

# Knowledge Representation Based On Interval Type-2 CFCM Clustering

Myung-Won Lee, Keun-Chang Kwak

**Abstract**—This paper is concerned with knowledge representation and extraction of fuzzy if-then rules using Interval Type-2 Context-based Fuzzy C-Means clustering (IT2-CFCM) with the aid of fuzzy granulation. This proposed clustering algorithm is based on information granulation in the form of IT2 based Fuzzy C-Means (IT2-FCM) clustering and estimates the cluster centers by preserving the homogeneity between the clustered patterns from the IT2 contexts produced in the output space. Furthermore, we can obtain the automatic knowledge representation in the design of Radial Basis Function Networks (RBFN), Linguistic Model (LM), and Adaptive Neuro-Fuzzy Networks (ANFN) from the numerical input-output data pairs. We shall focus on a design of ANFN in this paper. The experimental results on an estimation problem of energy performance reveal that the proposed method showed a good knowledge representation and performance in comparison with the previous works.

**Keywords**—IT2-FCM, IT2-CFCM, context-based fuzzy clustering, adaptive neuro-fuzzy network, knowledge representation.

## I. INTRODUCTION

ALTHOUGH the concept of the Type-2 (T2) fuzzy set was first introduced by [1], no one had extended the Type-1(T1) fuzzy system to the T2 fuzzy system until the literature of Marnik and Mendel has published [2]. During the past few decades, it has witnessed a rapid growth in the number and variety of applications of T2 fuzzy logic system and its extensions, ranging from pattern recognition and industrial process control to decision support systems and data clustering [3].

Especially pattern recognition and data clustering in real-world problems are subject to various sources of uncertainty. In order to solve these uncertainty problems, various T2 fuzzy clustering algorithms have proposed. Hwang [4] has proposed uncertain fuzzy clustering with Interval Type-2 Fuzzy C-Means (IT2-FCM) clustering approach. He has focused on the uncertainty associated with the fuzzifier parameter that controls the amount of fuzziness of the final  $c$ -partition in the design of fuzzy  $c$ -means clustering. Linda [5] has proposed general T2-FCM clustering algorithm for uncertain fuzzy clustering to solve the computation intensity of the processing of general T2 fuzzy sets. Qiu [6] has enhanced IT2-FCM algorithm with improved initial cluster centers. A new initialization method for cluster centers and an enhanced Karnik-Mendel (KM) algorithm [7] with optimization have been proposed. Yao [8] has developed Type-2 Gustafson-

Kessel (T2-GK) clustering algorithm with fuzzy covariance matrix.

However, although these clustering algorithms have their potential to deal with the uncertainty, these techniques are performed by context-free clustering method without considering the homogeneity between input and output spaces. In contrast to these context-free clustering algorithms such as FCM clustering, the objective of context-based fuzzy clustering approach is to estimate cluster centers preserving homogeneity of the clustered patterns in connection with their similarity in both the input and output variable based on linguistic contexts. The effectiveness of this Context-based Fuzzy C-Means (CFCM) clustering proposed by [9] has been demonstrated on RBFN [10], and Linguistic Models (LM) [11]. These models with uncertain output predicted by lower and upper bounds have represented the nonlinear characteristics more effectively than the well-known models with the aid of context-free clustering.

Therefore, in order to enhance the uncertainty processing and homogeneity characteristics, a new method for integrating the concept of Interval Type-2 fuzzy set with CFCM clustering algorithm, so called IT2-CFCM is proposed in this paper. For this purpose, the linguistic contexts generated by output space are transformed into T2 fuzzy form to deal with uncertainty degree. Finally, the experiments are performed to show the effectiveness and good performance in comparison to the previous works. Section II describes the basic concept of conventional FCM and IT2-FCM clustering. In the Section III, the IT2 context-based clustering technique for extracting fuzzy if-then rules in the design of Adaptive Neuro-Fuzzy Network (ANFN) [12] is proposed. The proposed model is applied to estimation problem of energy performance of residential buildings [13] with input-output data points in Section IV. Finally, conclusions are given in Section V.

