

# 타당성 척도를 내제한 머조리티 보팅 FCM

## Majority-Voting FCM with Implied Validity Measure

이강화 · 이동일 · 이석규

Gang-Hwa Lee, Dong-Il Lee and Suk-Gyu Lee

영남대학교 전자정보공학부

### 요 약

퍼지 클러스터링은 전통적인 FCM을 이용하여 수행될 수 있다. 그러나 FCM을 사용하는데 있어서 고려해야 할 문제점은 크게 두가지이다. 첫째는 FCM은 초기 멤버십 함수의 지정에 민감하고, 둘째는 FCM 알고리즘의 수행에는 클러스터의 개수를 미리 지정 해 주어야 한다는 것이다. 따라서 FCM과 동시에 수행하여야 할 과정은 클러스터의 개수를 찾기 위하여 타당성 척도를 이용한 시물레이션을 수행하여야 한다. 본 논문에서는 위의 두 가지 문제점을 동시에 해결 할 수 있는 FCM 알고리즘을 제시하고, 그 적용성을 시물레이션을 통하여 검증한다. 본 논문의 기여도는 MV-FCM 알고리즘의 제시와 이 알고리즘의 타당성 척도로써의 효율성이다.

### ABSTRACT

It is well known that FCM is an indispensable tool for fuzzy clustering. The problems of using FCM are 1) it is sensitive to the initial random membership functions and 2) FCM inherently requires the number of clusters. Hence we need to run FCM algorithms with an appropriate validity measure until we find a suitable number of clusters. In this paper, we suggest the 'Majority-Voting FCM with implied validity measure.' With this algorithm, we can solve the aforementioned problems. The working simulation results are provided. The contributions are 1) MV-FCM algorithm and 2) its definitive capability of being an excellent validity measure.

**Key Words** : FCM, MV-FCM, 머조리티 보팅, 타당성 척도

### 1. 서 론

퍼지 문제에 있어서의 필수적인 요소는 입력 데이터들의 공간상의 분포를 효과적으로 구분하는 것이다. 일반적으로 입력 대 출력의 사상관계는 명확하게 기본 함수 꼴들의 조합으로 표현되지 않으므로 입력 공간의 자의적인 분할과 퍼지 추론이 필요하게 된다. 이러한 퍼지 클러스터링을 위하여 FCM(Fuzzy C-Means) 알고리즘이 널리 사용되고 있다 [1-4]. FCM 알고리즘은 퍼지 범함수 (functional)를 최소화 하는 클러스터의 소속도 함수를 구하는 것이다. 그러나 FCM을 이용하여 클러스터링을 하면 구하여진 클러스터들은 소속도 함수의 초기 랜덤화에 많은 영향을 받게 된다.

FCM 알고리즘을 사용하려면 우선 클러스터의 개수를 미리 정해 놓아야 한다. 따라서 기존의 FCM 알고리즘 자체로는 적절한 클러스터의 개수를 찾아 낼 수 없으므로 통상 타당성 척도(validity measure)를 정의 한 후 구해지는 최적 척도 값으로부터 클러스터의 개수를 구한다 [4, 5]. 즉, FCM과 별개로 타당성에 대한 시물레이션이 필요하게 되는 것이다.

본 논문에서는 위에 언급한 FCM의 두가지 문제점 [가) FCM의 초기 조건 중속성, 나) 적절한 타당성 척도의 필요

성]을 한꺼번에 해결 할 수 있는 FCM 알고리즘을 제시하고 알고리즘의 효율성을 시물레이션을 통하여 검증하고자 한다.

제안된 MV-FCM (Majority-Voting FCM) [6]의 기본 아이디어는 가)의 문제점을 해결하기 위하여 복수의 초기화를 사용한 FCM의 병렬 수행, 나)의 문제점을 해결하기 위한 타당성 척도로써 1-norm을 이용한 최대 클러스터 거리를 사용하는 것이다. MV-FCM은 복수의 병렬 FCM 시물레이션 중에 위의 언급된 1-norm들을 선택적으로 치환하여 궁극적으로는 (예를 들면) 세 개의 병렬 FCM이 각각 구할 클러스터의 중심을 일치시키게 만드는 것이다. 이때 사용된 1-norm 스와핑(swapping) 방법은 기본적으로 머조리티 보팅(majority-voting) 방식이다.

