

국부 확률을 이용한 데이터 분류에 관한 연구

A Study on Data Clustering Method Using Local Probability

이재국*, 손창호, 최원호
(Jae-Kook Lee, Chang-Ho Son, and Won-Ho Choi)

Abstract : In this paper, we propose a new data clustering method using local probability and hypothesis theory. To cluster the test data set, we analyze the local area of the test data set using local probability distribution and decide the candidate class of the data set using mean, standard deviation and variance etc. To decide each class of the test data, statistical hypothesis theory is applied to the decided candidate class of the test data set. For evaluating, the proposed classification method is compared to the conventional fuzzy c-mean method, k-means algorithm and Discriminator analysis algorithm. The simulation results show more accuracy than results of fuzzy c-mean method, k-means algorithm and Discriminator analysis algorithm.

Keywords : local probability, classification, data clustering, hypothesis theory

I. 서론

데이터 클러스터링이란 다양한 클래스를 가지는 데이터를 유사한 특성을 가진 하나의 데이터 클래스영역으로 그룹화 하는데 있다. 공통된 특성을 추출하거나 정의하여 각 그룹별로 분류하는 작업 또한 데이터 클러스터링의 과정에 속한다. 이러한 기존의 데이터 클러스터링의 방법으로는 데이터의 기하학적 특성을 고려한 최소거리를 이용한 방법과 최대-최소법, 퍼지이론을 이용한 C-평균법[1,2], 그리고 학습벡터 양자화와 같은 방법이 있으며, 또한 데이터의 분포형태를 정규 분포라 가정하여 접근하는 통계적인 데이터 분류법 등을 이용하고 있다[3-5]. 그러나 기하학적 특성인 최소거리를 이용한 방법들은 데이터의 처리가 비교적 간단하고 데이터 그룹의 수에 관계없이 쉽게 데이터 클러스터링을 수행하기에 많이 이용되고 있지만 데이터의 특성과 중요도 및 각 변수들간의 상호 상관관계를 고려하지 않음으로 인해 임의로 데이터가 분류되는 단점을 가지고 있다. 또한 확률에 근거한 통계적 방법들은 클러스터링 인식률을 높이기 위해 데이터 구성요소의 척도를 통일시키는 정규화 과정을 가지고 있으며 또한 데이터에 관한 모든 확률 값을 미리 알고 있다고 가정하여 보다 정확한 데이터 분류가 어렵고, 전문가적 지식이 쉽게 사용되지 못하거나 수학적 모델에 대한 지식이 부족한 데이터의 경우 그 적용이 매우 어렵다[3]. 또한 데이터의 경계가 모호한 패턴이나 데이터의 경우 그 그룹화는 상당히 어렵거나 다른 클러스터에 그룹화 되는 문제점들을 가지고 있다 [1,6].

Ming와 Pal 등은 실험 데이터의 기하학적 특성을 고려하여 초기 클러스터의 중심을 실제 데이터에서 결정하여 시스템의 훈련과정을 줄이는 Fuzzy C mean 알고리즘을 제안하는 방법과 Fuzzy C means 알고리즘의 소속함수의 중요성을 고려하

여 클러스터링 하는 방법을 제안하였다[1,2]. 또한 Frakt와 Kato 등은 계층적 검정이론을 이용하여 비정상적인 데이터를 검출하는 방법을 제안하였으며 검정이론에 퍼지이론을 적용한 방법을 제안하였다[3,4,7].

본 논문에서는 기하학적 특성을 이용한 데이터간의 유클리드 거리(euclidean distance) 분포도를 만들고 이를 이용하여 각 데이터의 상호 관계 및 개별 데이터의 중요성을 인식하여 국부확률이론 및 검정이론(hypotheses theory)을 통하여 보다 높은 신뢰성을 가지는 데이터 분류 방법을 제안한다.

또한 실험에서는 K-means 기법과 Fuzzy C-mean 방법 및 판별식 분석 알고리즘(discriminant analysis algorithm)들을 사용하여 Iris 데이터와 실험을 위해 만들어진 임의의 데이터를 이용하여 모의 실험을 행함으로써 국부확률 및 검정이론을 기반으로 한 데이터 분류 성능을 서로 비교하였다.

본 논문의 구성은 2장에서는 논문에 적용된 국부확률이론 및 검정이론에 대하여 기술하였으며 3장에서는 제안한 알고리즘에 대하여 기술하였고, 4장에서는 실험데이터의 특성 및 결과에 대하여 기술하였으며, 그리고 5장에서는 본 논문의 결론 및 향후 과제에 대하여 기술하였다.

