

붓스트랩 기법과 유전자 알고리즘을 이용한 최적 군집 수 결정

Determination of Optimal Cluster Size Using Bootstrap and Genetic Algorithm

박민재 · 전성해 · 오경환

Min-Jae Park, Sung-Hae Jun and Kyung-Whan Oh

서강대학교 컴퓨터학과

요 약

데이터의 군집화를 수행할 때 최적 군집수 결정은 군집 결과의 성능에 많은 영향을 미친다. 특히 K-means 방법에서는 초기 군집수 K에 따라 군집결과의 성능 차이가 많이 나타난다. 하지만 대다수의 군집분석에서 초기 군집수의 결정은 경험을 바탕으로 하여 주관적으로 결정된다. 이때 개체수와 속성수가 증가하면 이러한 결정은 더욱 어려워지며 이때 결정된 군집수가 최적인 된다는 보장도 없다. 본 논문에서는 군집의 수를 자동으로 결정하고 그 결과의 유효성을 보장하기 위해 유전자 알고리즘에 기반한 최적 군집수 결정 방안을 제안한다. 데이터의 속성에 근거한 초기 해 집단이 생성되고, 해 집단 내에서 최적화된 군집수를 찾기 위해 교차 연산이 이루어진다. 적합도 값은 전체 군집화의 비 유사성의 합의 역으로 결정되어 전체적인 군집화 성능이 향상되는 방향으로 수렴된다. 또한 지역 국소값을 해결하기 위해 돌연변이 연산이 사용된다. 그리고 유전자 알고리즘의 학습 시간의 비용을 줄이기 위해 붓스트랩 기법이 적용된다.

Abstract

Optimal determination of cluster size has an effect on the result of clustering. In K-means algorithm, the difference of clustering performance is large by initial K. But the initial cluster size is determined by prior knowledge or subjectivity in most clustering process. This subjective determination may not be optimal. In this paper, the genetic algorithm based optimal determination approach of cluster size is proposed for automatic determination of cluster size and performance upgrading of its result. The initial population based on attribution is generated for searching optimal cluster size. The fitness value is defined the inverse of dissimilarity summation. So this is converged to upgraded total performance. The mutation operation is used for local minima problem. Finally, the re-sampling of bootstrapping is used for computational time cost.

Key Words : Optimal Cluster Size, Genetic Algorithm, Bootstrap, Variance Criterion

1. 서 론

일반적으로 군집 분석시에 군집수 결정은 주어진 학습 데이터와 사전경험에 의해 분석가가 주관적으로 결정한다. 그러나 데이터에 대한 정확한 정보가 없는 상태에서 다수의 개체수와 속성수를 갖는 데이터에 대해 주관에 의존해 군집수를 결정하는 것은 쉬운 작업이 아니다. 본 논문에서는 자동적인 군집수 결정 방안과 최적 결정에 대한 접근 방법에 대하여 연구한다. 제안되는 방법은 유전자 알고리즘을 응용한 최적 군집수 결정이다. 군집의 성능을 결정해주는 적합도 함수에 근거하여 반복적인 유전자 연산을 통해 군집수를 최적화에 근접시키며 이에 따른 결과는 우수한 군집 수행 결과를

보장해 준다. 본 논문 2절에서는 전체적인 유전자 알고리즘 기반의 결정 방법에 대해서 알아보고, 3절에서는 성능향상을 위해 데이터 속성 기반의 초기 해집단 결정과 유전자 알고리즘의 학습 시간 비용을 줄이기 위해 적용되는 붓스트랩 기법을 알아본다. 제안된 방법에 대한 실험 및 결과는 4절에서 알아보고, 마지막으로 5절에서는 본 논문의 결론 및 향후 연구과제에 대해 논의한다.

2. 유전자 알고리즘 기반 군집수 결정 방법

2.1 유전자 알고리즘

유전 알고리즘(genetic algorithm)은 진화형 유성 생식의 최적 전략에 기반을 둔다. 즉 개체군에서 적합한 개체는 자손의 번식에 많은 영향을 미친다. 개체군에서 각 개체는 주어진 문제를 해결하는 잠재적인 해로 표현된다. 유전자 알고리즘의 목표는 해공간(solution space)에서 통계적 전역 탐색의 평균에 의해 최대의 적합도를 가진 개체를 찾는 것이다[1].

