

# 클러스터링 기법을 이용한 비선형 공정의 병렬구조 모델링

## Parallel Structure Modeling of Nonlinear Process Using Clustering Method

\*박춘성\*, 최재호\*, 오성권\*, 안태천\*

\* 원광대학교 제어계측공학과, 전라북도 익산시 신용동 344-2 ☎570-749

\*Chunseong Park\*, Jaeho Choi\*, Sungkwun Oh\*, Taechon Ahn\*

\* Dept. of Control & Instrumentation Engineering, Wonkwang Univ., Iksan, KOREA

### Abstract

In this paper, We proposed a parallel structure of the Neural Network model to nonlinear complex system. Neural Network was used as basic model which has learning ability and high tolerance level. This paper, we used Neural Network which has BP(Error Back Propagation Algorithm) model. But it sometimes has difficulty to append characteristic of input data to nonlinear system. So that, I used HCM(hard c-Means) method of clustering technique to append property of input data. Clustering Algorithms are used extensively not only to organized categorize data, but are also useful for data compression and model construction. Gas furance, a sewage treatment process are used to evaluate the performance of the proposed model and then obtained higher accuracy than other previous medels.

### I. 서론

비선형 시스템의 모델을 구축하기 위해서는 시스템의 특성을 정확히 파악해야 한다. 이렇게 파악된 특성들은 여러 가지 모델링 방법에 적용되어지고, 최종모델을 구축하게 된다. 시스템의 특성을 파악하는 방법은 여러 가지 방법이 있는데 비선형 공정에서는 입출력 특성을 많이 사용한다. 이런 방법은 시스템의 내부적인 구조는 모르더라도 시스템의 동작특성을 대략적으로 파악할 수 있기 때문에 매우 널리 사용되고 있다. 비선형 시스템의 경우는 선형시스템보다 시스템의 내부 동작 특성을 파악하기 어렵기 때문에, 모델을 구축하기 위한 유일한 정보원으로 사용하는 경우가 많다. 이런 이유 때문에 GMDH, 퍼지, 신경망등 단순히 입출력 데이터만을 이용하여 비선형 시스템을 모델링하는 방법들이 많이 연구되고 있다. 이들 연구는 내계 모델 자체의 구조를 개선시키는 쪽에 초점을 맞추고 진행되고 있다. 본 논문의 기본 모델은 오류 역전파알고리즘으로 학습하는 신경망을 사용하고, 입출력 데이터의 특성별 분류는 유클리드 거리를 기준으로 데이터를 분류하는 HCM방법을 이용한다. 신경망은 학습능력 및 높은 융통성을 발휘하고 직관적으로 모델링할 수 있다. 클러스터링 기법을 사용하여 입출력 데이터의 분석을 통해 기본 모델은 그대로 유지하고, 입출력 데이터간의 유사한 특성을 가지는 데이터들끼리 몇 개의 집합으로 나누어 모델을 구축하는 것이다. 이것은 병렬구조를 가지

는데 데이터의 특성에 따라 신경망의 문턱치와 가중치를 각각 변화시켜 모델을 구축한다. 이 과정을 가스로 공정과 활성오니공정의 비선형 공정에 적용하여 단일 구조의 모델보다 우수한 모델을 구축할 수 있음을 보인다.

## II. 신경망의 구조

신경망은 기본적으로 인간의 뇌의 구조와 동작방식을 따른다. 물론, 신경망의 구조나 동작방식은 뇌의 그것에 비해 매우 단순화 되어 있다. 신경망모델은 하나의 처리단자에 여러개의 입력 패턴과 하나의 출력 패턴으로 연결된 구조를 갖는다. 처리단자는 여러개의 입력패턴으로부터 하나의 출력을 생성하며, 시그모이드(sigmoid)의 활성화함수로 표현된다. 그리고 각 노드에는 연결가중치가 있고 이 연결가중치는 처리단자의 출력과 출력패턴과의 오차에 의해 보정된다. 입력층, 은닉층, 출력층으로 구성된 신경 회로망을 사용하고 오류 역전파 알고리즘(Error Back Propagation Algorithm)을 사용하여 각 연결 노드의 연결가중치 및 문턱치를 보정한다.

