

개미 군락 시스템을 이용한 계층적 클러스터 분석[☆]

Ant Colony Hierarchical Cluster Analysis

강 문 수¹ 최 영 식^{*}
Mun-Su Kang Young-Sik Choi

요 약

본 논문에서는 방향그래프에서 개미가 한 노드에서 다른 노드들로 이동하는 새로운 개미 기반계층적 클러스터링 알고리즘을 제안한다. 노드페로몬은 로컬영역에서 상대 밀도값으로 간주될 수 있는 값으로 노드로 들어오는 에지들의 페로몬 양을 합한 것이다. 일정한 횟수만큼 개미들을 이동시킨 후 방향 그래프로부터 소량의 노드페로몬 값을 가진 노드들을 제거하고, 강하게 연결되어 있는 요소들을 하나의 클러스터로 구성한다. 반복적으로 낮은 값부터 높은 값까지 제거작업을 하여 계층적 클러스터들을 구축한다. 다양한 실험을 통해 제안하는 알고리즘과 기존 클러스터링 알고리즘을 비교하고 제안하는 알고리즘의 우수성을 실험을 통해 입증한다.

☞ 주제어 : 개미기반 클러스터링, 개미기반 계층적 클러스터링, 클러스터링, 집단지성

ABSTRACT

In this paper, we present a novel ant-based hierarchical clustering algorithm, where ants repeatedly hop from one node to another over a weighted directed graph of k -nearest neighborhood obtained from a given dataset. We introduce a notion of node pheromone, which is the summation of amount of pheromone on incoming arcs to a node. The node pheromone can be regarded as a relative density measure in a local region. After a finite number of ants' hopping, we remove nodes with a small amount of node pheromone from the directed graph, and obtain a group of strongly connected components as clusters. We iteratively do this removing process from a low value of threshold to a high value, yielding a hierarchy of clusters. We demonstrate the performance of the proposed algorithm with synthetic and real data sets, comparing with traditional clustering methods. Experimental results show the superiority of the proposed method to the traditional methods.

☞ keyword : Ant-based clustering, ant-based hierarchical clustering, clustering, swarm intelligence

1. 서 론

클러스터링 분석은 최근 생물정보학 및 웹 문서 검색 및 마이닝 등 다양한 분야에서 크게 성장하고 있다 [1][17]. 클러스터링의 목적은 자연의 숨겨진 데이터 구조의 이산 유한 집합으로 레이블되지 않은 데이터 집합으로 분리하는 것이다[1]. 클러스터링을 위해 많은 연구가 있었으며 그 중 개미 군락 시스템을 이용한 클러스터링 문제 해결은 현재에도 많이 연구되고 있다.

개미 기반의 클러스터링 알고리즘은 먹이 (시체와 애벌레)를 정렬하고 클러스터링하는 개미 군집의 동작에서 영감을 얻었다[2][6]. 개미 기반 클러스터링 및 정렬 알고리즘은 Deneubourg[3]에 의해 소개되었고, 개미가 무작위로 비슷한 이웃 객체의 수에 따라 객체를 이동하고, 집고, 떨어뜨리는 모델로 제안했다. 주로 집단 로봇에 적용 가능한 모델을 유도하는 목적이었으나 곧 데이터 분석에 적용되었다[6]. Lumer와 Faieta는 Deneubourg의 모델을 수정하여 탐색 및 데이터 분석에 적용하여 개미 기반 클러스터링 및 정렬 알고리즘의 기본 모델이되었다[4].

Lumer 및 Faieta(LF)의 작업은 개미 기반의 데이터 클러스터링 및 정렬에 대한 일련의 작업이다. [5]와 [11]은 주제 지도의 형태로 문서 컬렉션의 이차원 시각화에 적용할 수 있도록 LF 알고리즘을 수정하였고 [12]와 [13]는 그래프 분할에 적용 가능하도록 하였다. [6]은 기본적인 개미 기반의 클러스터링[4]을 향상시켰으며 전통적인 방법과 비교하여 다양한 실험 결과를 제공했다. [2]는 멀티

¹ Computer Engineering, Korea Aerospace University, Goyang, Gyeonggi-do, 412-791, Korea.

^{*} Corresponding author(choimail@kau.ac.kr)

[Received 21April 2014, Reviewed 23April 2014(R2 11 July 2014), Accepted 15September 2014]

[☆] This research was supported by the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the C-ITRC(Convergence Information Technology Research Center) support program (NIPA-2014-H0401-14-1021) supervised by the NIPA(National IT Industry Promotion Agency)

개미 군락 클러스터링 알고리즘을 사용하였고 하나의 개미 기반 클러스터링에 비해 더 나은 성능을 보여 주었다. 개미 기반의 접근 방식과 기존 방식의 하이브리드화하는 여러 시도가 있다. [8]은 개미 기반 알고리즘에서 얻은 클러스터링 결과를 향상시키기 위해 K-means 알고리즘을 적용했고 [7]은 DBSCAN의 아이디어를 사용했으며 [9]는 클러스터링 문제 해결을 위해 주어진 데이터 집합을 방향 그래프 분할 문제로 보았다.