II. 국부 확률 및 검정이론

1. 국부확률(local probability)

국부확률이란 두 가지 사건이 동시에 일어날 경우 각 사건에 대한 확률 비를 각각 계산하는 것을 말한다[8].

사건 A와 B가 동시에 일어날 경우 각 모수 x, y 에 대한 확률은 (1)와 같이 구할 수 있다.

$$Prob[x = x_i | y = y_j] = \frac{Prob[x = x_i, y = y_j]}{Prob[y = y_j]} \quad (1)$$

여기서 $Prob[x = x_i | y = y_j]$ 는 모수 y 에 대한 조건부 확률이다.

위 식으로부터 전체 확률을 (2)와 같이 구할 수 있다.

$$P_x(x_i) = \sum_{j=1}^d P_{(x|y)}(x_i | y_j) P_y(y_j) \quad (2)$$

각 확률 P_x 는 (3)에서 나타낸 바와 같이 두 가지 조건을

* 책임저자(Corresponding Author)

논문접수 : 2005. 11. 11., 채택확정 : 2006. 9. 26.

이재국, 최원호 : 울산대학교 전기전자정보시스템공학부

(alf037@hanmail.net/whchoi@ulsan.ac.kr)

손창호 : 삼창기업(주)(sohnch@korea.com)

※ 본 연구는 산업자원부 · 울산광역시 지원 울산대학교 네트워크 기반 자동화연구센터와 울산대학교 전기전자정보시스템공학부 BK 21사업단의 지원에 의한 것임.

만족하여야 한다.

$$P_x \geq 0, \text{ 그리고 } \sum_{x=1}^m P_{xk} = 1 \quad (3)$$

여기서 k는 분할된 구간을 의미하며, P_{xk} 는 확률 P_x 의 국부 확률을 의미한다.

본 논문에서는 (4)와 같은 소속 정도를 판별할 수 있는 소속함수를 사용하였다.

$$MF(\mu_i) = \frac{\sum_i^k \tilde{\mu}_i(y_i) \left(\frac{1}{\|x - y_i\|^2} \right)}{\sum_j^k \left(\frac{1}{\|x - y_j\|^2} \right)} \quad (4)$$

2. 검정이론(hypothesis theory)

검정이론이란 표본으로부터 주어진 정보를 이용하여 모수에 대한 예상, 주장 또는 단순한 가정의 가부를 판정하는 과정이다. 그림 1은 검정이론의 식별원리를 위한 1차원 두 클래스 식별에서의 최적 경계를 결정하는 그림이다. 전체 평균 오차확률은 아래 (5)에서 보인바와 같이 구할 수 있다.

$$P_e = P_{e1}P_{w1} + P_{e2}P_{w2} = \int_{D2} P(x|w_1)P(w_1)dx + \int_{D1} P(x|w_2)P(w_2)dx \quad (5)$$

여기서 D1 및 D2는 클래스 w1과 w2에 대응되는 영역임을 나타내고 결정경계를 구분 지을 수 있는 threshold값을 설정하여야 한다.

각 데이터의 표준편차(σ), 평균(μ)을 구하고 이를 이용하여 확률분포의 검정은 즉, 표본평균 (\bar{X})을 기준으로 유의수준인 T를 다음과 같은 (6)에 의하여

$$T = P(\bar{X} \geq \text{Threshold} | \mu = \mu_0) = P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \geq \frac{\text{threshold} - \mu_0}{\sigma/\sqrt{n}} \right) \quad (6)$$

구하고 일정 값 이상일 때 결정하도록 한다. 여기서 Threshold는 실험적 결과에 따른 적절한 상수 값이 되고, n은 데이터의 개수, 그리고 μ_0 는 데이터의 평균값을 의미한다.

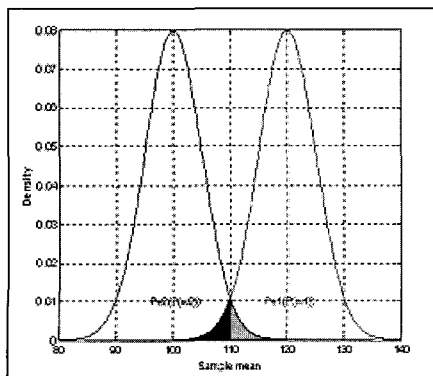


그림 1. 클래스 구분을 위한 경계 결정.
Fig. 1. Threshold decision for classifying.