유전자 알고리즘은 교배(crossover), 돌연변이, (mutation),

접수일자 : 2002년 11월 2일

완료일자 : 2003년 2월 1일

본 연구는 과학 기술부 주관 뇌신경 정보학 사업에 의해 지원 되었음.

역치(inversion)와 같이 유전학에 의해 고안된 연산자들과 함께 일종의 '자연의 선택'을 사용하여 "염색체(예를 들어 비트, 즉 1과 0의 문자열)의 한 개체집단에서 새로운 집단으로 이동시키는 방법이다. 각 염색체는 유전인자(gene)들로 이루어져 있고, 각 유전인자는 특별한 '대립유전자'의 한 예이다. 선택 유전자는 개체집단에서 재생산하도록 허용할 염색체들은 선택하고, 적합한 염색체들은 적합하지 않은 염색체들보다 평균적으로 더 많은 자손들은 생산한다. 교배는 개략적으로 두 개의 단순 염색체 유기체들간의 생물학적 재결합을 흉내내어 두 염색체의 일부분을 교환하고, 돌연변이는 염색체내의 어떤 위치의 대립 유전자 값을 임의적으로 변화시킨다. 그리고 역치는 염색체의 인접한 순서를 역전시켜서 유전인자들이 배열되어 있는 순서를 재배열한다. 그림 1은 일반적인 유전자 알고리즘의 동작 순서를 나타내고 있다.

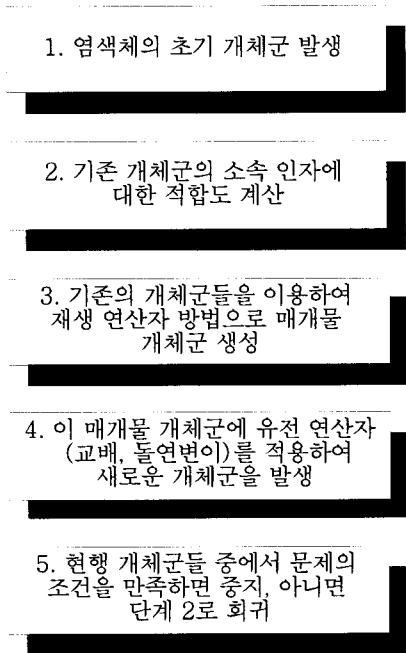


그림 1. 유전자 알고리즘의 동작 순서.
Fig. 1. General outline of GA

2.2 유전자 알고리즘 기반 군집수 결정

본 논문에서 제안된 군집수 결정을 위한 유전자 알고리즘은 다음의 주요 요소로 구성되어 있다.

■ 개체 표현: 각 개체들은 가능한 후보해를 나타내며 일련의 유전인자들(string of genes)로 구성되어 있다. 본 논문에서는 군집수를 2진수의 값으로 표현한 6비트의 이진 문자열(binary string)로 표현되었다. 예를 들어 군집수 5는 '000101'로 표현된다. 군집수는 1보다 큰 정수이며 일반적인 경우에 $111111_2=63$ 개를 초과하지 않기 때문에 6비트의 2진수로 모든 가능한 후보해를 표현 할 수 있다.

■ 초기 해집단: 초기 해집단은 통상적으로 무작위로 선택되지만 빠른 최적값 수렴을 위해 데이터에 의해 얻어지는 수렴 군집수를 기반으로 생성된다. 수렴 군집수를 중심으로 초기 해집단의 크기는 '11'로 생성된다.

그림 2는 수렴 군집수 '7'을 기반으로 형성된 초기 해집단의 예를 보이고 있다.

