

Interval Type-2 퍼지접근방법을 이용한 Possibilistic Fuzzy c-Means Clustering Algorithm

Interval Type-2 Fuzzy Approach to Possibilistic Fuzzy c-Means Clustering Algorithm

저자 (Authors)	이상수, 신기순, 이정훈 Sang-Soo Lee, Ki-Soon Shin, Chung-Hoon Rhee
출처 (Source)	한국지능시스템학회 학술발표 논문집 22(2) , 2012.11, 185-188 (4 pages)
발행처 (Publisher)	한국지능시스템학회 Korean Institute of Intelligent Systems
URL	http://www.dbpia.co.kr/Article/NODE02011633
APA Style	이상수, 신기순, 이정훈 (2012). Interval Type-2 퍼지접근방법을 이용한 Possibilistic Fuzzy c-Means Clustering Algorithm. 한국지능시스템학회 학술발표 논문집, 22(2), 185-188.
이용정보 (Accessed)	신라대학교 61.100.225.*** 2019/03/20 01:21 (KST)

저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독 계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

Interval Type-2 퍼지접근방법을 이용한 Possibilistic Fuzzy c-Means Clustering Algorithm

Interval Type-2 Fuzzy Approach to Possibilistic Fuzzy c-Means Clustering Algorithm

이상수 · 신기순 · 이정훈

Sang-Soo Lee, Ki-Soon Shin and Frank Chung-Hoon Rhee

한양대학교 전자통신공학과

요 약

Possibilistic Fuzzy C-Means(PFCM) 알고리즘은 Fuzzy C-Means(FCM) 알고리즘의 단점인 노이즈 민감성 및 특이점 문제와 Possibilistic C-Means(PCM) 알고리즘의 초기 클러스터의 prototype에 따라 위치가 겹치는 문제를 해결하기 위해 제안되었다. 하지만 이 방법 역시 퍼지화 파라미터 값에 따라 위와 같은 문제를 여전히 가지고 있다. 본 논문에서는 이와 같은 단점을 보완하기 위해 PFCM 알고리즘에 Interval Type-2 퍼지 접근방법을 적용하는 Interval Type-2 Possibilistic Fuzzy C-Means(IT2PFCM) 알고리즘을 제안 하였다. 실험을 통해 제안하는 알고리즘과 기존의 PFCM 알고리즘의 성능을 비교 결과 인식률이 향상 되었다.

키워드 : PFCM, Type-2 Fuzzy Sets, Interval Type-2 Fuzzy Sets, Fuzzy Clustering.

1. 서 론

Fuzzy clustering 에 일반적으로 많이 사용되는 알고리즘은 prototype 과의 유클리디안 거리를 이용해 멤버쉽을 할당하는 Fuzzy C-Mean(FCM) 알고리즘이다[1]. 하지만 FCM 알고리즘은 모든 클러스터의 멤버쉽 합이 항상 1이 되어야 한다는 조건 때문에 잘못된 클러스터링을 하거나 노이즈가 있는 데이터에서는 좋지 않은 결과를 보인다. 이러한 FCM 의 단점을 극복하기 위해 typicality 을 사용하는 Possibilistic C-Means(PCM) 알고리즘이 제안되었다[2]. PCM 은 잡음에는 강하지만 초기 파라미터 값에 민감하고, 상대적으로 가까이에 위치한 중심 값 들을 형성하는 패턴들의 경우에는 최종 중심 값의 위치가 겹치는 결과가 나올 수 있다는 단점이 있다. 이러한 문제점을 해결하기 위해 FCM 과 PCM 을 가중치 합으로 표현한 Possibilistic Fuzzy C-Means (PFCM) 알고리즘이 연구 되었다[3]. 하지만 이 역시 퍼지화 파라미터 값에 따라 여전히 상기의 문제를 가지고 있다. 그래서 본 논문에서는 PFCM 의 퍼지화 파라미터에 uncertainty를 부여하여 성능을 개선하는 Interval Type-2 Possibilistic Fuzzy C-Means(IT2PFCM) 알고리즘을 제안 하였다. 본 논문은 다음과 같이 구성된다. 두 번째 절과 세 번째 절에서는 PFCM 알고리즘과 Interval Type-2 Fuzzy Set 에 대해 설명하고, 네 번째 절에서는 본 논문에서 제안하는 IT2PFCM 알고리즘에 대해 소개한다. 다섯 번째 절에서는 제안하는 Interval Type-2 PFCM 알고리즘과 기존의 알고리즘들 간의 성능을 비교

해보고 마지막으로 여섯 번째 절에서 결론을 맺겠다.

