

성장곡선을 이용한 퍼지군집분석 기법의 연구

김응환¹⁾ 이석훈²⁾

요약

본 연구는 시간자료(Longitudinal data)의 분석을 위하여 Fuzzy k-means 군집분석 방법을 확장한 알고리즘을 제안한다. 이 논문에서 제안하는 군집분석방법은 각각의 개체에 대응하는 성장곡선에 Fuzzy k-means 군집분석의 알고리즘을 결합하는 것을 핵심 아이디어로 한다. 분석결과는 생성된 군집을 성장곡선모형으로 표현할 수 있고 또한 추정된 모형의 식을 활용하여 새로운 개체를 분류도 할 수 있음을 보인다. 그리고 이 군집분석방법은 아직 자라지 않은 나이 어린 개체가 미래에 어느 군집에 속할 것인가 하는 분류와 함께 이 개체의 향후 성장상태를 예측을 하는 데에도 적용이 가능하다. 제안된 알고리즘을 원숭이(macaque)의 상악동(maxillary sinus)의 자료에 적용한 실험을 보인다.

주요용어: 퍼지군집분석, 성장곡선, 시간자료.

1. 서론

우리가 관심을 가지고 있는 대상들을 어떤 관점에서 분류한다는 것은 인간의 판단과 인식체계에서 나타나는 가장 근원적이며 자연적인 현상중의 하나이다. Wolf(1926)는 “분류의 핵심은 어떤 사물들이 어떤 연유로 서로 관련되어져 있다라고 생각되어지는 바로 그것이다.”라고 말하면서 “이러한 분류는 과학분야에서 연구를 시작하려는 처음과 그 결론을 추론하기 위한 마지막에 사용되어지는 방법이며, 과학법칙의 증명은 이 분류가 완벽하게 이루어진 후에야 행해질 수 있다.”라고 분류의 중요성을 역설하였다.

통계학에서 군집분석(cluster analysis)은 일반적으로 대상들을 분류하는 기법들의 총칭으로 사용되는데 이 기법들은 Anderberg(1973), Cormark(1971)와 Gorden(1981) 등에 의하면 대상들간의 유사성 측정방법들과 대상들의 군집화 방법의 관점에서 계속적으로 개발 연구되어오고 있으며 이들의 연구에서는 특별히 다음 두 가지를 주목하였다.

첫째는, Hartigan(1975)와 Everitt(1977)의 설명과 같이 군집분석의 기존 연구들이 군집형성만을 주목적으로 하고 있기 때문에, 분류된 군집들의 특성을 활용하여 군집형성의 배후에 있는 군집형성 규칙을 수립하고자 하는 의도까지 포함하는 연구는 그렇게 활발하게 진행되지는 않고 있다. 이에 대하여 혹자는 군집분석을 필요로 하는 일반적인 상황이 군집화의 규칙을 요구하지 않는다고 주장할 수도 있지만 현실적으로는 그렇지 않은 경우를 쉽게 발견할 수 있다. 예컨대 정신병의 분류 같은 분야를 본다면 군집분석에서 나타난 군집

1) (314-701) 충남 공주시 신관동 182, 공주대학교 사범대학 수학교육과, 조교수

E-mail: yhkim@kongju.ac.kr

2) (305-764) 대전광역시 유성구 공동 220, 충남대학교 자연대학 통계학과, 교수

E-mail: shlee@stat.chungnam.ac.kr

을 명명(naming)하여 그 기준을 갖고 다시 환자를 구별하고 있는 경우나 이와 유사한 상황을 생각할 수 있다. 이에 대한 연구의 시도는 이석훈 등(1995)이 있다.

둘째는, 대부분의 시간자료(Longitudinal data)들의 군집분석이 마치 횡단자료인 것처럼 또는 통상의 다변량자료로 간주되며 수행된다는 점이다. 그 이유는 기본적으로 앞에서 언급한 바와 같이 군집분석을 단순한 군집형성이나 군집의 추출만을 목적으로 한 것이었기 때문에 종단(시간)적인 자료의 특성을 쉽게 무시하는데 연유한다고 생각된다. 물론 음성자료나, EEG(Electro Encephalo Graphic)자료의 분석에서 판별분석을 수행하고는 있으나 이들 방법론을 군집분석에서 사용하는 연구는 거의 없었던 것으로 알고 있다.