[27]에서 보듯이 최근에는 개미 군락 시스템을 이용한 계층적 클러스터링 연구도 많이 이루어지고 있으며 다양한 응용에도 적용되고 있다. [22]는 라우팅 알고리즘을 계층적 개미 시스템을 이용하였으며, [23]는 개미 기반 계층적 클러스터링 알고리즘을 응용에 적용한 예이다. [24]와 [25]는 멀티플 페로몬을 이용하였으며 [26]는 모바일 애드 혹 네트워크에 적용한 예이고 [g]는 하이브리드 하였다. 본 논문에서는 개미 군락 시스템을 이용하여 클러스터의 계층 구조를 생성할 수 있는 새로운 개미 군락 시스템을 이용한 계층적 클러스터링(Ant Colony Hierarchical Clustering) 알고리즘을 제안한다.

본 논문에서 제안하는 ACHC(Ant Colony Hierarchical Clustering) 알고리즘은 개미들이 그래프의 한 노드에서 다른 노드로 반복적으로 노드 페로몬 값이 높은 노드를 선호하여 선택하여 이동한다. 노드 페로몬은 노드로 유입되는 에지들과 관련된 페로몬량의 합으로 정의한다.

노드 페로몬 값이 높은 노드는 그래프에서 개미들이 자주 방문하였다는 것을 의미한다. 이러한 관점에서 노드 페로몬은 그래프에서 개미 확률적 이동을 통한 상대 밀도 값으로 간주할 수 있다.

일정한 횟수만큼 개미들을 이동시킨 후 노드 페로몬 값이 지정된 임계 값 이하인 노드를 제거하고 클러스터의 집합인 강하게 연결된 구성 요소(Strongly Connected Component)의 그룹을 얻는다. 노드 페로몬의 임계값을 낮은 값부터 시작하여 높은 값으로 반복적으로 설정하면서 노드 제거를 수행하면 클러스터의 계층 구조를 구축할 수 있다.

제안하는 방법은 [10]과 [14]에 제시된 밀도 기반의 계층적 클러스터링 HDS (Hierarchical Density Shaving)와 관련이 있다. HDS 클러스터링은 밀도가 주어진 임계 값보다 작은 노드의 특정 퍼센트를 제거하고 감소된 데이터 포인트로 DBSCAN[20]을 적용한다. 따라서 제안하는 ACHC와 HDS 알고리즘은 둘 다 데이터 포인트의 밀도에 따라 데이터 포인트를 제거(Shaving)하는 방법이고 각 제거(Shaving)레벨로 클러스터 계층 구조를 구축한다.

하지만 본 논문에서 제안하는 노드 페로몬은 로컬 지역에서의 상대적 밀도 측정법으로 만약 주변의 노드들보다 더 높은 밀도를 가지고 있다면 주변 노드들이 더 작은 양의 노드 페로몬을 가진다. 상대 밀도로 간주되는 노드 페로몬의 특성은 전통적인 밀도 기반 클러스터링 알고리즘[20]과 차별화되며 더 좋은 성능을 보여준다.

2장에서는 그래프에서의 노드 페로몬을 소개하고 3장에서는 ACHC 알고리즘을 제안한다. 4장에서는 합성 데이터(synthetic data)와 다양한 실제 데이터 집합들을 사용하여 제안하는 ACHC의 성능을 보여주고 기존의 알고리즘인 HDS와 비교한다. 5장은 본 논문의 결론이다.

2. 개미 군락 클러스터링

이 섹션에서는 개미 기반의 클러스터링 알고리즘을 설명한다. 첫째, 주어진 데이터 집합에서 가중치를 가진 방향 그래프(Weighted Directed Graph)를 생성하는 방법에 대해 설명한다. 다음은 제안하는 ACHC (Ant Colony Hierarchical Clustering) 알고리즘에 매우 중요한 노드 페로몬을 설명한다. 마지막으로 노드 페로몬을 사용하여 기본적인 개미 기반 클러스터링 알고리즘 절차를 설명한다.

2.1 가중치방향그래프

클러스터된 데이터 집합 $X = \{x_i, i = 1 \dots n\} \subseteq \mathbb{R}^d$ 라고 하자. 유클리드 거리와 같은 거리측정 $d(x_i, x_j)$ 는 데이터 X에서 x_i 와 x_j 로 정의되어 있다고 가정한다. D는 $n \times n$ 대칭적인 거리 행렬로 $D(i, j) = d(x_i, x_j)$ 표현한다.

