

제2종 퍼지 집합을 이용한 퍼지 C-means

A Type 2 Fuzzy C-means

황철*, 이정훈**

Hwang Cheul, Frank Chung-Hoon Rhee
Computational Vision and Fuzzy Systems Laboratory
한양대학교 전자공학과
425-791 경기도 안산시 사1동 1271 번지
{chwang, frhee}@fuzzy.hanyang.ac.kr

Abstract

This paper presents a type-2 fuzzy C-means (FCM) algorithm that is an extension of the conventional fuzzy C-means algorithm. In our proposed method, the membership values for each pattern are extended as type-2 fuzzy memberships by assigning membership grades to the type-1 memberships. In doing so, cluster centers that are estimated by type-2 memberships may converge to a more desirable location than cluster centers obtained by a type-1 FCM method in the presence of noise.

Key Words : 제2종 퍼지 집합, 퍼지 C-means, 불확실성.

1. 서론

기존의 FCM방법은 squared error function을 최소화 시키는 방법에 의해, 패턴들을 여러 개의 클러스터들로 나누게 된다.

$$J_m(U, V; X) = \sum_{i=1}^c \sum_{j=1}^n (u_{ij})^m \|x_j - v_i\|^2 \quad (1)$$

위의 식으로부터 유도된 소속도 함수 (membership function)는 클러스터 센터 (cluster center)와 패턴과의 거리측정에 의해 해당패턴 소속도 값을 할당한다. 각 패턴이 할당 받은 소속도 값은 각 입력 패턴이 해당 클러스터에 얼마나 소속되어 있는지를 나타내게 된다. 또한, 이는 각 클러스터 센터의 위치를 갱신시킬때, 각 패턴이 새로운 클러스터 센터의 위치결정에 얼마나 기여할 것인지를 나타낸다[1]. 그러나 기존의 제1종 FCM에서 사용된 소속도함수에 의해 할당된 각 패턴의 소속도 값은 각 클러스터에 대한 해당 패턴의 전형성 (typicality)을 고려한 것이 아니다. 이는 클러스터에 대한 전형성 보다는 패턴과 해당 클러스터의 센터들간의 상대적인 거리의 개념이다. 따라서 이러한 상대적인 거리개념의 소속도 값을 통해 각 패턴의 클러스터 센터의 위치결정에 대한 기여도를 결정하는 것은 합리적이지 못한 클러스터 센터의 위치를 생성해 낼 수도 있다 [6]. 이는 클러스터 센터의 위치를 결정할

때, 각 패턴이 센터의 위치 변화에 얼마나 기여할 것인가를 결정하는 패턴의 소속도 값이 해당 클러스터에 대한 해당 패턴의 전형성을 나타내지 못하기 때문이다. 본 논문에서는 기존의 FCM방법에서 사용되는 각 패턴의 소속도 값들을 다시 한번 퍼지화 (fuzzify)시켜서 제2종 퍼지 집합으로 확장하게 된다[2]-[5]. 이는 클러스터 센터의 위치를 결정할 때 각 클러스터내에서의 패턴이 할당 받은 소속도 값의 상대성을 감소시키기 위한 것이다. 즉, 제2종 퍼지 집합으로 확장된 소속도 값들은 그 값이 클수록 클러스터 센터의 위치 결정에 대한 해당 패턴의 기여도에 있어 불확실성이 줄어들고, 값이 작을수록 불확실성이 커지게 된다. 즉, 해당 패턴의 소속도 값이 크면 클수록, 해당 클러스터 센터위치 변화에 많은 기여하고, 그 값이 작으면 작을수록, 해당 클러스터 센터의 위치 변화에 대한 기여도는 훨씬 감소 한다는 것이다. 이러한 접근 방법은 제1종 FCM의 소속도 값의 클러스터 센터의 위치 결정에서의 역할과도 일치한다. 본 논문에서는 이러한 제2종 퍼지 집합으로의 소속도 값의 확장을 통해 다음과 측면을 향상시킬 수 있었다.

- 소속도 값의 상대성을 감소시킴으로 인해, 해당 클러스터에 상대적으로 관계가 적은 패턴들의 영향을 감소시킬 수 있다. 그 결과 좀더 합리적인 각 클러스터 센터의 위치를 얻을 수 있다.