이와 같이 군집화 규칙의 수립과 자료의 종단성 활용이라는 두 가지 방면에 초점을 맞춘 본 연구는 종단자료를 군집화하여 각 군집을 대표하는 모형을 제시하고 이 모형을 이용하여 부분적인 종단자료를 갖고 있는 임의의 개체를 기존에 형성된 군집에 할당하는 알고리즘을 제안하였다. 그리고 이 알고리즘을 8년동안 관찰된 원숭이의 상악동(maxillary sinus)에 관한 실제 자료에 구체적으로 적용하는 방법과 초기자료만 있는 개체의 향후 성장양태를 예측하는 방법을 고찰하였다.

2절에서는 퍼지군집방법에 대한 기본 알고리즘과 자료형태를 살펴보고, 모형과 추론을 다루는 3절에서는 성장곡선을 따르는 시간자료에 관한 군집분석 알고리즘과 모수 추정과정을 제안한다. 끝으로 4절에서 46마리의 원숭이의 자료를 예제로 하여 우리의 새로운 기법을 적용하고 결과를 토론한다.

2. Fuzzy 군집분석과 자료

Ruspini(1970)가 처음 도입한 Fuzzy군집분석은 Zadeh(1977)와 Bezdek(1981)에 의하여 Fuzzy k-partition으로 불려지면서 집단의 각 개체들에게 그 집단 내에 존재하리라고 생각되는 k개의 군집에 각각 소속될 정도를 0과 1 사이의 실수로 하여 나타내어 그 개체에 소속 정도 벡터를 할당하는 것을 그 핵심으로 하는데 본 연구에서는 이 방법론을 종단자료에 적용하고자 한다. Rousseeuws(1995)는 0과 1 사이의 소속정도의 값을 확률로 이해하여도 좋다는 견해를 밝힌바 있다.

먼저 자료를 다음과 같이 나타내자. 개체 $i(i = 1, 2, \dots, N)$ 가 시점 $t_{ij}(j = 1, 2, \dots, p_i)$ 에서 관찰된 반응을 $y_{ij}(j = 1, 2, \dots, p_i)$ 라 하고, 시점과 관찰값 사이의 모형을 $y = f(t; \theta)$ 로 나타낸다. 여기에서 p_i 는 개체 i 가 관찰된 횟수이다. 이들 자료와 개체 i 가 군집 $m(m = 1, 2, \dots, K)$ 에 속하는 소속정도를 u_{im} 라고 하면 종단자료는 다음과 같이 표현된다.

개체	종단자료	소속정도
1	$(t_{11}, y_{11}) \cdots (t_{1p_1}, y_{1p_1})$	$(u_{11}, u_{12}, \dots, u_{1K})$
2	$(t_{21}, y_{21}) \cdots (t_{2p_2}, y_{2p_2})$	$(u_{21}, u_{22}, \dots, u_{2K})$
\vdots	\vdots	\vdots
N	$(t_{N1}, y_{N1}) \cdots (t_{Np_N}, y_{Np_N})$	$(u_{N1}, u_{N2}, \dots, u_{NK})$

이들 자료에 대하여 퍼지군집분석을 수행하기 위해서는 특별히 두 개의 모형이 정의되는데 하나는 앞에서 언급한 군집 m 에 소속된 개체들의 시간과 관찰값 사이의 관계를 나타내는 것으로,

$$y = f(t; \theta_m) \text{과} \quad \theta_m = \theta_{m1}, \theta_{m2}, \dots, \theta_{ml} \quad (m = 1, 2, \dots, K)$$

이고, 다른 하나는 군집분석의 목적함수 (Objective Function)로서 다음과 같다.

$$I(\underline{u}_1, \dots, \underline{u}_N, \theta_1, \dots, \theta_K) = \sum_{i=1}^N \sum_{m=1}^K \left[u_{im}^\alpha \left\{ \sum_{j=1}^{p_i} (y_{ij} - f(t_{ij}, \theta_m))^2 \right\} \right]$$

