

RBF 네트워크에서의 기저함수의 최적위치 추정방법

이종필 김성수
충북대학교 전기전자컴퓨터공학부

Estimation of Basis Functions in RBF Networks

J. P. Lee S. S. Kim
School of Electrical and Computer Engineering, Chungbuk National University

Abstract - RBF 네트워크에서 기저함수의 위치는 네트워크의 성능에 매우 큰 영향을 미친다. 몇몇 응용들에서 교사학습을 이용한 기저함수의 위치 선정이 비교사학습에 비해 우수함을 보인다. 그러나 교사학습에 의한 네트워크는 시그모이드 네트워크와 같은 긴 학습시간을 필요로 한다. 본 논문에서는 오차함수의 gradient와 Hessian을 이용해 교사학습에서 요구하는 학습시간을 단축시키면서 기저함수의 최적위치를 추정하였다.

1. 서 론

방사 기저 함수(Radial Basis Function:RBF) 네트워크는 보편적인 함수의 근, ill-posed 문제의 일반화 된 해를 제공하고 보간법의 이론을 잘 처리할 수 있는 능력을 가지고 있고 그 구조가 비교적 간단하므로 학습 시간이 단축되는 특징이 있기에 패턴인식, 시계열 예측이나 함수의 근사 등의 분야에 계속적으로 이용되어왔다. 일반적으로 RBF 네트워크에서 조정 가능한 변수는 은닉층 뉴런의 숫자, 기저함수의 폭과 중심의 위치 그리고, 출력의 가중치 등이 있다[1].

은닉층 뉴런의 개수나 출력 가중치의 최적화에 관한 문제는 여러 논문에서 계속적으로 연구되어왔다. 네트워크의 최적화 된 크기를 얻는 방법으로는 필요한 정확도에 도달하기 위해 네트워크의 크기를 증가시키는 방법이나 불필요한 뉴런을 삭제해 나가는 방법 등이 있고, 출력의 가중치는 델타규칙 등의 방법에 의해 결정된다.

일반적으로 보다 빠르게 학습시킬 수 있는 작은 크기의 네트워크를 구성하는 것이 바람직하다고 볼 수 있는데, 이것을 기저함수의 최적위치라 부른다. 본 논문은 RBF 네트워크의 중심위치의 학습에 관한 문제를 다룬다. 2장에서는 기존의 기저함수 위치 추정방법에 대해 설명하고, 3장에서는 본 논문에서 제안한 적정위치 추정 방법에 대해 논의하고, 4장에서는 제안된 적정위치 추정 방법에 대한 시뮬레이션 결과를 보인다.

2. RBF 신경회로망

2.1 RBF 신경회로망의 구조

RBF 네트워크는 다층형 신경회로망에 속하며, 일반적인 RBF 네트워크는 그림 1과 같이 두 개의 층을 가지고 있다. 은닉층의 뉴런들은 가우시안 함수와 같은 방사형 기저함수를 활성화함수로 갖는다. 각 뉴런이 갖는 방사형 기저함수의 중심은 그 뉴런이 갖는 연결강도에 의해 결정되고, 그 위치와 함수의 폭은 학습을 통해 구한다. 출력은 모든 방사형 기저함수의 출력을 선형 조합을 해서 얻어진다.

식(2.1)은 기저함수로 가우시안 함수를 가지는 RBF 네트워크의 출력이다.

$$y_i = \sum w_j \exp \left[-\frac{(x_i - c_j)^2}{2\sigma^2} \right] \quad (2.1)$$

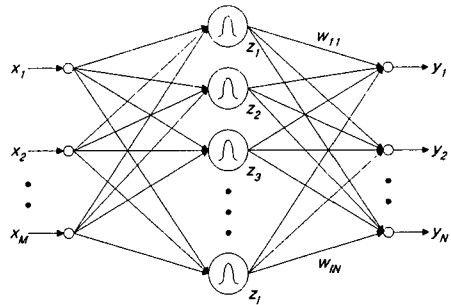


그림 1 RBF네트워크의 일반적 구조

2.2 중심 위치 선정 및 학습

일반적으로 RBF 네트워크 기저함수의 위치는 입력공간에서의 grid point의 집합으로 선정하거나[2] 학습자료의 임의의 부분집합으로 선정할 수 있는데, 이 경우에 기저함수의 수가 불필요하게 커질 수 있다. 또 다른 기저함수의 위치 결정 방법은 k-평균 클러스터링 알고리즘을 이용하여 학습의 입력자료에 대해 중심의 위치를 최적화 하는 것이다[3,4]. 진술한 방법들은 학습자료 중 RBF 네트워크의 입력자료에만 근거를 두고있고 출력값에 대해서는 고려하지 않는다. 입력과 출력데이터 모두를 고려한 방법으로는 직교최소제곱법(orthogonal least squares method) [5,6], 임출력값을 모두 포함하는 k-평균 클러스터링[7] 방법 등이 있다.