III. 제안한 알고리즘

국부 확률 및 검정이론을 이용하여 데이터를 분류 하기 위해 본 논문에서는 데이터 간의 거리분포를 이용하였다. 데이터의 거리분포를 이용하여 데이터를 임의의 그룹으로 그룹화 한 후 임의의 그룹과 다른 그룹들 간의 확률 비를 각각 계산하여 그 소속 정도를 구하였다.

그림 2는 실험에 사용 되어진 하나의 데이터의 거리 분포를 나타낸 것으로 거리를 측정하기 위하여 (1)에서 보여준 유클리드 거리로 구하였다.

임의의 d차원의 경우 두 데이터의 집합을 $x_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{id})^T$ 와 $x_j = (x_{j1}, x_{j2}, x_{j3}, \dots, x_{jd})^T$ 일 경우, 유클리드 거리 (Euclidean Distance)는 (7)과 같이 구할 수 있다.

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^d |x_{ik} - x_{jk}|^2} \quad (7)$$

이 거리는 두 데이터 간의 벡터와 벡터간의 거리이다. 전체 데이터의 거리분포를 구하고 이를 바탕으로 클러스터가 세 개인 경우 처음 데이터와 가장 가까운 데이터와 가장 먼 데이터 그리고 중간 데이터를 찾아서 후보그룹을 설정하였다.

그림 2는 실험에 사용되어진 임의의 데이터 집합 중 (7)을 이용하여 데이터 거리를 구한 그림을 나타내었다. x축은 데이터의 수를 의미하며 y축은 데이터 간의 거리를 나타내었다.

그림 3은 데이터 분류를 위한 전체적인 절차를 나타내었다. 데이터를 분류 하기 위해 먼저 데이터 간의 거리를 유클리드 거리를 이용하여 만들고, 데이터의 클래스를 결정한 후

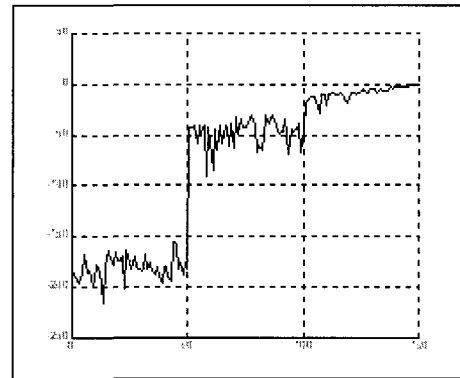


그림 2. 데이터의 거리 분포.
Fig. 2. The distance distribution of data set.

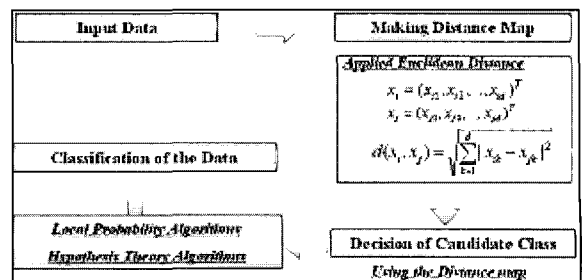


그림 3. 데이터분류를 위한 알고리즘.
Fig. 3. The algorithm of data clustering.

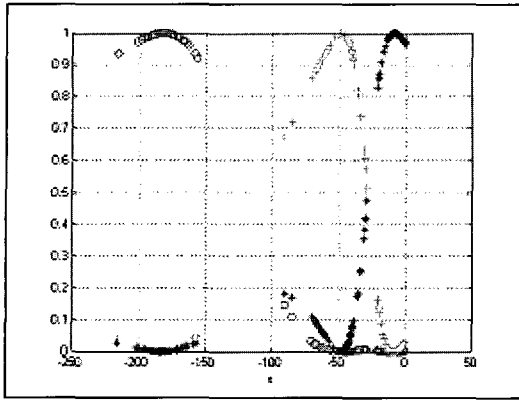


그림 4. 데이터의 국부확률분포.
Fig. 4. The local probability distribution of data set.

최적 경계를 결정하기 위해 각 개별 데이터의 중요도와 데이터 간의 상호관계를 고려 할 수 있는 후보 영역을 설정하였다. 후보 영역을 결정하기 위해 최소-최대 거리를 기준으로 하여 두 클래스를 만들고 이를 기준으로 하여 중앙의 클래스를 설정하는 방법으로 후보 영역을 결정하였으며 국부 확률 분포를 (4)와 (5)에 나타난 확률 분포와 소속 함수를 이용하여 확률 분포도를 계산하였다. 오차를 최소화 할 수 있도록 결정 경계를 (6)에서 나타난 바와 같이 경계를 결정하고 검정이론을 이용하여 각 클래스를 분류하도록 하였다.

