

# 점증적 입자 모델의 진화론적 설계에 근거한 에너지효율 예측

## Energy Efficiency Prediction Based on an Evolutionary Design of Incremental Granular Model

염 찬 옥\* · 곽 근 창†

(Chan-Uk Yeom · Keun-Chang Kwak)

**Abstract** - This paper is concerned with an optimization design of Incremental Granular Model(IGM) based Genetic Algorithm (GA) as an evolutionary approach. The performance of IGM has been successfully demonstrated to various examples. However, the problem of IGM is that the same number of cluster in each context is determined. Also, fuzzification factor is set as typical value. In order to solve these problems, we develop a design method for optimizing the IGM to optimize the number of cluster centers in each context and the fuzzification factor. We perform energy analysis using 12 different building shapes simulated in Ecotect. The experimental results on energy efficiency data set of building revealed that the proposed GA-based IGM showed good performance in comparison with LR and IGM.

**Key Words** : Incremental granular model, Genetic algorithm, Energy efficiency data, Linear regression

### 1. 서 론

실 세계 다양한 응용문제에 대해서 퍼지 모델에 관한 수많은 연구들이 수행되어져 왔다. 복잡하고 비선형 특성을 가진 실 세계 문제들은 다양한 모델들로부터 방법들, 구조, 기술들을 통합하는 하이브리드 지능 시스템 모델을 필요로 하고 있다. 보완적인 하이브리드 지능 시스템의 설계를 수행함에 있어서 개별적인 모델보다는 여러 계산적인 기법들을 통합하는 것이 이점이 있다. 하이브리드 지능 시스템과 관련해서 자주 사용되는 대표적인 방법은 뉴로-퍼지 추론 모델링이다. 신경회로망은 변화하는 환경에 대해 대처하도록 적응성과 학습 특징을 가진다. 퍼지추론 시스템은 인간 전문가의 지식을 결합하고, 퍼지 추론과 지식기반 의사결정을 수행한다[1].

또한, 유전알고리즘(GA: Genetic Algorithm)과 파티클 군집 최적화 (PSO: Particle Swarm Optimization)와 같은 진화론적 최적화 기법을 통해 퍼지 혹은 뉴로-퍼지 모델의 최적화에 관한 여러 연구들이 있다. Lin[2]은 TSK (Takagi-Sugeno-Kang) 타입의 뉴로-퍼지 네트워크의 설계를 위해 효과적인 번역기반 PSO 학습 알고리즘을 제안하였다. Oh[3]는 PSO를 통해 다항식기반 방사기저함수 신경회로망(RBFNN: Radial Basis Function Neural Network)의 최적화하는 방법을 제안하였다.

한편, 입자 모델 (GM: Granular Model)의 설계에서 퍼지 클러스터링의 근본적인 아이디어를 이용하는 퍼지 모델링의

새로운 방법에 관해 여러 연구들이 진행되어져왔다[4]. 이 입자모델의 근본적인 개념은 정보입자(information granules)들 사이의 관계로 표현하는 것이다. 즉, 입력변수와 출력변수 모두 퍼지 집합들을 자동적으로 생성한다. 정보입자들은 컨텍스트기반 퍼지 c-means 클러스터링 (CFCM: Context-based Fuzzy C-Means)을 이용함으로써 생성되어진다[5]. 일반적인 퍼지 c-means 클러스터링방법들과는 대조적으로 CFCM 클러스터링은 입력과 출력공간들 사이의 동질성을 연결하는 클러스터들을 추정한다. 먼저 컨텍스트들이 출력공간에서 생성되면, CFCM 클러스터링은 제공된 각 컨텍스트안에서 클러스터링이 수행되어진다. 직접적인 설계절차 때문에 입자 모델은 특히 빠른 시스템 프로토타입에서 유용하다. Pedrycz는 CFCM 클러스터링에 근거하여 각각 RBFNN[6], 언어적인 모델 (LM: Linguistic Model)[7], 동적 데이터 입자화를 설계하는 방법[8] 등을 제안하였다.

