

PSO 기반 최적화 다항식 RBF 뉴럴 네트워크

백진열*, 오성권*
 수원대학교*

Optimized Polynomial RBF Neural Networks Based on PSO Algorithm

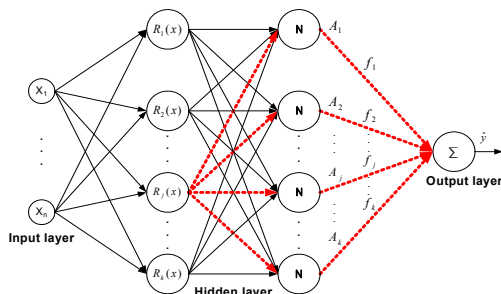
Jin-Yeol Baek*, Sung-Kwon Oh*
 Dept. of Electrical Engineering, The University of Suwon*

Abstract - 본 논문에서는 퍼지 추론 기반의 다항식 RBF 뉴럴네트워크(Polynomial Radial Basis Function Neural Network; pRBFNN)를 설계하고 PSO(Particle Swarm Optimization) 알고리즘을 이용하여 모델의 파라미터를 동정한다. 제안된 모델은 "IF-THEN" 형식으로 기술되는 퍼지 규칙에 의해 조건부, 결론부, 추론부의 기능적 모듈로 표현된다. 조건부의 입력공간 분할에는 HCM 클러스터링에 기반을 두어 구조가 결정되며, 기존에 주로 사용된 가우시안 함수를 RBF로 이용하고, 원뿔형태의 선형 함수를 제안한다. 또한 입력공간 분할시 데이터 집합의 특성을 반영하기 위해 분포상수를 각 입력마다 고려하여 설계함으로써 공간 분할의 정밀성을 높인다. 결론부에서는 기존 상수항의 연결가중치를 다항식 형태로 표현하는 pRBFNN을 제안한다. 제안한 모델의 성능을 평가하기 위해 Box와 Jenkins가 사용한 가스로 시계열 데이터를 적용하고, 기존 모델과의 근사화와 일반화 능력에 대하여 토의한다.

1. 서 론

입력층, 은닉층, 출력층, 세 층의 네트워크로 구성된 RBF 뉴럴네트워크는 퍼지추론의 추론 메커니즘과 유사한 기능을 수행할 수 있으며 최근 Jang에 의해 증명되었다[1]. 퍼지추론에 기반을 둔 모델의 규칙들은 최적구조 형성을 위해 전반부 동정과 후반부 동정으로 나뉘지며 각 동정은 다시 구조 동정과 파라미터 동정으로 나뉜다. 전반부의 구조 동정은 멤버쉽 함수의 형태와 수를 결정하는 문제로 입력공간을 퍼지 부분공간(fuzzy subspace)으로 분할하는 문제이며, 파라미터 동정은 구조로써 선택된 멤버쉽 함수의 파라미터를 결정하는 문제이다. 이는 시스템의 특성을 얼마나 잘 나타낼 수 있는가 하는 중요한 문제로 비선형 시스템의 성능을 결정하는데 많은 영향을 미친다. 이와 같이 퍼지추론 메커니즘 관점에서의 RBF 뉴럴네트워크 설계를 위한 전반부 동정의 핵심적인 문제는 RBF의 형태, RBF 중심의 위치 그리고 분포상수를 결정하는 것이다. 기존의 연구에서는 입력 데이터 중 임의의 중심 선택법[2], 군집화 기법을 이용한 중심 결정법[3] 등이 있다. 제안된 모델은 조건부, 결론부, 추론부의 기능적 모듈로 표현되며, 조건부의 RBF 개수와 중심의 위치를 결정하기 위해 HCM 클러스터링 알고리즘을 사용한다. 여기에, 입력변수에 대하여 동일하게 적용된 분포상수(표준편차)를 입력변수마다 독립된 분포상수로 사용하여 RBF를 통한 고차원 평면으로의 비선형 변환을 개선하고 공간 분할의 정밀성을 높인다. 또한 커널함수로 주로 사용되는 가우시안 이외에 원뿔형태의 선형 함수식을 제안한다. 결론부에서는 은닉층과 출력층 사이의 상수항 연결가중치를 다항식으로 표현함으로써 최종 출력을 비선형 결합으로 확장한다. 연결 가중치의 학습에 필요한 파라미터 즉, 학습을, 모멘텀 계수, 분포상수 가중치(RBF의 폭)를 사회적 행동양식을 바탕으로 한 PSO 최적화 알고리즘[4]을 사용하여 결정한다.