가중치를 가진 방향 그래프(Weighted Directed Graph) $G = \{V, E\}$ 를 생성한다. 각 데이터 포인트 x_i 는 각 노드 $v_i \in V$ 에 매핑되고 x_j 는 x_i 의 k개의 가장 가까운 이웃이면 $e_{ij} \in E$ 는 v_i 에서 v_j 의 방향 에지로 매핑된다. 가중치를 가진 방향 그래프에서 x_i 의 k개의 가장 가까운 이웃은 v_i 의 인접 노드로 구성된다. 노드에서 나가는 에지 수는 같지만 들어오는 에지 수는 다를 수 있다.

정규화된 거리 ξ_{ij} 에 비례하는 가중치를 가진 에지를 연결한다. 다양한 종류의 데이터 집합에 적용할 수 있도록 ξ_{ij} 를 다음과 같이 정규화된 거리를 사용한다.

$$\xi_{ij} = \frac{d(x_i - x_j) - d_{\min}}{d_{\max} - d_{\min}} Q + 1 \quad (1)$$

Q는 상수이고, d_{\max} 와 d_{\min} 는 각각 주어진 가중 방향 그래프에서 최소와 최대 거리를 나타낸다. 상수 Q는 정규

화된 거리 측정의 세부 사항을 결정한다. ACHC 알고리즘은 절대적인 거리가 중요한 것이 아니라 상대적인 거리가 중요하기 때문에 본 논문의 실험에서는 Q의 값을 100으로 설정하여 거리를 1에서 101 사이로 정규화 한다.

또한 데이터의 영역이 얼마나 고밀도 인지를 나타낸다. 밀도를 계산하기 위한 다양한 방법[10][14][20]이 있는데 [15]에서는 밀도 정도를 k개의 가장 가까운 이웃을 연결하는 가중치를 가진 방향 그래프 (k-Nearest Neighborhood Directed Graph)로 정의한다. x_i 의 밀도 ρ_i 는 다음과 같이 정의한다.

$$\rho_i = \exp\left(\omega \frac{\mu^k}{d(x_i, x_i^k)}\right), \quad i = 1, \dots, n(2)$$

$d(x_i, x_i^k)$ 는 x_i 와 x_i^k (the k-th Nearest Neighbor of x_i)의 거리이고, μ^k 는 k개의 가장 가까운 이웃의 평균이다. $0 \leq \omega \leq 1$ 는 가중치이며 본 논문에서는 0.5로 설정한다.

데이터 클러스터링 문제해결에 실제 개미의 행동을 적용하기 위해 그래프 G의 각 에지 e_{ij} 를 연결한다. 개미가 e_{ij} 를 통해 통과할 때 페로몬 흔적을 남긴다.

2.2 노드페로몬(Node Pheromone)

제안하는 개미 기반 클러스터링 모델에서 다음과 같은 노화 메커니즘을 제안한다. 수명을 가진 개미는 각 노드에서 일정한 수 만큼 생성하고 그런 다음 개미는 한 노드에서 인접노드로 에지 e_{ij} 를 통해 이동할 때 에지 e_{ij} 와 관련된 페로몬 흔적 τ_{ij} 에 페로몬의 일부를 남긴다. 개미 이동 후 수명을 감소시키고 수명이 0이 되면 개미를 제거한다. 각 노드에서 주어진 수만큼 개미를 재생성하고 이 과정을 일정한 수만큼 반복한다.

개미는 밀도가 높은 노드를 선호하기 때문에 이런 노화 시스템이 없다면 결국 고밀도 노드의 영역에 갇혀버린다.

개미는 방문할 다음 노드를 선택함에 있어서 다른 개미에 의해 더 자주 방문된 노드를 선택하는 것을 선호하기 때문에 로컬영역에서 상대 밀도 값으로 간주되는 노드 페로몬 값으로부터 다음 노드를 선택하게 된다.

본 논문에서는 노드 페로몬 π_j 를 식(3)과 같이 정의한다. 노드 페로몬 π_j 는 노드 v_j 로 들어오는 모든 에지 페로몬의 합으로 정의된다.

$$\pi_j = \sum_{i \in \Omega_j} \tau_{ij} \quad (3)$$

식(3)에서 Ω_j 는 에지 e_{ij} 의 소스노드들의 집합이고 τ_{ij}

는 e_{ij} 를 통해 들어오는 개미들의 페로몬 흔적, 즉 에지 페로몬이다. 노드페로몬 π_j 값이 높다는 것은 노드 v_j 에 이웃노드의 개미들이 자주 방문 하였다는 것을 의미한다. 따라서 π_j 를 사용하는 것이 하나의 페로몬흔적 τ_{ij} 를 사용하는 것보다 개미학습에 더 효과적이다.