위의 증괄호는 개체 i 가 m 번째 군집에 소속되었다고 할 때 m 번째 군집의 성장모형에 적합된 값 $f(t_{ij}; \theta_m)$ 과 시점 t_{ij} 에서의 관찰값 y_{ij} 의 차이(error) (또는 잔차(residual))의 제곱합이기 때문에 위에서 정의된 목적함수는 소속정도 u_{im} 의 α 제곱값을 가중치로 갖는 잔차 제곱합의 가중합이라고 할 수 있다. 그리고 목적함수 I 를 최소화하는 방법은 일반적으로 Zadeh(1977)가 제안한 알고리즘을 기본으로 하여 이를 다양하게 변형하여 사용한다.

3. 모형과 추론

2절에서 언급한 시간과 관찰값 사이의 관계식 f 를 위한 모형으로는 성장곡선 분석에서 Lee. etc(1998)에서 활용하고 있는 로지스틱 성장곡선 모형을 사용한다. 특별히 이 모형은 본 연구에서 고찰하는 원숭이 자료로부터 성장속도의 이상치의 발견을 연구했던 Koppe(1999)등이 사용한 바 있다.

군집 m 의 성장곡선의 로지스틱 모형은 다음과 같이 정의된다.

$$f_m(t_{ij}) = \frac{a_m}{1 + b_m e^{-k_m t_{ij}}}$$

여기서 주목할 것은 모형에 나타난 모수인데, a_m 은 군집 m 의 반응변수값의 상한의 의미를 갖게 되고 b_m 은 군집 m 의 초기의 크기와 관계되고, k_m 은 군집 m 의 성장곡선의 경사(steeptness)와 관계가 된다. 따라서 2절에서 제시한 목적함수 I 는 u_{im} 의 지수 α 를 퍼지군집분석에서 일반적으로 사용하고 있는 2를 이용하면 다음과 같이 나타낼 수가 있다.

$$I(\underline{u}_1, \dots, \underline{u}_N, \theta_1, \dots, \theta_K) = \sum_{i=1}^N \sum_{m=1}^K \left[u_{im}^2 \left\{ \sum_{j=1}^{p_i} \left[y_{ij} - \frac{a_m}{1 + b_m e^{-k_m t_{ij}}} \right]^2 \right\} \right]$$

이때 한가지 추가적인 가정은 각 군집의 성장곡선이 시간 0 에서는 동일한 값을 갖는다는 것이다. 즉, 모든 개체는 시작점에서 동일한 값을 갖는다고 가정하기로 한다. 이와 같이 가정하면,

$$\frac{a_1}{1 + b_1} = \frac{a_m}{1 + b_m} \text{로부터} \quad a_m = a_1 \frac{1 + b_m}{1 + b_1} \quad (m = 1, 2, \dots, K)$$

이 되어 $(K - 1)$ 개의 모수가 줄어들게 되며 목적함수 I 는 다음과 같이 나타난다.

$$I(\underline{u}_1, \dots, \underline{u}_N, \underline{\theta}_1, \dots, \underline{\theta}_K) = \sum_{i=1}^N \sum_{m=1}^K \left[u_{im}^2 \left\{ \sum_{j=1}^{p_i} [y_{ij} - a_1 \frac{1+b_m}{1+b_1} \frac{1}{1+b_m e^{-k_m t_{ij}}}]^2 \right\} \right]$$

목적함수 I 를 최소로 하는 과정을 $K = 2$ 인 경우에서 보면, 이것은 목적함수를 최소화 하는 a_1, b_1, k_1, b_2, k_2 와 u_{i1} ($u_{i2} = 1 - u_{i1}$ 임, $i = 1, 2, \dots, N$)들을 구하는 다음과 같은 과정이다.

단계1) u_{i1} ($i = 1, 2, \dots, N$)의 초기값 결정단계:

$(-\frac{1}{10}, \frac{1}{10})$ 사이의 일양분포를 따르는 난수 r 을 발생하여 $(0.5 + r)$ 을 u_{i1} 의 초기값으로 한다. 즉 각 개체의 초기 소속정도는 0.5 근처의 값으로 정한다. 이때 u_{i2} 는 $1 - u_{i1}$ 이 되도록 한다.