2. 다항식 RBF 뉴럴네트워크(pRBFNN)의 구조



〈그림 1〉 퍼지 추론 기반의 pRBFNN의 구조

본 논문에서 제안한 pRBFNN은 입력층, 은닉층, 출력층으로 구성되는 일반적인 RBF 뉴럴네트워크의 구조를 기반으로 한다. 하지만 기존의 기법과 달리, 퍼지 추론의 추론부에서 필요한 정규화 부분을 그림 1과 같이 은닉층에 추가하여 퍼지 추론 메커니즘의 기능적 모듈 동작 특성을 갖게 하였다.

3. 다항식 RBF 뉴럴네트워크(pRBFNN)의 설계

3.1 pRBFNN 커널함수(RBF)

일반적인 RBF의 입력변수 r은 중심점과 데이터간의 거리를 나타낸 것으로 중심에서 멀어질수록 큰 값을 갖게 된다. 식(1)은 Gaussian 함수를 정리한 것이다.

$$R_j = \exp(-r^2/2) = e^{-\frac{\|\mathbf{X}-\mathbf{u}_j\|^2}{2(\sigma_j)^2}}, \quad r = (\|\mathbf{X}-\mathbf{u}_j\|)/\sigma \quad (1)$$

식(1)은 식(2)로 표현되며, 이는 입력벡터에 동일한 분포상수를 갖는 형태가 된다.

$$R_j = e^{-\left(\frac{(x_1-u_{j1})^2}{2(\sigma_j)^2} + \frac{(x_2-u_{j2})^2}{2(\sigma_j)^2} + \dots + \frac{(x_n-u_{jn})^2}{2(\sigma_j)^2}\right)} \quad (2)$$

본 논문에서는 입력공간에 대한 각 특성을 모델에 반영하기 위해 동일한 분포상수를 갖는 RBF를 개선한다. 즉, 각 입력벡터에 대하여 다른 분포상수를 적용하면 식(3)과 같이 쓸 수 있다.

$$R_j = e^{-\frac{1}{2}\left(\frac{(x_1-u_{j1})^2}{(\sigma_{j1})^2} + \frac{(x_2-u_{j2})^2}{(\sigma_{j2})^2} + \dots + \frac{(x_n-u_{jn})^2}{(\sigma_{jn})^2}\right)} = e^{-\frac{1}{2}\left(\sum_{i=1}^n \frac{(x_i-u_{ji})^2}{(\sigma_{ji})^2}\right)} \quad (3)$$

식(3)을 다시 정리하면, 식(4)의 Weighted Euclidean Distance 정의에 의하여 각 입력에 대하여 독립적인 분포상수를 갖는 식(5)를 얻게 된다.

$$\|\mathbf{X}-\mathbf{u}_j\| = \sum_{i=1}^n \frac{\sqrt{(x_i-u_{ji})^2}}{\sigma_{ji}} \quad (4)$$

$$R_j = e^{-\frac{1}{2}\left(\sum_{i=1}^n \frac{(x_i-u_{ji})^2}{(\sigma_{ji})^2}\right)} = e^{-\left(\frac{\|\mathbf{X}-\mathbf{u}_j\|^2}{2}\right)}; \quad j = 1, 2, \dots, k \quad (5)$$

