Interval 제 2종 퍼지 Possibilistic C-Means 알고리즘

An Interval Type 2 Fuzzy Possibilistic C-Means Algorithm

저자 (Authors)	민지희, 이정훈 Ji-Hee Min, Frank Chung-Hoon Rhee
출처 (Source)	한국지능시스템학회 학술발표 논문집 18(2), 2008.10, 3-6(4 pages)
발행처 (Publisher)	한국지능시스템학회 Korean Institute of Intelligent Systems
URL	http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE01086290
APA Style	민지희, 이정훈 (2008). Interval 제 2종 퍼지 Possibilistic C-Means 알고리즘. 한국지능시스템학회 학술발표 논문집, 18(2), 3-6
이용정보 (Accessed)	신라대학교 61.100.225.*** 2020/08/11 00:01 (KST)

Interval 제 2종 퍼지 Possibilistic C-Means 알고리즘

An Interval Type 2 Fuzzy Possibilistic C-Means Algorithm

민지희 · 이정훈
Ji-Hee Min and Frank Chung-Hoon Rhee

한양대학교 전자전기제어계측공학과

요 약

Fuzzy C-Means(FCM)의 단점을 극복하기 위해 제안되었던 PCM은 잡음에는 강하지만 초기 파라미터 값에 민감하고, 상대적으로 가까이에 위치한 prototype들을 형성하는 패턴들의 경우에는 최종 prototype의 위치가 겹치는(동일한) 결과가 나올 수 있다는 단점이 있다. 이러한 PCM의 단점을 극복하기 위해 여러 방법이 제안되었지만, 본 논문에서는 PCM 알고리즘에 Interval Type 2 Fuzzy 접근 방법을 적용하여 PCM 알고리즘의 파라미터에 존재하는 Uncertainty를 제어함으로써 성능을 향상시키는 방법을 제안한다.

키워드 : PCM, Type 2 Fuzzy Sets, Interval Type 2 Fuzzy Sets, Fuzzy Clustering.

1. 서 론

Fuzzy clustering에 일반적으로 많이 사용되는 알고리즘은 prototype과의 상대적인 거리를 이용해 멤버쉽을 할당하는 Fuzzy C-Mean(FCM)이다[1]. 하지만 FCM 알고리즘은 모든 클러스터의 멤버쉽 합이 항상 1이 되어야 한다는 조건 때문에 잘못된 클러스터링을 하거나 노이즈가 있는 데이터에서는 좋지 못한 결과를 보인다. 이러한 FCM의 단점을 극복하기 위해 typicality membership을 사용하는 Possibilistic C-Means(PCM)이 제안되었다[2]. PCM은 잡음에는 강하지만 초기 파라미터 값에 민감하고, 상대적으로 가까이에 위치한 prototype들을 형성하는 패턴들의 경우에는 최종 prototype의 위치가 겹치는(동일한) 결과가 나올 수 있다는 단점이 있다[3]. 이러한 단점을 개선하기 위해 여러 알고리즘들이 제안되었지만 [4-6], 새로운 수식을 만들거나 새로운 알고리즘을 만들어 PCM의 단점을 개선하였다. 하지만 본 논문에서는 PCM에 있는 파라미터의 uncertainty를 효율적으로 다룸으로써 성능을 개선하는 방법을 제안하겠다.

대부분의 패턴인식이나 클러스터링 문제에서, 패턴에 대한 모든 정보를 알 수 없기 때문에 파라미터 값에 대한 uncertainty를 가지고 있다. Type 1 fuzzy set은 패턴에 대한 uncertainty를 표현하기 위해 사용되어져 왔는데 [1], type 1 fuzzy set으로 모든 uncertainty를 나타내지 못할 경우에는 대부분 type 2 fuzzy set을 통해 효율적으로 조절할 수 있다[7-15]. 하지만 type 2 fuzzy set을 수행하기 위해서는 너무 많은 계산량을 필요로 한다. 따라서 계산량을 줄이기 위해서 secondary degree 값을 모두 똑같이 함으로써 type 1 fuzzy set처럼 사용할 수 있는 interval type 2 fuzzy set이 제안되었다[7].