개미들이 그래프에서 다음 노드로 이동할 때 노드 페로몬에 따라 확률적으로 선택하여 이동하며 노드 v_i 에서 개미가 다음 노드로 이동할 확률은 다음과 같다.

$$p_{ij} = \begin{cases} \frac{\pi_j^\beta}{\sum_{k \in N_i} \pi_k^\beta}, & \text{if } j \in N_i; \\ 0, & \text{if } j \notin N_i; \end{cases} \quad (4)$$

N_i 는 v_i 의 인접노드의 집합이고 β 는 상수를 나타낸다. 개미가 다음 노드를 선택할 확률은 [16]에 따라 식(4)에 휴리스틱 정보를 포함할 수 있으며 수식은 다음과 같다.

$$p_{ij} = \begin{cases} \frac{\rho_j^\alpha \pi_j^\beta}{\sum_{k \in N_i} \rho_k^\alpha \pi_k^\beta}, & \text{if } j \in N_i; \\ 0, & \text{if } j \notin N_i; \end{cases} \quad (5)$$

N_i 는 v_i 의 인접노드의 집합이고 α 와 β 는 발견정보와 노드 페로몬의 상대적 영향력 계수를 나타낸다.

모든 개미가 다음 노드로 이동 후 페로몬 자취를 업데이트하는데 수식은 다음과 같다.

$$\tau_{ij} \leftarrow (1 - \theta)\tau_{ij} + \Delta\tau_{ij}, \quad (6)$$

$0 \leq \theta \leq 1$ 은 증발률[16]이고 τ_{ij} 는 페로몬 흔적이고 $\Delta\tau_{ij}$ 는 페로몬 흔적의 증가분이며 본 논문에서 θ 는 0.95로 설정한다.

최적화 문제[16]에서 증가분 $\Delta\tau_{ij}$ 는 일반적으로 페로몬의 양에 선형비례한다. 많은 수의 개미들이 노드 v_i 에서 노드 v_j 로 이동하면 페로몬 τ_{ij} 에 쌓이는 페로몬의 양은 그 수에 비례하여 매우 많아진다. 에지 e_{ij} 에 전달하는 개미의 숫자에 로그를 취해 다음과 같이 페로몬을 증가하는 방법을 적용하면 특정 에지의 영향을 줄일 수 있다.

$$\Delta\tau_{ij} = \log_2(O_{ij}) \frac{1}{\zeta_{ij}}, \quad (7)$$

O_{ij} 는 에지 e_{ij} 로 이동한 개미의 수이며, ζ_{ij} 는 식(1)에 정의된 표준화된 거리이다. 식(6)과 식(7)를 이용하여 식(3)을 다시 정리하면 노드페로몬 π_j 는 다음과 같다.

$$\pi_j \leftarrow (1 - \theta)\pi_j + \sum_{i \in \Omega_j} \log_2(O_{ij}) \frac{1}{\zeta_{ij}} \quad (8)$$

식(8)에서 들어오는 에지 $|\Omega_j|$, 각 에지로 들어오는 개미 수에 로그 값 $\log_2(o_{ij})$, 소스 노드로 들어오는 에지 정규화거리 ζ_{ij} 는 노드 v_j 에서 노드페로몬 π_j 값을 결정하는 요인이다.

각 노드에서 노드 페로몬 π_j 는 개미가 다음 이동할 노드를 선택할 확률에 사용되며 로컬 지역에서 상대 밀도로 간주된다. 예를들어 노드 v_j 가 들어오는 링크가 많을 때 노드 페로몬 π_j 은 상대적으로 큰값을 가질 것이고, 그 이웃 노드의 로컬 영역에서 더 높은 밀도 ρ_j 를 가진다. o_{ij} 의 값이 크다는 것은 그 주변의 이웃 노드들에 대한 밀도 ρ_j 가 상대적으로 높다는 것이다.

다음은 노드 페로몬을 사용하여 제안하는 개미 기반 클러스터링 알고리즘의 근거를 설명한다. 노드 v_j 가 클러스터의 경계영역에 존재하고 클러스터 내부보다 낮은 밀도 ρ_j 를 가진다고 가정하자. 그러면 결과적으로 노드 v_j 의 노드 페로몬은 비교적 작은 값을 가지며 이는 특이한 사실이다. 클러스터 내부 노드는 경계 또는 그 이웃 보다 높은 밀도를 가질 것이고 이러한 노드에 있는 개미는 영역내에서 서로를 방문할 확률이 매우 높다. 결과적으로 전체 반복후 클러스터 내부 노드는 경계 혹은 그 외부 노드들 보다 비교적 높은 노드 페로몬값을 가진다. 그래프에서 노드 페로몬값이 낮은 노드를 제거한다면 클러스터 내부의 노드만 구할 수 있고 해당 클러스터는 서로 분리된다. 그러면 쉽게 SCC (Strongly Connected Component) 를 찾아 클러스터의 그룹을 얻을 수 있다.