그러나, 복잡하고 비선형 특성을 가진 문제에 대해서는 근사화와 일반화 능력 모두 우수한 성능을 얻기가 어려워진다. Kwak[9, 10]은 퍼지 입자화를 가진 GA기반의 최적화된 언어적인 모델을 설계하는 방법을 제안하였다. 진화알고리즘을 근거로 한 입자 모델의 최적화에 관한 다양한 연구들이 수행되어져 왔다. 이러한 다양한 연구들이 수행되어졌음에도 불구하고, 모델링 오차를 예측하는 방법은 고려되어지지 않고 있다.

이러한 모델링 오차를 예측하기 위해 Pedrycz는 전역적인 모델로써 선형회귀(LR: Linear Regression)와 국소적인 모델로써 LM을 결합한 점증적 입자모델 (IGM: Incremental Granular Model)을 제안하였다[11]. 이 모델은 먼저 선형적인 부분에 대해 LR을 이용하여 설계한 후, LR에 의해 얻어진 모델링 오차는 국소적인 비선형성을 다루기 위해 LM에 의해 생성된 퍼지 규칙을 이용하여 예측되어진다. 그러나, 이 방법은 각 컨텍스트에서 데이터의 특성과 크기가 서로 다름에도 불구하고, 각 컨텍스트에서 클러스터의 수가 같도

† Corresponding Author : Dept. of Electronics Engineering, Chosun University, Gwangju, Korea

E-mail : kwak@chosun.ac.kr

\* Dept. of Electronics Engineering, Chosun University, Korea  
접수일자 : 2018년 2월 20일

최종완료 : 2018년 2월 22일

록 설계하는 문제점을 가지고 있다. 게다가, 퍼지화계수는 비선형성의 특성에 관한 중요한 영향력을 나타남에도 불구하고 일반적으로 고정된 값을 사용한다.

그러므로 본 논문에서는 GA기반의 점증적 입자 모델에 대해 설계하는 방법을 제시하고, 실세계 응용문제를 해결하고자 한다[12]. 여기서 각 컨텍스트에 대응되는 클러스터 수와 퍼지화 계수가 진화알고리즘인 GA에 의해 최적화되어진다. 제안된 방법은 빌딩의 냉난방 에너지 효율문제[13]에 적용하여 이전 연구된 방법들과 예측성능을 비교하여 우수성을 증명하고자 한다.

## 2. 선형회귀(LR)와 언어적인 모델(LM)의 설계

IGM의 구성요소는 전역적인 모델로서의 LR, 컨텍스트 기반 퍼지 C-Means 클러스터링, 국소적인 모델로서의 LM으로 구성되어진다. 이들 구성요소들은 각각 고유의 장점을 가지고 있으며, 통합된 형태는 IGM 설계의 핵심을 형성하고 있다.

### 2.1 선형회귀(LR)에 의한 전역적인 설계

간소성을 위해 설계하고자 하는 실험적인 데이터는 두 개의 입력과 하나의 출력으로 구성된다고 가정하자. 이들 데이터 포인트는  $\{x_k, y_k\}, k=1,2,\dots,N$ 의 형태를 갖고,  $x_k$ 는 입력벡터이고,  $y_k$ 는 출력이다. LR 모델은 다음과 같은 식으로 표현되어진다.

$$z_k = a^T x_k + a_0 \quad (1)$$

여기서,  $a^T$ 는 선형회귀의 계수들을 나타낸다. 이들 계수들은 최소자승법에 의해 결정되어진다. 여기서 선형모델의 오차인  $e_k = y_k - z_k$ 는 언어적인 규칙들에 의해 표현되고 있는 국소적인 LM에 의해서 예측되어진다.

### 2.2 언어적인 모델(LM)에 의한 국소적인 설계

국소적인 모델로써 언어적인 모델은 컨텍스트 기반 퍼지 클러스터링(CFCM)을 통해 설계되어진다. 이러한 CFCM 클러스터링에 대해서 간략히 설명하면 다음과 같다. 이 클러스터링의 소속행렬  $U$ 는 다음 식과 같이 계산되어진다.

$$u_{ik} = \frac{f_k}{\sum_{j=1}^c \left( \frac{\|x_k - c_i\|}{\|x_k - c_j\|} \right)^{\frac{2}{m-1}}} \quad (2)$$