본 논문에서는 PCM 알고리즘에 interval type 2 fuzzy set을 적용하여 PCM 알고리즘의 파라미터에 존재하는 uncertainty를 제어함으로써 성능을 향상시키는 방법을 제안한다. 2절에서는 PCM 알고리즘에 존재하는 uncertainty에 대해 알아보고, 3절에서는 이해를 돕기 위해 interval type 2 fuzzy set의 개념에 대해 간단하게 알아보겠다. 4절에서는 본 논문에서 제안하는 interval type 2 fuzzy set을 PCM 알고리즘에 적용하는 방법에 대해 알아보고, 5절에서는 알고리즘의 타당성을 보이기 위해서 몇 가지 간단한 실험 결과를 제시한 다음 마지막으로 6절에서 결론을 맺겠다.

감사의 글 : 본 연구는 한국과학기술원 영상정보특화연구센터를 통한 국방과학연구소의 연구비 지원으로 수행되었습니다.

2. PCM 알고리즘의 Uncertainty

Typicality membership을 이용해 클러스터링을 하는 PCM 알고리즘은 다음과 같은 식으로 각 클러스터에 대한 멤버쉽 값을 구할 수 있다.

$$u_j(X_i) = \frac{1}{1+\left(\frac{d_{ji}^2}{\eta_j}\right)^{1/(m-1)}} \tag{1}$$

식 (1)에서 X_i는 i번째 입력 패턴이고 $u_j(X_i)$는 i번째 입력패턴이 클러스터 j에 속하는 멤버쉽 값을 나타낸다. 그리고 d_{ji}는 입력패턴과 prototype간의 거리, m은 퍼지화 상수, 마지막으로 η_j는 j번째 클러스터의 멤버쉽 값이 0.5일 때의 거리를 가르킨다. PCM 알고리즘은 초기 파라미터 값에 민감하기 때문에, m과 η_j값을 잘 설정하는 것이 알고리즘의 성능에 중요한 역할을 한다.

퍼지화 상수 m이 멤버쉽 값에 어떻게 영향을 주는지 알아보기 위해, 다양한 m값에 대한 멤버쉽 값을 d_{ji}^2/η_j거리에 따라 나타내면 다음과 같다. 그림 1에서 볼 수 있듯이, m값이 커질수록 멤버쉽 함수의 기울기가 완만해져 0으로 수렴하지 않고, 0이나 1이 아닌 멤버쉽 값을 가지는 거리(퍼지 지역)도 점점 넓어진다.

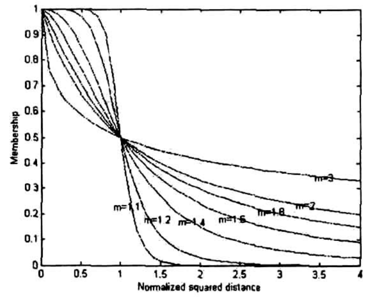
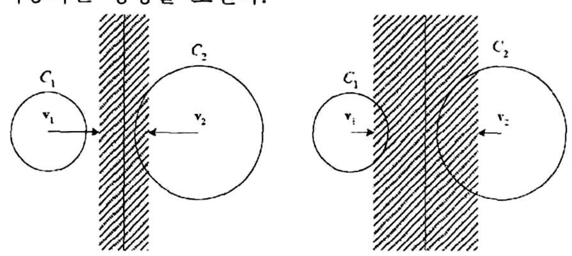


그림 1. 다양한 m값에 따른 멤버쉽 함수.
Fig 1. A plot of the membership function for various values of the fuzzifier parameter m.

클러스터들의 부피가 비슷하거나 같을 경우에는 m값에 따라 좋은 성능을 기대할 수 있지만, 아래 그림과 같이 클러스터들의 부피가 다를 경우에는 하나의 m으로는 좋은 성능을 보일 수 없다. 그림 2(a)는 부피가 다른 2개의 클러스터에서 작은 m값을 설정했을 때를 보여준다. C_2 클러스터에 퍼지 멤버쉽 값을 가지는 구간이 걸쳐 있기 때문에 상대적으로 C_1 클러스터에 패턴들이 많이 할당이 된다. 그림 2(b)와 같이 큰 m값을 설정했을 때에는 비슷한 멤버쉽 값이 할당되기 때문에 좋은 성능을 보일 것 같지만, C_1 클러스터의 센터 값이 C_2 클러스터로 이동하는 경향을 보인다.