이 MV-FCM을 이용하면 클러스터의 개수를 효과적으로 찾아 낼 수가 있다. 기본 아이디어는 FCM이 초기화 조건에 따라서 다른 클러스터의 모양을 찾아내게 되지만, 궁극적으로 적절한 클러스터의 개수가 시물레이션 초기에 주어진다면 복수의 초기화에 의한 클러스터링 결과는 거의 일치하여야 한다는 것이다. 따라서 복수의 각각 다른 초기화에도 불구하고 MV-FCM 수행 결과 클러스터 중심의 요동이 가장 작은 클러스터의 개수가 최적화 된 클러스터가 된다. 결과적으로 타당성 척도를 사용하여 FCM을 수행 한 후 그 척도의 최적성을 일일이 판별 할 필요성이 없는 FCM 알고리즘을 제공하게 된다.

접수일자 : 2002년 5월 2일  
완료일자 : 2002년 11월 6일

## 2. Fuzzy c-means 알고리즘

FCM 알고리즘은 잘 알려져 있으며 이하 제시하는 알고리즘들은 Babuskal[4]의 용어와 프로세싱에 따른다. 아래의 알고리즘에서 Z는 데이터 집합, N은 데이터 개수, V는 클러스터,  $\mu$ 는 소속도 함수, c는 클러스터의 개수, m은 가중치 지수,  $\epsilon$ 는 시뮬레이션 종료할 위한 매 스텝의 소속도 함수 수렴치의 상한, 그리고 A는 norm-구성 행렬이다. 초기의 랜덤화된 소속도 함수로부터 각 시뮬레이션 스텝마다 다음의 프로세스를 반복 수행한다.

### 2.1 FCM 알고리즘 [4] :

Repeat for  $l = 1, 2, 3, \dots$

스텝 1 ; 클러스터 중심의 계산

$$v_i^{(l)} = \frac{\sum_{k=1}^N (\mu_{ik}^{(l-1)})^m z_k}{\sum_{k=1}^N (\mu_{ik}^{(l-1)})^m}, \quad 1 \leq i \leq c. \quad (1)$$

스텝 2 ; 데이터와 클러스터 중심간의 거리 계산

$$D_{ikA}^2 = (z_k - v_i^{(l)})^T A (z_k - v_i^{(l)}), \quad (2)$$

스텝 3 ; 파티션(partition) 행렬의 갱신

가) If  $D_{ikA} > 0$  for  $1 \leq i \leq c, 1 \leq k \leq N$ ,

$$\mu_{ik}^{(l)} = \frac{1}{\sum_{j=1}^c (D_{ikA} / D_{jkA})^{2/(m-1)}}, \quad (3)$$

나) otherwise

$$\mu_{ik}^{(l)} = 0 \quad (4)$$

if  $D_{ikA} > 0$  with  $\sum_{i=1}^c \mu_{ik}^{(l)} = 1.$

스텝 4 ; 종료 조건

$$\|U^{(l)} - U^{(l-1)}\| < \epsilon, \quad (5)$$

### 2.2 FCM의 초기화 종속성

기존 FCM의 소속도 함수 초기화 종속성을 살펴보기 위하여 각각 서로 다른 초기화를 사용하여 기존의 FCM을 30회 반복 수행하였다. 시뮬레이션에 사용된 입력 데이터는 잘 알려진 나비 모양(butterfly)의 데이터이다. 아래의 그림 1에서 'x' 표시는 나비 모양 입력 데이터의 위치를, '+' 표시는 구해진 클러스터의 중심을 나타낸다. 시뮬레이션 파라미터는 FCM의 시작 초기에 주어지는 클러스터의 수이다. FCM을 수행시킬 때 클러스터의 수가 두 개(Nc=2)이면 30회 반복의 FCM 결과는 거의 일치한다. 반면 그림 1에서와 같이 클러스터의 수가 3개(Nc=3)이면 구해진 중심들이 옮겨 다니는 것을 알 수가 있다. 위의 시뮬레이션 결과 클러스터의 수가 적절하지 못하면 FCM으로 구한 클러스터의 중심은 소속도 함수의 초기화에 강하게 종속됨을 알 수가 있다.