여기서  $m$ 은 퍼지화 계수이며,  $f_k$ 의 값은 0과 1사이의 소속도 값을 나타낸다.  $f_k = W(d_k)$ 는 출력공간에서 생성된 임의의 퍼지 집합에서  $k$ 번째 데이터의 포함정도를 표현한다. 출력공간의 퍼지 집합은  $W: D \rightarrow [0,1]$ 로 정의되며,  $D$ 는 출력 변수의 전체집합이다. 여기서 이러한 특성에 의한 소속행렬은 다음과 같다.

$$U(f) = \left\{ \begin{array}{l} u_{ik} \in [0,1] \mid \sum_{i=1}^c u_{ik} = f_k \forall k \\ \text{and } 0 < \sum_{k=1}^N u_{ik} < N \end{array} \right\} \quad (3)$$

출력 변수의 확률적인 분포에 의해 컨텍스트를 발생시킨다[12]. 각 컨텍스트에서 생성된 클러스터의 중심은 기존 FCM 클러스터링과 같이 계산한다.

$$c_i = \frac{\sum_{k=1}^N u_{ik}^m x_k}{\sum_{k=1}^N u_{ik}^m} \quad (4)$$

그림 1은 언어적인 모델의 구조를 보여주고 있다. 그림 1에서 보는 바와 같이 원형 노드에 포함하고 있는 전체부의 파라미터는 CFCM 클러스터링을 통해 얻어진다. 출력공간에서 생성된 컨텍스트  $W_t$ 는 결론부의 파라미터가 된다. 연결가중치가 일반적인 수치값을 갖는 다층 퍼셉트론이나 RBFNN과는 다른 독특한 형태를 가지고 있음을 알 수 있다. 또한, 신경회로망에서는 지식생성이 어렵지만 그림 1의 LM은 퍼지규칙을 생성할 수 있다는 장점을 가지고 있다. 결과적인 추론값  $Y$ 는 다음과 같이 퍼지 수(fuzzy number)에 의해 계산되어진다.

$$Y = \sum_{\oplus} W_t \otimes z_t \quad (5)$$

언어적인 모델의 출력 값은 0과 1사이의 신뢰도를 포함하고 있으며 구간 값으로 표현할 수 있다. 이러한 불확실성을 갖는 출력 경계는 다음과 같이 계산되어진다.

- lower bound:  $y_- = \sum_{t=1}^p z_t w_{t-} + w_0$
- upper bound:  $y_+ = \sum_{t=1}^p z_t w_{t+} + w_0$

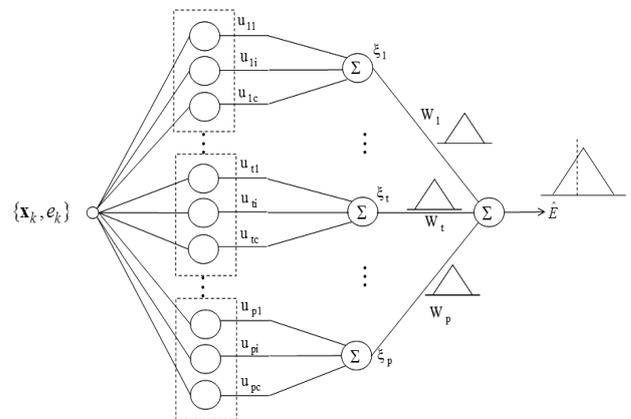


그림 1 언어적인 모델(LM)의 구조  
Fig. 1 Architecture of LM

### 3. GA에 의한 점증적 입자모델의 최적화

#### 3.1 유전알고리즘 (GA)

본 절에서는 IGM의 설계에서 GA에 근거한 최적화 설계 방법을 제시한다. 기존의 IGM의 설계에서는 각 컨텍스트에서 고정된 클러스터 수와 전형적인 퍼지화계수를 사용하지 않, 본 논문에서는 진화론적인 방법을 이용하여 이들의 값을 최적화하고자 한다. GA는 이진비트 스트링으로 최적화 되도록 각 파라미터들을 인코딩한다. 각각의 세대에서 GA는 교배, 돌연변이와 같은 유전 연산자를 이용하여 새로운 집단을 생성한다. GA는 인코딩 기법, 적합함수 평가, 부모 개체선택, 교차연산, 엘리트즘 기법, 돌연변이연산 등을 포함하고 있다.