ACC(Ant Colony Clustering) 알고리즘의 절차는 시작에서 수명을 가진 개미들이 각 노드에서 생성되고 각 노드에서의 각 개미는 식(5)의 확률에 따라 인접한 노드들중 하나로 이동하고 식(6)과 식(7)에 따라 노드 페로몬을 업데이트 한다. 이과정을 반복후 노드 페로몬 π_j 값이 주어진 임계값 보다 작은 값을 갖는 노드를 제거한다. 노드 v_j 를 제거하면 v_j 에서 나가는 에지들과 v_j 로 들어오는 에지들 모두를 제거한다. 그러면 SCC의 집합을 얻을 수 있다. 제거된 노드는 k -NN 규칙[18]을 이용하여 얻어진 클러스터 중 하나로 할당한다. 수행 절차는 그림 1과 같다.

알고리즘 1 Ant Colony Clustering

```

1: procedure compute_np (G)
   /* Initialization Phase*/
2:   for each node  $v_j$  do
        $\pi_j = 1$ 
   end for
   /* Main Loop */
5:   for  $m = 1$  to #iteration do

```

```

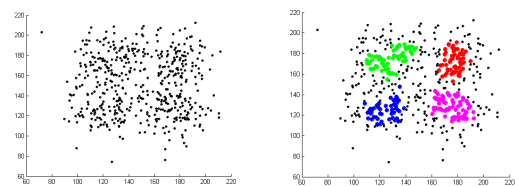
generate_ants (#ants)
for  $m = 1$  to lifetime do
  for each ant at each node  $v_i$  do
     $v_j = \text{select\_next\_node}(p_{ij})$  // see (3)
11:  move_to( $v_j$ )
    end for
    for each arc  $e_{ij}$  do
       $\tau_{ij} = \text{update\_pheromone\_trail}(o_{ij}, \xi_{ij})$  // see (4) and (5)
    end for
    for each node  $v_j$  do
16:       $\pi_j = \text{update\_node\_pheromone}(\tau_{ij})$  // see (6)
    end for
    end for
    remove_ants()
  end for
  return  $\pi$ 
21: end procedure

procedure remove_labeling( $G, \pi, \text{threshold}$ )
  /* remove nodes and label nodes */
  for each node  $v_j$  do
    if  $\pi_j < \text{threshold}$  then remove_node( $v_j, G$ ) end if
  end for
  SCC = label_node( $G$ )
  return (SCC)
end procedure

```

(그림 1) ACC 알고리즘
(Figure 1) ACC Algorithm

제안하는 ACC 알고리즘을 설명하기 위해 가우시안 분포를 따르는 4개의 클러스터를 만드는데 각 클러스터는 125개의 데이터로 이루어져 있으며 $[N(125,152), N(125,152)]$, $[N(175,152), N(125,152)]$, $[N(125,152), N(175,152)]$, $[N(175,152), N(175,152)]$ 를 중심으로 그림 1(a)과 같이 구성한다. 가중치를 가진 방향 그래프를 구축시 각 노드는 20개의 가까운 이웃을 연결한다. 알고리즘에서 개미의 수명은 10으로 반복 회수도 10으로 설정한다. 제안하는 알고리즘에서 사용하는 매개 변수는 4장에서 설명한다.



(그림 2) (a) 4개의 클러스터로 구성된 실험 데이터와
(b) ACC 알고리즘 클러스터링 결과
(Figure 2) Experiment of the ACC algorithm with
(a) a synthetic data set and
(b) the corresponding clustering result.

그림2(b)은 ACC 알고리즘에서 노드 중 낮은 값의 57%를 제거한 결과이다. 그림2(b)에서 검은 점은 제거 아웃 데이터 포인트이며, 다른 색상은 다른 클래스 라벨을 나타낸다. 제안하는 ACC 알고리즘은 클러스터가 매우 가까운 위치에도 불구하고 네 개의 클러스터를 찾는다.

3. 계층 클러스터 분석

이 섹션에서는 제안하는 ACC(Ant Colony Clustering) 알고리즘을 기반으로 계층적 클러스터링 분석을 위한 프레임 워크를 설명한다.

알고리즘 1에서 기재된 ACC 알고리즘은 두 부분으로 구성된다.:(a) 노드 페로몬을 계산하고 (b) 노드 제거 및 라벨링 한다. 노드 페로몬 계산후 임계값을 바꾸면서 노드 제거와 라벨링을 한다. ACC 알고리즘은 노드 페로몬 π_i 의 다른 임계값에 따라 서로 다른 클러스터링 결과를 준다. 알고리즘 1에기재된 `compute_np` 절차로부터 얻어지는 노드 페로몬 π_j 의양에 따라 클러스터의 계층 구조를 생성한다.알고리즘 1에서 임계값 t_c 와 t_D 를 사용하여 `remove_labeling` 과정에서 얻은클러스터의집합을 $C = \{C_i, i = 1 \dots n\}$ 과 $D = \{D_j, j = 1 \dots m\}$ 로 가정하자. 그러면, $t_c \leq t_D$ 인 경우 각 D_j 가 포함되는 클러스터 C_i 가 다음 식과 같이 항상 존재한다.