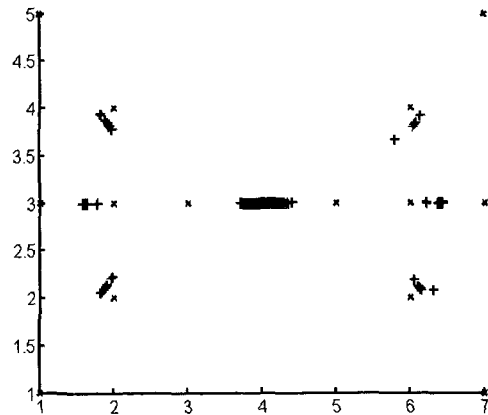


그림 1. 30회 반복 수행 FCM 결과 (Nc=3)  
Fig. 1. Results of 30 FCM (Nc=3)

## 3. 머조리티 보팅 FCM (MV-FCM)

### 3.1 MV-FCM [6]

MV-FCM은 기존의 FCM 알고리즘을 크게 세가지로 변형한 것이다. 그 변형은 1) 각 FCM 스텝마다 세 조의 각각 다른 소속도 함수의 초기화로부터 병렬로 클러스터의 중심  $V_1, V_2, V_3$ 를 계산 (FCM의 스텝 1 변형), 2) 이때 구해진 세 조의 클러스터 중심들을 가지고 그 세 조의 병렬 클러스터 중심 중 1-norm이 가장 작은 조의 것을 나머지 두 조의 평균 값으로 치환 (이 스와핑 전략은 세 조의 병렬 FCM에서 구해진 클러스터 중심들 간의 최대 거리(maximum inter-cluster distance)를 보장 함), 그리고 3) 시뮬레이션의 종료 조건을 조정하는 것이다. 따라서 MV-FCM에서 사용된 1-norm은 병렬 FCM 간의 스위칭 전략으로 사용되는 동시에 최대 클러스터 간 거리를 제공하게 된다.

아래에 FCM의 스텝 1과 스텝 4의 변형을 상술한다.

스텝 1 ;

가) 세 조의 클러스터 중심들을 계산

$$v_{1i}^{(l)}, v_{2i}^{(l)}, v_{3i}^{(l)}. \quad (6)$$

나) 스와핑

$$\text{if } \|v_{1i}\| < \|v_{2i}\|, v_{1i} = \frac{v_{2i} + v_{3i}}{2}; \quad (7)$$

$$\text{elseif } \|v_{3i}\| < \|v_{2i}\|, v_{3i} = \frac{v_{1i} + v_{2i}}{2}; \quad (8)$$

$$\text{else } v_{2i} = \frac{v_{1i} + v_{3i}}{2}; \quad (9)$$

이때  $\|\cdot\|$ 은 어떤 norm도 허용되나 시뮬레이션에서는 1-norm을 사용하였다.

스텝 4 ; 시뮬레이션 종료 조건 변형

$$\max(Abs1Err, Abs2Err, Abs3Err) < \epsilon . \quad (10)$$

### 3.2 MV-FCM 시뮬레이션

다음은 세 조의 각각 다른 소속도 함수의 초기화에 따른 MV-FCM의 시뮬레이션 결과를 나타내는 것이다. 그림 2는 앞의 나비 구조 데이터로 클러스터 개수를 세 개로 하여 MV-FCM을 수행했을 경우 그림 1의 클러스터 중심을 향하여 세 조(◇, ○, +)의 클러스터 중심들이 수렴하는 궤적을 보여 주고 있다. MV-FCM에 사용된 세 조의 서로 다른 초기화에 따른 클러스터 중심들이 스와핑에 의하여 단일 중심들로 수렴되고 있다.