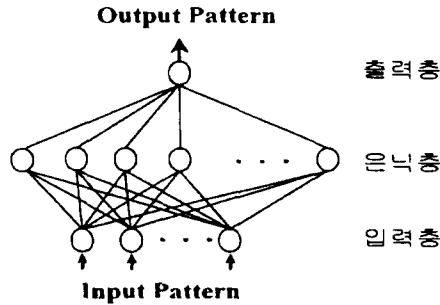


그림 2.1. 신경망의 구조

## III. 클러스터링 알고리즘

데이터의 분류를 위해서 클러스터링 기법을 많이 사용된다. 클러스터링이란 데이터의 내부의 비슷한 패턴, 속성, 형태 등의 기준을 통해 데이터를 분류하여 내부의 구조를 찾아내는 것이다. 클러스터링 기법은 HCM(Hard C-means), FCM(Fuzzy C-means), mountain Clustering 방법 등이 있다. 본 논문에서 사용되는 방법은 데이터들간의 거리를 기준으로하여 근접한 정도를 측정하고, 이를 바탕으로 데이터를 분류하는 HCM(Hard C-mean)을 사용한다. HCM 알고리즘은  $n$ 개의 데이터를  $c$ 개의 그룹으로 분할하고 데이터의 거리가 최소인 각 그룹의 중심을 찾는다. 또한, 클러스터의 소속을 "0"과 "1"로 이치 논리를 사용한다. 클러스터링 기법에서 가장 먼저 결정해야 될 것이 클러스터의 개수이다. 일반적으로  $n$ 개의 데이터로 구성된 집합안에 클러스터의 개수는 클러스터가 1개일 때와  $n$ 개일 때는 클러스터링의 의미가 없다. 그러므로 2보다 크거나 같고  $n$ 보다 작아야 한다. 적절한 클러스터의 개수는 데이터에 따라 적절한 개수를 선택해주어야 한다. 이렇게 결정된 클러스터의 수는 사용될 모델의 개수와 같다. 적절하지 못한 클러스터의 개수는 모델의 효율성을 떨어뜨리기 때문에 가장 적절한 클러스터 개수를 구한다는 것은 매우 중요하다. 클러스터링의 절차는 다음과 같다.

step 1. 클러스터 중심  $c_i, i=1, \dots, c$ 을 랜덤하게 선택하여 초기화한다.

step 2. 소속행렬  $U$ 를 결정한다.

$$u_{ij} = \begin{cases} 1 & \text{if } \|x_i - c_j\|^2 \leq \|x_i - c_k\|^2, \text{ for } k \neq j \\ 0 & \text{otherwise} \end{cases} \quad (3.1)$$

step 3. 식 (3.2)을 계산한다. 중심 근처에 충분히 다가갔는지 확인한다.

$$J = \sum_{i=1}^c J_i = \sum_{i=1}^c \left( \sum_{k, x_k \in C_i} d(x_k - c_i) \right) \quad (3.2)$$

step 4. 새로운 클러스터 중심( $c_i$ )을 생성하고 step 2로 간다.

$$c_i = \frac{1}{|C_i|} \sum_{k, x_k \in C_i} x_k \quad (3.3)$$

#### IV. 제안된 비선형 공정의 병렬구조 모델

클러스터링 기법을 통해 데이터는 적절하게 분할된 집합을 갖는다. 이렇게 결정된 클러스터 개수는 병렬모델을 구성하는 기본모델의 수가 되고, 각 기본모델은 특정 클러스터에 포함되는 데이터들에 의해 학습하게 된다. 제안된 모델구조는 병렬로 구성되는 모델안의 신경망들은 모두 같은 구조를 가지게 된다. 단지 학습되는 데이터가 다르기 때문에 구조 내부의 연결 가중치만 달라질 뿐이다. 분류기는 모델의 학습시에는 입력된 데이터가 어떤 모델을 이용했을 때 가장 좋은 결과를 낼지 분석하여 모델을 선택하고, 선택된 모델에 데이터를 전달하는 역할을 하게 된다.

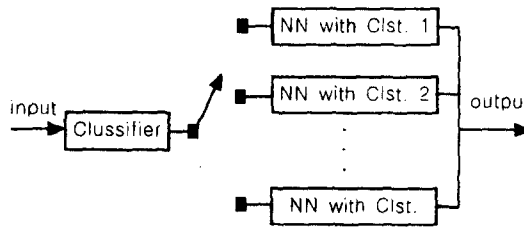


그림 4.1 병렬구조 모델

#### V. 시뮬레이션

##### V-1 가스로 공정 모델링

제안한 모델링 기법을 가스로 공정에 적용해 본다. Box와 Jenkins의 Gas Furnace 데이터는 296쌍의 입출력으로 이루어졌으며 입력( $u$ )은 Gas의 흐름을 나타내고 출력( $y$ )은 이산화탄소( $CO_2$ )의 농도를 나타낸다. 본 논문에서는 모델링 연구에 많이 사용되는 아래의 식 (5.1)을 사용한다.

$$\hat{y}(t) = F(u(t-3), y(t-1)) \quad (5.1)$$

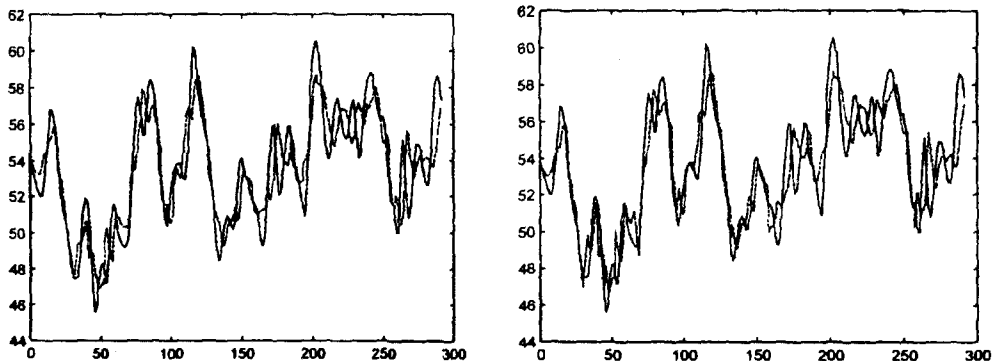


그림 5.1 기본방법과 제안된 방법의 성능 비교 (좌: 단일구조, 우: 병렬구조)

