

정보입자기반 RBFNNs에 의한 하수처리공정 시뮬레이터의 설계

이승주, 오성권
수원대학교 전기공학과

Design of Sewage Treatment Process Simulator with the Aid of IG-based RBFNNs.

Seung-Joo Lee and Sung-Kwun Oh
Department of Electrical Engineering, The University of Suwon

Abstract - RBFNNs(Radial Basis Function Neural Networks) 모델의 경우 Min-Max, HCM(Hard C-means)클러스터링 그리고 FCM(Fuzzy C-means)클러스터링 중 한가지를 통해 데이터 입자는 로드 규칙을 생성한 후 퍼지 공간을 분할 및 가우시안 함수의 정점을 정의한다. 본 논문은 기존의 방법과는 다르게 Min-Max와 FCM클러스터링을 혼합하여 로드의 규칙을 생성한 후 퍼지 공간을 분할 및 가우시안 함수의 정점을 정의하는 방법으로 사용하고자 한다. PSO최적화 알고리즘을 이용하여 같은조건에서 최적화한 기존의 방법으로 모델링된 RBFNNs와 Min-Max와 FCM 클러스터링을 혼합하여 사용한 방법의 비교를 통하여 어떤 모델의 성능이 더욱 좋은지 비교하고자 한다.

1. 서 론

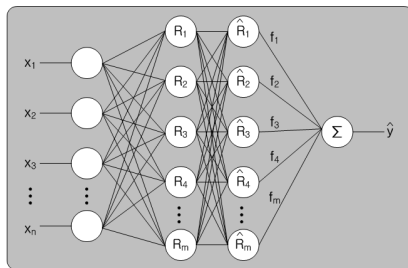
본 논문에서는 RBFNNs를 사용하여 모델을 구성하는데 있어서 전반부의 퍼지 공간 분할 방법에 변화를 주고자 한다. 기존 RBFNNs의 모델의 경우 전반부의 퍼지 공간의 분할 시에 입력데이터의 최대값과 최소값을 기준으로 사용하거나 HCM(Hard C-means)이나 FCM(Fuzzy C-means)를 사용하여 중심 값을 구하는 방법을 사용하게 된다. 하지만 이와 같은 방식에는 문제가 발생한다. 학습 데이터를 사용하여 모델을 만들고 테스트 데이터를 사용하여 성능을 평가하고자 할 때 은닉층의 증가에 따라 학습데이터의 성능이 나아지지만 테스트 데이터에 대한 성능은 급격하게 안좋아지는 경향을 보인다는 것이다.

그리하여 본 논문은 기존 FCM을 사용하는 방법과 Min-Max를 혼합하여 사용하고 이것을 같은 조건하에 PSO알고리즘을 통하여 어떠한 모델이 성능이 더 뛰어나게 나오지는 비교해 보도록 하겠다.

2. 본 론

2.1 RBFNNs

RBFNNs의 구조는 입력층, 은닉층, 출력층으로 구성되며 은닉층의 기본 구조는 <그림 1>과 같다. 여기서 n은 입력공간의 차원 수를 나타내며 m은 은닉층의 수를 나타낸다. 은닉층은 활성화 함수로서 방사형 기저 함수(Radial Basis Function)을 사용하여 은닉층의 출력 값을 결정한다.



<그림 1> MISO RBFNNs

전반부는 다중 입력, 후반부는 단일 출력을 기반으로 한 단일 규칙을 가지고 그 규칙은 식 (1)과 같으며,

$$\text{If } x_1 \text{ and } \dots \text{ and } x_n \text{ is } R_i \text{ then } y_i = f(x_1, \dots, x_n) \quad (1)$$

후반부 추론식은 식 (2),(3)과 같다.

간략 추론(case 1) :

$$f = w_{j0}^s \quad (2)$$

선형 추론(case 2) :

$$f = w_{j0}^s + \sum_{k=1}^d w_{jk}^s x_k \quad (3)$$

여기서, s는 출력변수의 수를 의미한다.

설계된 RBFNNs의 각 층의 연산과정은 다음과 같다.

[1층] 입력층

입력 신호가 다음 층의 해당 노드로 전달된다.

[2층] 적합도 계산 : R^i

가우시안 함수를 사용하여 은닉층의 출력 값을 결정

$$R^i = e^{-\sum_{j=1}^n \frac{(x_{ij} - v_{ij})^2}{2\sigma_j^2}} \quad (4)$$

[3층] 적합도 정규화 : \hat{R}^i

$$\hat{R}^i = \frac{R^i}{\sum_{i=1}^m R^i} \quad (5)$$

[4층] 최종 출력 : \hat{y}

LSE를 사용하여 전체 퍼지 규칙의 파라미터 계수를 동시 계산

$$\hat{y} = \sum_{i=1}^m \hat{R}^i \cdot f_i \quad (6)$$

2.1.1 FCM 클러스터링

FCM클러스터링은 각 클러스터의 점에 대하여 데이터의 소속 정도를 소속함수로 나타낸 데이터 분류 알고리즘으로 소속정도는 클러스터의 중심점과 각 데이터들의 거리로 계산한다.