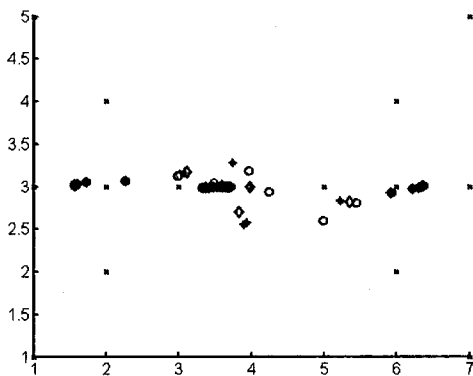


그림 2. MV-FCM에서 세 조의 FCM 중심이 수렴하는 모양 (세 개의 중심을 보임)

Fig. 2. MV-FCM convergence trajectory of three parallel FCM (showing all three centers)

### 4.2 타당성 척도 시뮬레이션

그림 3과 4는 3-클러스터의 입력 데이터를 사용하여 MV-FCM을 5회 반복하였을 경우의 클러스터 중심들의 요동 상황을 보여주고 있다. 이때 그림 3은 클러스터의 개수를 3으로, 그림 4에서는 4로 선택 한 경우이다. 클러스터의 수가 3인 경우는 5회 반복에도 불구하고 구해진 클러스터의 중심이 거의 변동이 없지만 클러스터의 수가 4가 되면 MV-FCM은 주어진 소속도 함수의 초기화 조건에 따라서 불명확한(ambiguous) 클러스터링을 나타내게 되고, 따라서 MV-FCM 자체가 타당성 척도의 좋은 수단이 됨을 알 수가 있다.

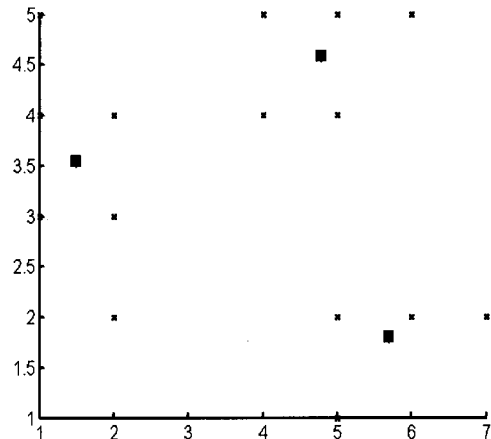


그림 3. MV-FCM을 5회 반복한 후의 클러스터 중심의 요동. (3-클러스터 데이터, Nc=3)

Fig. 3. The variation of cluster centers after 5 MV-FCM. (Three-cluster data, Nc=3 given)

## 4. MV-FCM과 타당성 척도

### 4.1 타당성 척도

FCM으로 구해진 퍼지 클러스터링의 타당성을 확인하기 위하여 여러 가지의 모델들이 제시되었으나 [4, 5], 공통적으로 이들 수학적 모델들은 클러스터의 개수가 증가함에 따라 그 값이 단조 감소하는 경향을 지닌다. 또한 구하여진 타당성 척도는 사용하는 척도의 형태에 따라서 구해진 클러스터의 개수가 반드시 같은 값을 나타내는 것은 아니다. 뿐만 아니라 최저(혹은 최고)점의 선택성이 나쁜 경우가 종종 있다.

반면 MV-FCM에 의한 시뮬레이션은, 항상 병렬의 소속도 함수 초기화를 사용하여 이터레이션(iteration) 도중에 서로 스와핑을 하기 때문에, 주어진 클러스터의 개수에 따른 최적의 클러스터 분류를 제공한다. 반면 주어진 데이터의 구조에 따른 클러스터의 개수가 잘못 주어지면 반복적인 MV-FCM 시뮬레이션에서 클러스터의 위치가 계속해서 변화하게 된다. 따라서 세 조의 클러스터 중심들이 미리 지정된 중심 수렴 반경까지 수렴되는 이터레이션 회수를 사용하면 MV-FCM의 타당성 척도로서의 유용성이 증명되어 질 수가 있다.

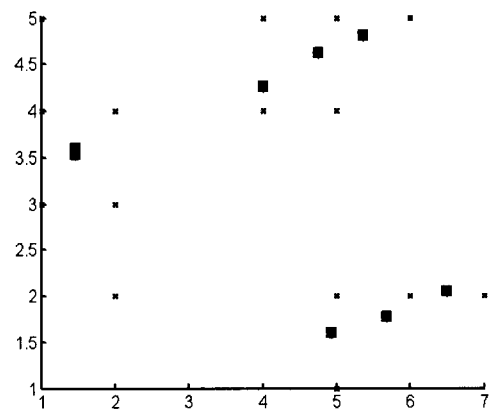


그림 4. MV-FCM을 5회 반복한 후의 클러스터 중심의 요동. (3-클러스터 데이터, Nc=4)

Fig. 4. The variation of cluster centers after 5 MV-FCM. (Three-cluster data, Nc=4 given)

아래의 그림 5와 6은 다섯 개의 클러스터를 가지는 입력 데이터를 사용한 시뮬레이션 결과를 보여주고 있다. 시뮬레이션 조건은 MV-FCM을 3회 반복 한 것이다. 그림 5는 클러스트의 개수를 정확히 5개로, 그림 6은 6개로 주었을 경우의 클러스터링 결과를 보여주고 있다.