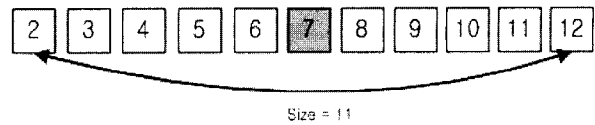


그림 2. 초기해집단의 예.
Fig. 2. An Example of initial population

■ 적합도 함수: 본 논문에서 제안된 적합도 함수는 다음과 같다.

$$F = \begin{cases} 0 & \text{if } K=0 \\ \frac{1}{\sqrt{K}} \sum_{i=1}^n \frac{1}{(e_i - C_i)^2} & \text{otherwise} \end{cases}$$

Where, e_i = *ith* element in data set,
 C_i = centroid of cluster that holds *ith* element,
 K = Number of clusters

적합도 함수는 두개의 목적함수(objective function)의 조합으로 구성되며, 각 목적함수는 다음과 같다.

$$\sum_{i=1}^n \frac{1}{(e_i - C_i)^2} : \text{전체 군집의 밀도}$$

$$\frac{1}{\sqrt{k}} : \text{군집수에 대한 불이익}$$

즉, 이 적합도함수는 동일한 군집내의 원소간의 유사도가 최대를 이루는 군집의 수로 수렴하는 것을 보장하지만 동시에 군집수가 계속 증가하여 발생하는 무의미한 군집의 생산을 억제한다. 군집수에 대한 불이익이 주어지지 않는다면, N개의 원소로 구성된 데이터에 대해 N개의 군집수로 수렴하기 때문이다.

■ 재생 연산: 본 논문에서는 다음세대의 새로운 후보해를 선택하기 위해 Roulette wheel 연산자를 이용해 두 부모 후보해를 선택한다. 적합도가 높은 개체는 그 비율만큼 다음세대로 자신의 유전형을 전달할 비율이 높아진다. 이때 전 세대의 후보해 중 가장 적합도가 높은 후보해를 다음세대로 전이시키기 위해 일정 비율의 엘리트 이론(elitism)이 적용된다.

■ 교배 연산: 새로운 후보해를 생성하기 위해 0.5-균일교배(uniform crossover)를 사용한다. 0.5-균일교배는 0.5의 확률로 선택된 임의의 교차점에서 두 부모의 유전인자가 교환된다. 그림 3은 0.5-균일교배의 예를 보이고 있다. 어둡게 표시된 부분이 교차점으로 선택된 부분이며 선택된 두 부모로부터 교배연산에 의해 새로운 자식이 생성됨을 알 수 있다.

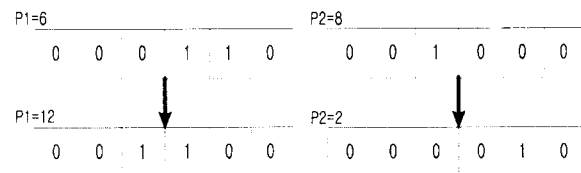


그림 3. 0.5-균일교배연산의 예.
Fig. 3. An Example of 0.5-uniform crossover

■ 돌연변이 연산: 후보해 탐색과정에서 적합도 값이 지역 극대점(local maximum)머무는 것을 방지하기 위해 돌연변이 연산을 사용한다. 돌연변이는 6비트 이진 문자열 중에

서 임의(random)의 위치에 해당하는 한 비트를 반전 시킨다. 그림 4는 돌연변이 연산의 예를 보이고 있다. 어둡게 표시된 유전자가 임의적으로 선택된 돌연변이 위치이다.

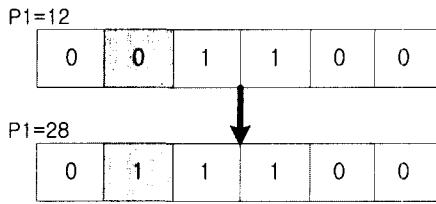


그림 4. 돌연변이 연산의 예.
Fig. 4. An Example of mutation

이상에서 제안한 군집수 결정을 위한 유전자 알고리즘의 전체적인 개요는 그림 5에 나타나 있다.

```

Determination of Optimal Cluster Size Based on Genetic Algorithm
Begin
Generate population P0
  Begin
  if (k' > 6)
    P0 = k', k'+1, ..., k'+5, k'-1, ..., k'-5
  else
    P0 = 1, 2, ..., 10, 11
  End
Evaluate P0
For i:=1 to MaxIteration do
  For j:=1 to 11 do
    Select two parents p1 and p2 from Pi-1
    Offspring ← (p1,p2)
    Crossover(p1,p2)
    Mutate offspring
    Evaluate offspring and assign it to fitness
    if ( fitness = StopCriterion)
      Return fitness
    Add offspring to Pi
  Done
Return the best value from PMaxIteration
End
    
```