2. PFCM 알고리즘

2.1 Possibilistic Fuzzy c-Means Clustering Algorithm(PFCM)

FCM 의 경우 하나의 패턴과 각 중심값 간의 상대거리로 멤버쉽을 할당하고 PCM 의 경우 오직 하나의 패턴과 하나의 중심 값 간의 절대거리로 typicality 를 할당한다. PFCM 은 위 두 가지의 특징을 전부 가지고 있음으로 다음과 같은 목적함수를 최소화 하는 방향으로 클러스터링 한다.

$$J_{m,\eta}(U, T, V; X) = \sum_{k=1}^n \sum_{i=1}^c (au_{ik}^m + bt_{ik}^\eta) \times \|x_k - v_i\|^2 + \sum_{i=1}^c \gamma_i \sum_{k=1}^n ((1-t_{ik})^\eta) \quad (1)$$

$$\sum_{i=1}^c u_{ik} = 1, 0 \leq u_{ik}, t_{ik} \leq 1, m > 1, \eta > 1, \gamma_i > 0 \quad (2)$$

식 (1)에서 x_k 는 k 번째 입력 패턴이고, v_i 는 i 번째 클러스터의 중심값을 나타낸다. 그리고 u_{ik} 는 입력 패턴 k 가 클러스터 i 에 소속되는 멤버쉽을 나타내고 m 은 퍼지화의 정도를 나타내는 상수로서 $m \in (1, \infty)$ 의 조건을 만족시키는 상수이다. t_{ik} 는 입력 패턴 k 가 클러스터 i 에 소속되는 typicality 를 나타내고, 퍼지화의 정도를 나타내는 상수로서 $\eta \in (1, \infty)$ 의 조건을 만족시키는 상수이다. PCM 에서는 γ_i 는 i 번째 클러스터의 typicality 가 0.5 가 되는 지점의 거리를 가리키는데 이는 PFCM 에서도 동일하게 적용된다. 위의 식 (1)을 최소화하기 위한 멤버쉽은 다음과 같이 구한다.

접수일자 : 2012년 10월 24일

완료일자 : 2012년 10월 24일

감사의 글 : 이 논문은 2011년도 정부(과학기술부)의 재원으로 한국과학재단의 지원을 받아 수행된 연구임(No. 2011-0025655).

$$u_{ik} = \left(\sum_{j=1}^c \left(\frac{d_{ik}}{d_{jk}} \right)^{\frac{2}{m-1}} \right)^{-1} \quad (3)$$

위의 식 (1)을 최소화하기 위한 typicality 및 γ_i 는 다음과 같이 구한다. 여기서 η 는 학습률 파라미터 이다.

$$t_{ik} = \left(\left(1 + \frac{b}{\gamma_i} d_{ik}^2 \right)^{\frac{1}{\eta-1}} \right)^{-1} \quad \gamma_i = K \frac{\sum_{k=1}^n u_{ik}^m d_{ik}^2}{\sum_{k=1}^n u_{ik}^m} \quad (4)$$

식 (4)를 보면 $K=1$ 일 경우 클러스터 i 의 평균 내부 거리 값을 γ_i 로 할당한다. 이때 최적의 중심 값을 위한 필요조건은 다음과 같다. 식 (3), (4)를 사용하여 다음과 같이 중심 값을 구한다.

$$\mathbf{v}_i = \frac{\sum_{k=1}^n (a u_{ik}^m + b t_{ik}^\eta) \mathbf{x}_k}{\sum_{k=1}^n (a u_{ik}^m + b t_{ik}^\eta)} \quad (5)$$

PFCM 알고리즘은 멤버쉽과 typicality 의 가중치 합으로 a 와 b 의 값에 따라 가중치를 조절할 수 있다. 따라서 적절한 a 와 b 값을 이용하여 노이즈의 영향을 줄이고, 중심 값이 겹치는 문제도 해결할 수 있다.