단계2) u_{i1} ($i = 1, 2, \dots, N$)값을 $u_{i1}^{(0)}$ 라고 할 때 목적함수를 a_1, b_1, k_1, b_2, k_2 의 함수로 보고 Newton-Raphson의 반복알고리즘을 이용하여 함수를 최소화하는

$$a_1^{(0)}, b_1^{(0)}, k_1^{(0)}, b_2^{(0)}, k_2^{(0)}$$

를 구하고, 이때의 목적함수의 값을 I_0 라고 놓는다.

단계3) $m = 1, 2$ 에 대하여

$$D_{im}^{(0)} = \sum_{j=1}^{p_i} [y_{ij} - a_1^{(0)} \frac{1+b_m^{(0)}}{1+b_1^{(0)}} \frac{1}{1+b_m^{(0)} e^{-k_m^{(0)} t_{ij}}}]^2$$

을 구한다. 여기서 $D_{im}^{(0)}$ 은 단계2)에서 추정된 m 군집의 성장곡선과 개체 i 의 관찰치 사이의 잔차의 제곱합이다.

단계4) D_{i1} 과 D_{i2} 를 이용하여 $u_{i1}^{(0)}$ 을 $\frac{D_{i2}^{(0)}}{D_{i1}^{(0)} + D_{i2}^{(0)}}$ 로 놓는다.

단계5) 단계4)에서 구한 $u_{i1}^{(0)}$ 을 주어진 값으로 하여 목적함수 I 를 최소화하는 $a_1^{(1)}, b_1^{(1)}, k_1^{(1)}, b_2^{(1)}, k_2^{(1)}$ 를 Newton-Raphson의 반복알고리즘으로 구하고 이때의 목적함수 값을 I_1 이라고 한다.

단계6-1) 주어진 ϵ 값에 대하여 $\frac{|I_1 - I_0|}{|I_0|} < \epsilon$ 의 조건을 만족하면, 반복을 정지하여 $u_{i1}^{(1)}$ 과 $a_1^{(1)}, b_1^{(1)}, k_1^{(1)}, b_2^{(1)}, k_2^{(1)}$ 를 수렴된 값으로 하고, 이들 값으로부터 각 군집을 대표하는 성장곡선과 개체 i 의 소속정도 ($u_{i1}^{(1)}, u_{i2}^{(1)}$)을 결정한다.

단계6-2) 주어진 ϵ 값에 대하여 $\frac{|I_1 - I_0|}{|I_0|} < \epsilon$ 의 조건을 만족하지 못하면, 단계2), 3)과 같이 $m = 1, 2$ 에 대하여

$$D_{im}^{(1)} = \sum_{j=1}^{p_i} \left[y_{ij} - a_1^{(1)} \frac{1 + b_m^{(1)}}{1 + b_1^{(1)}} \frac{1}{1 + b_m^{(1)} e^{-k_m^{(1)} t_{ij}}} \right]^2$$

을 구하여 새로운 소속정도의 값 u_{ui} 을 $\frac{D_{i2}^{(1)}}{D_{i1}^{(1)} + D_{i2}^{(1)}}$ 로 놓고, 단계2)로 부터 단계6)까지를 반복 수행한다.

4. 예제

4.1. 모수의 추정

본 연구에 사용된 자료는 암수 각각 23마리인 총 46마리의 원숭이에 있어서 어금니와 광대뼈 사이에 있는 구멍-상악동-의 면적을 측정 한 자료이다. 이들은 출생에서부터 약 8세 까지에 대하여 p_i 번씩 반복 측정되었다. 이들 자료를 활용하여 제안한 군집분석 기법을 적용하여 이들 46마리 사이에 나타나는 대표적인 2개의 성장 유형을 탐색한 결과는 다음과 같이 2개의 성장곡선으로 나타났다.

$$f_1(t) = \frac{4.534}{1 + 4.694e^{-0.358t}}, \quad f_2(t) = \frac{2.662}{1 + 2.350e^{-0.525t}}$$

이 두 곡선과 원자료를 암수로 구별하여 그래프로 나타낸 것이 그림 4.1과 같다.