## II. CONTEXT-FREE AND CONTEXT-BASED FUZZY CLUSTERING

Context-free fuzzy clustering such as Fuzzy C-Means (FCM) partitions a collection of input vector into  $c$  fuzzy groups, and finds a cluster center in each group such that a objective function is minimized. The membership matrix is allowed to have elements with values between 0 and 1 to perform fuzzy partitioning. However, imposing normalization stipulates that the summation of degrees of belongingness for a data set always be equal to unity.

$$U(\mathbf{f}) = \left\{ u_{ik} \in [0,1] \mid \sum_{i=1}^c u_{ik} = 1 \forall k \text{ and } \forall i \right\} \quad (1)$$

M. Y. Lee and K. C. Kwak are with the Department of Control and Instrumentation Engineering, Department of Electronics Engineering Chosun University, Gwangju, Korea ( e-mail: kwak@chosun.ac.kr).

The objective function for FCM clustering is a generalization of (2)

$$J = \sum_{i=1}^c \sum_{k=1}^N u_{ik}^m d_{ik}^2 \quad (2)$$

where  $d_{ik}$  is the Euclidean distance between  $i$ 'th cluster center and  $k$ 'th data point. In a batch-mode operation, FCM clustering algorithm determines the cluster centers and the membership matrix using the following steps;

Step 1. Initialize the membership matrix with random values between 0 and 1 such that the constraints in (1) are satisfied.

Step 2. Compute  $c$  fuzzy cluster centers using (3).

$$c_i = \frac{\sum_{k=1}^N u_{ik}^m \mathbf{x}_k}{\sum_{k=1}^N u_{ik}^m} \quad (3)$$

Step 3. Compute the objective function according to (2). Stop if either it is below a certain tolerance value or its improvement over previous iteration is below a certain threshold.

$$|J^p - J^{p-1}| \leq \varepsilon \quad (4)$$

where  $p$  is an iteration number.

Step 4. Compute a new  $\mathbf{U}$  using (5). Go to step 2.

$$u_{ik} = \frac{1}{\sum_{j=1}^c \left( \frac{\|\mathbf{x}_k - \mathbf{c}_i\|}{\|\mathbf{x}_k - \mathbf{c}_j\|} \right)^{\frac{2}{m-1}}} \quad (5)$$

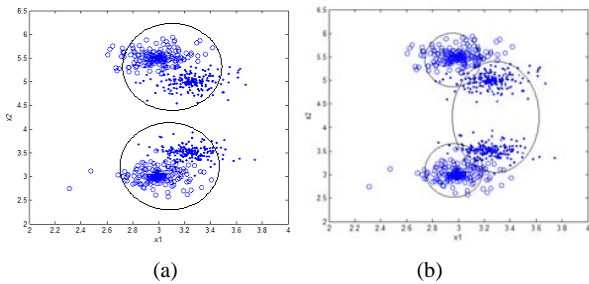


Fig. 1 Clusters obtained by FCM and CFCM clustering (a) context-free clustering (b) context-based fuzzy clustering

On the other hand, the Context-based Fuzzy C-Means (CFCM) clustering is an effective algorithm to generate the cluster centers preserving homogeneity on the basis of fuzzy granulation. In contrast to the FCM clustering, CFCM clustering algorithm is achieved with the aid of the linguistic contexts produced in output space. To compare with these two

clustering algorithms, let us briefly describe it for a simple training set  $(\mathbf{x}_k, z_k)$ ,  $\mathbf{x}_k = [x_{k1} x_{k2}]$ , as shown in Fig. 1.