$$D_j \subseteq C_i, \quad \text{for } \forall D_j \text{ if } t_c \leq t_D. \quad (9)$$

즉, 모든 클러스터 $D_j \in D$ 는 $C_i \in C$ 에서 나오고 D_j 는 C 클러스터 중 하나의 자식 클러스터(child cluster)이다. C_i 에 있는 모든 노드가 임계값 t_D 를 통해 제거되었을 수 있기 때문에 특정 클러스터 C_i 는 D 의 자식 클러스터가 없을 수 있다. 클러스터 C_i 가 여러 클러스터로 분할될 수 있기 때문에 다른 한편으로는 C_i 가 D 에 두 개 이상의 하위 클러스터로 있을 수 있다.

식(9)를 이용하여 노드 페로몬의 값을 낮은 값에서 높은 값으로 임계 값을 변화하여 클러스터의 계층 구조를 구축한다. 계층 구조 클러스터는 시작 및 종료 임계 값과 구성된 데이터 포인트로 구성된다. 시작 임계 값은 해당 클러스터가 생성되는 시점이며, 종료 임계 값은 해당 클러스터가 제거되는시점이다.

그림3에서는 ACHC과정을 보여준다. 그림1에 기술된 절차를 사용하여 각 노드에 대한 노드 페로몬을 얻고 반복적으로 임계 값을 낮은 값에서 높은 값으로 조절하여 `remove_labeling` 루틴을 적용한다.

알고리즘 2 Ant Colony Hierarchical Clustering

```

1: procedure ACHC
     $\pi = phi = \text{compute\_np}(G)$  // see Algorithm 1
2: sort(phi)
    root = {0, 0, X} // {start, end, set of data points}
    t.add_nodes(root) // tree for hierarchy
    q.add_nodes(root) // queue for set of leaf nodes
5: for s = 0 to h_level do // h_level = 99
    percentile = s x step // step = 1%
    t_value = phi[percentile]
    SCC = remove_labeling(G,  $\pi$ , t_value)
    // see Algorithm 1
    add_nodes(SCC, t_value)
    end for
11: end procedure

procedure add_nodes(C, th)
    size = q.size
    for i = 0 to size do
        leaf = q.remove_node()
16: leaf.end = th // update end threshold
        D = get_child_clusters(leaf, C) // D is a set of child clusters
        if (|D| = 1) then // if leaf has not been split
            q.add_nodes({leaf}) // put leaf back into queue
        end if
21: else if (|D| > 1) then // if leaf has been split
        L = make_nodes(D, th) // make a node for each child cluster
        t.add_nodes(leaf, L) // add new child nodes to this leaf
        q.add_nodes(L) // add new nodes in queue
    end else if
    end for
end procedure
    
```

(그림 3) ACHC 알고리즘
(Figure 3) ACHC Algorithm

[10][14]에서와 같이 여러 가지 한계 레벨을 결정할 수 있다. 그래서 먼저 노드 페로몬 π_j 을 정렬하고 대응하는 제거 퍼센트에 대한 임계값을 계산한다. 본 논문은 100개의 각 임계값 수준에서 다른 클러스터링 결과 C를 가질 수 있도록 주어진 데이터 노드의 1%씩을 제거한다. 제거 과정을 수행 하기 직전에 노드 페로몬은 한번 계산하고 그 다음 알고리즘2에 따라 임계 단계에 따라 클러스터의 계층 구조를 구축한다. 트리에서 분할하지 않으면 단순히 종료 임계 값을 업데이트 하지만 계층 구조의 시각화에 대한 클러스터링 결과는 유지한다.

4. 실험 결과

이 섹션에서는 제안하는 ACHC (Ant Colony Hierarchical Clustering) 알고리즘의 성능을 잘 보여 줄 수 있는 합성 데이터 집합을 사용하여 실험하고 [1], [2], [6], [23], [28]

과 같이 개미 기반 클러스터링에서 자주 사용하는 UCI 기계학습 저장소 데이터 집합[19]인 실제 데이터 집합을 사용하여 실험한다. 제안하는 알고리즘은 밀도를 기반으로 클러스터링하는 알고리즘으로 매우 유사한 HDS(Hierarchical Density Shaving)[14]와 비교 실험한다.

4.1 파라미터설정(Parameter Settings)

제안하는 ACHC (Ant Colony Hierarchical Clustering) 알고리즘은 사전에 식(5)에서 k개 이웃 연결, 개미수명, 반복횟수, 노드 제거량, α , β 매개 변수들에 대한 값을 결정한다.