3. Interval Type-2 Fuzzy Set

Type-1 fuzzy set은 pattern recognition에서 패턴의 uncertainty를 나타내는데 사용되어져왔다[4]-[8]. 하지만 Type-1 fuzzy set으로 좋은 결과를 나타낼 수 없는 경우에는 Type-2 fuzzy set을 사용함으로써 더 나은 결과를 나타낼 수 있다[9]-[13]. Type-2 fuzzy set으로 Type-1 fuzzy set 보다 더 패턴의 uncertainty를 다양하게 제어할 수 있기 때문이다. Type-2 fuzzy set \tilde{A} 를 나타내면 다음과 같다.

$$\tilde{A} = \int_{x \in X} \mu_{\tilde{A}}(x) / x = \int_{x \in X} \left[\int_{u \in J_x} f_x(u) / u \right] / x \quad J_x \subseteq [0,1] \quad (6)$$

$f_x(u)$: the secondary membership function,
 J_x : the primary membership of x which is the domain of the secondary membership function

Type-2 fuzzy set이 Type-1 fuzzy set보다 우수한 성능을 나타내지만, 많은 계산 량을 필요로 한다는 단점이 있다. 이러한 단점을 보완하기 위해 secondary membership degree가 1인 Interval Type-2 fuzzy set을 사용한다. Interval Type-2 fuzzy set \tilde{A} 를 나타내면 다음과 같다.

$$\tilde{A} = \left[\int_{u \in J_x} 1/u \right] / x \quad (7)$$

식 (7)를 보면 알 수 있듯이 secondary membership degree가 모두 같기 때문에 Type-1 fuzzy set 처럼 사용할 수 있다. 그림 1은 Interval Type-2의 예를 보여주고 있다. (a)는 Type-2 fuzzy set 의 모든 primary

membership의 집합인 FOU(footprint of uncertainty)를 회색으로 나타내고 있다. 이때 x' 에 대한 primary membership을 나타내면 그림과 같이 Upper and Lower membership function을 사용해 범위 $[\underline{u}(x'), \bar{u}(x')]$ 으로 나타낼 수 있다. (b)는 이 부분을 secondary membership domain 상에서 표현한 것으로 이 범위의 모든 secondary degree는 1이라는 것을 알 수 있다.

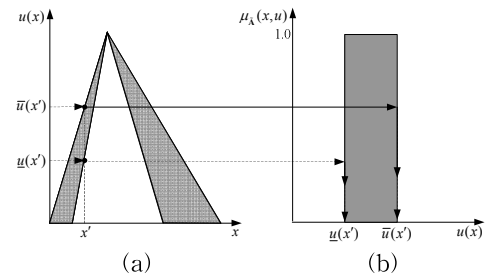


그림 1. Interval type-2의 예 : (a) Type-2 fuzzy set의 FOU. (b) Interval secondary function.

4. Interval Type-2 PFCM 알고리즘

PFCM의 목적함수에서 보면, 퍼지화 파라미터 m 의 값에 따라서 γ_i 값이 변화 한다. 본 논문에서는 Interval Type-2 fuzzy set을 이용해 m 에 대한 uncertainty를 표현하여 하나의 값만 사용하는 PFCM보다 나은 성능을 보이는 알고리즘을 제안 한다.

4.1 Extension to Interval Type-2 Fuzzy Set

PFCM에서는 Crisp인 입력패턴과 클러스터에 속하는 멤버쉽 값으로 primary fuzzy set을 만들지만, m 에 대한 uncertainty를 표현하기 위해서는 primary fuzzy set인 입력패턴을 Interval Type-2 Fuzzy Set으로 할당하여야 한다. 이를 위해서 primary membership 함수를 사용하여 Upper membership function과 Lower membership function을 만들면 된다. 본 논문에서는 m 에 따른 Upper and lower membership function을 구하겠다. m 은 퍼지화 파라미터로써 최종 클러스터링의 퍼지화 정도를 결정한다. 즉 m 이 무한대로 갈수록 멤버쉽 값은 점점 퍼지화 된다. m 은 1보다 큰 임의의 파라미터 값 2개로 uncertainty를 만들 수 있다. m 에 따른 Upper and lower membership function은 다음과 같이 만들 수 있다.

$$\bar{u}_{ik} = \begin{cases} \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{ik}} \right)^{\frac{2}{m_1-1}}}, & \text{if } \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{ik}} \right)^{\frac{2}{m_1-1}}} < \frac{1}{C} \\ \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{ik}} \right)^{\frac{2}{m_2-1}}}, & \text{otherwise} \end{cases}$$

$$\underline{u}_{ik} = \begin{cases} \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{ik}} \right)^{\frac{2}{m_1-1}}}, & \text{if } \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{ik}} \right)^{\frac{2}{m_1-1}}} \geq \frac{1}{C} \\ \frac{1}{\sum_{k=1}^c \left(\frac{D_{ij}}{D_{ik}} \right)^{\frac{2}{m_2-1}}}, & \text{otherwise} \end{cases} \quad (8)$$

그림 2 는 입력 패턴 x' 에 Interval type-2 을 이용하여 m 값의 변화에 따른 membership 의 변화를 보여 주고 있다.