본 논문에서 제안한 국부 확률 및 검정이론을 이용한 데이터 분류방법의 알고리즘을 정리하면 다음과 같다.

- 단계 1: 일반화된 입력 데이터의 유클리드 거리 (euclidean distance) 분포도를 (1)를 이용하여 계산한다.
- 단계 2: 계산되어진 거리 분포도를 이용하여 후보 클래스의 중앙값을 결정한다.
- 단계 3: 각 후보클래스의 중앙값을 기준으로 하여 국부확률 분포를 구하고 소속 함수를 이용하여 확률 비를 (3)과 (4)를 이용하여 계산한다.
- 단계 4: 클래스 구분을 위한 경계를 결정하기 위하여 검정이론을 이용하여 결정경계를 결정한다.
- 단계 5: 개별 클래스를 구분하여 데이터 분류를 수행한다.

그림 4는 임의의 실험 데이터의 소속함수(member-ship function)를 (4)를 이용하여 그 정도를 나타낸 그림이다.

각 데이터의 국부확률 값은 (3)에서 나타난 바와 같은 소속 정도가 0보다는 커야 되며 전체 소속의 합이 1이 되어야 한다는 두 조건을 만족해야 한다.

IV. 실험 및 결과

그림 5는 데이터 분류와 패턴 인식에 많이 연구되어지는 Iris data를 이용한 데이터 분류 결과를 제안한 알고리즘을 이용하여 분류한 결과를 나타내었다. Iris data는 Iris setosa, Iris versicolor, 그리고 Iris virginica의 3개의 종으로 나뉘어지며 4개의 특징을 가지는 데이터로 각각 클래스마다 50개의 변화 값이 있다.

표 1은 Iris데이터의 평균과 표준편차를 나타내며 그림 5에서 보듯이 클래스 1인 경우 다른 클래스들과 확연히 구별할 수 있는 데이터 분포를 가지고 있기에 K-means이나 다른 분

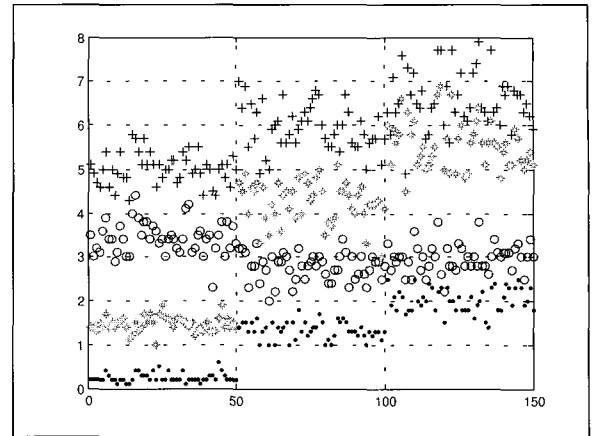


그림 5. 아이리쉬 데이터 집합.
Fig. 5. Iris class data set.

표 1. 아이리쉬 데이터 집합의 특성.

Table 1. Features of the iris data set.

Class 1: Iris Setosa	Mean	Std Deviation
Sepal Length	5.01	0.35
Sepal Width	3.43	0.38
Petal Length	1.46	0.17
Petal Width	0.30	0.13
Class 2: Iris Versicolor	Mean	Std Deviation
Sepal Length	5.94	0.52
Sepal Width	2.77	0.31
Petal Length	4.29	0.47
Petal Width	1.33	0.20
Class 3: Iris Virginica	Mean	Std Deviation
Sepal Length	6.59	0.64
Sepal Width	2.98	0.32
Petal Length	5.55	0.55
Petal Width	2.03	0.27

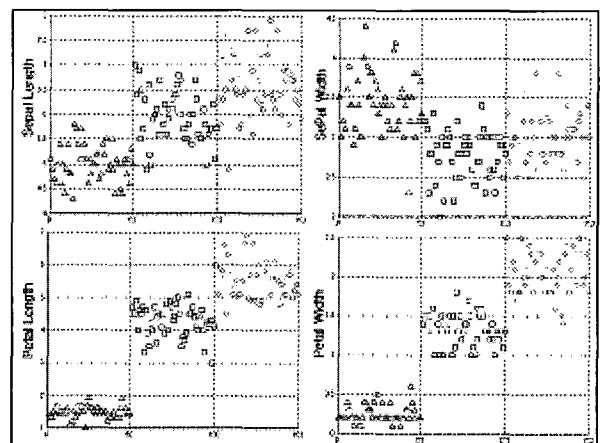


그림 6. Iris 데이터의 분류 결과: setosa(Δ), versicolor(\square), 그리고 virginica(\circ).