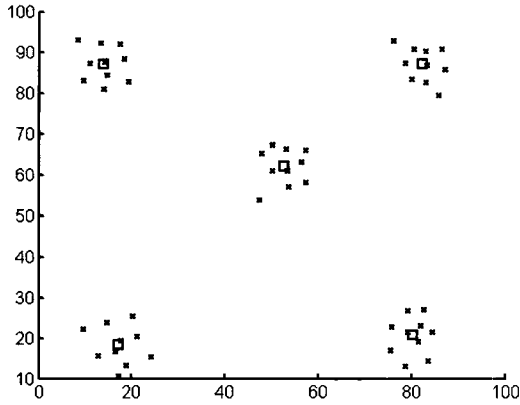


그림 5. MV-FCM을 3회 반복한 후의 클러스트 중심의 요동. (5-클러스트 데이터, Nc=5)  
 Fig. 5. The variation of cluster centers after 3 MV-FCM. (Five-cluster data, Nc=5 given)

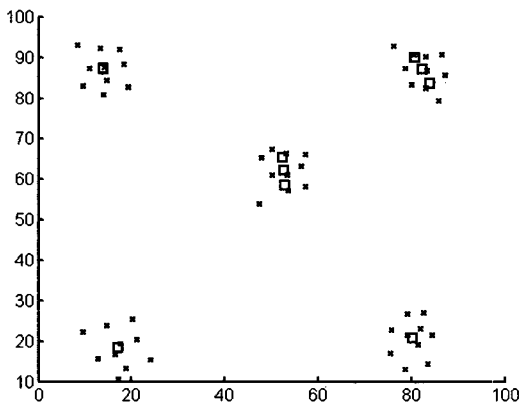


그림 6. MV-FCM을 3회 반복한 후의 클러스트 중심의 요동. (5-클러스트 데이터, Nc=6)  
 Fig. 6. The variation of cluster centers after 3 MV-FCM. (Five-cluster data, Nc=6 given)

다음의 표 1에 위의 두 가지 클러스터 데이터를 사용한 MV-FCM 시뮬레이션에서 클러스터 중심 수렴 속도를 타당성 척도로 사용한 결과를 요약하였다. 각 클러스터 입력 구조에 대한 양호한 선택성을 잘 보여주고 있다.

표 1. 3-, 5-클러스터 데이터에서의 MV-FCM 수렴 속도 (각 5 샘플)

Table 1. The MV-FCM convergence speed with 3-, 5-cluster data. (5 samples each)

# of clusters	3-cluster data					5-cluster data				
	#1	#2	#3	#4	#5	#1	#2	#3	#4	#5
Nc=2	21	22	25	26	18	100	93	53	100	43
Nc=3	13	13	14	15	15	100	100	100	100	32
Nc=4	25	25	31	30	24	20	17	23	20	18
Nc=5	32	32	32	27	50	14	11	12	11	11
Nc=6	-	-	-	-	-	37	31	36	26	42
Nc=7	-	-	-	-	-	37	37	100	59	25
Nc=8	-	-	-	-	-	44	34	27	36	100

### 4.3 Xie-Beny[5] 데이터에서의 비교 분석

기존의 사용 중인 타당성 척도의 특성과 제안된 MV-FCM의 타당성 척도 능력을 비교하기 위하여 논문[5]의 데이터 구조를 사용하여 시뮬레이션을 하였다. 그 입력 데이터는 그림 7에 'x' 표시로 그려진 '북극성 근처의 밝은 51개 별자리'이다. 클러스트의 개수를 Nc=8로 하여 MV-FCM을 수행한 클러스터 중심을 'o'로 같은 그림에 표기하였다. 그림 8은 Nc=9인 경우이다.