위와 같은 비교사 학습방법 이외에 식(2.2)의 평가함수를 최소화하도록 오차 역전파(Back propagation :BP) 학습법을 이용해 학습시키는 방법이 있다.

$$E(X_i) = \frac{1}{2} [d_i - y_i]^2 \quad (2.2)$$

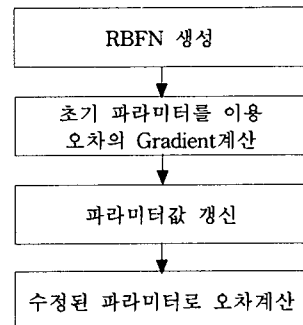


그림 2 오차 역전파 학습 알고리즘

BP 학습법의 절차를 그림 2에 나타내었다. BP 학습 알고리즘은 네트워크를 생성하고 초기 파라미터가 주어지면 초기 파라미터에서의 오차의 gradient를 계산하여 오차를 감소하는 방향으로 파라미터들을 gradient값 만큼씩 수정하는 방법이다.

RBF 네트워크에서 식(2.2)의 $E(X_t)$ 는 식(2.3)과 같이 놓을 수 있다.

$$E(X_t) = \frac{1}{2} \mathbf{r}^T(X_t) \mathbf{r}(X_t) \quad (2.3)$$

여기서, $\mathbf{r}(X_t) = [r_1(X_t), r_2(X_t), \dots, r_i(X_t)]$ 이고 기저함수로 가우시안 함수를 사용하는 RBF 네트워크에 대해서는 식(2.4)와 같이 표현된다.

$$r_i(X_t) = d_i - \sum_{j=1}^M w_j \exp\left[-\frac{(x_t - c_j)^2}{2\sigma_j^2}\right] \quad (2.4)$$

이때, X_t 에서의 gradient $\nabla E(X_t)$ 는 다음과 같이 계산된다.

$$\nabla E(X_t) = \mathcal{J}(\mathbf{r}(X_t))^T \mathbf{r}(X_t) \quad (2.5)$$

여기서, $\mathcal{J}(\mathbf{r}(X_t))$ 는 $\mathbf{r}(X_t)$ 의 자코비안이고 파라미터의 수정은 식(2.6)에 의해 수행된다.

$$X_{t+1} = X_t - \nabla E(X_t) \quad (2.6)$$

교사학습에 의해 기저함수를 선정하는 경우 비교사학습에 의한 네트워크에 비해 오차 면에서 우수함을 보인다. 이런 네트워크는 시그모이드 네트워크의 학습에서와 같이 매우 많은 반복학습을 필요로 한다. BP 학습법에서는 이 문제를 해결하기 위해 학습률(Learning rate)이라는 변수를 이용하여 보다 효과적인 학습이 되도록 유도하고 있다.

3. 제안된 적정위치 추정방법

본 논문에서는 학습의 속도를 향상시키기 위한 학습률로 에러평선의 Hessian을 이용하였다. 초기의 네트워크가 주어지면 BP 학습법에서와 같은 방법으로 gradient를 구하고 아울러 Hessian행렬을 구성해서 식(3.2)의 방법으로 학습률을 계산한다. 학습률은 gradient에 곱해서 기저함수의 위치를 조정하게 된다. 그림 3에 제안된 학습 절차를 나타내었다.

t 시점에서의 Hessian 행렬의 각 요소는 식(3.1)에 의해 구해진다.

$$\nabla^2 E(X_t) = \sum_{i=1}^N r_i(X_t) G_i(X_t) \quad (3.1)$$

여기서, $X_t = (c, \sigma, w)$ 이고 $G_i(X_t)$ 는 t 시점에서의 에러평선을 각 파라미터에 대해 2번 편미분한 것으로 식(3.2)와 같다.

$$G_i(X_t) = \begin{bmatrix} \frac{\partial^2 r_i(X_t)}{\partial c^2} & \frac{\partial^2 r_i(X_t)}{\partial \sigma \partial c} & \frac{\partial^2 r_i(X_t)}{\partial w \partial c} \\ \frac{\partial^2 r_i(X_t)}{\partial c \partial \sigma} & \frac{\partial^2 r_i(X_t)}{\partial \sigma^2} & \frac{\partial^2 r_i(X_t)}{\partial w \partial \sigma} \\ \frac{\partial^2 r_i(X_t)}{\partial c \partial w} & \frac{\partial^2 r_i(X_t)}{\partial \sigma \partial w} & \frac{\partial^2 r_i(X_t)}{\partial w^2} \end{bmatrix} \quad (3.2)$$

식(3.1)의 Hessian이 구해지면 식(3.3)과 같이 Hessian 행렬의 놈(norm)값으로 학습률을 정하여 식(2.5)에서 구한 gradient와 곱해서 t 시점에서의 파라미터값을 식(3.4)와 같이 수정하여 t+1시점의 파라미터를 구한다. 이러한 과정의 반복을 통해 네트워크의 학습이 진행하게 된다.