(a) (b)

그림 2. 부피가 다른 2개 클러스터에서의 m에 따른 퍼지 지역 : (a) 작은 m값 (b) 큰 m값.
Fig 2. Fuzzy region for two clusters of different volume by fuzzifier m of (a) low degree (b) high degree.

그래서 다른 부피를 가진 클러스터의 경우에는 다음 그림과 같이 m을 설정하는 것이 가장 이상적이다 [12~13].

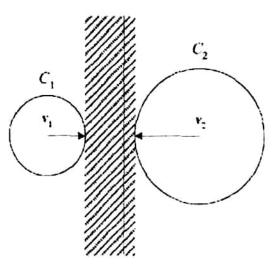


그림 3. 적합한 퍼지 지역.
Fig 3. Desirable fuzzy region.

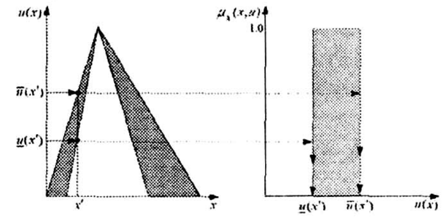
하지만 위와 같이 퍼지 지역을 형성하기 위해서는 두 개의 m값을 가져야 하지만, PCM 알고리즘에서는 불가능하다. 따라서 interval type 2 fuzzy set을 PCM 알고리즘에 적용하여 비슷한 성능을 보이게 하겠다.

3. Interval Type 2 Fuzzy Set

Type 2 fuzzy set에서 secondary 멤버쉽 값이 모두 1인 interval type 2 fuzzy set $\widetilde{A}$를 식으로 나타내면 다음과 같다[15].

$$\widetilde{A}=\int_{x\in X}\left[\int_{u\in J_x} 1/u\right]\Big/x \qquad J_x \subseteq [0,1] \tag{2}$$

Interval type 2 fuzzy set에 대한 이해를 돕기 위해 아래 그림과 같은 예를 보겠다. 그림 4(a)와 같은 FOU를 가질 때, x'에 대한 primary 멤버쉽 값은 그림 4(b)와 같이 secondary 멤버쉽 값이 모두 같기 때문에 $[\underline{u}(x'),\overline{u}(x')]$와 같이 구간으로 표현할 수 있다.



(a) (b)

그림 4. Interval type 2의 예 : (a) Type 2 fuzzy set의 FOU (b) Interval secondary function.
Fig 4. Example of (a) an interval type 2 fuzzy set (b) the vertical slice of sample x'.

이와 같은 interval type 2 membership set의 특징을 이용하여 m값에 따른 퍼지 지역을 설정하면 다음과 같다. 그림 3과 같이 클러스터 크기에 맞게 퍼지 지역을 설정할 순 없기 때문에, 2개의 m값을 이용하여 uncertainty를 줌으로써 클러스터 부피에 맞는 적당한 퍼지 지역을 형성할 수 있다.

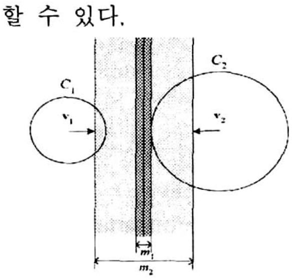


그림 5. 2개의 m값으로 만든 퍼지 지역.
Fig 5. Uncertain fuzzy region by two fuzzifiers.

4. Interval Type 2 퍼지 PCM 알고리즘

4.1 Extension to Interval Type 2 Fuzzy Sets

PCM 알고리즘에 interval type 2 fuzzy set 개념을 적용하여 m에 따른 uncertainty를 나타내기 위해서는, 2개의 m값에 따른 멤버쉽 함수를 만들어야 한다. m_1, m_2을 이용하여 만든 interval type 2 멤버쉽 함수는 다음과 같다($m_1 \neq m_2$). 만약 m_1과 m_2가 같으면 하나의 m을 사용하는 것과 같기 때문에 PCM 알고리즘과 동일하다.