그림 5. 군집수 결정을 위한 유전자 알고리즘
Fig. 5 GA for Determination of optimal cluster size

3. 제안된 알고리즘의 개선

본 논문에서 제안된 알고리즘을 개선은 크게 두 가지 관점에서 진행되었다. 첫 번째는 최적화 군집수에 근접하는 수렴 군집수 K'의 결정이고, 두 번째는 붓스트랩(bootstrap)을 사용한 데이터의 표본 추출이다

3.1 수렴 군집수 결정

초기 해집단은 수렴 군집수에 의해 생성되므로 적절한 수렴 군집수의 결정은 보다 빠른 속도로 최적화 값에 접근하는 것을 보장해준다.

따라서 본 논문에서는 데이터의 주성분 분석(Principle Component Analysis: PCA)에 의해 최적화 군집수에 근접한 수렴 군집수를 얻는다[2]. 이러한 수렴 군집수는 유전자 알고리즘의 초기해 집단을 위해 사용되어 짧은 세대 안에 최적화 값에 수렴하는 것을 보장해 준다. 다음 장에서 주성분 분석에 의해 수렴군집수를 결정하는 과정을 설명한다.

3.1 주성분 분석

3.1.1 주성분의 개념

여러개 ($p \geq 2$)의 반응변수에 대하여 얻어진 다변량 자료를 분석의 대상으로 하는 주성분 분석은 다차원적인 변수들을 축소하는 차원의 단순화와 서로 상관되어있는 변수들 상호간의 복잡한 구조를 분석한다. 즉, 반응변수들을 선형 변환시켜, 주성분이라는 서로 독립적인 새로운 변수들을 유도한다. 이때 각 주성분이 보유하는 변이의 크기를 기준으로 그 중요도의 순서를 생각할 수 있는데, 그들 중 상위 m 개 ($m \ll p$)의 주성분에 의해 원래자료에 내재하는 전체변이 중 가능한 한 많은 부분이 보유되도록 변환시킴으로서 정보의 손실을 최소화하는 차원의 축소(dimension reduction)가 이루어진다.

서로 상관되어 있는 $p(\geq 2)$ 개의 확률변수 X_1, X_2, \dots, X_p 를 원소로 하는 확률벡터 X 가 평균벡터 μ 와 공분산행렬 Σ 를 다음과 같이 표현한다.

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_i \\ \vdots \\ X_p \end{bmatrix} \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_i \\ \vdots \\ \mu_p \end{bmatrix} \quad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1j} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2j} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \sigma_{i1} & \sigma_{i2} & \cdots & \sigma_{ij} & \cdots & \sigma_{ip} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pj} & \cdots & \sigma_{pp} \end{bmatrix}$$

여기서 $\sigma_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)] = \sigma_{ji}$ 는 X_i 와 X_j 의 공분산이다. 그리고 $\mu_i = E[X_i]$ 는 X_i 의 평균이다. 또한 p 는 변수의 개수를 나타낸다. σ_{ii} 는 i 번째 변수의 분산이고 σ_{ij} 는 i 번째 변수와 j 번째 변수의 공분산을 나타낸다. 주성분분석은 입력 변수벡터 X 를 선형변환시켜 정보의 손실을 최소화하면서 p 보다 매우 작은 m 개의 새로운 인공변수를 생성함으로써, p 차원 변이를 m 차원으로 축소하여 전체 데이터의 특성을 요약하여 전체 변수들간의 복잡한 구조를 파악하고자 하는 것이다. 이 분석은 X 의 원소들 간의 상관구조 관계를 나타내는 Σ 를 분석대상으로 한다.

Σ 의 p 개의 고유값(eigen value), δ_j 들을 크기순으로 배열하고 각각의 고유값에 대응되는 고유벡터(eigen vector), r_j 의 짝들을 $(\delta_1, r_1), (\delta_2, r_2), \dots, (\delta_p, r_p)$ 라 하고, δ_j 들을 크기순으로 배열하면 다음 식과 같이 표현된다.

$$\sum r_j = \delta_j r_j, j = 1, 2, \dots, p$$

$\delta_1 \geq \delta_2 \geq \delta_3 \cdots \geq \delta_p$ 의 관계를 갖게 된다. 이 값에 의해 p 개의 입력변수를 m 개의 주성분으로 차원을 축소하여 데이터의 특성을 파악하게 된다. 이 때 주성분의 개수는 다음 절에서와 같이 몇 가지 판정 기준을 통해 결정하게 된다.