$$\eta = \|\nabla^2 E(X_t)\| \quad (3.3)$$

$$X_{t+1} = X_t - \eta \nabla E(X_t) \quad (3.4)$$

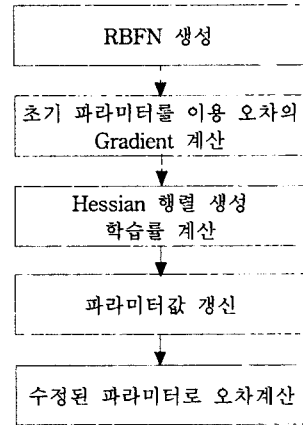


그림 3 제안된 학습 알고리즘

4. 시뮬레이션

3장에서 논의한 기저함수의 위치추정 방법을 기존의 BP 학습법과 비교하였다. RBF 네트워크를 이용하여 근사할 함수는 $f(x_1, x_2) = x_1 \times x_2$ 이다. 오차 역전파 방법과 제안된 방법에 의한 기저함수의 위치 추정에 따른 RBF 네트워크의 학습 오차와 속도를 비교하였다. 기저함수의 위치에 따른 오차만을 고려하기 위해 두 가지 모두 기저함수의 위치를 제외한 함수의 폭과 출력 가중치는 수정하지 않고 시뮬레이션 하였다.

실험에 사용된 초기 파라미터와 학습자료 입출력쌍은 각각 표1, 표2와 같다. 표1에는 임의로 선정된 초기 파라미터로 RBF 네트워크 기저함수의 위치(c_1, c_2), 폭, 출력 가중치가 나와있다. 파라미터의 학습을 위해 임의로 선정된 네트워크의 입출력 학습자료를 표2에 나타내었다.

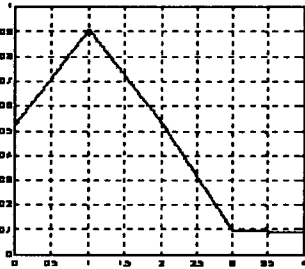
표 1 초기 파라미터

c_1 위치	c_2 위치	폭(σ)	출력 가중치(w)
-1	-1	1	1

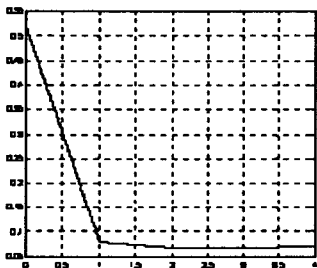
표 2 입출력 학습자료

x_1 입력	x_2 입력	목표값
-0.2886	-0.9335	0.2694
-0.3297	-0.9921	0.3271
-0.5500	-0.1000	0.0550
-0.3000	-0.9000	0.2700

각각의 방법에 의해 학습된 RBF 네트워크의 학습 횟수에 따른 오차의 변화를 그림 4에 나타내었다. 그림 4의 가로축은 학습횟수이고 세로축은 네트워크 출력의 오차 제곱합(Sum-squared error)이다. 기저함수의 위치에 따른 오차의 변화를 보이기 위해 양쪽의 경우 모두 기저함수의 위치를 제외한 폭과 출력가중치에 대한 학습을 시키지 않아 오차의 크기가 비교적 크게 나타났다. 그림 4의 a)와 b)를 비교한 결과 기존의 BP 학습법을 사용한 경우보다 제안된 적정위치 추정방법이 학습 속도 면에서 좀더 빠를 뿐만 아니라 오차의 수렴 면에서도 안정적으로 감소하는 것을 알 수 있다.



a) 기존의 BP학습법에 의한 결과



b) 제안된 학습법에 의한 결과

그림 4 시뮬레이션 결과

5. 결 론

RBF 네트워크에서 기저함수의 최적위치는 네트워크의 크기를 작고 효율적으로 만드는데 있어서 필수적이다. 출력오차의 관점에서 볼 때, 기저함수의 최적위치는 비교사 학습방법에 비해 교사학습 방법이 비교적 우수한 성능을 보이나, 그 학습속도가 느려지는 단점을 보인다.

본 논문에서는 교사학습방법을 이용하여 기저함수의 최적위치를 추정하면서, 교사학습 방법에서 발생하는 학습시간의 감소를 위해 오차함수의 gradient와 Hessian을 이용하여 학습 속도의 감소를 피하고 정확도 역시 향상시킬 수 있었다.

[참 고 문 헌]

[1] C. Panchapakesan, et al., "Effects of moving the centers in an RBF network", IEEE Trans. Neural Networks, vol. 13, No. 6, pp. 1299-1307, November, 2002.
 [2] J. Platt, "A resource allocating network for function interpolation", Neural Comput., vol. 3, pp. 213-225, 1991.

[3] J. Moody and C. J. Darken, "Fast learning in networks of locally-tuned processing units", Neural Comput., vol. 1, pp.281-294, 1989.
 [4] M. Vogt, "Combination of radial basis function neural networks with optimized learning vector quantization", Proc. IEEE, vol.83, pp. 1841-1846, Dec. 1993.
 [5] M. J. L. Orr, "Regularization in the selection of radial basis function centers", Neural Comput., vol. 7, pp. 606-623, 1995.
 [6] S. Chen, C. F. N. Cowan, and P. M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks", IEEE Trans. Neural Networks, vol. 2, pp.302-309, Mar. 1991.
 [7] Y. Zhang et al., A New Clustering and Training Method for Radial Basis Function Networks. New York:IEEE, vol.1, pp.311-316, 1996.