$$\overline{u}_j(X_i)=\begin{cases}\dfrac{1}{1+\left(d_{ji}^2/\eta_j\right)^{1/(m_1-1)}} & if\ \dfrac{1}{1+\left(d_{ji}^2/\eta_j\right)^{1/(m_1-1)}}>\dfrac{1}{1+\left(d_{ji}^2/\eta_j\right)^{1/(m_2-1)}}\\ \dfrac{1}{1+\left(d_{ji}^2/\eta_j\right)^{1/(m_2-1)}} & otherwise\end{cases}$$

$$\underline{u}_j(X_i)=\begin{cases}\dfrac{1}{1+\left(d_{ji}^2/\eta_j\right)^{1/(m_1-1)}} & if\ \dfrac{1}{1+\left(d_{ji}^2/\eta_j\right)^{1/(m_1-1)}}<\dfrac{1}{1+\left(d_{ji}^2/\eta_j\right)^{1/(m_2-1)}}\\ \dfrac{1}{1+\left(d_{ji}^2/\eta_j\right)^{1/(m_2-1)}} & otherwise\end{cases}$$

(3)

그림 6은 m_1, m_2값이 1.2와 1.7일 때의 interval type 2 멤버쉽 함수에 대한 uncertainty를 보여주고 있다. 이와 같은 uncertainty를 통해 하나의 m값을 사용하는 PCM 알고리즘보다 좋은 성능을 보일 수 있다.

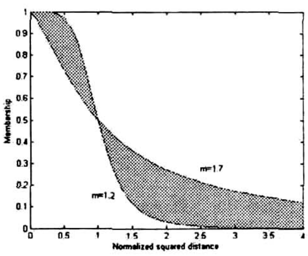


그림 6. 2개의 m에 따른 멤버쉽 함수의 FOU.
Fig 6. FOU of an interval type 2 membership function.

4.2 Type Reduction and Defuzzification

각 클러스터에 대한 interval type 2 멤버쉽 값을 구하고 나면, PCM 알고리즘에 따라 center 값을 갱신해야 한다. 그런데 멤버쉽 값은 interval type 2 fuzzy set이고, center는 crisp한 값이기 때문에 갱신이 불가능하다. 따라서 center의 갱신을 위해 interval type 2 fuzzy 멤버쉽 값을 type 1 fuzzy set으로 바꾸는 type reduction과 type 1 fuzzy set에서 crisp한 값으로 바꾸는 defuzzification 과정이 필요하다.

Type reduction을 위한 여러 알고리즘들이 제안되었지만, 본 논문에서는 KM 알고리즘을 사용하였다[7]. KM 알고리즘을 통해 가장 작은 왼쪽 값과 가장 큰 오른쪽 값을 구하여 이 두 개의 값으로 다음과 같이 interval type 1 fuzzy set을 만들 수 있다.

$V_j = 1.0/[V_L, V_R]$
V_L : The least value of left side. (4)
V_R : The most value of right side.

식 (4)로 interval type 1 fuzzy set을 구하고 나면, 다음과 같은 식으로 defuzzification 하여 crisp한 center 값을 구한다.

$$\text{Crisp } V_j = \frac{\sum_{V \in J_{Y_i}} (u(V))\, V}{\sum_{V \in J_{Y_i}} (u(V))} = \frac{V_L + V_R}{2} \quad (5)$$

4.3 Hard partitioning

알고리즘에 의해 최종 center값을 구하고 나면, 주어진 패턴들을 멤버쉽 값에 따라 클러스터별로 구분할 수 있다. 하지만 uncertainty를 위해 interval type 2 fuzzy set 멤버쉽 함수를 사용했기 때문에 hard partitioning을 위해서는 type reduction을 수행해야한다. Center의 type reduction을 위해 KM알고리즘을 수행할 때 구했던 각 패턴들의 왼쪽 멤버쉽(u_j^L)과 오른쪽 멤버쉽(u_j^R)값을 이용하여 type reduction을 구할 수 있다.

$$u_j(X_i) = \frac{u_j^L(X_i) + u_j^R(X_i)}{2} \quad j = 1, \dots, C \quad (6)$$

마지막으로 위에서 구해진 멤버쉽 값으로 다음과 같이 hard partition 할 수 있다.