본 논문의 나머지 부분은 다음과 같이 구성되어 있다. 2절에서는 제 2종 퍼지 집합을 이용한 FCM방법을 구현하기 위한 제 1종 퍼지 소속도 값에 대한 제2종 퍼지 집합으로의 확장을 제안한다. 3절에서는 제안된 제2종 퍼지 집합을 제1종 FCM에 적용해 클러스터 센터를 갱신하는 방법을 설명한다. 그리고 4 절에서는 노이즈가 존재하는 패턴 데이터대 해서 제1종 FCM과 제2종 FCM의 클러스터링 결과를 비교한다. 또한 제2종 FCM을 적용하여, 영상에 대한 분할(segmentation)을 수행하고, 그 결과를 제1종 퍼지 집합의 결과와 비교해 보인다

2. 본 론

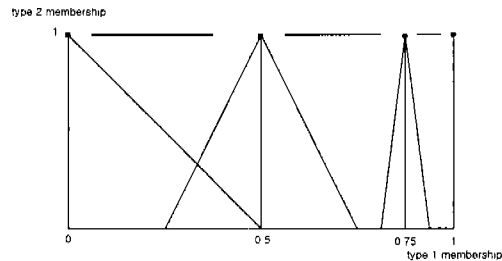
2.1 제2종 퍼지 집합으로의 확장

제1종 FCM은 각 클러스터 센터의 위치를 결정하기 위해, 패턴 공간(pattern space)내의 모든 패턴들과 각 클러스터 센터들간의 거리에 기초해 모든 패턴들에 소속도 값을 할당한다. 이 소속도 값들은 클러스터 센터의 새로운 위치를 결정할 때, 각 패턴이 클러스터 센터의 위치 변화에 얼마나 기여할 것인지를 결정하는 역할을 한다[1].

$$v_{ik} = \frac{\sum_{k=1}^n (u_{ik})^m x_{ik}}{\sum_{k=1}^n (u_{ik})^m} \quad (2)$$

즉, 식(2)에서처럼, 패턴들의 위치의 평균 값에 의해 클러스터 센터의 위치를 결정할 때 각 패턴의 위치에 대한 가중치(weight)로서 작용하게 된다. 그러나 이러한 소속도 값들은 클러스터에 대한 패턴의 전형성에 기초한 소속도 값이 아닌 해당 패턴과 각 클러스터 센터들과 거리에 대한 비율에 의해 발생된 값들이기 때문에, 매우 상대적인 특성을 가지게 된다. 즉, 클러스터 센터의 위치를 결정할 때, 실제 각 패턴이 할당 받은 소속도 값에 비례해서 해당 클러스터의 센터의 위치 변화에 기여할지는 불확실하게 된다. 본 논문에서는 이러한 가중치로서의 각 패턴 소속도 값의 불확실성을 보완하기 위해 각 소속도 값을 제2종 퍼지 집합으로 확장을 시도한다[4]. 이러한 시도에 대한 원리는 매우 간단하다. 각 패턴과 클러스터들간의 상대적인 거리에 기초해 생성된 소속도 값들은 그 값이 크면 클수록, 해당 클러스터 센터의 위치를 결정할 때 해당 패턴의 위치는 큰 역

합을 하게 된다. 반대로 그 값이 작으면 작을수록, 더 작은 역할을 하게 된다. 그러나 소속도 값의 상대성 때문에 해당 패턴이 클러스터의 센터 위치를 결정할 때 얼마나 기여할지는 다소 불확실성을 갖게 된다. 이는 소속도 값이 작으면 작을수록 클러스터 센터 위치 결정에서의 패턴의 위치에 대한 가중치로서의 불확실성이 더욱 높아지게 된다. 이를 기초로 소속도 값의 불확실성을 표현하면 다음과 같은 퍼지 집합을 구성할 수 있다.



[그림 1] 삼각형 함수(triangular function)에 의해 소속도 값의 불확실성을 퍼지화한 모습
그림 1의 퍼지 집합을 통해 알 수 있는 것처럼, 그 소속도 값이 1에 가까울수록, 클러스터 센터의 위치 결정에서 패턴의 기여도에 대한 불확실성은 작아진다. 반대로, 0에 가까울수록 해당패턴의 기여도는 불확실해진다. 이를 통해, 해당 클러스터 센터의 위치 결정에서, 상대적으로 관계가 적은 패턴의 역할을 줄일 수 있다. 본 논문에서는 계산의 편의를 위해, 삼각형 모델을 이용해 소속도 값에 불확실성을 부여했지만, 다른 어떠한 기존의 퍼지 모델을 사용해서도 확장이 가능하다.