적합함수는 세대가 생성된 후 집단에서 각 개체에 대한 적합도를 계산함으로써 평가되어진다. 각 개체의 적합함수는 최대화문제로 표현된 목적함수에 의해 계산되어진다. 적합함수는 주어진 해가 문제를 해결하는데 얼마나 적합한지를 정량적으로 측정한다. 여기서 GA와 관련하여 다양한 방법 가운데 자주 사용되는 적합함수를 다음과 같이 사용하고자 한다.

$$f = \frac{1}{F_t + F_c} \quad (6)$$

$$F_t = \sqrt{\frac{1}{N} \sum_{k=1}^N (y_k - \hat{Y}_k)^2} \quad (7)$$

$$F_c = \sqrt{\frac{1}{N} \sum_{k=1}^N (y_k - \hat{Y}_k)^2} \quad (8)$$

여기서  $y_k$ 와  $y_k$ 는 각각 학습과 검증데이터에 대한 실제 출력값이며,  $\hat{Y}_k$ 와  $\hat{Y}_k$ 는 각각 학습과 검증데이터에 대한 모델 출력값들이다. IGM에 대한 예측성능을 평가하기 위해 RMSE(root mean square error) 성능척도를 사용한다.

#### 3.2 GA기반 점증적 입자모델의 설계

GA기반의 점증적 입자 모델의 설계과정을 간략히 설명하면 다음과 같다.

[단계 1] 먼저 수치적인 입출력데이터로부터 전역적인 모델인 LR을 이용하여 선형예측을 수행한다. 그리고, 실제 원하는 출력과 LR 모델 출력사이의 모델링 오차를 얻는다. 초기 입출력데이터로부터 새로운 형태의 입력-오차 데이터  $(x_k, e_k)$ 를 형성한다.

[단계 2] 새롭게 형성된 오차공간에서 컨텍스트들을 생성한다. 이들 컨텍스트는 오차분포의 통계적인 특성에 의해 얻어진다[12].

[단계 3] 오차공간에서 생성된 컨텍스트와 대응되는 입력공간에서 CFCM클러스터링을 수행한다. 전형적인 LM의 설계에서는 컨텍스트의 수  $p$ 와 각 컨텍스트에서 추정된 클러스터의 수  $c$ 를 곱한  $p \times c$ 개의 클러스터들을 얻는다. 그러나, 각 컨텍스트로부터 할당된 입력 데이터들의 고유특성을 반영하는 클러스터들을 결정할 필요가 있다.

그러므로, GA에 근거하여 각 컨텍스트에 대응하는 최적화된 클러스터의 수를 찾는다.

[단계 4] 대응하는 컨텍스트에 의해 생성된 클러스터들의 활성함수 값을 얻고, 컨텍스트를 통한 가중치에 의해 출력값을 계산한다. 결과적으로 IGM 출력은 삼각형 형태의 퍼지 수를 얻는다.

[단계 5] LR의 출력과 LM의 입출력을 결합한다. 결과적으로 최종적인 예측은  $\hat{Y} = z \oplus \hat{E}$ 에 의해 계산되어진다.

그림 2는 GA기반 IGM의 최적화 과정을 보여주고 있다. 그림 2에서 보는 바와 같이 각 컨텍스트에 대응되는 클러스터의 수와 퍼지화 계수는 이진비트의 형태로 표현되어짐을 알 수 있다. 그림과 같이 컨텍스트의 수가 6개일 경우, 최적화하고자하는 파라미터들은 6개의 클러스터의 수와 1개의 퍼지화 계수이다. 이들 파라미터들의 탐색범위는 응용문제에 맞게 한정되어진다.