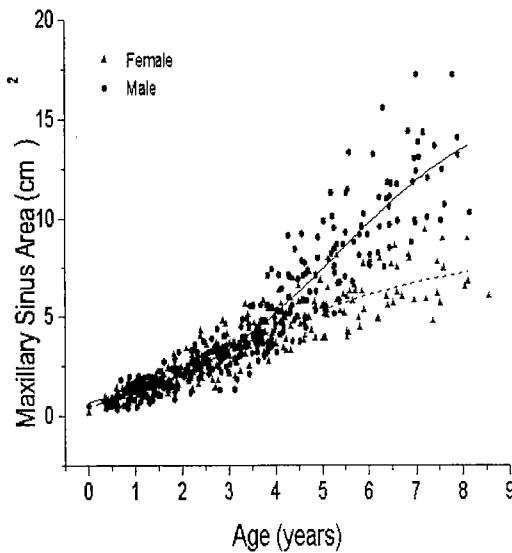


그림 4.1: 중단자료와 적합된 성장곡선

4.2. 군집의 특성

실제로 원숭이의 자료는 23마리의 암컷과 23마리의 수컷의 관찰값이었으나, 이들의 성별에 관한 정보는 모르는 것으로 하여 사용하지 않고, 시점 $t = j$ 와 관찰치 y_{ij} 만을 사용하여 우리가 제안한 알고리즘에 따라 구한 각 개체의 소속정도 u_{im} 의 값을, 기존에 알고있는 성별 정보인 암컷과 수컷으로 분류하여 분석한 결과는 다음의 표와 같다. (u_{i2} 는 $1 - u_{i1}$ 이므로 u_{i1} 에 관한 요약치만을 제시한다.)

표 4.1: 소속 정도 u_{i1} 의 기술통계

u_{i1}	개체수	평균	표준편차
수컷	23	0.7393	0.1808
암컷	23	0.2389	0.1773

위 표에서 수컷인 개체들이 집단1에 소속되는 정도 u_{i1} 값의 평균이 0.73이고 표준편차는 0.1808인데 비하여 암컷의 평균은 0.23이고 표준편차는 0.17을 갖는 것으로 나타났다. 따라서 군집1은 수컷과 비슷한 성장 유형을 갖는 개체들로 구성되어 있다고 할 수 있다.

여기서 우리는 소속 정도의 경계를 0.5로 정하고 각 개체의 소속정도가 0.5보다 크면 군집 1에 할당하고 0.5보다 작으면 군집2에 할당을 하였다. 그리고 그 분류된 개체들을 알려져 있는 성별에 따라서 분류한 결과를 다음 표 4.2에 나타내었다.

표 4.2: 성별과 군집과의 관계

	$u_1 < 0.5$	$u_1 > 0.5$	합 계
수컷	4(8.7%)	19(41.30%)	23(50%)
암컷	20(43.87%)	3(6.52%)	23(50%)
합계	24(52.17%)	22(42.83%)	46(100%)

이 결과를 바탕으로 해석을 하면, 군집2는 23마리의 암컷 중에 20마리의 암컷을 포함한 것을 알 수 있다. 또한 수컷은 23마리 중 19마리가 군집1에 속하는 것으로 나타났다. 이것은 u_{i1} 값의 요약에서 언급한 바와 같이 우리의 새로운 군집분류 알고리즘이 원숭이의 자료를 암컷과 수컷의 성장유형을 중심으로 하여 두 군집으로 분류하고 있는 것을 알 수 있다.

4.3. 개체의 미래예측의 가능성 탐색

우리는 여기서 제안한 알고리즘을 가지고 자료를 각 군집으로 분류만 하는 것이 아니고, 군집을 형성할 때 사용된 군집화 규칙을 이용하여 새로운 개체를 분류 예측하는 시도를 하여 그 가능성을 다음과 같이 확인하였다.