The circle and asterisk of the data points represent the corresponding input data points assumed by the output variable. Fig. 1 (a) shows two cluster centers estimated by the FCM clustering algorithm. However, these clusters change as shown in Fig. 1 (b) when we use the corresponding output value. Fig. 1 (b) shows an extra cluster to preserve homogeneity between the clustered patterns from the linguistic contexts produced in the output space. We can recognize from Fig. 1 that the clusters obtained from context-based fuzzy clustering have the more homogeneity than those produced by context-free fuzzy clustering. The fuzzy set of context is defined as;

$$T : \mathbf{Z} \rightarrow [0,1] \quad (6)$$

where  $\mathbf{Z}$  is a universe of discourse of output variable. The membership matrix in CFCM clustering is expressed as;

$$u_{ik} = \frac{f_k}{\sum_{j=1}^c \left( \frac{\|\mathbf{x}_k - \mathbf{c}_i\|}{\|\mathbf{x}_k - \mathbf{c}_j\|} \right)^{\frac{2}{m-1}}} \quad (7)$$

where  $m \in [1, \infty)$  is a fuzzification factor. Here,  $f_k$  is the membership degree of the  $k$ 'th data in the certain context. For this reason, we can change the requirements of the membership matrix as;

$$\mathbf{U}(\mathbf{f}) = \left\{ u_{ik} \in [0,1] \mid \sum_{i=1}^c u_{ik} = f_k \forall k \text{ and } 0 < \sum_{k=1}^N u_{ik} < N \forall i \right\} \quad (8)$$

The contexts are generated through a series of triangle membership functions with equally spaced along the domain of an output variable. CFCM fuzzy clustering algorithm determines the clusters and the membership matrix using the following steps;

Step 1. Select the number of linguistic context and the number of cluster in each context, and initialize the membership matrix  $\mathbf{U}$  with random values between 0 and 1. For simplicity, we assume that the number of the cluster center in each context is equal.

Step 2. Generate the contexts with triangle membership function using equally partitioning method in the output space.

Step 3. Compute cluster centers in each context by using (3).

Step 4. Compute the objective function using (2). Stop if its improvement over previous iteration is below a certain threshold as follows

Step 5. Compute a new membership matrix using (5). Go to Step 3.

### III. IT2-CFCM CLUSTERING AND DESIGN OF ANFN

In this Section, the procedure of the T2-CFCM clustering is described. Fig. 2 shows T1 context with general fuzzy set and

IT2 context with lower and upper bound. Fig. 3 visualizes IT2 contexts produced in the same manner as in CFCM clustering. The estimation method of cluster center is similar to the procedure of CFCM clustering.

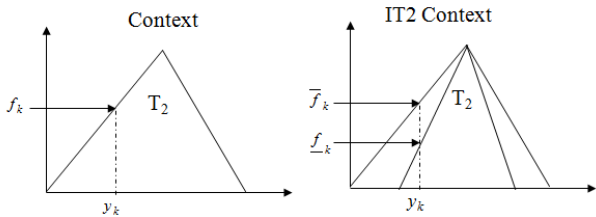


Fig. 2 T1 and IT2 context in the output space

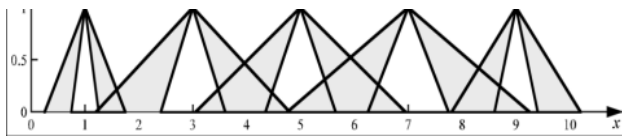


Fig. 3 IT2 contexts obtained in the output space

IT2-CFCM clustering is performed by the following steps:

- Step 1. Select the number of context and cluster per context. For simplicity, we assume that the number of cluster per context is equal. Initialize the membership matrix with random values between 0 and 1.
- Step 2. Generate IT2 contexts with triangular membership function using equally partitioning method in the output space. The upper bound is generated by a 1/2 overlap between successive fuzzy sets. The left and right lower bounds determine 1/3 value from the center of each triangular context, respectively.
- Step 3. Compute upper and lower partition matrices as:

$$\begin{aligned} \bar{u}_{ik} &= \max \left( \bar{f}_k / \sum_{j=1}^c \left( \frac{\|\mathbf{x}_k - \mathbf{c}_i\|}{\|\mathbf{x}_k - \mathbf{c}_j\|} \right)^{\frac{2}{m-1}}, f_k / \sum_{j=1}^c \left( \frac{\|\mathbf{x}_k - \mathbf{c}_i\|}{\|\mathbf{x}_k - \mathbf{c}_j\|} \right)^{\frac{2}{m-1}} \right) \\ u_{ik} &= \min \left( \bar{f}_k / \sum_{j=1}^c \left( \frac{\|\mathbf{x}_k - \mathbf{c}_i\|}{\|\mathbf{x}_k - \mathbf{c}_j\|} \right)^{\frac{2}{m-1}}, f_k / \sum_{j=1}^c \left( \frac{\|\mathbf{x}_k - \mathbf{c}_i\|}{\|\mathbf{x}_k - \mathbf{c}_j\|} \right)^{\frac{2}{m-1}} \right) \end{aligned} \quad (9)$$

We focus on uncertainty of  $f_k$  without considering the uncertainty of defuzzification factor  $m$  in this paper. The uncertainty of both  $f_k$  and  $m$  will be covered in the future research.

- Step 4. Update the cluster center. The individual values of the left and right cluster boundaries in each dimension can be computed by sorting the order of patterns in particular dimension and then applying Karnik-Mendel (KM) iterative procedure [7]. Here KM algorithm is used to update the interval set of cluster centers. The new cluster center is computed by a defuzzification method as:

$$c_i = \frac{\bar{c}_i + c_i}{2} \quad (10)$$

- Step 5. Compute the cost function. Stop if either it is below a certain tolerance value or its improvement over previous iteration is below a certain threshold.

- Step 6. Compute a new membership function based on average of lower and upper bound as type-reduce step (11). Go to Step 3.

$$u_{ik} = \frac{\bar{u}_{ik} + u_{ik}}{2} \quad (11)$$

Next, we describe the design method of the Adaptive Neuro-Fuzzy Network (ANFN) based on IT2-CFCM clustering. The Tagaki-Sugeno-Kang (TSK) fuzzy types are by far the most popular method for data-driven fuzzy modeling. Fig. 4 shows the architecture of the proposed ANFN. As shown in Fig. 4, this network consists of five layers. While ANFIS with grid partition often encounter ‘‘curse of dimensionality’’ problem that the number of fuzzy rules exponentially increases due to the grid partitioning of the input space, the proposed ANFN can solve such a problem due to the flexible scatter partitioning of IT2-CFCM clustering.

A typical fuzzy rule in this model has the following form with first-order linear function.

$$\begin{aligned} R^i : & \text{If } x_1 \text{ is } A_1^i \text{ and } x_2 \text{ is } A_2^i, \dots, x_m \text{ is } A_m^i \\ & \text{then } y^i = a_0^i + a_1^i x_1 + a_2^i x_2 + \dots + a_m^i x_m \end{aligned} \quad (12)$$

where  $R^i, i=1,2,\dots,r$ , denotes  $i$ 'th fuzzy rule. ‘‘ $A_j^i$ ’’ is that fuzzy sets in the antecedent of  $i$ 'th rule and  $a_m^i$  is consequent parameters. As shown in Fig. 4, the linguistic labels in the first layer are constructed by Gaussian membership function. Each of the cluster centers generated by IT2-CFCM clustering represents a prototype that exhibits certain characteristics of the system to be modeled. The final inference output of the ANFN is computed as the following weighted average method,