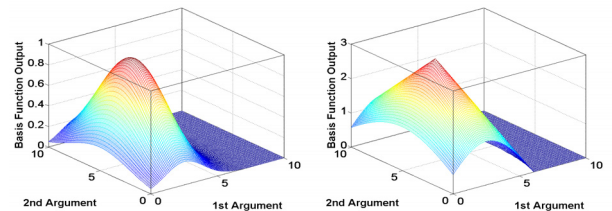
본 논문에서는 삼각형 형태와 유사한 멤버쉽함수의 구현을 위해 원뿔 형태의 RBF를 제안한다. 제안된 원뿔 형태의 RBF는 기존의 Linear type의 RBF[식(6)]를 이용하여 구현한다. 기존 Linear type의 경우 활성화 영역이 반전되었다는 점만 다를 뿐 그 특성이 가우시안 형태와 동일하다. 따라서 본 논문에서는 기존 Linear type을 중심에서 멀어질수록 작은 활성화 값을 갖도록 하여 원뿔형태의 RBF를 생성한다. 이를 위해 Linear type 커널함수에 식(4)를 적용하면, 식(6)의 Linear type 함수는 식(7)과 같이 쓸 수 있다.

$$R_j = r = \frac{\|\mathbf{X}-\mathbf{u}_j\|}{\sigma} \quad (6)$$

$$R_j = w_j \|\mathbf{X}-\mathbf{u}_j\|; \quad j = 1, 2, \dots, k \quad (7)$$

$$R_j = c - w_j \|\mathbf{X}-\mathbf{u}_j\| \quad (\text{단, } R_j < 0 \text{ 이면 } R_j = 0) \quad (8)$$

제안된 원뿔형태의 커널함수는 식(7)을 이용한 식(8)에 의해 생성된다. 여기서 c는 기본적으로 1의 값을 가지며, 활성화 영역의 조절을 위해 1보다 큰 값이 선택될 수 있다. 그림 2는 식(5)와 식(8)에 의한 커널함수의 활성화 영역이다.



(a) Gaussian RBF (b) Modified Linear RBF
 〈그림 2〉 커널함수(RBF)에 의한 활성화 영역

3.2 커널함수(RBF)의 수 및 중심점 결정

본 논문에서는 RBF의 수와 중심을 결정하기 위해 데이터들 간의 거리를 기준으로 근접한 정도를 측정하고, 이를 바탕으로 데이터를 특성별로 분류하는 HCM 클러스터링을 이용한다. 주어진 데이터는 먼저 HCM 클러스터링에 의해 학습 데이터를 분류하고, 분류 되어진 클러스터에 대하여 주어진 입력공간을 로컬영역으로 분할한다. 즉 HCM에 의한 클러스터의 수와 클러스터의 중심은 RBF의 수와 중심이 된다.

3.3 커널함수(RBF)의 분포상수 결정

분포상수(spread constant) σ 는 입력에 대해 어떻게 응답하는지를 결정하는 파라미터로 σ 가 클수록 RBF 활성화영역은 증가한다. RBF의 활성화 영역에 영향을 주는 분포상수는 HCM에 의해 결정된 클러스터에 속한 데이터의 표준편차를 이용하여, 식(9)와 같이 수학적으로 정의된다.

$$\text{std}(\sigma) = \sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 / (N - 1)} \quad (9)$$

x_i 는 클러스터에 속한 입력이고 N 은 클러스터에 속한 데이터의 수, \bar{x} 는 x 의 평균값이다. 본 논문에서는 모든 입력변수에 대하여 동일하게 적용된 활성화 영역의 개선을 위해 식(10)과 같이 각 클러스터에 속한 입력벡터 각각의 표준편차를 분포상수 σ 로써 사용한다.

$$\sigma_{ji} = \Delta\sigma_j + \text{std}_{ji}(x) \quad (10)$$

여기서 j 는 j 번째 클러스터를 나타내며, i 는 i 번째 입력변수이다. $\Delta\sigma_j$ 는 분포상수 가중치로 초기값은 1이다. 또한 분포상수 가중치 $\Delta\sigma_j$ 를 조절함으로써 각각의 입력변수에 적용된 최적 활성화영역을 결정하고 모델의 출력력을 향상시킬 수 있다. 본 논문에서의 최적 활성화영역 형성은 PSO 알고리즘을 이용하여 분포상수 가중치의 동조를 수행한다.