다양한 데이터 집합과 일련의 실험에서 개미의 수명과 반복횟수가 노드 페로몬에 크게 영향을 미치지 않는 것을 발견해서 실험을 통해 수명은 10으로 반복 횟수는 100으로 설정한다.

한편, k개의 가장 가까운 이웃의 개수는 노드 페로몬의 계산에 영향을 미친다. k값이 작으면 노드 페로몬은 주어진 노드의 주변지역의 밀도분포를 반영하지 않을 수 있고 너무 크면 노드 페로몬 로컬영역에서 밀도 분포의 내용을 대표하지 않을 수도 있다. 따라서 데이터 포인트의 집합에서 주어진 가중방향 그래프를 구축할 때 적절한 방식으로 가장 가까운 이웃들의 수를 결정하는 것이 중요하다. [9]에서 제시하는 방법을 이용하여 K값은 데이터 포인트의 수의 제곱근을 사용한다.

제안하는 ACHC의 다른 중요한 요인은 식(5)에 정의된 휴리스틱 정보와 노드 페로몬, α , β 이다. 본 논문에서는 α , β 를 다음과 같은 세가지 경우로 실험한다. - Case 1: ($\alpha = 1, \beta = 0$), Case 2: ($\alpha = 1, \beta = 1$), Case 3: ($\alpha = 0, \beta = 1$). 이 실험은 극단적인 경우를 비교할 수 있다.

클러스터의 계층 구조 시각화는 데이터 구조를 이해하고 탐구하는데 매우 도움이 된다. 제안하는 ACHC 알고리즘으로 부터 얻어진 클러스터의 계층 구조를 시각화하기 위해 [14]에 기재된 시각화 기법을 이용한다. 이 방식에서 x축은 제거 퍼센트이며 y축은 그에 따른 각 클러스터에서의 데이터 포인트의 수를 나타낸다. 클러스터는 서로 다른 색으로 표시하고 부모로부터 분리되지 않은 하위 클러스터는 자신의 부모와 함께 같은 색을 공유하며 새로운 색상은 부모로부터 분리되는 자식 클러스터이다. 이 방식에서는 같은 색의 폭은 계층 구조에서 노드의 시작과 끝 사이의 임계 지속 시간을 나타내고 신장은 각 클러스터에서의 데이터 포인트 수를 나타낸다.

4.2 평가함수(Evaluation Functions)

제안하는 알고리즘과 가장 유사한 계층적 클러스터링인 HDS(Hierarchical Density Shaving) 알고리즘[14]과 비교한다. 분할 클러스터링 알고리즘인 K-means는 분석을 보다 보다 쉽게 하기 위해 비교 대상에 추가한다. 클러스터링 알고리즘의 성능을 비교하기 위해 F-측정(F-Measure), Dunn 지수(Dunn Index), 클러스터 내 분산(Intra-cluster variance)의 세 가지 분석 측정 방법을 사용한다[2][6].

F-측정(F-measure)는 정보 검색 문헌 [17]에서 정밀도(precision)와 리콜(recall)을 결합한 것이다. $p(i, j) = n_{ij}/n_i$ 와 $r(i, j) = n_{ij}/n_j$ 이며 n_{ij} 는 클러스터 j에 있는 클래스 i의 개수이고, n_j 는 클러스터 j에 멤버 개수이다.

클러스터 J와 클래스 I의 F-측정은 다음과 같다.

$$F(i, j) = \frac{2 \times p(i, j) \times r(i, j)}{p(i, j) + r(i, j)}$$

클러스터링 결과에 대한 전체 F-측정에 의해 주어진 F-측정에 대한 모든 값의 가중 평균을 고려하여 계산된다.

$$F = \sum_i \frac{n_i}{n} \max_j \{F(i, j)\}$$

N은 데이터 집합의 전체 크기이다. F 측정은 값이 클수록 클러스터링 결과가 좋다.

Dunn 지수(Dunn Index)는 특정 파티션을 클러스터 간 거리와 클러스터 직경의 최소 비율을 결정한다. 하나의 클러스터 내의 데이터 요소가 서로 다른 클러스터 내의 데이터 보다 훨씬 가깝게 되어야 한다는 개념이다. 이것은 다음과 같이 정의되며, 값이 클수록 클러스터링 결과가 좋다.

$$D = \min_{c, d \in C} \left[\frac{\delta(\mu_c, \mu_d)}{\max_{e \in C} [diam(e)]} \right]$$

C는 모든 클러스터들의 집합이고, 클러스터 C의 직경 $diam(c)$ 는 최대 클러스터 간 거리로서 계산하고 $\delta(\mu_c, \mu_d)$ 는 클러스터 c와 d의 중심사이의 거리이며 최대화되어야 한다.