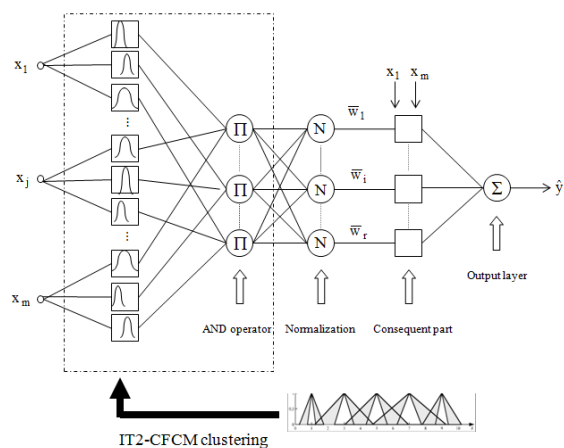


Fig. 4 Architecture of the proposed ANFN

$$\hat{y} = \sum_{i=1}^r \bar{w}_i y^i = \frac{\sum_{i=1}^r w_i (a_0^i + a_1^i x_1 + \dots + a_m^i x_m)}{\sum_{i=1}^r w_i} \quad (13)$$

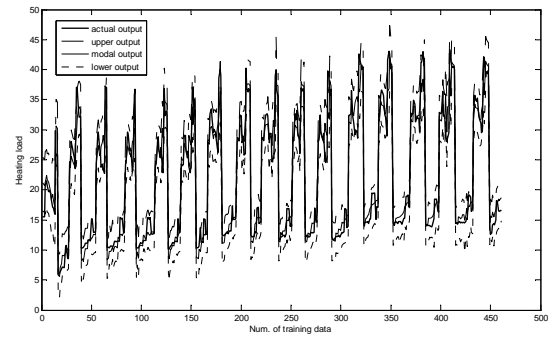
where  $\bar{w}_i$  is a normalized firing strength of  $i$ 'th rule. These values are obtained by the ratio of  $i$ 'th rule's firing strength to the sum of all rule's firing strengths. The learning scheme of the proposed ANFN is realized by hybrid scheme based on Back-Propagation (BP) and Least Square Estimate (LSE) [12].

#### IV. EXPERIMENTAL RESULTS

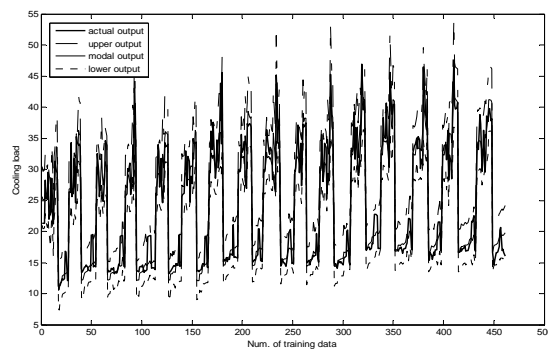
This section performs the experiments for predicting the heating load and cooling load requirements of residential buildings (that is, energy efficiency) as a function of building parameters [13]. The energy analysis is performed by using 12 different building shapes simulated in Ecotect. The buildings differ with respect to the glazing area, the glazing area distribution, the orientation, relative compactness, surface area, roof area, and overall height. The experiments are simulated with various settings as functions of the afore-mentioned characteristics to obtain 768 building shapes.

The dataset comprises 768 samples and 8 input variables, aiming to predict two real valued responses. The training and test data set are randomly selected by 60%-40% split in the normalized space between 0 and 1, respectively. The experiment is performed by 10 iterations. The training data set is used for model construction, while the test set is used for model validation. Fig. 5 shows the prediction performance for heating and cooling load outputs with upper and lower bounds representing uncertainty of the network output. Table I lists the mean of RMSE (root mean square error) results regarding approximation and generalization capability, respectively. In the design of LM, we used six contexts and six clusters in each context for IT2-CFCM as the best results. Although the LM has a structured knowledge representation in the form of fuzzy if-then rules, it lacked the adaptability to deal with nonlinear model. Moreover, we constructed the RBFN-CFCM based on six contexts and six clusters in the same manner. As listed in Table I, the experimental results revealed that the proposed networks yielded a better performance in comparison with the previous works.