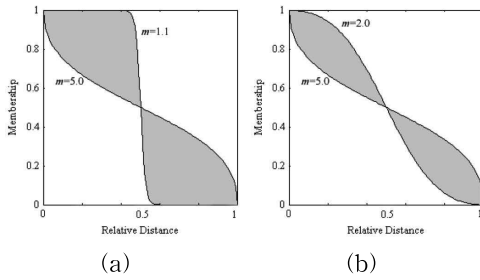


그림 2. m 값의 변화에 따른 membership 값의 변화 예
(a) $m_1 = 1.1, m_2 = 5.0$ (b) $m_1 = 2.0, m_2 = 5.0$

또한 Upper and lower membership 에 따라 γ_i 의 값이 달라진다.

$$\gamma_{i,m_1} = K \frac{\sum_{k=1}^n u_{ik,m_1}^{m_1} d_{ik}^2}{\sum_{k=1}^n u_{ik,m_1}^{m_1}}, u_{ik,m_1} = \left(\sum_{k=1}^c \left(\frac{d_{ij}}{d_{ik}} \right)^{\frac{2}{m_1-1}} \right)^{-1}$$

$$\gamma_{i,m_2} = K \frac{\sum_{k=1}^n u_{ik,m_2}^{m_2} d_{ik}^2}{\sum_{k=1}^n u_{ik,m_2}^{m_2}}, u_{ik,m_2} = \left(\sum_{k=1}^c \left(\frac{d_{ij}}{d_{ik}} \right)^{\frac{2}{m_2-1}} \right)^{-1} \quad (9)$$

γ_i 를 사용하여 Upper and lower typicalty function 은 다음과 같이 만들어진다.

$$\bar{t}_{ik}(x_i) = \begin{cases} \left(1 + \frac{b}{\gamma_{i,m_1}} d_{ik}^2 \right)^{\frac{1}{\eta-1}}, & \text{if } m_1 < m_2 \\ \left(1 + \frac{b}{\gamma_{i,m_2}} d_{ik}^2 \right)^{\frac{1}{\eta-1}}, & \text{otherwise} \end{cases}$$

$$\underline{t}_{ik}(x_i) = \begin{cases} \left(1 + \frac{b}{\gamma_{i,m_1}} d_{ik}^2 \right)^{\frac{1}{\eta-1}}, & \text{if } m_1 \geq m_2 \\ \left(1 + \frac{b}{\gamma_{i,m_2}} d_{ik}^2 \right)^{\frac{1}{\eta-1}}, & \text{otherwise} \end{cases} \quad (10)$$

그림 3 는 입력 패턴 x' 에 Interval type-2 을 이용하여 m 값의 변화에 따른 typicality 의 변화를 보여 주고 있다.

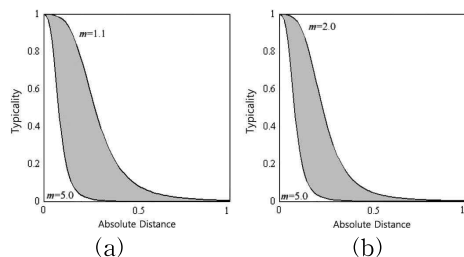


그림 3. m 값의 변화에 따른 typicality 값의 변화 예
(a) $m_1 = 1.1, m_2 = 5.0$ (b) $m_1 = 2.0, m_2 = 5.0$

4.2 Type Reduction and Defuzzification

입력 패턴들이 각 클러스터에 대한 upper와 lower 멤버쉽을 구하고 나면 각 클러스터들의 중심 값을 갱신해야 한다. 하지만 이때 멤버쉽은 Type-2 fuzzy set이지만, 중심값은 crisp한 값이기 때문에 갱신이 불가능하다. 따라서 중심값 갱신을 위해 Type-2 fuzzy set을 Type-1 fuzzy set으로 바꾸는 type reduction과 type-1 fuzzy set을 crisp한 값으로 바꾸는 defuzzification이 필요하다. Type reduction을 수행하는 여러 알고리즘이 있지만, 본 논문에서는 KM 알고리즘을 사용하겠다. KM 알고리즘은 다음과 같다. 알고리즘을 보면 알 수 있듯이, 가장 작은 왼쪽 값과 가장 큰 오른쪽 값을 구한다. 이 두 개의 값으로 다음과 같이 Interval Type-1 fuzzy set을 만들 수 있다.