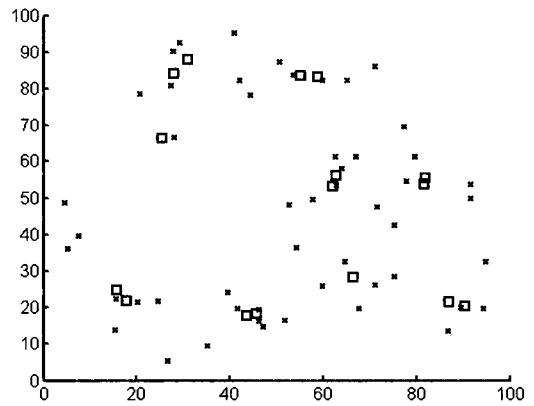


그림 7. MV-FCM을 3회 반복한 후의 클러스트 중심의 요동. (Xie-Beny 데이터, Nc=8)  
 Fig. 7. The variation of cluster centers after 3 MV-FCM. (Xie-Beny data, Nc=8 given)

논문 [5]에서는 '밀집, 분리형 (compactness and separation)' 타당성 척도 s를 제안하여 사용하였으며,

$$s = \frac{\sum_{i=1}^n \sum_{j=1}^n \mu_{ij}^2 \|V_i - z_j\|^2}{n \cdot \min_{ij} \|V_i - V_j\|^2} \quad (11)$$

로 정의하였다. 이때 n은 총 데이터 수이다.

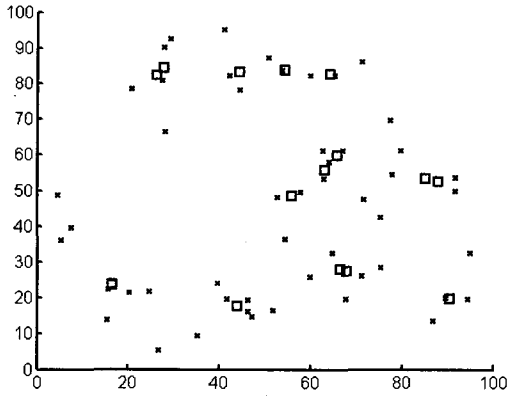


그림 8. MV-FCM을 3회 반복한 후의 클러스터 중심의 변동. (Xie-Beny 데이터, Nc=9)

Fig. 8. The variation of cluster centers after 3 MV-FCM. (Xie-Beny data, Nc=9 given)

아래의 표 2에는 (11)의 타당성 척도를 사용하여 계산한 s 값들과 MV-FCM의 수렴속도를 사용한 타당성 척도의 값을 비교하였다. 표 2의 첫 번째 열은 시뮬레이션 파라미터인 클러스터의 개수 Nc, 두 번째 열은 MV-FCM으로 세 조의 클러스터 중심을 수렴시키는 반복횟수 그리고 마지막 열은 논문 [5]의 s 타당성 척도에 의하여 계산된 값이다.

표 2. MV-FCM과 Xie-Beni 타당성 척도 비교 (Xie-Beni[5] 데이터 사용)

Table 2. The comparison of MV-FCM and s-validity measure. (Xie-Beni[5] data)

# of clusters	MV-FCM 수렴 속도	Xie-Beni s-Validity
Nc=3	32	0.12
Nc=4	53	0.18
Nc=5	39	0.14
Nc=6	71	0.21
Nc=7	59	0.19
Nc=8	51	0.17
Nc=9	51	0.17
Nc=10	61	0.18
Nc=11	63	0.22
Nc=12	76	0.21

표 2의 결과를 보면, s에 의한 적정 클러스터의 수는 Nc=3, 5, 8, 9, 4의 순서가 적당하나 그 선택성은 양호하지 못한 편이다. 반면에 MV-FCM의 결과는 Xie-Beni의 s 척도와 동일한 결과를 나타내지만 특히 Nc=7, 8, 9, 10에서 상당히 양호한 선택성을 나타내고 있다.