3.1.2 보유주성분의 수에 관한 판정기준

우선 전체 변이에 대한 공헌 정도로 결정 할수 있다. 즉, 보유 주성분들이 전체 분산에 대해 주어진 일정비율 (예를 들어 80~90%) 이상을 설명할 수 있기 위해 필요한 최소 개수의 주성분을 보유하는 방법이다. 다음으로 고유값의 크기를 이용할 수 있다. 이 기준은 주성분으로 보유되기 위해서 대응되는 고유값은 적어도 1 이상이 되어야 한다는 것으로서, 이는 Kaiser의 규칙(Kaiser, 1960)으로 알려져 있다.

3.2 붓스트랩을 사용한 데이터의 표본 추출

유전자 알고리즘에서 각 후보해에 대한 적합도 계산에 대한 시간 비용을 줄이기 위해 붓스트랩기법에 의한 데이터의 표본추출을 하였다[3]. 붓스트랩은 통계적 추론을 위한 방법으로 분포에 대한 정보없이 자료만을 사용하여 재표본(re-sampling)전략을 통하여 컴퓨터 모의실험으로 수행된다. 붓스트랩 추정량은 추정량 자체가 하나의 통계량(표본에 의한 식)으로 표현된다는 것이 아니라 컴퓨터 알고리즘의 형태로 표현된다. 이는 자료만을 가지고 자료의 분포에 대한 특성을 찾아낸다는 방법을 나타내는 것을 의미한다[5.]. 재표본 전략은 표본의 불편성(unbiasness)을 유지하기 위함이다. 즉, 편향된 표본의 추출을 방지함에 따라 표본과 원 자료가 유사한 분포로 구성될 수 있음을 보장하는 것이다. 따라서 본 논문에서는 표본 추출시 추출된 표본의 재사용을 허용했으며 총 4개의 표본을 추출하여 각 표본의 결과값을 토대로 최종적인 군집수를 결정하였다. 데이터의 표본 추출과 그에 따른 결과값 통합의 과정이 그림 3에 나타나있다.

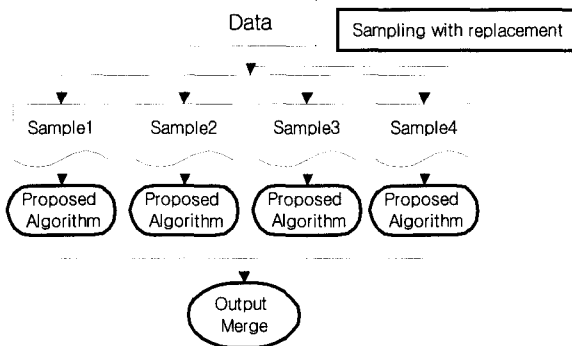


그림 3. 데이터의 표본 추출 및 결과 통합 과정
Fig. 3 Extraction samples and merging outputs

각 표본에 대한 출력 값의 평균에 의해 전체 결과 값이 결정된다. 이상에서 언급한 군집화 결정과정의 전체적인 과정은 그림 4와 같다.

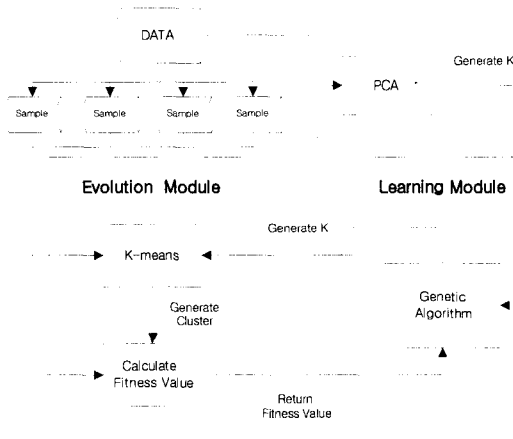


그림 4. 유전자 알고리즘 기반 군집화 결정 과정
Fig. 3 Process of Clustering Using Genetic Algorithm

전체과정은 크게 학습모델(learning model)과 평가모델

(evaluation model)로 구성된다. 학습 모델에서는 주성분 분석을 통한 수렴 군집수를 기반으로 초기해 집단이 생성되고 교배, 돌연변이 연산을 통해 새로운 군집수를 생산한다. 학습 모델에 의해 생성된 군집수에 대해서 평가모델의 K-평균(means) 방법을 통해 생성된 데이터의 군집을 기반으로 성능이 평가되며, 이 평가는 다시 학습 모델로 전달되어 새로운 개체를 생성하는 지표로 사용하게 된다.