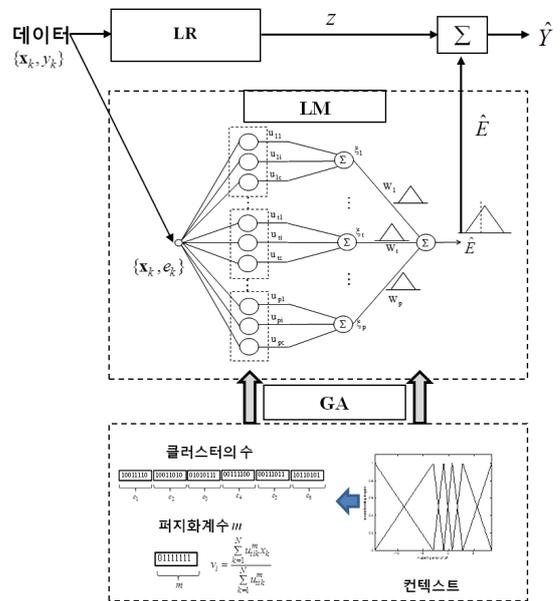


그림 2 GA기반 IGM의 최적화 과정  
Fig. 2 Optimization process of GA-based IGM

### 4. 실험 결과

본 절에서는 제안된 방법의 예측성능을 보이기 위해 빌딩의 냉난방 에너지 효율문제를 다루고자 한다. 이 예제는 8개의 입력특성과 2개의 출력변수들을 포함하고 있다. 8개의 건물 특성은 상대 간결성, 표면적, 벽으로 된 공간, 지붕으로 된 공간, 전체 높이, 방향, 유약 지역, 유약 지역 분포도로 이루어져 있고 출력은 난방부하와 냉각 부하로 구성되어 있다. 여기서 출력은 난방부하로 한정하여 실험을 수행한다. 본 실험에서는 GA기반 IGM에 대해 모델을 구축하기 위해 384개의 학습데이터를 이용하고, 나머지 384개는 모델평가를 위해 검증데이터로 사용되어진다.

GA를 이용하여 파라미터를 최적화하기 위해서 클러스터 수의 탐색공간을 2개에서 9개까지로 설정하고 퍼지화계수는 1.5에서 3까지 범위를 한정한다. GA의 파라미터는 개체 30,

교차율 0.97, 돌연변이율 0.01로 하고, 엘리티즘을 이용하여 가장 좋은 적합도를 갖는 두 개의 개체들은 다음 세대로 교차나 돌연변이가 변화없이 유지된다.

그림 3은 LR에 의해 얻어진 오차에 대해서 6개의 컨텍스트를 생성한 그림을 보여주고 있다. 그림 3에서 보는 바와 같이 오차가 0인 근처에 컨텍스트가 집중되어 있는 것을 알 수 있다. 이러한 컨텍스트는 오차에 대한 언어적인 정보입자의 형태로 표현할 수 있다. 예를 들면, 오차가 NE (Negative Error)라면, 이 컨텍스트에 해당되는 퍼지 규칙들이 생성된다. 그림 4는 LR에 의해 얻어진 오차를 국소적인 LM에 의해서 예측되고 있음을 확인할 수 있다. 그림 4에서 보는 바와 같이 일부 값을 제외하고는 오차들을 잘 예측하고 있음을 알 수 있다.

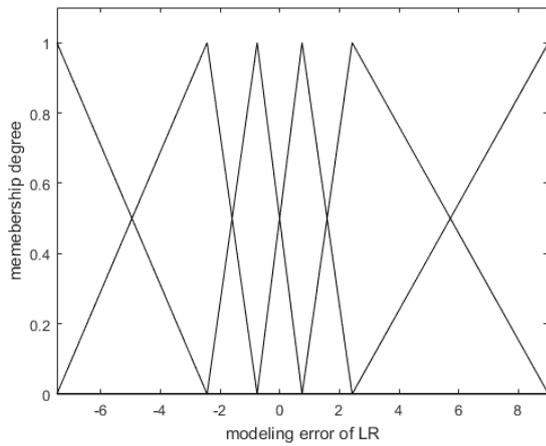


그림 3 컨텍스트의 생성 (p=6)  
Fig. 3 Generation of contexts (p=6)

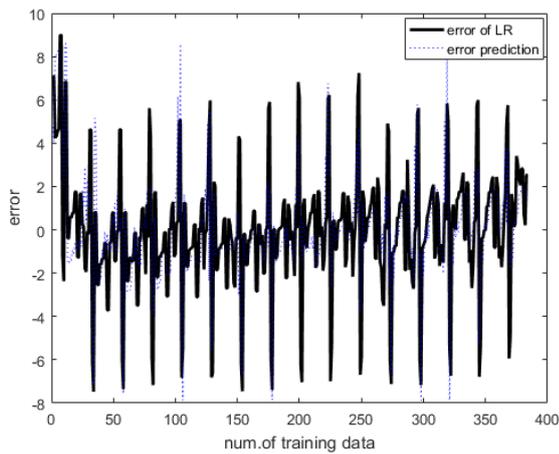


그림 4 LR 오차에 대한 예측  
Fig. 4 Prediction of LR error

그림 5와 6은 난방부하에 대한 학습데이터와 검증데이터에 대해서 근사화와 일반화 성능을 각각 보여주고 있다. 그림에서 보는 바와 같이 제안된 방법은 우수한 근사화와 일반화 능력을 가지고 있음을 보여주고 있다.