3.4 pRBFNN 결론부의 다항식 연결가중치 학습

퍼지추론 기반 pRBFNN의 결론부에 존재한 연결가중치의 학습은 경사하강법과 모멘텀을 통하여 이루어진다. 네트워크의 연결가중치, 즉 다항식 계수 동정을 위한 오차식 정의는 식(11)로 표현되며, 전체 학습데이터에 대한 모델의 최소 오차를 얻기위해 학습이 이루어진다.

$$E_p = (y_p - \hat{y}_p)^2 \quad (p = 1, 2, \dots, N) \quad (11)$$

여기서 E_p 는 p 번째 데이터에 대한 모델의 오차이며, N 은 입력 데이터 수이다. y_p 는 p 번째 입력에 대한 원 출력이며, \hat{y}_p 는 p 번째 입력에 대해 추론된 모델의 결과로 식(12)에 의한다.

$$\hat{y}_p = \left(\sum_{j=1}^k R_j \cdot f_j \right) / \left(\sum_{j=1}^k R_j \right) = \sum_{j=1}^k (A_j f_j) \quad (12)$$

경사하강법에 의한 다항식 계수의 학습은 식(11)로 표현되는 오차가 최소가 되도록 식(13)과 같이 계수의 변화분 Δa_{ji} 값을 이용하여 현재의 계수값을 갱신한다.

$$a_{ji}(\text{new}) = a_{ji}(\text{present}) + \Delta a_{ji} + \alpha \cdot \Delta a_{ji}(\text{old}) \quad (13)$$

여기서, η 는 학습율이고, α 는 모멘텀 계수이다. pRBFNN의 가중치 학습은 다음의 3가지로 나뉘어 진행된다.

1) $f_j = a_{j0}$: E_p 와 a_{j0} 사이의 관계를 Chain rule로 나타내면 식(14)와 같다. 식(14)의 우변항을 정리하면 식(15)와 같이 정리된다.

$$\Delta a_{j0} = -\eta_w \cdot \frac{\partial E_p}{\partial a_{j0}} = -\eta_w \left(\frac{\partial E_p}{\partial \hat{y}_p} \cdot \frac{\partial \hat{y}_p}{\partial f_j} \cdot \frac{\partial f_j}{\partial a_{j0}} \right) \quad (14)$$

$$\Delta a_{j0} = 2 \cdot \eta_w \cdot (y_p - \hat{y}_p) \cdot A_j \quad (15)$$

모멘텀을 고려한 새로운 연결가중치의 최종식은 식(16)과 같다.

$$a_{j0}(\text{new}) = a_{j0}(\text{present}) + \Delta a_{j0} + \alpha \cdot \Delta a_{j0}(\text{old}) \quad (16)$$

2) $f_j = a_{j0} + \sum_{i=1}^n a_{ji} \cdot x_i$: pRBFNN 결론부의 다항식 형태가 선형식의 경우 상수항인 a_{j0} 의 학습은 식(16)에 의한다. a_{ji} 의 변화분은 Chain rule에 의하여 식(17)과 같이 계산되며, 모멘텀을 고려한 새로운 a_{ji} 는 식(18)로부터 결정된다.

$$\begin{aligned} \Delta a_{ji} &= -\eta_w \cdot \frac{\partial E_p}{\partial a_{ji}} = -\eta_w \left(\frac{\partial E_p}{\partial \hat{y}_p} \cdot \frac{\partial \hat{y}_p}{\partial f_j} \cdot \frac{\partial f_j}{\partial a_{ji}} \right) \\ &= 2 \cdot \eta_w \cdot (y_p - \hat{y}_p) \cdot A_j \cdot x_j \end{aligned} \quad (17)$$

$$a_{ji}(\text{new}) = a_{ji}(\text{present}) + \Delta a_{ji} + \alpha \cdot \Delta a_{ji}(\text{old}) \quad (18)$$

