「經營情報研究」 第9號
大韓經營情報學會
2002年 6月

확장된 Fuzzy 집락분석방법에 관한 연구

임 대 혁*

〈목 차〉

Ⅰ. 서 론

집락분석의 기본목적은 개체들을 집락으로 가능한 한 명확히 분류하는 것이다. 이 목적을 달성하기 위한 한 방법인 Fuzzy집락분석은 각 개체에 집락에 속할 확률을 할당하는 방법인데 이 확률적인 소속정도는 때로는 분석결과를 해석하는 연구가들에게 동일집단으로 인식되어야 하는 개체들을 차별하도록 하는 선입견을 줄 수 있다. 또한 분류하기 어려운 개체들이 많은 경우에는 분류를 1차 목적으로 하는 분석기법으로서의 기능을 상실할 수 있는 문제가 있을 수가 있다.

본 연구에서는 기존의 Fuzzy 집락분석에서 사용하는 목적함수에 Entropy 개

* 혜천대학 경영정보시스템 교수

념을 도입한 새로운 목적함수를 만들게 되었다. 이를 바탕으로 집락분석을 한 결과 위에서 제기한 문제점을 해결해 줄 수 있음을 알게 되었다. 한편 새로운 목적함수의 도입으로 인하여 종래의 목적함수를 위한 최적화 기법인 Fuzzy ISODATA 알고리즘이 새로운 목적함수의 최적화에는 적절하지 않음을 발견하여 신경회로망 등에서 최근 활발히 연구 활용하는 MFA알고리즘을 사용하였다. 그 결과 Fuzzy ISODATA 알고리즘이 지역적 최적해를 제공하는 단점을 극복하였다.

본 연구의 구성은 다음과 같다. 2장에서는 Fuzzy 집락분석의 기본개념에 대해서 살펴보고 3장에서는 Fuzzy 목적함수에 Entropy 개념을 삽입시켜서 개체들의 분별력을 향상시킨 목적함수를 제안하고 이 목적함수를 최적화하기 위한 기법인 MFA알고리즘에 대하여 살펴보았다. 4장에서는 자료를 통하여 Fuzzy-Entropy 목적함수와 기존의 Fuzzy 목적함수를 사용한 결과를 비교하였고, 최적화 방법으로는 MFA알고리즘과 Fuzzy ISODATA 방법을 함께 사용하였다. 5장에서는 앞장들의 결과를 종합하였고 앞으로의 연구과제를 토의하였다.

Ⅱ. Fuzzy 집락분석

1. Fuzzy 집락분석의 정의

고전적인 집락분석 방법을 사용하면 각 개체는 오직 하나의 집락에만 속하게 된다. 즉, 각 개체들의 소속함수 값은 0 혹은 1로 주어진다. 그러나 확실하게 집락을 결정하기 어려운 개체들에 대해서는 소속함수의 값을 0 또는 1로 주어 분류한다는 것은 문제가 있다. 이러한 문제를 해결하는 방법으로 Fuzzy집합론에 근거를 둔 Fuzzy 집락분석 방법이 제안되었다.

Fuzzy 집락분석 방법은 Bezdek(1976)에 의해서 제안되었고 최소화하는 방법으로 Fuzzy K-means 방법이 많이 사용되는데 다음과 같은 목적함수를 갖는다.

$$S(g,r,y) = \sum_{i=1}^{n} \sum_{m=1}^{g} Y_{im}^{r} \sum_{k=1}^{p} (X_{ik} - Z_{mk})^2 \qquad \text{...(2.1)}$$

$$1 < r < \infty$$

위에서, $X_i'=(X_{i1}, ..., X_{ip})$ $(i=1, 2, ..., n)$는 I번째의 개체이고 $Y_i'=(Y_{i1}, ..., Y_{ig})$는 I번째 개체가 각 집락에 속할 확률이며 $Z_m'=(Z_{m1}, ..., Z_{mp})(m=1, ..., g)$는 m번째 집락의 중심을 나타낸다.

$$\text{여기서,}\quad Z_{mk} = \sum_{i=1}^{n} Y_{im} * X_{ik} \Big/ \sum_{i=1}^{n} Y_{im}$$

$$(k=1, 2, ..., p)$$

이다.