Method	Training data	Testing data
RBFN-CFCM [10] (36 nodes)	2.767	3.106
LM[11] (p=c=6)	4.084	4.388
Proposed method (36 rules)	2.201	2.563



(a)



(b)

Fig. 5 Prediction performance for heating and cooling loads (a) heating load (b) cooling load

#### V. CONCLUSIONS

We developed ANFN based on IT2-CFCM clustering with the aid of fuzzy granulation. The proposed clustering algorithm includes the characteristics with respect to uncertainty of contexts and the homogeneity between the clustered patterns from the IT2 contexts produced in the output space. For further future research, we will concentrate on the uncertainty of both defuzzification factor and context and develop the general IT2-CFCM clustering algorithm. We demonstrated the effectiveness and superiority of the proposed network through energy efficiency prediction example.

#### ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (NRF-2013R1A1A2012127).

#### REFERENCES

- [1] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning-1", *Information Sciences*, Vol. 8, pp.199-249, 1971.
- [2] N. N. Karnik, J. M. Mendel, *An Introduction to Type-2 Fuzzy Logic Systems*, Univ. of Southern California, Los Angeles, CA, June, 1998.
- [3] J. M. Mendel, *Uncertain Rule-Based Fuzzy Logic Systems*, Prentice Hall, 2001.
- [4] C. Hwang, F. C. H. Rhee, "Uncertain fuzzy clustering: Interval type-2 fuzzy approach to c-means", *IEEE Trans. on Fuzzy Systems*, Vol. 15, No. 1, pp. 107-120, 2007.

- [5] O. Linda, M. Manic, "General type-2 fuzzy c-means algorithm for uncertain fuzzy clustering", *IEEE Trans. on Fuzzy Systems*, Vol. 20, No. 5, pp. 883-897, 2012.
- [6] C. Qiu, J. Xiao, L. Han, M. N. Iqbal, "Enhanced interval type-2 fuzzy c-means algorithm with improved initial center", *Pattern Recognition Letter*, Vol. 38, pp. 86-92, 2014.
- [7] N. N. Karnik, J. M. Mendel, "Centroid of a type-2 fuzzy set", *Information Sciences*, Vol. 132, No. 1, pp.195-220, 2001.
- [8] L. Yao, K. S. Weng, "On a type-2 fuzzy clustering algorithm", *The Fourth International Conference on Advanced Cognitive Technologies and Applications*, pp. 45-50, 2012.
- [9] W. Pedrycz, "Conditional fuzzy c-means", *Pattern Recognition Letter*, Vol.17, pp.625-632, 1996.
- [10] 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.
- [11] W. Pedrycz and A. V. Vasilakos, "Linguistic models and linguistic modeling", *IEEE Trans. on Systems, Man, and Cybernetics-Part C*, Vol. 29, No.6, pp.745-757, 1999.
- [12] K. C. Kwak, D. H. Kim, "Adaptive neuro-fuzzy networks with the aid of fuzzy granulation", *IEICE Trans. Information & Systems*, Vol. E88D, No. 9, pp. 2189-2196, 2005.
- [13] A. Tsanas, A. Xifara, "Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools", *Energy and Buildings*, Vol. 49, pp. 560-567, 2012.

**Myung-Won Lee** received the B.Sc. and M.Sc. degrees from Chosun University, Gwangju, Korea, in 2010 and 2012, respectively. He is currently studying for Ph.D. degree. His interest areas are computational intelligence, motion capture and analysis, and human-robot interaction.

**Keun-Chang Kwak** (M'04) received the B.Sc., M.Sc., and Ph.D. degrees from Chungbuk National University, Cheongju, Korea, in 1996, 1998, and 2002, respectively. During 2003–2005, he was a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada. From 2005 to 2007, he was a Senior Researcher with the Human–Robot Interaction Team, Intelligent Robot Division, Electronics and Telecommunications Research Institute, Daejeon, Korea. He is currently the Associative Professor with the Department of Control & Instrumentation, Engineering and Department of Electronics Engineering, Chosun University, Gwangju, Korea. His research interests include computational intelligence, human–robot interaction, and pattern recognition. Dr. Kwak is a member of IEEE, IEICE, KFIS, KRS, ICROS, KIPS, and IEEK.