3) $f_j = a_{j0} + \sum_{i=1}^n a_{ji} \cdot x_i + \sum_{i=1}^n \left(\sum_{k=1}^n a_{jik} \cdot x_i \cdot x_k \right)$: 모델 오차를 최소화하기 위한 a_{j0} 와 a_{ji} 의 학습은 식(16)과 식(18)에 의해 각각 이루어진다. 뿐만 아니라, a_{jik} 의 변화분은 구하기 위해 위에서 적용된 동일한 방법을 이용하여 식(19)를 유도한다. 모멘텀을 고려한 새로운 a_{jik} 의 최종식은 식(20)과 같이 정리된다.

$$\Delta a_{jik} = 2 \cdot \eta_w \cdot (y_p - \hat{y}_p) \cdot A_j \cdot (x_i \cdot x_k) \quad (19)$$

$$a_{jik}(\text{new}) = a_{jik}(\text{present}) + \Delta a_{jik} + \alpha \cdot \Delta a_{jik}(\text{old}) \quad (20)$$

3.5 PSO를 이용한 pRBFNN의 최적화

PSO(Particle Swarm Optimization) 알고리즘은 1995년 Eberhart와 Kennedy에 의해서 제안되었으며, 개체의 군집이론을 근본으로 사회적인 행동양식에 기반을 두고 있다. 본 논문에서 제안된 pRBFNN의 파라미터로는 가중치 학습에 관련된 학습율 η 와 모멘텀 계수 α 가 있으며, 은닉층 뉴런의 구조를 결정하는 분포상수의 가중치가 있다. 분포상수 가중치는 식(10)에서 언급된 $\Delta\sigma$ 이며, 분포상수 벡터는 규칙의 수와 동일한 수의 정보를 갖게 된다. 즉, 각각의 은닉층 뉴런은 입력의 수와 무관하게 동일한 분포상수 가중치를 갖게 된다. 이러한 분포상수 가중치는 입력 벡터가 늘어날 경우 계산량의 증가를 어느 정도 방지할 수 있다.

4. 시뮬레이션 및 결과고찰

본 논문에서 제안된 pRBFNN의 성능평가를 위해 Box와 Jenkins의 가스로 시계열 데이터를 이용한다. 근사화 성능과 일반화 성능을 평가하기 위한 지표로 식(21)의 mean square error(MSE)를 사용하였다.

$$PI \text{ or } EPI = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2 \quad (21)$$

여기서 PI(Performance Index)는 학습데이터에 대한 성능지수를, EPI(Evaluated Performance Index)는 테스트데이터에 대한 성능지수를 나타낸다. 제안된 pRBFNN의 최적화를 위해 사용된 PSO 알고리즘의 파라미터들은 표 1과 같이 요약된다.

〈표 1〉 PSO 파라미터의 요약

파라미터	세대 수	particle 수	v_{\max}	$[w_{\min} w_{\max}]$	c_1, c_2	탐색범위		
						학습율	모멘텀 계수	분포상수 가중치
값	60	50	탐색공간의 20%	[0.4 0.9]	2.0 2.0	[0 1.0]	[0 1.0]	[1.0 9.0]