Fuzzy K-means 방법은 다음과 같은 제약조건하에서 목적함수 $S(g,r,Y)$를 최소화하는 $Y_i'=(Y_{i1}, Y_{i2}, ..., Y_{im})$을 구하는 것이다.

1) $\sum_{m=1}^{g} Y_{im} = 1 \ (i=1, 2, ..., n)$

2) $0 \leq Y_{im} \leq 1$

$(i=1, 2, ..., n : m=1, 2, ..., g)$

위 방법을 수행하기 위하여 처음 제안된 알고리즘이 Fuzzy ISODATA 알고리즘이다. 이 알고리즘을 사용하기 위해서는 Y_i'들의 초기 값을 주어야 하는데, 주어진 초기 값에 따라 결과가 달라지게 된다. 이러한 문제점이 생기는 이유는 이 방법을 사용하면 전역적이 아닌 지역적 최소치에 수렴하는 경우가 많이 발생하기 때문이다. 이 단점을 극복하기 위해 본 논문에서는 근사 전역적인 최적해를 주는 SA 알고리즘 도입을 검토하게 되었고, 이 과정중 SA 알고리즘이 수행시간이 많이 걸린다는 단점을 보완하고 효율성을 향상시킨 MFA 알고리즘을 수행 알고리즘으로 사용하게 되었다.

2. Fuzzy ISODATA 알고리즘

Fuzzy K-means 방법으로 결과를 얻기 위해 Bezdek(1976)이 제안한 Fuzzy ISODATA 알고리즘은 다음과 같다.

$$S(g,r,y) = \sum_{i=1}^{n} \sum_{m=1}^{g} Y_{im}^{r} \sum_{k=1}^{p} (X_{ik} - Z_{mk})^2 \qquad ...(2.2)$$

여기서, $1 < r < \infty$

1) 집락의 수 g를 선택하고 $Y_i'(i=1, 2, ..., n)$들의 초기치를 준다.
2) 각 집락의 중심인 $Z_m'(m=1, 2, ..., g)$을 계산한다.
3) 다음 식에 의해서 $Y_i'(i=1, 2, ..., n)$을 계산한다.

$$Y_{im} = 1 / \sum_{j=1}^{g} [\sum_{k=1}^{p} (X_{ik} - Z_{mk})^2 / \sum_{k=1}^{p} (X_{ik} - Z_{jk})^2]^{1/(r-1)}$$

$$(1 \le m \le g,\ 1 \le I \le n)$$

4) Y_i'과 Z_m'들이 주어진 수렴조건을 만족할 때까지 2)와 3)을 반복수행한다.

일반적으로 최적화 문제에 있어서 반복 개선(iterative improvement) 알고리즘이 초기치 선택에 따라 지역적 최소치에 수렴하는 단점이 있는 것과 마찬가지로 Fuzzy ISODATA 알고리즘도 소속함수의 초기치 선택에 따라 전역적이 아닌 지역적 최소치에 수렴하는 단점이 있다. 그래서 본 논문에서는 최적화 방법으로 근사 전역적 최적해를 제공해 주는 MFA 알고리즘을 사용하게 되었다.