암컷 중에서 2마리와 수컷 중에서 랜덤하게 2마리를 제외하고 나머지 42마리를 가지고 군집분석을 수행한 후에, 이때 추정된 2개의 성장곡선모형을 활용하여 미리 제외했던 4마리의 개체를 할당한다. 구체적으로는 4마리 개체 각각의 D_{i1} 과 D_{i2} 를 계산하여 u_{i1} 을 구하고, 이 값이 0.5보다 큰 값인지를 확인한다. 이러한 과정을 100번 반복한 결과가 다음 표 4.3과 같다.

표 4.3: 모형의 예측가능성 평가

	$u_1 < 0.5$	$u_1 > 0.5$	합 계
수컷	25(6.25%)	175(43.75%)	200(50.0%)
암컷	164(41.00%)	36(9.00%)	200(50.0%)
합계	189(47.25%)	211(52.75%)	400(100%)

수컷 200마리 중 175마리를 수컷으로, 암컷 200마리 중 164마리를 암컷으로 분류하여 오분류율은 15.25%을 얻었다. 이것은 표본의 개체가 자신이 소속된 집단을 알면서 분류규칙을 생성한 판별규칙이 아니라, 군집분석 모형을 직접 활용하였다는 면에서 좋은 결과라고 판단된다.

엄격한 검증의 결과는 아니지만 이 실험을 통하여 우리의 새로운 군집분석방법이 예측을 위한 기초로서 사용 가능하다고 보고 다음과 같은 실험을 시도하였다.

예를 들어 아직 성장하지 않은 어린 원숭이가 몇 개밖에 안 되는 관찰 자료를 가지고 있다고 생각해보자. 이 어린 원숭이가 미래에는 과연 어느 군집에 속할 것인지를 예측해 보고자 하는 것이다. 우리의 아이디어는 이들 원숭이의 잔차의 제곱합을 계산하여 소속 정도를 얻는 것이다.

실제로 2살짜리 원숭이의 어린 시절의 처음 5개의 관찰치를 가지고 예측을 시도하였다. 그 결과 이 개체의 예측된 소속 정도 $u_1 = 0.61$ 이었다. 우리는 이 수치를 이 어린 원숭이가 미래에 군집 1에 소속할 정도로 보고 이 개체의 향후 성장을 군집1의 성장곡선 f_1 을 따라서 예측하고자 한다. 이 결과의 그림은 다음과 같다.(그림 4.2에 나타난 점선은 이 원숭이의 실제 측정자료를 나타낸 것이다.)

5. 결론

우리가 제안한 새로운 군집분석 알고리즘은 시간자료에 대한 분석을 위하여 Fuzzy k-means 방법을 확장한 군집분석방법이다. 이 연구에서 우리는 거리에 대한 척도로서 군집 중심으로부터의 개체간의 유클리디안 거리 대신에 성장곡선으로부터의 잔차 제곱합을 개체와 성장곡선간의 거리로 이용하였다는 것이다. 그리고 이를 바탕으로 시간자료에 관한 새로운 군집분석 알고리즘을 제시하였다.

우리는 이 새로운 방법을 원숭이의 상악동 자료에 적용하였는데 여기서 로지스틱 성장곡선 모델을 사용하였고 군집의 수를 사전에 2개로 고정하였다.