Fig. 6. Classification result of Iris data: setosa(Δ), versicolor(\square), and virginica(\circ).

류 알고리즘을 이용하더라도 구분이 용이하지만 클래스 2와

표 2. 실험데이터 집합의 특성.

Table 2. Features of the test data set.

실험데이터	클래스		Mean	Mean
실험데이터 1	Class 1	X-axis	176.6	16.41
		Y-axis	-81.76	5.93
	Class 2	X-axis	382.2	39.13
		Y-axis	-109.1	9.31
	Class 3	X-axis	502	43.8
		Y-axis	-128.4	10.85
실험데이터 2	Class 1	X-axis	28.61	2.21
		Y-axis	-55.71	4.67
	Class 2	X-axis	60.39	5.96
		Y-axis	-50.82	4.31
	Class 3	X-axis	76.17	7.05
		Y-axis	-58.81	4.83
실험데이터 3	Class 1	X-axis	119.3	21.46
		Y-axis	37.97	19.09
	Class 2	X-axis	41.03	19.48
		Y-axis	119.5	19.19
	Class 3	X-axis	159.1	19.31
		Y-axis	158.0	21.59

클래스 3의 경우 여러 변수가 중복되는 영역을 가지기에 그 분류가 용이하지 않다.

본 논문에서는 그림 6에서 나타난 바와 같이 Sepal length, Sepal width, Petal length, 그리고 Petal Width에 대하여 각 클래스마다 중앙값의 변화를 K-means 알고리즘의 결과와 비교하였다. 이는 중앙값의 변화가 실제 데이터의 최종 분류에 영향을 주는 중요한 요소임을 알 수 있다. 모의 실험을 위하여 Iris 데이터 뿐만 아니라 세 개의 임의의 데이터 집합을 만들었다. 세 개의 데이터 집합은 그 그룹을 각각 세 개의 클래스로 구성되게 하였으며, 성능 분석을 위하여 각 데이터의 클래스를 미리 알고 있도록 하였다. 또한 데이터의 형태를 데이터 그룹이 명확히 분리되어 있는 것과 나머지 두 개의 데이터 집합은 일정 영역에서 그 데이터 분류가 불분명하게 되어있도록 일정 영역에서 서로 혼합되어진 2차원의 데이터이다.

표 2는 실험에 사용되어진 데이터 집합의 특성인 평균(mean)과 표준편차(std)를 나타내었다.

데이터의 형태를 미리 설정되어진 데이터로 만들기 위해 2차원의 행렬 형태로 만들었으며 또한 x 축과 y축의 경우는 각각의 변위 값을 의미한다.

그림 7과 그림 8, 그리고 그림 9는 실험을 위해 사용되어진 데이터 집합의 분포 형태를 나타낸다.

각 그림에서 볼 수 있듯이 실험데이터 1과 실험데이터 2는 하나의 클래스는 명확하게 다른 두 클래스와 구별될 수 있도록 2차원 데이터 형태로 만들었으며, 나머지 두 클래스의 경우 경계선이 모호하도록 임의로 만들었다. 실험에 사용되어진 데이터는 실제 산업현장에서 고장 검출을 위해 획득되어지는 센서 데이터들의 경우 정상상태의 경우와 비정상상태

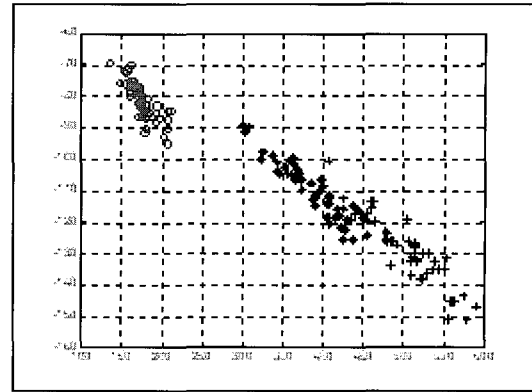


그림 7. 실험데이터 집합 1.

Fig. 7. The test data set 1.