표 1은 제안된 방법의 난방부하 예측성능을 RMSE 성능

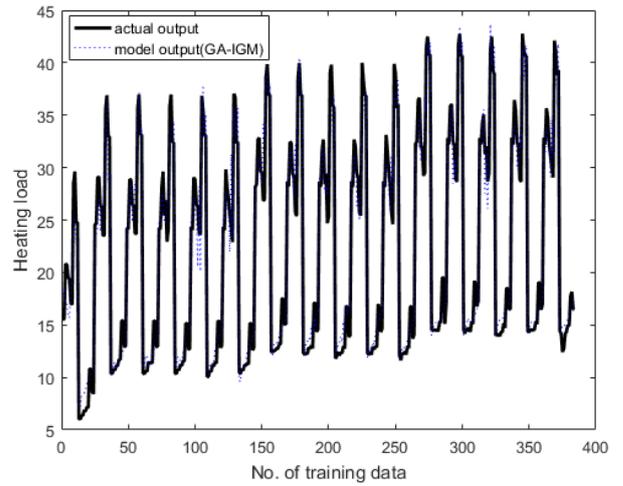


그림 5 학습데이터에 대한 근사화 성능  
Fig. 5 Approximation performance for training data set

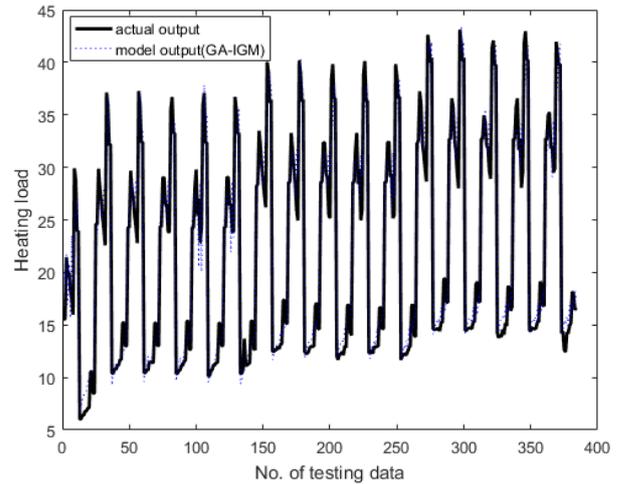


그림 6 검증데이터에 대한 일반화 성능  
Fig. 6 Generalization performance for testing data set

표 1 난방부하 예측성능비교

Table 1 Performance comparison for heating load prediction

	p, c	No. rule	RMSE (training)	RMSE (testing)
LR	-	-	2.90	2.93
IGM [11]	p=c=6	36	2.46	2.63
	p=6,c=7	42	2.36	2.62
	p=6,c=8	48	2.33	2.56
	p=6,c=9	54	2.26	2.54
	p=6,c=10	60	2.21	2.53
GA-IGM	p=6, c=[8 7 6 8 9 9] m=1.5	47	2.12	2.36

척도로 비교하고 있다. 표 1에서 보는 바와 같이 제안된 방법으로 구한 최적의 파라미터는  $p=6$ 일 경우  $c=[8\ 7\ 6\ 8\ 9]$ ,  $m=1.5$ 임을 알 수 있다. 또한 효율적인 규칙의 수도 함께 구해짐을 알 수 있었다. 기존 IGM에서 같은 클러스터의 수를 사용하는 방법보다는 GA에 의해 최적화된 클러스터 수와 퍼지화 계수를 갖는 것이 더 우수한 성능을 보임을 알 수 있었다.