2.2 제2종 퍼지 집합을 이용한 FCM

2절에서 확장된 퍼지 집합을 FCM에 적용하기 위해서는, 각 패턴이 할당받은 소속도 값의 클러스터 센터 갱신과정에서의 역할을 이해해야 한다. 이는 명확하게 클러스터 센터 위치 변화에 있어서, 각 패턴의 위치가 미치는 영향에 대한 가중치로서 작용을 한다. 따라서 클러스터 센터 갱신 방법에 있어서 개선이 필요하게 된다. 그 방법은 가중치로서 소속도 값의 역할에 불확실성을 부여하는 방향으로 개선되어야 한다. 즉, 클러스터 센터 갱신과정에서의 각 패턴의 소속도 값은 그 값의 크기에 따라 감소시킨다. 제2종 퍼지 소속도 함수는 그림 1에 사용된 삼각형 모델을 사용하였을 때, 제2종 퍼지 소속도 값은 다음과 같은 식으로 표현할 수 있다.

$$a_{ik} = u_{ik} - \frac{(1-u_{ik})}{2} \quad (3)$$

여기서 a_{ik} 는 제2종 퍼지 소속도 값
 제2종 퍼지 소속도 함수들은 그 면적을 계산하여, 함수의 최고 높이에 해당하는 제1종 소속도 값에서 불확실성이 부여된 만큼 빼게 된다. 이때 삼각형 함수의 밑변은 (1 - 삼각형의 높이 (= 제1종 소속도 값))으로 그 길이를 지정하였다. 이는 소속도 값이 1에 가까워 질수록 그 불확실성이 줄어들고, 반대로 0에 가까워 질수록 그 불확실성이 커지기 때문이다. 제2종 소속도 함수의 면적은 제1종 퍼지 소속도 값이 클러스터 센터를 갱신시킬때 해당패턴의 기여도에 대한 불확실성을 나타낸다.

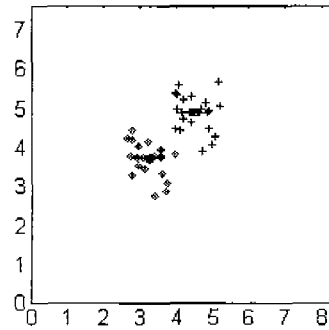
식 3을 FCM의 클러스터 센터 갱신식에 대입하면, 다음과 같은 제2종 FCM의 클러스터 센터 갱신식을 유도할 수 있다.

$$v_{jk} = \frac{\sum_{i=1}^n (a_{ik})^m x_{ik}}{\sum_{i=1}^n (a_{ik})^m} \quad (4)$$

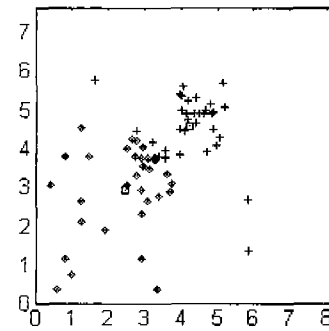
식 4에서 얻어진 클러스터 센터의 위치에 따라 클러스터링을 수행하는 알고리즘에 있어서는, 제1종 FCM 알고리즘과 동일한 방법으로 수행된다. 즉 클러스터의 센터들의 최종 위치가 결정되고 나면, 각 패턴의 클러스터들에 대한 분할에는 클러스터 센터의 최종 위치로부터 각 패턴이 할당 받은 제1종 퍼지 소속도 값에 따라 퍼지 분할을 수행한다.

3. 실험 및 결과

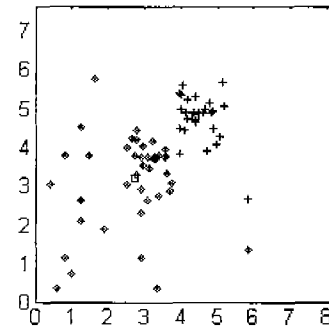
본 논문에 제시된 제2종 퍼지 집합을 이용한 FCM은 기본적으로 제1종 FCM의 특성을 따르면서도, 좀더 합리적인 클러스터 센터의 위치를 얻기 위해서, 소속도 값을 제2종 퍼지 집합으로 확장하였다. 제2종 퍼지 집합으로 확장에 의해, 클러스터 센터의 위치 결정에 가중치로서의 소속도 값의 역할을 그 크기에 따라 감소시킨 효과를 확인하기 위해 다음과 같은 실험을 하였다. 우선 첫번째로, 잘 분리되어 있는 패턴 데이터들에 대해 제1종 FCM과 제2종 퍼지 집합을 이용한 FCM에 대한 클러스터링 결과를 비교하고, 패턴 공간 상에 노이즈를 분포시켜, 같은 실험을 수행하고, 그 결과를 비교하였다. 그 결과는 아래와 같다.