4. 실험 및 결과

본 논문에서 사용된 실험 데이터는 UCL Machine Repository의 Iris데이터와 arrhythmia데이터이다[6]. 각 데이터의 속성은 표 1에 정리되어 있다.

표 1. 데이터의 속성 값.
Table 1. Properties of datum

	Iris	Arrhythmia
Instance	150	457
Attribute	4	279
Class	3	16

4.1 수렴 군집수 결정

유전자 알고리즘의 초기 해집단을 형성하기 위해 데이터의 주성분 분석을 통해 수렴 군집수를 결정하였다. 각 데이터에 대하여 3개의 보유 주성분을 각각 축으로 하는 3-D산점도(3-D plot)와 이를 통해 얻어진 수렴 군집수 Kprior는 그림 5와 표 2에 각각 정리되어 있다.

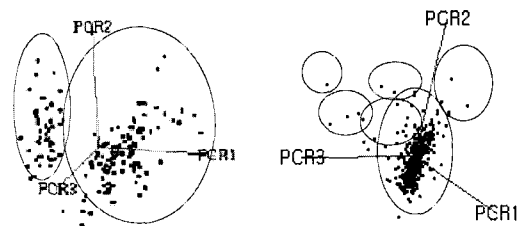


그림 5. 데이터의 주성분 분석에 의한 분포
Fig. 3 3-D Scatters of datum

표 2. 데이터의 속성에 의해 추정된 근사 군집수
Table 2. Approximated cluster numbers

	Iris	Arrhythmia
K_{prior}	2	6

4.2 제안된 알고리즘에 의한 실험 결과

실험에 사용되기 위한 표본은 각 데이터에 대해서 30개의 크기로 4번씩 추출되었다. 실험을 위한 유전자 알고리즘에 대한 변수 값은 표 3과 같다.

실험은 두 데이터에 대해 각각 20회씩 반복하여 이루어졌다. 본 논문에서 제안하고 있는 알고리즘에 의해 결정된 군

집수의 평균값(mean)과 표준편차(standard deviation)가 표 4에 나타나 있다.

표 3. 유전자 알고리즘 변수
Table 3. Parameters for Genetic Algorithm

Crossover Rate	0.9
Mutation Rate	0.1
MAX Iteration	30

표 4. 군집수의 평균값과 표준편차
Table 4. Means and Standard derivation of cluster numbers

	Mean	SD
Iris	3.35	0.3479
Arrhythmia	8.05	1.3028

4.3 결정된 군집수에 대한 평가

결정된 군집수에 대한 성능평가의 기준은 다음의 두 가지 조건을 포함해야 한다.

- 1) 각 군집의 평균 밀도(average density)
- 2) 군집수 증가에 대한 불이익(penalty)

위의 두 조건을 만족하기 위해 군집수에 대한 성능평가의 기준으로 Variance Criterion(VC)를 사용한다.[2] VC측도는 다음 식과 같이 정의된다.

$$VC_M = \sum_{i=1}^M V_i / M + 0.1 * M$$

위의 식에서 M은 군집의 수를, V_i 는 i번째 군집의 평균 분산을 의미한다. 앞의 분산의 평균은 각 군집의 평균밀도를 의미하고, 뒤의 군집수는 군집수 증가에 따른 불이익을 의미한다. 따라서 위의 식은 좋은 성능의 군집에 대해서 보다 작은 값을 가지게 된다. 두 항(term)의 균형을 위한 상수는 명확한 군집의 구분을 가지는 인공 데이터들에 대한 실험을 통해 0.1로 결정되었다.

Iris와 arrhythmia데이터의 각 군집수에 대한 VC값은 표 5에 나타나 있다. 표에 나타나 있듯이 본 논문에서 제안한 유전자 기반 군집수 결정 방식이 판단한 군집수는 최적 군집수에 수렴함을 알 수 있다.