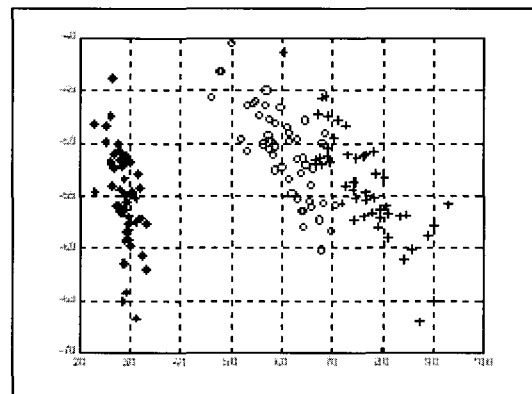


그림 8. 실험데이터 집합 2.

Fig. 8. The test data set 2.

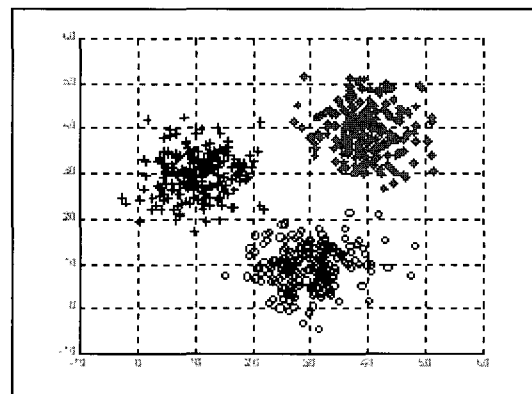


그림 9. 실험데이터 집합 3.

Fig. 9. The test data set 3.

의 경우를 고려하여 각 경계선을 구분 지을 수 있는 경우와 명확히 할 수 없는 경우를 고려하여 실험데이터를 작성하였다. 그리고 그림 9에서는 실험데이터 집합은 세 그룹의 데이터가 두 개의 데이터만이 혼합되어진 형태로 구분되어져 있도록 만들었다.

또한 제안한 데이터 클러스터링 알고리즘의 성능을 확인 및 비교하기 위하여 데이터 분류에 많이 사용되어지는 기하학적 특성을 이용한 Fuzzy C-mean 알고리즘과 K-means 알고리즘 및 입력 데이터의 분포가 정규분포를 따른다는 가정하

에 판별함수를 이용한 판별식 분석알고리즘을 각 실험 데이터에 적용하여 컴퓨터 시뮬레이션을 하였으며 제안한 알고리즘과 비교 분석하였다.

k-means클러스터링 알고리즘의 경우 각각의 요소가 반드시 하나의 클러스터에 속해야 하도록 최소거리를 가질 수 있도록 반복 수행을 하도록 되어있고 Fuzzy C-mean알고리즘 경우 또한 최소 거리를 기준으로 반복적으로 어느 데이터의 요소가 클러스터에 속할 가능성이 크기에 따라 데이터 분류가 이루어지므로 거의 같은 성능을 가지기에 Fuzzy C-mean알고리즘과 판별식을 이용한 분류 결과를 제안한 알고리즘과의 비교를 하였다.

다음의 그림 10, 그리고 그림 11은 Fuzzy C-mean 알고리즘을 이용한 분류 결과와 제안한 알고리즘을 이용하여 데이터 클러스터링 결과를 나타내었다.

그림 10에서 보듯이 Fuzzy C-mean알고리즘과 K-means알고리즘의 경우 기준이 되어지는 일정 데이터의 단순한 데이터 간의 거리를 이용하여 데이터를 분류하기에 그림 5에서 보인 실험데이터 1에서 경계선이 모호할 경우 정확한 분류가 어려움을 알 수 있고 그림 11에서 보는 바와 같은 판별식 분석 알고리즘의 경우 입력패턴에 대하여 각 클래스 별로 미리 정하여 놓은 판별함수의 값을 이용하여 가장 큰 값의 함수에 대응되도록 하였기에 그림에서 나타낸 바와 같이 결정 경계면에서 잘못된 분류를 발생하게 된다.

그러나 그림 12에서 보인 각 후보 클래스의 국부확률과 검정이론을 이용한 경우 각 데이터의 거리와 일정 판별함수에 의한 결정보다 경계선 부분에서 보다 높은 분류 성능을 보임을 알 수 있다.

표 3은 제안된 클러스터링 알고리즘과 Fuzzy C-mean 알고리즘과 그리고 K-means알고리즘과 그리고 판별식 분석알고리즘의 데이터 분류 성능을 비교한 표이다.