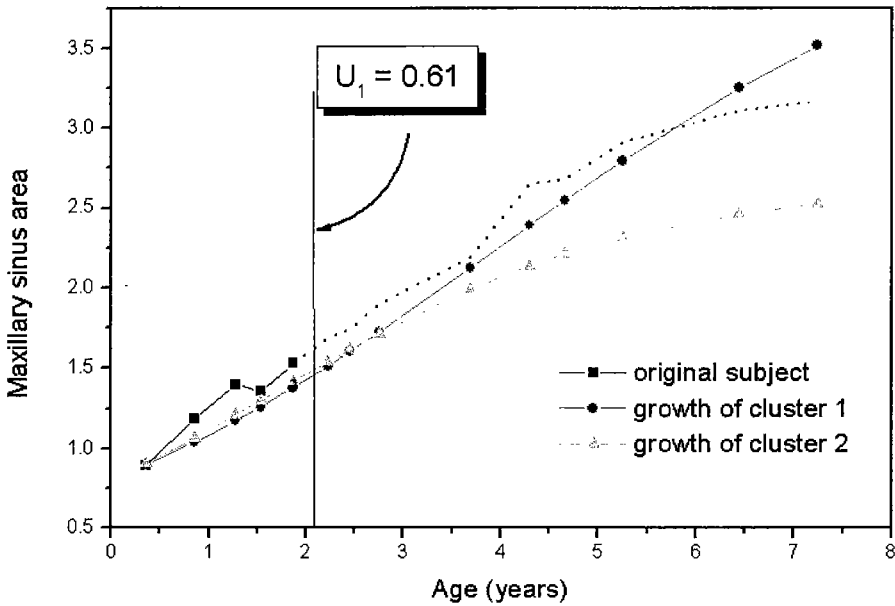


그림 4.2: 2세 원송이의 실제값과 예측값

군집분석 결과 나타난 두 개의 군집은 암컷 집단과 수컷집단에 가까운 것으로 판명되었다. 특히 우리는 새로운 성장곡선을 이용한 군집분석방법은 아직 자라지 않은 어린 시절의 초기의 몇 개의 관측치를 통하여 미래에 어느 군집에 속할 것인지를 예측하고 또한 이를 활용하여 그 개체의 향후 성장 상태를 예측하는 하나의 방안을 제공하였다.

참고문헌

- [1] 이석훈, 박래현, 김응환 (1995). 클론네트워크를 이용한 집락분석, <응용통계연구>, 제 8권 1호 pp.39-50, 한국통계학회
- [2] Anderberg, M.R. (1973). *Cluster analysis for applications*, Academic Press, New York.
- [3] Bezdek, J.C (1981). *pattern recognition with fuzzy objective function algorithm*, Plenum, New York.
- [4] Cormack, R.M. (1971). A review of classification, *Journal of the Royal statistical society*, series A, **134**, 321-367.
- [5] Everitt, B.S. (1977). *cluster analysis*, Heinemann Educational Books.
- [6] Gorden, A.D. (1981). *classification*, Chapman and Hall, New York.

- [7] Hartigan, J.A. (1975). *clustering algorithms*, John Wiley & Sons, New York.
- [8] Koppe, T., Swindler, D.R. and Lee, S.H. (1999). A Longitudinal Study of the growth pattern of the Maxillary Sinus in the Pig-Tailed Macaque(*Macaca nemestrina*), *Folia Primatologica*, vol **70**, no.6, pp301-312.
- [9] Ruspini, E.H. (1970). Numerical methods for fuzzy clustering, *Information Sciences*, **2**, 319-350.
- [10] Lee, S.H., Koppe, T. and Otake, M. (1998). A longitudinal study on the growth pattern of cranial structures of the pig-tailed macaque, *Proceedings of the Japanese society of applied statistics*, pp.141-144.
- [11] Rousseeuw, P.J. (1995). Discussion : Fuzzy Clustering at the Intersection, *Technometrics*, Vol. **37**, No. 3, pp. 283-286
- [12] Wolf, A. (1926). *Essentials of scientific method*, Macmillan company, New York.
- [13] Zadeh, L.A. (1977). *Fuzzy sets and their Application to pattern Classification and Clustering analysis*, In *Classification and Clustering*, ed., J. Van Ryzin, New York, Academic Press, 251-299

[2001년 4월 접수, 2001년 8월 채택]

A Study of the Fuzzy Clustering Algorithm using a Growth Curve Model

Yung-hwan Kim¹⁾ Sukhoon Lee²⁾

ABSTRACT

This paper proposed an extended algorithm of the fuzzy k-means clustering for longitudinal data analysis. The basic idea we had is to apply the fuzzy clustering algorithm to the growth curves corresponding to the individual subjects.

The algorithm proposed provided an expression of the groups with growth curves which enable a new subject to be classified. And also we used this method to predict future growing pattern of a subject who has only a few early observations.

As an illustration, an analysis of maxillary sinus data from monkey macaque has been performed.

Keywords: Fuzzy clustering; Growth curve; Longitudinal data.

1) Associated Professor, Department of mathematics education, Kongju National University.

E-mail: yhkim@kongju.ac.kr

2) Professor, Department of Statistics, Chungnam National University.

E-mail: shlee@stat.chungnam.ac.kr