입력 $u(t)$ 는 가스 흐름율이고 출력 $y(t)$ 는 이산화탄소 농도인 1입력 1출력의 데이터 296쌍으로부터 $u(t-3), u(t-2), u(t-1), y(t-3), y(t-2), y(t-1)$ 의 6개 입력변수와 한 개의 출력변수 $y(t)$ 를 설정하고 학습데이터로 145개, 테스트 데이터로써 각각 145개의 데이터 쌍을 사용한다. 데이터 집합 $[u(t-3), y(t-1); y(t)]$ 이외의 경우, 그 분포가 균일하지 못하고 흩어진 분포 구조를 보여줌으로 가스로 공정의 모델링에 적절하지 못하다. 따라서 본 논문에서는 pRBFNN 모델링을 위해 $[u(t-3), y(t-1); y(t)]$ 입력력 구조를 이용한다. 시뮬레이션 결과로부터, 2가지 형태의 RBF 모두에서 퍼지규칙 수가 증가 할수록 모델의 오차가 감소하는 특징을 보였으며, 후반부 다항식은 고차원 형태에서 좀 더 좋은 성능을 보였다. 규칙수에 따른 최적 모델의 구조를 결정하기 위하여 규칙수를 2~11개로 제한하였으며, 테스트 데이터에 대하여 가장 적은 오차를 갖는 규칙수를 최적의 모델로 선정하였다. 표 3은 각 RBF 사용시 선정된 최적 모델에 대한 오차와 기존 연구된 모델과의 성능비교를 보여준다.

〈표 3〉 최적 pRBFNN과 기존 모델과의 성능비교

모델	규칙수	PI	EPI	Ratio(EPI/PI)
Neural Network	15(mode)	0.018	0.411	22.833
Fuzzy model[5]	8	0.020	0.333	16.650
George's model[6]	7	0.022	0.266	12.090
pRBFNN				
Gaussian	8	0.031	0.234	7.5480
Modified Linear	8	0.041	0.226	6.4870

5. 결론

제안된 퍼지 추론 기반 pRBFNN은 RBF를 사용하여 퍼지공간 분할과 멤버십 함수를 결정한다. RBF에 의한 멤버십 함수의 구조를 결정하기 위한 파라미터로, 멤버십의 수, 각 멤버십의 정점, 그리고 활성화 정도를 결정하는 분포상수는 HCM에 기반을 두어 이루어진다. 여기서 기존에 주로 사용하는 가우시안 타입의 RBF 이외에 그 특징이 멤버십 함수로 적당한 Modified Linear 타입의 원뿔형 RBF를 제안하고, 이를 적용함으로써 사용함으로써 플랜트의 특성을 알 수 없는 데이터에 대하여도 일반화 성능을 향상 시켰다. 가스로 공정 데이터를 적용하여 얻어진 실험 결과로부터 제안된 pRBFNN 모델이 다른 모델과 비슷한 근사화 성능과 일반화 성능을 확인하였으며 기존의 모델이 과적화 되는 특성이 강했던 반면, 제안된 모델은 overfitting 측면에서 유연한 특성을 보여 줌으로써 모델의 우수성과 타당성을 입증하였다.

감사의 글

이 논문은 2008년도 한국전자통신연구원원의 재원으로 지원을 받아 수행된 연구임. (과제번호:2008-35MIC)

[참고 문헌]

[1] J.S Roger Jang, C.T. Sun, "Functional equivalence between radial basis function networks and fuzzy inference systems", IEEE Trans. Neural Networks, Vol. 4, No. 1, pp. 156-158, 1993.
 [2] K. Z. Mao, "RBF Neural Network Center Selection Based on Fisher Ratio Class Separability Measure", IEEE Trans. Neural Network, Vol. 13, No. 5, pp. 1211-1217, 2002.
 [3] M. J. D. Powell, "Radial basis functions for multivariable interpolation: A review", in Proc. IMA Conf. Algorithms for the Approximation of Functions and Data, Shrivensham, U.K., 1985.
 [4] R. C. Eberhard, P. Simpson, R. Dobbins, "Computational Intelligence PC Tools", Boston: Academic Press Professional, 1996.
 [5] S.-K. Oh and W. Pedrycz, "Identification of Fuzzy Systems by means of an Auto-Tuning Algorithm and Its Application to Nonlinear Systems", Fuzzy Sets and Syst., Vol. 115, No. 2, pp. 205-230, 2000.
 [6] George E. Tsekouras, "On the use of the weighted fuzzy c-means in fuzzy modeling", Advances in Engineering Software, Vol. 36, pp. 287-300, 2005.