 또 다른 패턴들을 학습과정중에 두가지 가능성 즉, 학습 성공의 경우 또는 지역 최소값으로 수렴하는 경우중의 하나로 학습과정을 결합하면서 학습하였다. 전자는 학습과정이 지역 최소값에 빠질 때까지 하나씩 더 많은 학습패턴을 결합하면서 학습과정을 계속한다. 후자인 경우에는 은닉노드의 개수를 늘림으로써 네트워크의 모수를 증가시킨다. 최적 처리절차에서 새롭게 추가된 은닉노드의 가중값에 대한 초기값 결정은 선형계획 방법으로 구했다. 최적 처리절차를 이용하므로써, 제시된 알고리즘은 거의 유사하게 허용오차 내에서 시스템의 오차를 만족했다. 그리고 학습의 발산이 나타나는 경우에도 수렴으로 진행되고 있음을 알 수 있다.

참 고 문 헌

1. M.T.Hagan and M.B.Menhaj, "Training feedforward networks with the Marquardt algorithm, "IEEE Trans. Neural Networks, vol.5, pp. 989–993, Nov. 1994.
2. Y.Fukuoka, H.Matsuki, H.Minamitani, and A.Ishida,"A modified back-propagation method to avoid false local minima,"Neural Netw., vol.11, no.6, pp.1059–1072,1998.
3. R.Parekh, J.Yang, and V.Honavar,"Constructive neural-network learning algorithms for pattern classification," IEEE Trans. Neural Networks, vol. 11, pp.436–451, Mar. 2000.
4. P.G.Maghami and D.W.Sparks,"Design of neural networks for fast convergence and accuracy: Dynamics and control," IEEE Trans. Neural Networks, vol.11, pp.113–123, Jan. 2000.
5. R.Setiono and L.C.K.Hui,"Use of quasi-Newton method in a feedforward neural network construction algorithm," IEEE Trans. Neural Networks, vol.6, pp.273–277, Jan.1995.
6. T.S.Chang and K.A.S.Abdel-Ghaffar,"A universal neural net with guaranteed convergence to zero system error,"IEEE Trans. Signal Processing, vol. 40, pp.3022–3031, Dec.1992.
7. Y.Hirose, K.Yamashita, and S.Hijiya,"Back-propagation algorithm which varies the number of hidden units,"Neural Networks, vol.4, no.1, pp.61–66, 1991.
8. M.R.Azimi-Sadjadi,S.Sheedvash, and F.O.Trujillo, "Recursive dynamic node creation in multilayer neural networks," IEEE Trans. Neural Networks, vol.4, pp.242-256, Mar. 1993
9. T.S.Chang and J.N.Hwang,"A novel framework for neural net learning,"in Intelligent Engineering Systems Through Artificial Neural Networks. NewYork: ASME, 1996, vol.6, pp.169–174.
10. P.RoyChowdhury, Y.P.Singh, and R.A.Chansarkar, "Dynamic tunneling technique for efficient training of multilayer perceptrons,"IEEE Trans. Neural Networks, vol.10, pp.48–55, Jan.1999.
11. Y.Wu and L.Zhang, "The Effect of Initial Weight, Learning Rate and Regularization on Generalization Performance and Efficiency," Proceedings on ICSP, pp.1191-1194. 2002
12. D.Liu, T.S.Chang and Y.Zhang, "A New Learning Algorithm for Feedforward Neural Networks," Proceedings of IEEE, International Symposium on Intelligent Control, September 5-7, Mexico, pp.39-44. 2001