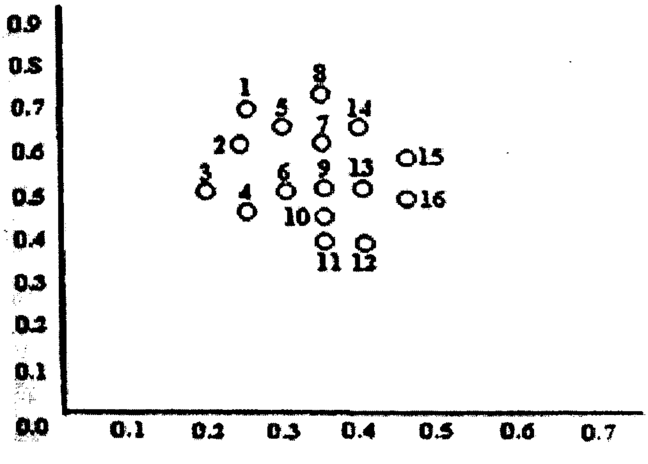
Ⅲ. Fuzzy-Entropy 집락분석

1. Fuzzy-Entropy 집락분석을 위한 목적함수

기존의 Fuzzy 집락분석의 목적함수는 개체들에게 각 집락에 속할 확률을 할당하여 주는 것을 궁극적인 목표로 하기 때문에 통계적인 또는 집단적인 의미에서 볼 때는 같은 집락에 소속되었다고 볼 수 있는 개체들마저도 수치적인 확률에서 차이가 드러나서 때로는 결과의 해석에 부정적인 요소로 작용될 가능성이 있을 수 있다. 엄밀히 말하면 소속인가 아닌가만 보여주는 기존의 집락분석이 지나치게 조악한 결론을 내리는 것을 보완하려는 의도에서 Fuzzy 집락분석

이 제안된 것이라고 생각할 때 이는 또한 긍정적인 면이라고도 할 수 있다. 그러나 집락분석의 본래 목적 중에 하나가 어떤 의미에서 동질적인 개체들이 형성하는 집락의 특성을 파악하는 것이라고 할 때 가능한한 분명히 개체들을 집락에 속하도록 하는 것이 바람직하다고 생각된다. 이에 본 논문에서는 다항분포에서 Entropy가 최소인 상황이 한 개체가 어느 집락에 확률 1로 소속되는 상태임을 고려하여 목적함수에 각 개체가 각 집락에 속하는 확률로 만들어지는 Entropy 항을 추가하였다. 따라서 목적함수는 두 가지 특성으로 구성되었는데, 목적함수의 최소화는 첫째 항에서 각 개체에 수리적으로 엄격하게 각 집락에 속하는 확률을 할당하는 역할을 하도록 하고, 동시에 둘째 항에서는 Entropy의 최소화를 통하여 집락분석의 본래 목적인 개체들을 가능한 한 한 집락에 분명히 속할 수 있게 한다. 이와 같이 일면 서로 상반된 두 성질이 주어진 자료에 따라서 균형을 갖고 결합할 수 있도록 하기 위하여 Entropy 항에 가중치를 사용하면서 다음과 같은 목적함수를 제안하였다.

$$S(g,r,y) = \sum_{i=1}^{n} \sum_{m=1}^{g} Y_{im}^{r} \sum_{k=1}^{p} (X_{ik}-Z_{mk})^{2} - \lambda \sum_{i=1}^{n} \sum_{m=1}^{g} Y_{im}*\log Y_{im} \qquad ...(3.1)$$



<그림 3-1> 개체들이 뭉쳐있는 경우

<그림 3-1>과 같은 자료들은 집락형성이 모호해서 분류가 어려운 자료인데 4장의 예에서 토의된 바와 같이 Fuzzy ISODATA 방법은 각 개체가 어느 집락

에 속할 확률의 정도가 모호한 데 비하여 Fuzzy-Entropy 목적함수를 이용한 확장된 MFA 알고리즘 방법으로는 집락에 속할 소속정도가 좀더 명확하게 나온다.

2. Fuzzy-Entropy 집락분석 알고리즘의 배경

Fuzzy ISODATA 알고리즘이 집락내의 각 자료들과 집락내의 중심값 간의 오차 제곱의 합을 근간으로 하는 Fuzzy 목적함수를 최소화시키는 좋은 알고리즘으로 사용되고 있고, 이 알고리즘에서는 각 집락의 중심값에 대한 초기화 설정이 매우 중요하며 이에 따라 전역적 해가 아닌 지역적 최소치에 수렴하는 단점이 있고, 밀집지역에서 멀리 떨어진 개체들이나 분리가 모호한 개체들에 대해서 어느 집락에 속할 정도가 명확하지 않을 때가 있다는 사실을 앞장에서 언급한바 있다.

본 연구에서는 기존의 Fuzzy 목적함수에 Entropy 항이 추가된 변형된 목적함수의 최소화가 요구되는데 Fuzzy ISODATA방법은 알고리즘의 특성상 추가된 Entropy항을 전혀 고려하지 않기 때문에 더 이상 사용될 수가 없다. 따라서 본 논문에서는 3.1절에서 제안된 목적함수의 최소화를 위한 알고리즘의 개발이 필요하게 되었다.