실험에 사용 되어진 Iris 데이터 및 임의로 만들어진 데이터의 수와 형태를 고려하였을 때 실제 클래스 구분이 힘든 영역에서의 데이터 존재 개수가 명확히 구분되어질 수 있는 데이터의 개수보다 작기 때문에 표 3에서 보듯이 전체적인 데이터 분류 성능은 높게 나왔다. 그러나 Iris데이터와 실험데이터 3의 경우 명확히 구분 할 수 있는 데이터의 경우 fuzzy c-mean 알고리즘의 결과와 K-means알고리즘, 그리고 판별식

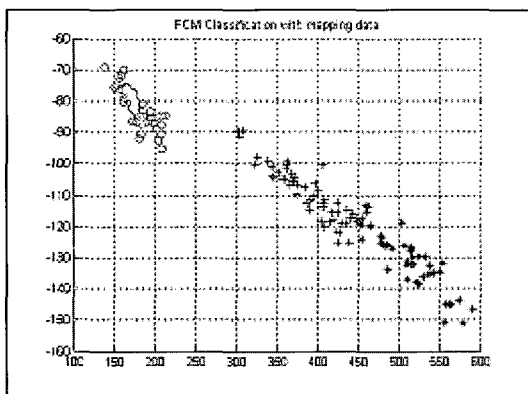


그림 10. FCM 알고리즘의 적용결과.
Fig. 10. The result of FCM algorithm.

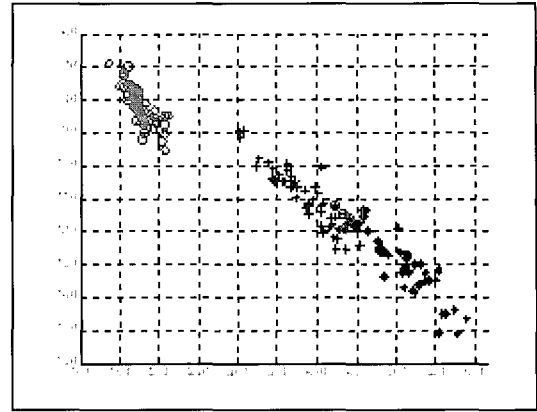


그림 11. 판별식 분석 알고리즘의 적용결과.
Fig. 11. The result of discriminant analysis.

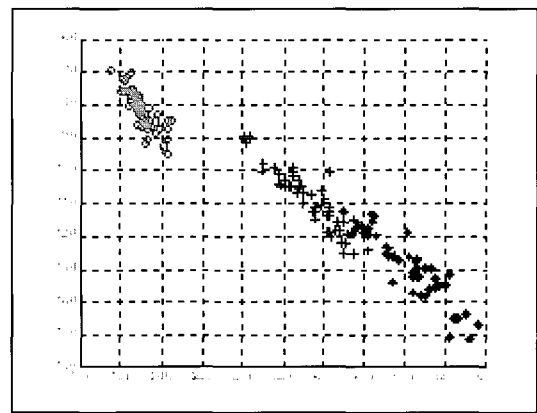


그림 12. 제안한 알고리즘의 적용결과.
Fig. 12. The result of proposed algorithm.

표 3. 데이터 분류를 비교표.

Table 3. The comparison table of classification.

No	Number	Fuzzy C-mean	K-means	Discriminant Analysis	Proposed Algorithm
Data 1	150	94	94	95	96
Data 2	150	95	95	96	97
Data 3	600	99	99	99	99
Sepal Length	150	98	99	99	99
Sepal Width	150	98	99	99	99
Petal Length	150	99	99	99	99
Petal Width	150	99	99	99	99

분석알고리즘의 성능은 99%로 같게 나왔지만, 실험데이터 1 과 실험 데이터 2의 경우 두 그룹의 경계가 명확하지 않는 데이터 집합의 경우 각각 96%, 97%의 데이터 분류 성능이 Fuzzy C-mean알고리즘과 K-means알고리즘의 경우인 94%, 95%인 결과 보다 정확한 데이터 분류를 할 수 있음을 알 수 있었다. 또한 판별함수를 이용한 데이터 분류 경우 95% 및 96%의 분류결과를 얻었고 제안한 알고리즘을 적용한 경우 높은 분류 성능을 가지게 됨을 알 수 있다.