$$v = 1.0 / [v_L, v_R]$$

v_L : The least value of left side
 v_R : The most value of left side

최종 crisp center는 다음과 같이 defuzzification 하여 구한다.

$$\text{Crisp } v = \frac{v_L + v_R}{2} \quad (12)$$

5. 실험 결과 및 분석

본 논문에서 제안한 알고리즘의 성능을 알아보기 위해서 여러 패턴과 이미지에 PFCM 알고리즘과 제안한 Interval type-2 PFCM 알고리즘 이용하여 실험해 보겠다. 모든 PFCM 실험에는 m 값을 2~9, η 값은 2 로 K 는 1 로 실험하였고, Interval type-2 FCL 실험에서도 m, η 값의 범위는 같지만 Interval type-2 FCL 에서는 두개의 초기 m_1 값 m_2 를 사용 하였다. 그리고 본 절에서는 각 m, η 값에 대해 PFCM 알고리즘과 제안한 Interval type-2 PFCM 알고리즘에 적용하여 실험 하고, 그 결과를 비교해 보겠다. PFCM 알고리즘과 Interval type-2 PFCM 알고리즘의 차이점을 비교하기 위해 a 와 b 는 1 로 고정하였다.

5.1 Iris56 데이터

Iris56 데이터는 세 개의 클러스터로 나뉘져 있고, 각 클러스터 마다 각 50 개의 데이터를 가지고 있다. 실험한 결과, PFCM 알고리즘보다 제안된 Interval Type-2 PFCM 알고리즘이 조금더 정확한 클러스터링을 하였다. Interval Type-2 PFCM 알고리즘의 경우 PFCM 보다 서로 다른 크기를 가진 클러스터를 클러스터링 하는데 더 좋은 성능을 보여주었다.

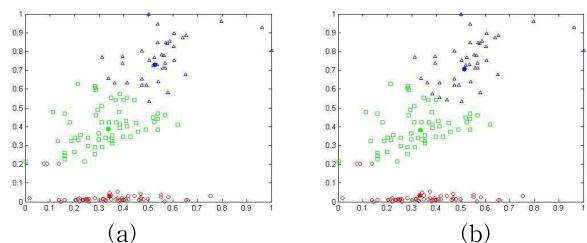


그림 4. Iris56 데이터의 클러스터링 결과
(a) PFCM (b) IT2PFCM

5.2 이미지 데이터(길, 숲, 하늘)

이번 실험으로 길, 숲, 하늘 영상을 분할해 보겠다. 길, 숲, 하늘 영상은 그림 5 (a)와 같이 길, 숲, 하늘 세부분으로 이루어져 있으며 200×200 의 사이즈를 가진다.

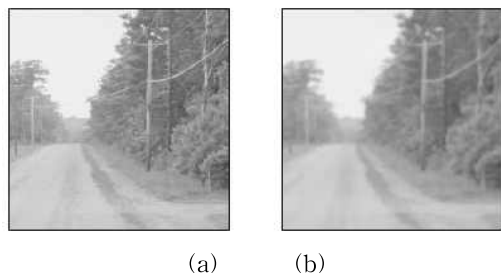


그림 5. 길, 숲 하늘 영상

(a) Intensity 속성, (b) Gaussian Blur 적용 속성

두 개의 속성 데이터를 모두 이용하여 영상을 클러스터링 하였고, 영상을 분할하기 위해서 두 개의 클러스터에서 100개씩 샘플 데이터를 추출하여 학습시킨 후 학습된 결과에 의해서 영상을 분할하도록 한다. 300개의 샘플 데이터로 클러스터링 한 결과와 영상 분할한 결과는 다음과 같다.