기존의 Fuzzy 목적함수의 최소화를 위한 Fuzzy ISODATA 알고리즘이 지닌 지역적 최소치의 문제를 극복하기 위하여 평균장 이론을 이용한 MFA 알고리즘을 적용하는 방법을 개발하여 두 알고리즘의 성능을 비교한 바 있는데 이 연구에서는 MFA 알고리즘이 Fuzzy ISODATA보다도 더 작은 함수 값을 제공하는 해(확률값)를 제공하였으나 그 차이가 그렇게 크게 나타나지는 않은 것으로 나타났다. 따라서 본 연구에서는 분류가 모호한 자료를 이용 MFA알고리즘의 적용을 고찰하고, 이 알고리즘을 3.1절에서 제안한 Fuzzy-Entropy 목적함수에 또한 적용시키는 방법을 개발하여 Fuzzy-Entropy 목적함수의 최소화를 위한 알고리즘으로 제안하였다. 최적화 문제의 관점에서 이 알고리즘의 특징은 반복수행 알고리즘(iterative improvement algorithm)이 지역적 최소값에 수렴할 수 있는 약점을 극복하는 방법을 포함하고 있는 것이다. 이 방법은 목적함수 값이 증가 될 수도 있는 약점을 극복하는 방법을 포함하고 있는 것이다. 이 방법은 목적함수 값이 증가될 수도 있는 변수값들을 적당한 확률을 갖고 택할 수 있도록 하는 생각을 알고리즘에 투입한 것이다. 따라서 반복 수행 알고리즘이 반복

적으로 작은 함수값을 제공하는 변수 값만을 택하기 때문에 지역적 최소해로 수렴하는 상황을 양의 확률값을 갖고 피할 수 있게 된다.

3. 최적화 연구

Fuzzy목적함수의 경우와 같이 연속적인 값을 갖는 변수들로 구성된 시스템에 적용하기 위한 확장된 MFA 알고리즘을 살펴보면, 변수값이 [0,1] 사이의 연속적인 값을 갖고 그 변수의 평균은 다음과 같이 계산된다.

$$\langle si \rangle = \frac{\int_0^1 s_i \exp(-H_i/T)\, ds_i}{\int_0^1 \exp(-H_i/T)\, ds_i} \qquad .....(3.2)$$

여기에서, H_i는 s_i를 제외한 나머지 변수들을 고정한 상태에서의 에너지

위 (3.2)식을 MFA 알고리즘을 적용하여 열평형 상태를 찾는 이완과정과 온도를 내리는 과정을 반복하여 수행하면 각 변수는 온도가 감소함에 따라 0부터 1 사이의 어느 한 값으로 수렴하게 된다. 그러나 실제로 (3.2)식을 계산하려면 연산과정에서 적분을 계산하기 어려울 뿐만 아니라, 연산수행시 부동소수점 오차가 발생되어 정확한 계산이 불가능하게 된다. 이와 같은 어려움을 극복하기 위하여 $\langle s_i \rangle$를 확률밀도함수

$$f(u) = \frac{\exp(-H_i(u)/T)\, du}{\int_0^1 \exp(-H_i(u)/T)\, du} \text{ 에서}$$

m개의 확률표본을 추출하여 그들의 표본평균으로 추정하는 방법을 사용하려고 한다. 그런데 이때 제기되는 문제인 f(u)를 따르는 표본을 추출하는 방법은 Metropolis가 다변량문제를 다룰 때 고려한 알고리즘을 일변량의 문제로 변형하여 사용하기로 하였다. 이 알고리즘을 정리하면 다음과 같다.

단계 1) 변수 I가 갖는 변수값 s_i를 변환하여 s_i'을 갖도록 한다.

단계 2) $H_i(s_i')$을 계산하고 이전 값 $H_i(s_i)$와 차이를 계산하여 변수 I가 변수

값 s_i에서 s_i'의 값을 다음과 같은 확률로 받아들이도록 한다.

$$Pr(s_i \dashrightarrow s_i') = 1 \qquad \text{만약, } H_i(s_i') \leq H_i(s_i)$$

$$= \exp\{-(Hi(s_i') - Hi(s_i))/T\} \qquad \text{만약, } Hi(s_i') < H_i(s_i)$$

단계 3) 단계 1)과 단계 2)를 여러 번 반복한다.