FCM 클러스터링의 방법은 다음과 같다.

[단계1] 클러스터의 개수c(2<c<n)를 정하고 퍼지화계수m(1<m<∞)을 선택한다. 초기 소속함수를 초기화 한다.

[단계2] FCM 클러스터 중심 $v_{ij}(i=1,2,\dots,c, j=1,2,\dots,n)$ 를 계산한다.

$$v_{ij} = \frac{\sum_{k=1}^n (u_{ik})^2 x_{kj}}{\sum_{k=1}^n (u_{ik})^2} \quad (7)$$

$m=2(\text{지수의가중})$

[단계3] FCM클러스터의 중심과 데이터의 거리를 구한다.

$$d_{ik} = \left(\sum_{j=1}^s (x_{kj} - v_{ij})^2 \right)^{1/2} \quad (8)$$

[단계4] 구한 거리를 바탕으로 새로운 소속함수를 구한다.

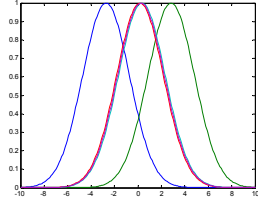
$$u_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{d_{jk}}{d_{ik}} \right)^{\frac{2}{m-1}}} \quad (9)$$

[단계5] 이전 소속함수와 새로운 소속함수의 차가 초기 설정치에 만족하면 종료하고 아니면 위의 과정을 반복한다.

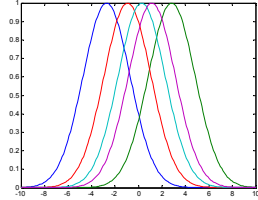
2.1.2 FCM 클러스터링과 Min-Max 혼합

FCM클러스터링과 Min-Max 혼합에는 여러 가지 방법을 생각해 볼 수 있겠지만 2가지 방법을 사용하도록 하겠다. 첫 번째 방법은 전체 은

닉층 중심 값의 50%는 FCM클러스터링을 통해서 나머지 50%는 Min-Max값의 균등 분할을 통해서 찾아내는 방법, 두 번째 방법은 FCM 클러스터링을 사용하되 클러스터링의 중심 값 중 두 개의 값을 입력데이터의 Min-Max값으로 고정해 시키는 방법이다. 두 번째 방법을 추가로 설명하면 은닉층이 5개가 존재할 때 2개의 중심 값을 데이터의 Min값과 Max값으로 고정해서 FCM클러스터링을 통해 구한 중심 값과 2개의 중심 값을 데이터의 Min값과 Max값을 넣고 3개만 FCM클러스터링을 통해 구한 중심 값을 확실히 차이가 있다. <그림 2>는 FCM클러스터링 시 Min-Max값을 함께 클러스터링에 사용한 것이며, <그림 3>은 Min-Max 값을 추가시킨 것이다.



<그림 2> Min-Max 혼합



<그림 3> Min-Max 추가

2.2 PSO알고리즘

PSO 알고리즘은 진화연산의 결합에 의한 기술이다. PSO는 물고기, 새 떼와 같은 행동양식을 바탕으로 이루어진다. RBFNNs에서 입력 데이터의 수, 입력 데이터의 선택, FCM의 퍼지화계수, 은닉층의 수와 후 반부 추론방법을 찾아낸다.

3. 실험 및 결과 고찰

본 논문의 사용하는 하수처리 시스템의 데이터를 주로 사용하되 비교를 돕기 위해 가스료 데이터도 사용한다. 데이터는 학습 : 트레이닝 : 검증으로 나누어 5 : 3 : 2 구성한다. 모델구성 및 최적화 시 학습데이터와 트레이닝 데이터를 사용하며 최적화된 파라미터를 적용하며 모델구성 시 모델의 평가는 검증데이터를 사용한다. 검증데이터는 모델의 구성 시 어떠한 가담 하지 않는다.

3.1 PSO알고리즘 최적화 조건 설정

앞 절에서 본 것과 같이 PSO최적화 알고리즘을 이용하여 찾아낸 파라미터들은 <표 1>과 같다. 데이터 선택에 있어서는 중복 선택이 가능하다.