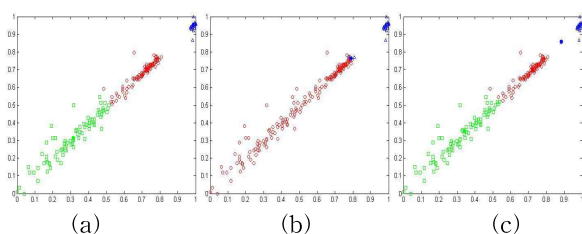
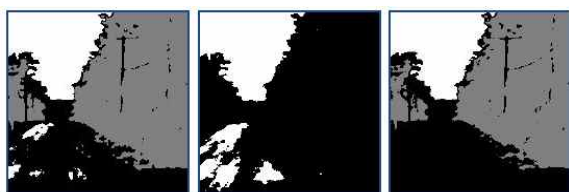


그림 5. 길, 숲, 하늘 영상 데이터의 클러스터링 결과

(a) IT2FCM (b) IT2PCM (c) IT2PFCM



(a) (b) (c)

그림 6. 길, 숲, 하늘 영상 데이터의 영역 분할 결과

(a) IT2FCM (b) IT2PCM (c) IT2PFCM

표 1. 길, 숲, 하늘 영상의 인식률 및 Iteration 비교

	IT2FCM	IT2PCM	IT2PFCM
인식률(%)	80.0	30.6	80.6
Iteration	36	44	20

6. 결론

본 논문에서는 PFCM 알고리즘의 성능을 개선하기 위해 Interval type-2 PFCM 알고리즘을 제안 하였다. 제안한 Interval type-2 PFCM 알고리즘에서는 m 에

uncertainty 를 부여함으로써 임의로 정해진 하나의 값만 사용하는 PFCM 알고리즘 보다 모든 데이터에서 더 나은 성능을 보였다. 또한 비슷한 알고리즘인 Interval type-2 FCM 와 Interval type-2 PCM 의 경우 패턴의 특성에 따라 성능이 달라지는데 반해 Interval type-2 PFCM 의 경우 일정하게 우수한 성능을 보여준다. 하지만 데이터 마다 최적의 성능을 찾아내는 파라미터의 값이 다르기 때문에, 적절한 파라미터 값을 찾는 것이 어렵다는 단점이 있다. 이번 실험에서는 a 와 b 및 K 를 1로 η 는 2로 고정시켰지만 향후에는 PFCM 알고리즘에 가중치를 나타내는 파라미터 a 와 b 및 η 에도 uncertainty 를 적용하여 본 논문에서 제안한 Interval type-2 PFCM 알고리즘과의 성능 비교를 하겠다.

참 고 문 헌

- [1] J. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum, 1981.
- [2] R. Krishnapuram and J. Keller, "The Possibilistic Approach to Clustering," *IEEE Trans. Fuzzy Syst.*, Vol. 1, No. 2, pp. 98-110, May 1993.
- [3] Pal, N.R., Bezdek, J.C., "A Possibilistic Fuzzy c-Means Clustering Algorithm," *Fuzzy Systems, IEEE Transactions on*, vol. 13, pp. 517 - 530, Aug. 2005.
- [4] R. Duda, P. Hart, and D. Stork, Pattern Classification, John Wiley & Sons, 2001.
- [5] J. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, New York: Plenum Press, 1981.
- [6] G. Klir, and B. Yuan, Fuzzy Sets and Fuzzy Logic: Theory and Applications, Prentice Hall, 1995.
- [7] R. Dave, "Fuzzy-shell clustering and applications to circle detection in digital images," *Int. J. General Systems*, vol. 16, pp.343-355, 1990.
- [8] F. Rhee, *Fuzzy rule generation and inference methods for pattern recognition and computer vision*, Ph.D. dissertation, University of Missouri, Columbia, MO, 1993.
- [9] C. Hwang and F. Rhee, "Uncertain fuzzy clustering: interval type-2 fuzzy approach to C-means," *IEEE Trans. Fuzzy Syst.*, vol. 15, no.1, pp. 107-120, Fed. 2007.
- [10] F. Rhee, "Uncertain fuzzy clustering: insights and recommendations," *IEEE Computational Intelligence Magazine*, vol. 2, no. 1, Fed. 2007.
- [11] F. Rhee and C. Hwang, "A type-2 fuzzy C-means clustering algorithm," in *Proc. 2001 Joint Conf. IFSA/NAFIPS, Fuzzy Syst.*, pp.1926-1919, Jul. 2001
- [12] C. Hwang and F. Rhee, "An interval type-2 fuzzy C spherical shells algorithm," in *Proc. 2004 Int. Conf. Fuzzy Syst.*, vol. 2, pp. 1117-1122, Jul. 2004.
- [13] F. Rhee and C. Hwang, "An interval type-2 fuzzy perceptron," in *Proc. 2002 Int. Conf. Fuzzy Syst.*, vol. 2, pp. 1331-1335, May. 2002.