단계 4) 단계 1)에서 단계 3)의 과정에서 받아들이게 된 변수값들의 평균을 계산하고 그것을 주어진 온도 T에서의 $\langle s_i \rangle$로 간주한다.

이 알고리즘은 앞의 SA알고리즘과 같고 다만 새로운 변수값은 모든 변수값을 고정한 상태에서 한 변수를 변환하여 그것의 채택 여부를 결정한다. 이렇게 한 변수를 평균장하에서 변환하게 되면, 모든 변수를 변환시키는 종래의 방법에 비해 전체적으로 병렬로 움직일 수 있게 되어 안정된 해를 구할 수 있는 장점이 있다.

이와 같이 연속적인 변수값을 갖는 시스템에서의 평균값을 구하는 알고리즘을 이용한 확장된 MFA알고리즘을 다시 요약하면 다음과 같다. 여기서 알파는 변수값을 변화시킬 때 T가 변함에 따라 변화폭을 조정하기 위하여 사용하는 모수로 최대값 1에서 감소되도록 하며, ξ 는 -1과 1 사이의 난수이다. 이때 변환된 스핀 i의 값이 허용된 구간내에서 연속적인 값으로 표현되어야 하며 이에 따라 변수가 취할 수 있는 값의 영역의 경계에는 반사면이 있다고 가정하여 변환방식에 따라 변환된 값이 영역을 벗어날 경우 그 경계를 기준으로 반대편의 영역내로 대칭된 값을 택하도록 정한다.

(단계 1)

1) 각 변수가 가질 수 있는 값의 중간값 부근에서 약간의 노이즈(noise)를 첨가하여 초기화한다.

2) 초기온도(T_s)를 정한다.

(단계 2) 평형점이 발견될 때까지 반복한다.

1) a.$\langle s \rangle$로부터 임의의 한 개의 변수평균 값 $\langle s_i \rangle$를 선택한다.

b.$\langle s_i \rangle^{new} = \langle s_i \rangle^{old} + alpha * \xi$

c. $\Delta E \leq 0$ 이면 , $\langle s_i \rangle^{new}$ 채택

$\triangle E > 0$ 이면 , $\exp(-E/T) > random[0,1]$이면 $\langle s_i \rangle^{new}$ 채택
그렇지 않으면 채택하지 않는다.

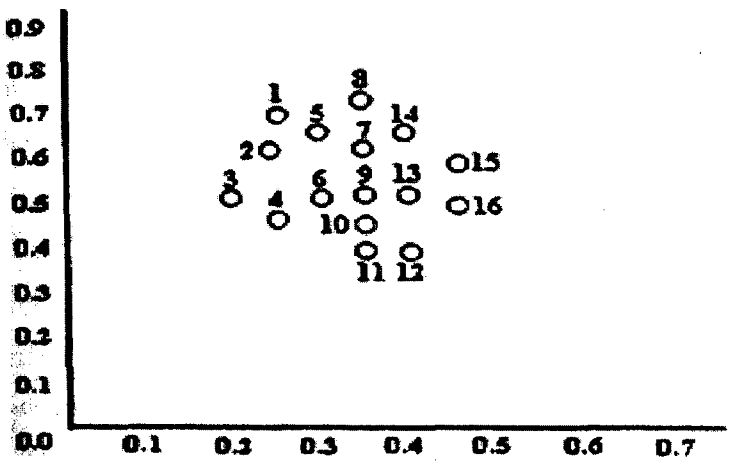
2) 1)을 반복하여 새로운 변수값의 평균 $\langle s_i \rangle^{new}$를 계산한다.

(단계 3) 온도 T를 감소시키고 최적해가 발견될 때까지 위의 (단계 2)를 반복한다.

Ⅳ. 비 교

3장에서 정의한 Fuzzy-Entropy 목적함수를 MFA 알고리즘으로 최소화(방법 1)하여 얻은 개체들이 각 집락에 속하는 소속확률값과 Fuzzy 목적함수를 MFA 알고리즘을 이용하여 최적화(방법 2)한 소속확률값 및 Fuzzy ISODATA 방법을 통하여(방법 3) 얻은 소속확률값을 개체들이 서로 여러 형태를 갖는 자료를 가지고 비교하였다.