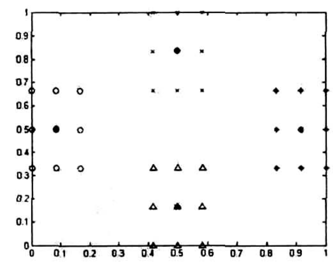
$$\text{If } u_j(X_i) > u_k(X_i) \;\; k = 1, \dots, C \text{ and } j \neq k \quad (7)$$
then, X_i is assigned j

5.실험 결과

제안한 interval type 2 퍼지 PCM 알고리즘과 PCM 알고리즘의 성능을 비교하기 위해 몇 가지 패턴에 대한 실험을 하였다. FCM 알고리즘을 이용해 PCM 알고리즘의 파라미터들의 초기 값을 설정 하였고, 모든 패턴과 이미지는 0~1 값으로 normalize 하였다.

5.1 Four Square 데이터

첫 번째 실험에서는 동일한 부피를 가진 패턴에서의 성능 비교를 위해 같은 크기의 정사각형 패턴에 대한 실험 결과를 보겠다.

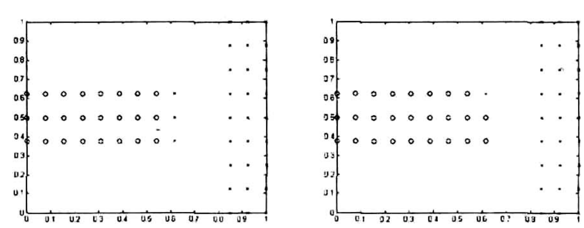


● : 최종 클러스터 center
그림 7. Four Square 데이터의 클러스터링 결과
Fig 7. Clustering result of four squares.

위 실험에서 볼 수 있듯이 같은 부피를 가진 패턴의 경우에는, PCM 알고리즘을 사용했을 때와 interval type 2 퍼지 PCM 알고리즘을 사용한 것과 차이 없이 정확한 center 값을 찾았고, 클러스터링도 정확하게 되었다.

5.2 Two Square 데이터

이번 실험에서는 같은 부피를 가지고 있지만, 다른 클러스터 모양을 가지고 있는 데이터에 대한 결과를 보겠다.



(a) (b)

그림 8. Two Square 데이터의 클러스터링 결과 : (a) PCM($m = 1.4$) (b) Interval type 2 PCM($m_1 = 1.1$, $m_2 = 1.6$).
Fig 8. Clustering results of two square : (a) PCM($m = 1.4$) (b) Interval type 2 PCM($m_1 = 1.1, m_2 = 1.6$).

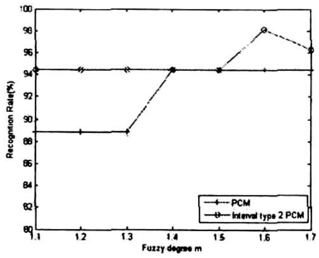
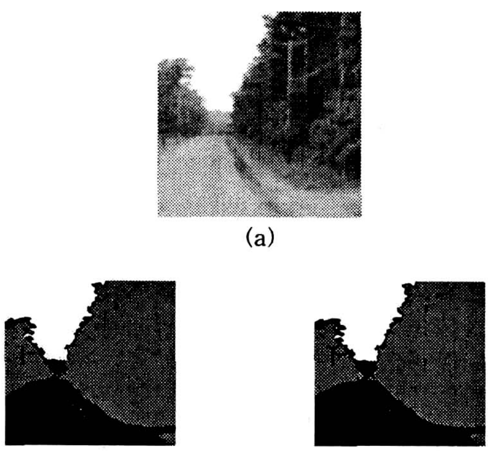


그림 9. Two Square 데이터에 대한 인식률 비교($m=m_1,m_2$ =가장 좋은 성능을 보이는 값).
Fig 9. Recognition plot($m=m_1,m_2$ =value for best result).

그림 8과 그림 9에서 볼 수 있듯이, interval type 2 PCM 알고리즘이 PCM 알고리즘보다 성능이 더 좋았다.

5.3 이미지 데이터(길, 숲, 하늘)

마지막 실험으로 길, 숲, 하늘 세 부분으로 이루어진 이미지를 분할해보겠다. 그림 10(a)는 200×200 사이즈의 실제 영상이고, 그림 10(b)와 (c)는 영상 분할한 결과이다.