표 5. 군집수와 VC값
Table 4. Cluster numbers and VC values

Iris		Arrhythmia	
군집수	VC	군집수	VC
2	0.7	7	1.98
3	0.64	8	1.73
4	0.69	9	2.33

4. 결론 및 향후 연구과제

본 논문에서는 최적화된 군집수 결정을 위해 유전자 알고리즘을 적용한 방법을 제시 하였다. 그리고 유전자 알고리즘의 성능을 향상 시키기 위한 주성분 분석을 통한 수렴 군집수 결정과 적합도 함수(fitness function) 계산 비용을 줄이기 위한 붓스트랩기법을 적용하였다.

군집수 결정의 문제를 일종의 탐색 문제(search problem)로 인식한다면, 탐색 공간이 제한적인 군집수 결정문제에 있어 유전자 알고리즘은 빠른 속도로 최적화 해 탐색을 수행할 수 있고, 그 결과 역시 어느 정도 최적화에 근접함을 실험을 통해 입증하였다. 더욱이 데이터에 대한 기본 지식(prior knowledge)를 사용하지 않고도, 최적화 군집수 수렴을 이루어 낼 수 있었다. 이러한 접근 방법은 데이터에 대한 사전 정보를 입수하기 힘든 생물정보학(bioinformatics) 데이터 등의 분석에 상당한 도움이 될 수 있을 것이다.

하지만 주성분 분석(PCA)을 사용하여 수렴 군집수를 결정하는 과정에서 분석가의 주관적 판단이 요구되며, 이러한 수렴 군집수의 결정에 요구되는 시간 역시 최적화 군집수 탐색에 비해 크다고 할 수 있다.

따라서 향후 연구에서는 신경망 방식과 같은 대안적인 방식을 사용한 수렴 군집수 결정을 통해 분석가의 판단 없이 최적화된 군집수 결정을 이루어야 할 것이다.

참 고 문 헌

- [1] E. Vonk, L. C. Jain, R. P. Johnson "Automatic Generation of Neural Network Architecture Using Evolutionary Computation", World Scientific Publishing, pp. 22~23, 1997,
- [2] 박정은, 전성해, 오경환, "자동 군집화를 위한 지능화된 데이터 마이닝 에이전트", 한국 지능정보시스템학회지, pp. 370~376, 11, 2002
- [3] Christian P. Robert, George Casella, "Monte Carlo Statistical Methods", Springer, pp. 33~34, 1999
- [4] M. Wall, "GAlib: A C++ library of genetic algorithm components", in Manual: Mechanical Eng. Dept., MIT, 1996
- [5] 손건태, "전산통계개론", 자유아카데미, pp. 209~210, 1996
- [6] <http://www.ics.uci.edu/~mllearn/MLRepository.html>

저 자 소 개



박민재 (Min-Jae Park)

1999년 : 서강대 컴퓨터학과 졸업(학사)
2001년~현재 : 동 대학원 컴퓨터학과 석사 과정

관심분야 : 기계학습, 바이오인포매틱스, 데이터 마이닝

Phone : 02-703-7626
Fax : 02-704-8278
E-mail : pmj219@ailab.sogang.ac.kr



전성해 (Sung-Hae Jun)

1993년 : 인하대 통계학과 (학사)
1996년 : 인하대 통계학과 (이학석사)
2001년 : 인하대 통계학과 (이학박사)
2001년 ~ 현재 : 서강대학교 대학원 컴퓨터학
과 박사과정

관심분야 : 데이터마이닝, 기계학습, 인지과학
Phone : 02-703-7626
Fax : 02-704-8278
E-mail : shjun@ailab..sogang.ac.kr



오경환 (Kyung-Whan Oh)

1978년 : 서강대학교 수학과 졸업(학사)
1985년 : Florida State University
Computer Science(공학석사)
1988년 : Florida State University
Computer Science(공학박사)
1989년 ~ 현재 : 서강대학교 컴퓨터학과 교수

관심분야 : 퍼지로지, 인공지능, 다중에이전트
Phone : 02-703-7626
Fax : 02-704-8278
E-mail : kwoh@ccs..sogang.ac.kr