1. 뭉쳐있는 개체들에 대하여 방법 1, 2, 3을 사용하여 두 집락으로 분류



<그림 4-1> 개체들이 뭉쳐 있는 경우

<표 4-1> <그림 4-1>의 개체들에 대하여 방법 1, 2, 3을 사용하여 두 집락으로 분류한 소속 정도

개체 번호	방법 / DATA		MFA(λ =0.002) Fuzzy=Entropy		MFA Fuzzy		Fuzzy ISODATA	
1	[0.250]	[0.670]	[0.914]	[0.085]	[0.892]	[0.107]	[0.611]	[0.388]
2	[0.250]	[0.600]	[0.902]	[0.097]	[0.873]	[0.126]	[0.623]	[0.376]
3	[0.200]	[0.530]	[0.585]	[0.414]	[0.588]	[0.411]	[0.520]	[0.479]
4	[0.250]	[0.470]	[0.307]	[0.692]	[00324]	[0.675]	[0.439]	[0.560]
5	[0.300]	[0.630]	[0.999]	[0.000]	[0.993]	[0.006]	[0.693]	[0.306]
6	[0.300]	[0.500]	[0.185]	[0.814]	[0.229]	[0.770]	[0.364]	[0.635]
7	[0.350]	[0.600]	[0.999]	[0.000]	[0.882]	[0.117]	[0.746]	[0.253]
8	[0.350]	[0.700]	[0.887]	[0.112]	[0.863]	[0.136]	[0.607]	[0.392]
9	[0.350]	[0.530]	[0.164]	[0.835]	[0.229]	[0.770]	[0.254]	[0.745]
10	[0.350]	[0.470]	[0.000]	[0.999]	[0.008]	[0.991]	[0.268]	[0.731]
11	[0.350]	[0.400]	[0.085]	[0.914]	[0.108]	[0.891]	[0.379]	[0.620]
12	[0.400]	[0.400]	[0.096]	[0.903]	[0.115]	[0.884]	[0.387]	[0.612]
13	[0.400]	[0.530]	[0.166]	[0.833]	[0.205]	[0.794]	[0.386]	[0.613]
14	[0.400]	[0.630]	[0.775]	[0.224]	[0.751]	[0.248]	[0.595]	[0.404]
15	[0.450]	[0.570]	[0.422]	[0.577]	[0.423]	[0.576]	[0.489]	[0.510]
16	[0.450]	[0.470]	[0.136]	[0.863]	[0.154]	[0.845]	[0.401]	[0.598]
	적함수 값		0.087		0.081		0.096	

2. λ값에 따른 변화

Fuzzy-Entropy 목적함수를 MFA알고리즘에 적용하여 소속정도를 알아볼 때 λ의 값에 따라 달라질 것임을 쉽게 예상할 수 있어 이에 대한 고찰을 해보고자 한다. 값에 따라 어떻게 소속정도가 달라지는 지를 알아보기 위해 <그림 3-1>의 자료를 바탕으로 아래와 같은 <표 4-2>를 만들었다. <표 4-2>에서 보듯이 λ의 값이 0.0001에서 구한 소속정도보다 0.001에서 구한 소속정도가 명확하고 0.01에서는 전통적인 집락알고리즘처럼 거의 0,1로 갈라지는 것을 볼 수 있다. 이것에 바탕을 두어 판단해 볼 때 불확실성의 정도를 나타내 주는 Entropy를 Fuzzy 목적함수에 추가해 줌으로써 불확실성을 해소시켜 개체들의 분류에 대하여 명확하게 해주었다.