## V-2 하수 처리 공정 모델링

하수처리를 위한 다양한 방법중 가장 많이 사용되는 활성오니(activated sludge)를 이용한 방법이다. 활성오니 공정은 일반적으로 침사지, 최초침전지, 폭기조 및 최종침전지로 구성된다. 침사지와 최초침전지에서는 하수중에 포함된 이물질들이 중력에 의해 침강되고, 폭기조에서는 하수에 공기를 흡입시킨다. 활성오니란 유기물 제거의 중심을 이루는 미생물로 폭기조에서 자연 발생되며, 응집한 미생물의 덩어리를 플록(floc)이라 한다. 활성오니는 폭기조내에서 공급되는 산소와 결합하여 유기물을 물과 탄산가스로 분해한다. 최종침전지에서는 플록을 침강, 제거하고 맑은 물을 외부로 흘려보낸다. 한편, 최종침전지에서 침강한 미생물은 폭기조 입구에서 재순환되어 다시 유기물 제거에 사용된다.

본 논문에서는 활성오니공정을 폭기조 및 최종침전지를 하나의 단일 모델로 설정하여, 실제 최초 침전지로 유입되는 생물학적 산소요구량(Biological Oxygen Demand;BOD) 및 유입수의 부유물 농도(Suspended Solid;SS)를 측정하고, 최종침전지에서 유출되는 생물학적 산소요구량 및 부유물 농도를 통하여 하수처리 정도를 비교한다. 사용된 데이터는 수도권 하수처리장 중의 하나를 모델로 선정하여 처리장의 1년치 수질 데이터이다. 아래의 표에서 다른 모델링 방법과 제안된 성능을 비교하였다.

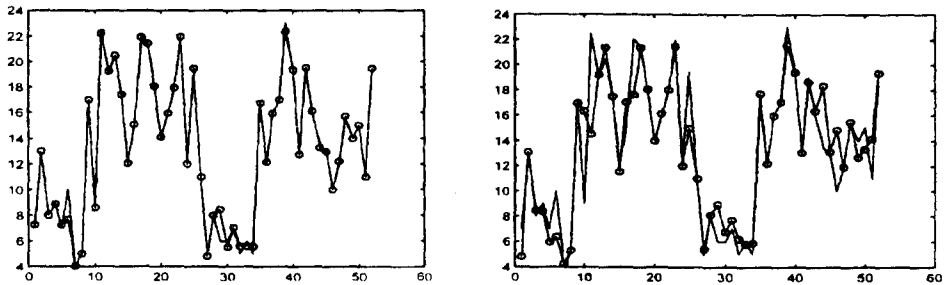


그림 5.2 기본 방법과 제안된 방법의 성능 비교(좌:단일구조, 우:병렬구조)

## VI. 결 론

본 논문에서는 클러스터링 기법을 이용하여 병렬구조를 가진 신경망으로 비선형 공정 시스템을 모델링 하였다. 비선형 시스템의 입출력 데이터가 어떤 작은 상호간의 특성이 있으면 제안된 구조에서는 확실히 성능향상을 볼 수 있다. 클러스터링 기법을 이용한 비선형 공정 시스템의 병렬구조 모델링을 단일 구조의 신경망과 병렬구조 신경망을 비교하였다. 성능평가를 위해 가스로 공정과 하수처리 공정에 적용하였고 성능평가측면에서 우수한 성능을 보였다. 제안된 방법은 비선형 시스템의 입출력 데이터를 분석하여 어떤 상호간의 연관성을 찾아 몇 개의 구조를 이루어 전체적으로 강한 구조를 만드는데 장점이 있다.

## VII. 참고문헌

- [1] Neural Networks - A Comprehensive Foundation, Simon Haykin, McMaster University, 1994
- [2] Neural Fuzzy Systems - A NeuroFuzzy Synergism to Intelligent systems, Chin-Teng Lin and C.S. George Lee. Prentice Hall, 1996
- [3] Neuro\_Fuzzy and Soft Computing - J.-S.R Jang, C.-T. Sun, E. Mizutani, Prentice Hall, pp423-pp425
- [4] R.M.Tong, "Synthesis of fuzzy models for industrial processes", Int. J. Gen. Syst., Vol.4, pp.143-162, 1978.
- [5] 오성원, 우광방, "퍼지추론 방법에 의한 퍼지동정과 하수처리공정시스템 응용", 대한전자공학회 논문집 제31권 B권 제6호, pp.43-52, 1994년 6월.
- [6] 오성원, 조석범, 남경문, "퍼지-뉴럴 네트워크 구조에 의한 비선형 공정시스템의 비선형 모델링", 한국퍼지 및 지능시스템학회 논문집 제5권 제4호, pp.41-55, 1995.
- [7] 하승하, 기계 신경망, 이상원 Ohm사, 1993