(a)

(b) (c)

그림 10. 이미지 분할 결과 : (a) 본래 이미지 (b) PCM(95.0022%) (c) Interval type 2 PCM(95.2431%).
Fig 10. Result of Image segmentation : (a) original image (b) PCM (c) Interval type 2 PCM.

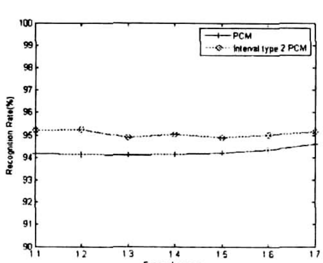


그림 11. 이미지 데이터에 대한 인식률 비교($m=m_1,m_2$ =가장 좋은 성능을 보이는 값).
Fig 11. Recognition plot($m=m_1,m_2$ =value for best result).

그림 11과 같이 이미지 데이터에 대한 인식률을 비교해보면 interval type 2 퍼지 PCM 알고리즘이 PCM 알고리즘보다 좋은 성능을 나타낸다는 것을 알 수 있다.

6. 결 론

본 논문에서는 PCM의 단점을 극복하기 위해, interval type 2 fuzzy set을 적용하여 2개의 m값에 대한 uncertainty를 표현함으로써 성능을 개선시키는 방법을 제안하였다. 여러 패턴에 대해 실험한 결과 PCM보다 좋은 성능을 나타냄을 알 수 있었다. 향후 PCM 알고리즘의 η에 대한 uncertainty를 표현하거나 type reduction 부분에서 속도를 개선시키는 연구가 이루어질 수 있을 것이다.

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] M. Barni, V. Cappellini, and A. Mecocci, “Comments on 'A possibilistic approach to clustering',” *IEEE Trans. Fuzzy Syst.*, Vol. 4, pp. 393-396, June 1996.
[4] R. Krishnapuram and J. Keller, “The Possibilistic C-means Algorithm:Insights and recommendations,” *IEEE Trans. Fuzzy Sys.*, Vol. 4, pp. 385-393, 1996.
[5] J. S. Zhang and Y. W. Leung, “Improved possibilistic c-means clustering algorithms,” *IEEE Trans. Fuzzy Sys,*. Vol. 12, No. 2, April 2004.
[6] Adam Schneider, "Weighted Possibilistic c-Means Clustering Algorithms," *IEEE Int. Conf. Fuzzy Systems*, Vol. 1, pp. 176-180, May 2000.
[7] J. Mendel, *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*, Prentice Hall, 2001.
[8] F. Rhee and C. Hwang, "A type-2 fuzzy *C*-means clustering algorithm," in *Proc.* 2001 *Joint Conf. IFSA/NAFIPS*, pp. 1926-1919, Jul 2001.
[9] F. Rhee and C. Hwang, “An interval type-2 fuzzy perceptron,” in *Proc.* 2002 *Int. Conf. Fuzzy Syst.*, Vol. 2, pp. 1331-1335, May 2001.
[10] F. Rhee and C. Hwang, “An interval type-2 fuzzy *K*-nearest neighbor,” in *Proc.* 2003 *Int. conf. Fuzzy Syst.*, Vol. 2, pp. 802-807, May 2003.
[11] F. Rhee and C. Hwang, “An interval type-2 fuzzy *C* spherical shells algorithm,” in *Proc.* 2004 *Int. conf. Fuzzy Syst.*, Vol. 2, pp. 1117-1122, Jul 2004.
[12] F. Rhee and C. Hwang, “Uncertain fuzzy clustering: interval type-2 fuzzy approach to *C*-means,” *IEEE trans. Fuzzy Syst.*, Vol. 15, No. 1, pp. 107-120, Feb 2007.
[13] F. Rhee, “Uncertain fuzzy clustering: insights and recommendations,” *IEEE Computational Intelligence Magazine*, Vol. 2, No. 1, Feb 2007.
[14] N. Karnik, J. Mendel, and Q. Liang, "Type-2 fuzzy logic systems," *IEEE Trans. Fuzzy Syst.*, Vol. 7, pp. 643-658, Dec 1999.
[15] J. Mendel, and R. John, “Type-2 fuzzy set made simple,” *IEEE Trans. Fuzzy Syst.*, Vol. 10, No. 2, April 2002.