<표 4-2> 값의 변화에 따른 소속정도의 비교표

data \ λ	λ = 0.0001	λ =0.001	λ = 0.01
[0.250][0.670]	[0.893][0.106]	[0.194][0.085]	[1.000][0.000]
[0.250][0.600]	[0.874][0.125]	[0.902][0.097]	[1.000][0.000]
[0.200][0.530]	[0.588][0.411]	[0.585][0.414]	[0.586][0.431]
[0.250][0.470]	[0.324][0.675]	[0.306][0.693]	[0.000][1.000]
[0.300][0.500]	[0.230][0.769]	[0.185][0.814]	[0.000][1.000]
[0.300][0.630]	[0.999][0.000]	[0.999][0.000]	[1.000][0.000]
[0.350][0.700]	[0.865][0.134]	[0.886][0.113]	[1.000][0.000]
[0.350][0.700]	[0.865][0.134]	[0.886][0.113]	[1.000][0.000]
[0.350][0.530]	[0.231][0.768]	[0.164][0.835]	[0.000][1.000]
[0.350][0.470]	[0.000][0.999]	[0.000][0.999]	[0.000][1.000]
[0.350][0.400]	[0.160][0.893]	[0.085][0.914]	[0.000][1.000]
[0.400][0.400]	[0.113][0.886]	[0.096][0.903]	[0.000][1.000]
[0.400][0.530]	[0.207][0.792]	[0.167][0.832]	[0.000][1.000]
[0.400][0.063]	[0.752][0.247]	[0.775][0.224]	[1.000][0.000]
[0.450][0.570]	[0.425][0.574]	[0.422][0.577]	[0.370][0.629]
[0.450][0.470]	[0.153][0.846]	[0.137][0.862]	[0.000][1.000]
목적함수의값	0.082	0.087	0.107

3. MFA방법으로 수행한 결과와 고전적인 K-means방법과의 비교

여기서는 <그림 3-1>의 자료를 가지고 고전적인 K-means방법으로 수행한 결과와 λ값을 0.01로 준 Fuzzy-Entropy 목적함수를 MFA알고리즘으로 수행한 결과를 비교하려 하는데 그 결과가 <표 4-3>이다.

<표 4-3>에서 보면 고전적인 K-means 방법은 1,0으로만 분리되고 Fuzzy-Entropy 목적함수를 MFA알고리즘으로 수행한 결과는 3,14번은 1,0으로 갈라지지 않음을 알수 있다. λ값을 0.01로 준 것은 불확실성을 많이 제거한 상태에서 집락분석을 수행한 것임에 비추어 볼 때 3,14번 개체는 소속정도가 원래 모호한 개체인데 본 논문에서 제안한 Fuzzy-Entropy 목적함수를 MFA알고리즘으로 수행한 방법이 이러한 고전적인 K-means 방법의 약점을 보완하여 보다 명확히 분류해 주는 방법임을 알 수 있게 해 주었다

<표 4-3> MFA방법으로 수행한 결과와 고전적인 K-means 방법과의 비교표

Data \ λ	MFA(λ =0.01) Fuzzy-Entropy	고전적인 K-means
[0.250][0.670]	[1.000][0.000]	[1.000][0.000]
[0.250][0.600]	[1.000][0.000]	[1.000][0.000]
[0.200][0.530]	[0.568][0.431]	[1.000][0.000]
[0.250][0.470]	[0.000][1.000]	[1.000][0.000]
[0.300][0.500]	[0.000][1.000]	[1.000][0.000]
[0.300][0.630]	[1.000][0.000]	[1.000][0.000]
[0.350][0.600]	[1.000][0.000]	[1.000][0.000]
[0.350][0.700]	[1.000][0.000]	[1.000][0.000]
[0.350][0.530]	[0.000][1.000]	[0.000][1.000]
[0.350][0.470]	[0.000][1.000]	[0.000][1.000]
[0.350][0.400]	[0.000][1.000]	[0.000][1.000]
[0.400][0.400]	[0.000][1.000]	[0.000][1.000]
[0.400][0.530]	[0.000][1.000]	[0.000][1.000]
[0.400][0.630]	[1.000][0.000]	[1.000][0.000]
[0.450][0.570]	[0.370][0.629]	[0.000][1.000]
[0.450][0.470]	[0.000][1.000]	[0.000][1.000]
목적함수의 값	0.170	0.588

V. 결　론

본 연구에서 제안한 Fuzzy-Entropy 집락분석은 Fuzzy집락분석보다도 더 분명히 개체가 어느 집락에 속하도록 분류하려는 것에 목적을 두고 있다. 또한 확장된 MFA알고리즘을 집락분석에 적용하여 지역적인 최적해에서 벗어나 근사적인 전역적 최적해가 되도록 하는 것도 또 하나의 목적이다. 이러한 목적하에서 나온 MFA방법에 바탕을 두어 해를 구한 Fuzzy-Entropy 집락분석의 특징은 다음과 같다.