V. 결론 및 향후 과제

본 논문에서는 데이터 클러스터링의 신뢰성을 높이기 위해 거리분포 및 확률이론 그리고 검정이론을 이용하여 데이터 분류 방법을 제안하였다. 먼저 각 데이터들의 기하학적 유사성을 측정하기 위해 유클리드 거리를 이용하여 데이터 거리분포도를 만들고 이 거리 분포도를 이용하여 후보 그룹을 설정한 후 국부확률이론을 적용하여 확률 분포를 계산하고 또한 각 개별데이터의 중요도 및 상관성을 고려한 검정이론을 적용하여 데이터 분류를 하였다. 제안된 알고리즘의 성능분석을 위해 Fuzzy C-mean 알고리즘과 K-means 알고리즘 및 판별식분석 알고리즘을 이용하여 비교 분석하였으며 그 결과를 컴퓨터 시뮬레이션을 통하여 확인하였다. 향후, 본 논문에서 사용되어진 실험을 위하여 만들어진 데이터가 아닌 실제 산업현장에서 고장검출 및 진단을 위해 사용되는 다양한 형태의 센서 데이터들을 이용하여 적용하는 연구를 수행할 예정이다. 또한 실험에서는 국부확률을 적용하기 위한 후보 클래스의 기준을 거리분포를 이용하여 결정하였으나 실제 산업현장에서 고장 검출 및 진단을 위해 사용되어지는 센서데이터는 작업환경이나 잡음 등에 영향을 받아 많은 오차를 발생시킬 수 있으므로 센서의 데이터 형태 및 데이터의 한계(threshold value) 결정 그리고 후보 클래스의 결정에 대한 연구도 병행하여 수행할 것이다.

참고문헌

[1] M. C. Hung and D. L. Yang, "An efficient fuzzy C-means

clustering algorithm," *Data Mining*, 2001. ICDM 2001. *Proc IEEE international Conf.*, 2001., pp. 225-232, Nov 29-Dec, 2001.

[2] N.R. Pal and J. C. Bezdek, "On cluster validity for the fuzzy C-means model," *IEEE Trans. on Fuzzy Systems*. vol. 3 pp. 370-379, 1995.

[3] A. B. Frakt, W. C. Karl, and A. S. Willsky, "A multiscale hypothesis testing approach from noisy tomographic data," *IEEE Trans. on Image Processing*. vol. 7 pp. 825-837, 1998.

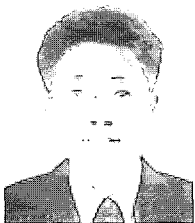
[4] Y. Kato, M. Takahashi and O. R. Yamaguchi., "A proposal of fuzzy test for statistical hypothesis," *Systems, Man, and Cybernetics, 2000 IEEE international Conf.*, pp. 2929-2934, Oct, 2000.

[5] D. Toth and T. Aach, "Improved minimum distance classification with gaussian outlier detection for industrial inspection," *Proc. 11th Image Analysis and Processing. Conf.*, pp. 584-588, Sept. 2001.

[6] J. H. Han and Y. K. Kim, "A fuzzy K-NN algorithm using weights from the variance of membership values," *Computer Vision and Pattern Recognition, 1999. Proc. IEEE Computer Society Conf.*, pp. 394-399, June 1999.

[7] M. Setnes and H. Roubos, "GA-fuzzy modeling and classification: complexity and performance," *IEEE Trans on*. vol. 8 pp. 509-522, 2000.

[8] P. Jeong, S. Nedvschi, and M. Daniliuc, "Local probability based safe region detection for autonomous driving" *2004 IEEE Sym on*. pp. 744-749, 2004.



이재국

1973년 11월 30일생. 1999년 울산대학교 전기전자정보시스템공학부 졸업. 2003년 동 대학원 석사. 2005년 동 대학원 박사 수료. 2005년~현재 ATM(주). 관심분야는 신경망, 시스템 고장 검출 및 진단, 영상처리 등.



최원호

1956년 2월 9일생. 1978년 연세대학교 전자공학과 졸업. 1980년 동 대학원 석사. 1990년 동 대학원 박사. 1979년~1984년 제일정밀공업(주). 1985년~1986년 삼성 휴레패커드. 1986년~현재 울산대학교 전기전자정보시스템공학부 교수. 관심

분야는 Image Processing, 고장 진단 등.



손창호

1952년 5월 5일생. 2002년 울산대학교 컴퓨터정보통신 대학원 석사. 2004년 동 대학원 박사수료. 1971년~1997년 한국전력공사. 1997년~현재 삼창기업(주). 관심 분야는 시스템 고장 검출 및 진단, 신호 처리 등.