(a) FCM과 type2 FCM의 clustering결과



(b) 노이즈가 첨가되었을때의 FCM결과



(c) 노이즈가 첨가되었을때의 type 2 FCM결과

[그림 2] 가우시안 난수 생성기에 의해 생성된 노이즈가 첨가된 패턴에 대한 FCM과 제2종 FCM의 결과

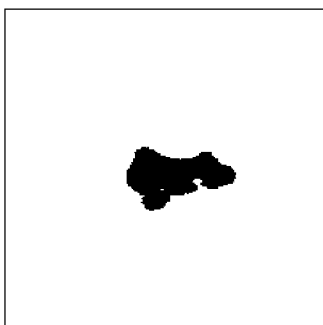
그림 2 (a)는 원래의 패턴 데이터를 각각 FCM과 제2종 FCM에 의해 퍼지 분할한 결과이다. 제1종 FCM과 2종 FCM 모두 같은 결과를 보였다. 그러나 (b)와 (c)에서는 FCM에 비해, 제2종 FCM 이 더 타당한 클러스터링 결과를 보이고 있음을 알 수 있다. 이러한 결과는 클러스터 센터의 갱신에 기여도가 낮은 즉, 소속도 값이 작은 패턴들일수록, 제2종

퍼지 집합으로의 확장에 의해 그 불확실성이 커지기 때문에, 클러스터로부터 넓게 분포된 노이즈 패턴들의 영향은 제1종 FCM에 비해, 제2종 FCM에 의한 퍼지 분할결과가 작게 영향 받게 되기 때문이다.

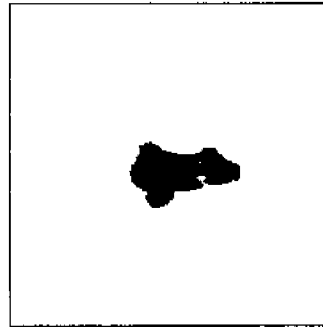
두번째 실험으로는 실제 영상 데이터를 분할하기 위해, 논문에 제시된 방법을 적용하여 보았다. 실험에 사용된 우주선영상으로부터, 생성된 5가지의 특징(feature)영상 (Homogeneity, Entropy, DV, DE, SV feature)으로부터 200개의 패턴을 추출하여 실험에 사용하였다. 각 패턴은 우주선과 배경으로부터 각각 임의의 위치로부터 100개씩의 데이터가 추출되었다. 우주선에 대한 그 레이스케일(grayscale)영상과 우주선과 배경 사이의 영상 분할 결과는 아래와 같다.



(a)우주선 영상



(b)제1종 FCM에 의한 영상분할 결과



(c) 제2종 FCM에 의한 영상분할 결과

[그림 3] original image와 제1종 과 제2종 FCM에 의해 영상 분할 결과

그림 3의 (b)와 (c) 결과 영상을 비교해 보면, 제2종 퍼지 집합에 의한 FCM의 영상분할 결과가 제1종 FCM 영상분할 결과에 비해서 우주선을 배경으로부터 더 명확히 구분해 내고 있음을 알 수 있다.

본 논문에서는 제1종 FCM에서 사용되는 소속도 값에 불확실성을 부여하였다. 이 불확실성은 클러스터 센터의 갱신과정에서 소속도 값의 가중치로서의 불확실성을 의미한다. 따라서 클러스터와 패턴간의 상대적인 거리에 기초한 소속도 값의 가중치로서의 영향을 그 크기에 따라 감소시킴으로 인해, 클러스터링 결과에서 좀더 합리적인 결과를 얻을 수 있었다. 본 논문에서는 삼각형 모델을 사용하여 제2종 퍼지 집합으로 확장을 시도했지만, 이는 일반적인 어떠한 퍼지 모델도 적용이 가능하다. 따라서 앞으로 여러가지 다른 퍼지 모델을 사용한 제2종 퍼지집합으로의 확장이 연구될 것이다.

4. 참고 문헌

- [1] J. Bezdek, Pattern Recognition with Fuzzy objective Function Algorithms. New York: Plenum, 1981
- [2] Nilesh N. Karnik, Jerry M. Mendel, "Introduction to type-2 fuzzy logic systems," IEEE Conference. Fuzzy Syst., pp.915-920, May. 1998
- [3] Qilian Liang, Jerry M. Mendel, "Interval type-2 fuzzy logic systems : theory and design," IEEE. Trans. Fuzzy Syst. Vol. 8, no. 5, October. 2000
- [4] Qilian Liang, Jerry M. Mendel, "An introduction to type-2 TSK fuzzy logic systems," IEEE Conference. Fuzzy Syst. pp. 1534-1539, August. 1999
- [5] Nilesh N. Karnik, Jerry M. Mendel, "Type-2 fuzzy logic systems," IEEE. Trans. Fuzzy Syst. vol. 7, no. 6, December. 1999.
- [6] R. Krishnapuram and James M. Keller, "A possibilistic approach to clustering," IEEE Trans. Fuzzy Syst. Vol.1 no. 2, May 1993