첫째, 기존의 Fuzzy집락분석은 주로 Fuzzy ISODATA 방법을 사용하여 왔는바 이 방법은 집락들의 중간에 있는 개체들과 이상치에 대해서는 애매한 경우가 많았고, r이 1.2 , 1.4일 때는 잘 분리가 되지 않았는데 본 연구의 방법은 λ

값의 변화에 따라 분류를 명확하게 해주었다.

둘째, 분석하고자 하는 집단의 성격에 따라 λ의 값을 어느 정도 해 주는 것이 좋은지를 사용자가 분석 자료의 특성에 맞게 주관적으로 결정할 수 있다는 것이다. 즉 이 주관적 결정에 따라 어느 개체가 어느 집락에 속할 확률이 더 적당하게 나타나 타당한 결과를 주는지를 선택할 수 있다는 장점이 있다. 이것은 유사성 거리에 대한 주관적인 값의 기준에 따라 집락결과가 달라지는 기존의 집락분석 방법과 대응된다고 할 수 있다.

셋째, MFA방법이 지역적 최적해에서 벗어나 근사 전역적 최적해로 가기 때문에 기존의 Fuzzy ISODATA 알고리즘의 단점을 극복하고 있다.

앞에서 요약한 특징을 가지는 Fuzzy-Entropy 집락분석에 근거한 차후의 연구과제로 다음과 같은 것이 있다. 하나는 집락의 수를 늘려가며 λ의 관계를 알아보며 가장 적절한 집락의 수를 정할 수 있는 방법과 다음은 λ를 어떻게 적절하게 선정하는지에 대한 고찰이 필요하다고 판단되고 Fuzzy-Entropy 집락분석의 집락결과와 다른 집락분석의 집락 결과의 비교를 통한 이론적 정립도 연구할만한 과제라 할 수 있다.

參 考 文 獻

1. 김기영 · 전명식(1990), SAS군집분석, 자유아카데미
2. 임대혁 · 김민희 · 이원돈(1992), "확장된 MFA를 이용한 Fuzzy 집락함수" Chungnam Journal of Sciences Vol. 19, No. 2. Dec.
3. 이원돈(1988), "신경회로망의 배경 및 응용," 정보과학지 제6권 제6호, pp. 66~77.
4. 이원돈 · 이석훈(1989), "신경회로망 최적화 기법의 배경 및 응용," 전기학회지 제38권 제2호, pp. 23~30.
5. Arabie, P. and Douglas Carrol, Wayne Desarbo, and Jerry Wind(1981), "Overlapping Clustering : ANew Method for Product Positioning," *Journal of Marketing Research*, Vol. 18 (August 1981), pp. 310~317.
6. Anderberg, M. R.(1973), "Clustering Method of Ambiguous Representation," *Academic Press*, New York.
7. Anderson, J. R.(1983), "A Mean Field Computation Model for PDP," *Connectionist Models*, pp. 217~223.
8. Bezd다, J. C.(1974), "Numerical Taxonomy with fuzzy sets," J. Math. Biol., 1. pp. 57~71.
9. Bezd다, J. C.(1976), "A Physical interpretation of Fuzzy ISODATA," *IEEE Trans. on SMC*, pp. 387~389.
10. Bilbro, G., Mann, M., Miller, T.(1989), "Optimization by mean field annealing," In Advances in Neural Information processing systems.
11. Bout, D. E. and Huntrey, C. L.(1992), "Apractical application of simulated annealing to clustering," *Pattern Recognition*, Vol. 25, No. 4, pp. 401~412.

Abstract

A Study on an Extended Fuzzy Cluster Analysis

Im, Dae-heug

We consider the Fuzzy clustering which is devised for partitioning a set of objects into a certain number of groups by assigning the membership probabilities to each object. The researches carried out in this field before show that the Fuzzy clustering concept is involved so much that for a certain set of data, the main purpose of the clustering cannot be attained as desired.

Thus we propose a new objective function, named as Fuzzy-Entroppy Function in order to satisfy the main motivation of the clustering which is classifying the data clearly. Also we suggest Mean Field Annealing Algorithm as an optimization algorithm rather than the ISODATA used traditionally in this field since the objective function is changed.

We show the Mean Field Annealing Algorithm works pretty well not only for the new objective function but also for the classical Fuzzy objective function by indicating that the local minimum problem resulted from the ISODATA can be improved.