<표 1> PSO알고리즘 최적화 범위

입력 수	데이터선택	퍼지화계수	은닉층수	추론방법
2-5	1-5	1-3	10-40	간략,선형

3.2 최적화에 따른 모델 파라미터

다음 <표 2>과 <표 3>는 최적화알고리즘을 이용하여 찾아낸 각 모델의 파라미터 값이다. 가스료데이터의 FCM'을 나머지는 은닉층의 수가 비교적 비슷하게 나타남을 볼 수 있다.

<표 2> 수처리 데이터 최적화 파라미터

	입력 수	데이터선택	퍼지화계수	은닉층수	추론방법
FCM	4	1,3,4,5	1	25	선형
FCM'	4	1,3,4,5	1	25	선형
FCM''	5	1,2,4,5,1	2	28	선형

<표 3> 가스료 데이터 최적화 파라미터

	입력 수	데이터선택	퍼지화계수	은닉층수	추론방법
FCM	2	1,5	1.054423933068	31	간략
FCM'	2	1,5	2.132545434500	28	간략
FCM''	2	1,5	1	10	선형

FCM : FCM클러스터링만을 사용

FCM' : FCM클러스터링 시 Min-Max값 고정

FCM'' : FCM클러스터링 시 총 은닉층의 반은 Min-Max 균등분할

3.3 각 모델의 성능

다음 <표 4>과 <표 5>는 수처리 데이터와 가스료 데이터라 모델링 시 성능을 보여주는 표이다. 그리고 <그림 4>와 <그림 5>는 <표 4>과 <표 5>의 값을 보기 쉽게 그래프로 나타낸 것이다. 수처리 성능과 가스료 성능의 공통점은 일반 FCM클러스터링을 사용하는 것 보다 FCM클

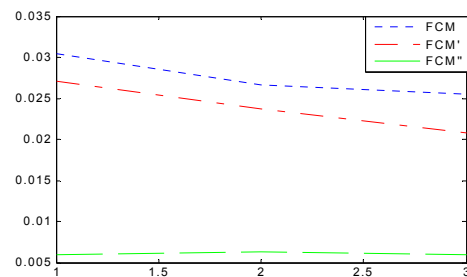
러스터링 시 데이터의 Min-Max값을 함께 넣고 클러스터링을 하는 것이 학습데이터, 트레이닝데이터, 검증데이터의 성능이 더 좋은 것을 볼 수 있다. 그리고 은닉층의 반은 FCM클러스터링으로 나머지 반은 Min-Max 값의 균등분할 값으로 사용한 경우는 데이터에 따라 차이가 있음이 볼 수 있다.

<표 4> 수처리 성능

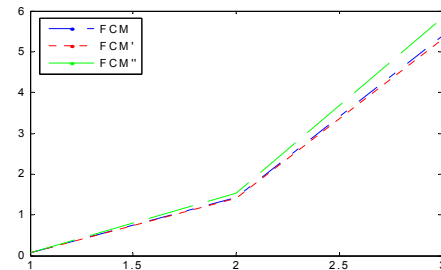
	FCM	FCM'	FCM''
PI	0.0304793609	0.0270630851	0.0059829215
EPI	0.0266828235	0.0236782147	0.0063280158
Validation	0.0255000616	0.0207776458	0.0059138139

<표 5> 가스료 성능

	FCM	FCM'	FCM''
PI	0.0702438090	0.0691243009	0.0745598019
EPI	1.4304777418	1.4135013230	1.5255778626
Validation	5.3911614702	5.3091817118	5.8479568155



<그림 4> 수처리 성능



<그림 5> 가스료 성능

x축 1: PI, 2: EPI, 3: Validation, y축 각 성능 값

4. 결 론

이번 논문은 FCM클러스터링으로 RBFNNs의 은닉층 중심 값을 찾을 때 데이터의 최대 값과 최소 값을 혼합함에 따른 효과를 알아보기 위한 실험이었다, FCM클러스터링 시 학습 데이터의 최대 값과 최소 값을 고정해 놓고 찾아낸 RBFNNs의 중심 값을 사용 시 일반적인 방법의 PI, EPI 와 Validation 보다 전반적으로 더 좋은 성능을 보여준다.

감사의 글

본 연구는 중소기업청에서 지원하는 2010년도 산학연 공동 기술개발 사업(No. 00043125)의 연구수행으로 인한 결과물임을 밝힙니다 그리고 이 논문은 2009년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2009-0074928)

[참 고 문 헌]

- [1] S. K Oh, W. Pedrycz, H. S. Park, "A granular-oriented development of functional radial basis function neural networks", Neurocomputing 72, pp. 420-435, 2008.
- [2] S. K Oh, W. Pedrycz, "genetic optimization-driven multi-layer hybrid fuzzy neural networks", Simulation Modelling Practice and Theory, Vol. 14, pp. 627-642, 2003.