## 5. 결 론

본 논문에서는 GA기반 IGM의 최적화 설계 방법을 제시하고 실세계응용문제를 다루었다. 기존 IGM은 각 컨텍스트에서 데이터의 고유 특성과 데이터의 규모가 서로 다름에도 불구하고, 각 컨텍스트에 대응되는 클러스터의 수가 같게 설계하는 문제점을 가지고 있다. 또한, 퍼지 클러스터링을 수행할 때 퍼지화계수는 비선형성의 특성을 반영하는 파라미터임에도 불구하고 전형적인 값을 사용하였다. 제안된 방법은 클러스터 수와 퍼지화 계수의 최적값을 찾음으로써 이러한 문제점들을 해결할 수 있었다. 빌딩의 에너지 효율데이터에 적용하여 실험한 결과 제안된 방법은 기존의 IGM보다 우수한 근사화와 일반화능을 보임을 알 수 있었다. 앞으로 클러스터 수와 퍼지화계수의 최적화 뿐만 아니라 컨텍스트의 수에 대한 최적화와 오토엔코더와의 연관성을 찾기 위한 연구를 진행할 계획이다.

## References

- [1] J. S. R. Jang, C. T. Sun, and E. Mizutani, "Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence," Prentice Hall, 1997.
- [2] C. J. Lin, "An efficient immune-based symbiotic particle swarm optimization learning algorithm for TSK-type neuro-fuzzy networks design," *Fuzzy Sets and Systems*, vol. 159, no. 21, pp. 2890-2909, 2008.
- [3] S. K. Oh, W. D. Kim, W. Pedrycz, and B. J. Park, "Polynomial-based radial basis function neural networks (P-RBF NNs) realized with the aid of particle swarm optimization," *Fuzzy Sets and Systems*, vol. 163, no. 1, pp. 54-77, 2011.
- [4] W. Pedrycz and A. V. Vasilakos, "Linguistic models and linguistic modeling," *IEEE Trans. on Systems, Man, and Cybernetics-Part B*, vol. 29, no. 6, pp. 745-757, 1999.
- [5] W. Pedrycz, "Conditional fuzzy c-mans," *Pattern Recognition Letters*, vol. 17, no. 6, pp. 625-631, 1996.
- [6] W. Pedrycz, "Conditional fuzzy clustering in the design of radial basis function neural networks," *IEEE Trans. on Neural Networks*, vol. 9, no. 4, pp. 601-612, 1998.
- [7] W. Pedrycz and K. C. Kwak, "Linguistic models as a framework of user-centric system modeling," *IEEE Trans. on Systems, Man, and Cybernetics-*

*Part A: Systems and Humans*, vol. 36, no. 4, pp. 727-745, 2006.

- [8] W. Pedrycz, "A dynamic data granulation through adjustable fuzzy clustering," *Pattern Recognition Letters*, vol. 29, no. 16, pp. 2059-2066, 2008.
- [9] K. C. Kwak, W. Pedrycz, "A design of genetically oriented linguistic model with the aid of fuzzy granulation," *2010 IEEE World Congress on Computational Intelligence*, Barcelona, Spain, pp. 18-2, July, 2010.
- [10] K. C. Kwak, "A design of genetically optimized linguistic models," *IEICE Trans. on Information and Systems*, vol. E95D, no. 12, pp. 3117-3120, 2012.
- [11] W. Pedrycz, K. C. Kwak, "The development of incremental models," *IEEE Trans. on Fuzzy Systems*, vol. 15, no. 3, pp. 507-518, 2007.
- [12] Y. H. Byeon and K. C. Kwak, "A design of genetically oriented rules-based incremental granular models and its application," *Symmetry*, vol. 9, no. 12, Article ID 324, 2017.
- [13] UCI Machine Learning Repository, Available online: <https://archive.ics.uci.edu/ml/datasets>.

## 저 자 소 개



### 염 찬 욱 (Chan-Uk Yeom)

2010년 조선대학교 제어계측공학과 졸업 (공학사), 2017년 8월 동대학원 석사과정 졸업(공학석사), 현재 동대학원 박사과정중  
E-mail : walt18@naver.com



### 곽 근 창 (Keun-Chang Kwak)

1996년 충북대학교 대학원 전기공학과 석사과정 졸업(공학석사), 2002년 동대학원 박사 졸업(공학박사), 2003-2005년 캐나다 앨버타대학교 전기 및 컴퓨터공학과 박사후연구원, 2005-2007년 한국전자통신연구원 지능형로봇연구단 선임연구원  
현재 조선대학교 전자공학과 교수  
E-mail : kwak@chosun.ac.kr