이상의 여러 가지 입력 데이터 구조를 통한 시뮬레이션과 논문 [5]와의 결과를 비교 분석해 보면, 제시한 MV-FCM은

1) 주어진 병렬 초기화 조건에 대하여 최적의 클러스터를 제공하는 효율적인 알고리즘이고, 2) 이 논문의 가장 큰 기여인 타당성 척도로써의 역할을 동시에 만족시킬 수 있다는 것이다.

### 5. 결론

FCM은 그 수렴 특성 [7, 8]이 잘 알려져 있으나 클러스터의 개수를 정확하게 찾아 낼 수는 없다. 그리고 클러스터의 개수가 적절하게 주어지지 못하면 FCM은 소속도 함수의 초기 랜덤화에 강한 종속성을 보이게 된다.

반면 MV-FCM은 복수의 병렬 랜덤화를 사용하여 주어진 클러스터의 개수에 따른 최선의 클러스터링을 제공한다. 세 조의 FCM을 각 이터레이션마다 머조리티 보팅의 스와핑을 수행시킴으로써 MV-FCM을 수렴시키는 동시에 MV-FCM을 유용한 타당성 척도로써 사용할 수가 있게 한다.

제시한 주장을 검증하기 위하여 수행한 시뮬레이션 결과 MV-FCM의 유용성과 특히 타당성 척도로써의 능력을 확연히 보일 수가 있었다.

본 시뮬레이션에 사용된 1-norm 이외의 norm을 사용한 고찰과 그 분석 모델, MV-FCM의 수렴 특성에 대한 수학적 모델, 그리고 MV-FCM의 유용성 및 그 한계 등에 대한 고찰은 향후 과제로 남긴다.

### 참고 문헌

- [1] Ruspini, "Numerical Methods for Fuzzy Clustering," *Information Science*, vol. 2, pp. 319-350, 1970.
- [2] J. Bezdek, *Pattern Recognition with Fuzzy Objective Function*, Plenum Press, New York, 1981.
- [3] I. Gath and A. Geva, "Unsupervised optimal fuzzy clustering," *IEEE Trans. PAMI*, vol. 7, pp. 773-781, 1989.
- [4] R. Babuska, *Fuzzy Modeling for Control*, Kluwer Academic Publishers, Boston, 1998.
- [5] X. L. Xie and G. Beni, "A Validity Measure for Fuzzy Clustering," *IEEE Trans. PAMI*, vol. 13, no. 8, pp. 841-847, 1991.
- [6] GangHwa Lee, YoonChul Lee, SoonHak Kwon and SukGyu Lee, "Majority-Voting FCM Algorithm in the Vague Fuzzy Classification," *The 10th IEEE International Fuzzy Systems Conference*, vol. 2, 2001.
- [7] J. C. Bezdek, "A Convergence Theorem for the Fuzzy ISODATA Clustering Algorithms," *IEEE Trans. PAMI*, vol. 2, no. 1, pp. 1-8, January, 1980.
- [8] M. J. Sabin, "Convergence and Consistency of Fuzzy c-means/ISODATA Algorithms," *IEEE Trans. PAMI*, vol. 9, no. 5, pp. 661-668, Sept., 1987.

저 자 소개



**이강화 (Gang-Hwa Lee)**

1981년 서울대 전기공학과 졸업  
1984년 서울대 대학원 전기공 석사  
(84)-89 금성사 중앙연구소 선임  
1994년 Oklahoma S. 전기공 석사  
1997년 UT Arlington 전기공 박사수료  
1997년 National Semiconductor co-op  
2000년~현 영남대 대학원 전기공 박사  
수료. 대구대 정보통신공 겸임교수

관심분야 : 시스템, 제어, VLSI, 통신



**이동일 (Dong-II Lee)**

2001년 영남대 전기전자공학부 졸업  
2001년~현재 영남대 대학원  
전기공학과 석사과정

관심분야 : 제어 및 시스템, 이동 로봇,  
정보 가전, 내장형 시스템



**이석규 (Suk-Gyu Lee)**

1979년 서울대 전기공학과 졸업  
1981년 서울대 전기공학과 석사  
1990년 UCLA 전기공학과 박사  
1982년~현재 영남대학교  
전자정보공학부 교수

관심분야 : 퍼지 뉴럴 이론, 로봇 공학, 제어 시스템

Phone : 053-810-2487  
Fax : 053-813-8230  
E-mail : sglee@yu.ac.kr