클러스터 내 분산(Intra-cluster variance)은 모든 데이터 항목과 개별 클러스터 내의 데이터 요소가 유사해야 한다. 연관된 클러스터 중심 사이의 편차의 제곱의 합을 계산하며 다음과 같다.

$$I = \sqrt{\frac{1}{n} \sum_{c \in C} \sum_{i \in c} \delta(i, \mu_c)^2}$$

C는 모든 클러스터의 집합이고, μ_c 는 클러스터 c의 중심이다. $\delta(i, \mu_c)$ 는 각 데이터 I와 관련된 클러스터 중심 사이의 편차를 계산하기 위해 사용된 거리함수이며 최소화 되어야 한다. 클러스터 내 분산은 값이 작을수록 클러스터링 결과가 좋다.

제안하는 ACHC(Ant Colony Hierarchical Clustering) 및 HDS(Hierarchical Density Shaving)의 성능을 비교 평가하기 위해 계층 구조에서 클러스터를 식별할 필요가 있다. 본 논문에서 계층 구조를 사용하고 제거된 노드(shaved-out nodes)를 K개의 인접 이웃 분류 체계를 적용하여 클러스터 레이블을 할당한 다음 다양한 평가 수단을 계산한다. 이 제할당은 모든 응용 프로그램에 항상 필요한 것은 아니다.

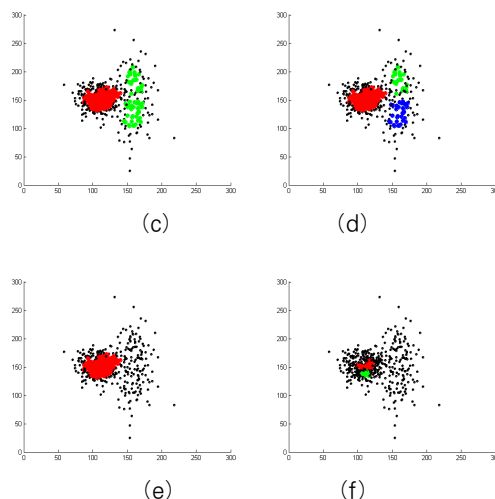
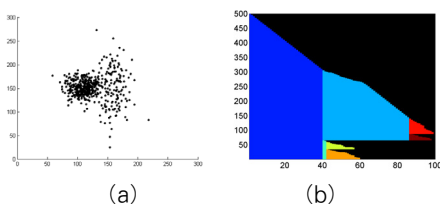
4.3 합성데이터집합(Synthetic Data Set)

제안하는 ACHC(Ant Colony Hierarchical Clustering)와 HDS(Hierarchical Density Shaving)의 알고리즘의 성능을 비교 평가하기 위해 그림4(a)에서와 같이 서로 다른 밀도를 갖는 A, B, C세 개의 클러스터로 구성된 인공 데이터 집합을 사용한다.

클러스터 A는 $[N(110,152), N(150,152)]$ 의 가우시안 분포에서 생성된 350개의 데이터 포인트를 가지고 다른 두 개의 클러스터 B와 C는 각각 $[N(160,152), N(120,302)]$ 와 $[N(160,152), N(180,302)]$ 에서 각각 125의 데이터 포인트를 가지며 클러스터 A는 클러스터 B와 C보다 밀도가 높다.

그림4(b)에서 $\alpha=1, \beta=0$ 으로 제안하는 ACHC를 사용하여 클러스터의 계층구조를 시각화한다. 그 결과 제안하는 ACHC는 4개의 클러스터로 분할 한다. 그림4 (c), (d), (e) 및 (f)는 각각 40%, 42%, 62% 및 87%의 퍼센트를 제거하여 대응하는 클러스터링 결과를 나타낸다.

제안하는 ACHC에서 클러스터 A는 클러스터 B와 C 사이의 경계에서 보다 클러스터 A의 경계의 ρ_j 밀도가 높기 때문에 B와 C보다 쉽게 분할 된다. 식(5)에 따라 개미가 이웃하는 노드들에 대해 더 높은 밀도의 노드를 선호하기 때문이고 클러스터 A의 경계에 있는 노드의 밀도가 클러스터 B 및 C의 경계에 있는 노드의 밀도보다 상대적으로 낮기 때문이다.



(그림 4) Case 1($\alpha=1, \beta=0$)의 ACHC 실험결과이며 (a)는 밀도가 다른 3개의 클러스터로 이루어진 데이터이고 (c) 40%, (d) 42%, (e) 62%, (f) 87%를 제거한 결과이며 (b)는 계층적 클러스터 구조시각화

(Figure 4) Experimental results obtained from the ACHC of Case 1 ($\alpha=1$ and $\beta=0$) with (a) a synthetic data set with a different removing percentile of (c) 40% (d) 42% (e) 62%, and (f) 87%, (b) and the visualization of hierarchy of clusters

표1에서의 ACHC, HDS의 성능 측정을 비교한 것이다. K-Means는 ACHC 알고리즘 성능이 우수하다는 것을 보이기 위해 단순 비교를 목적으로 측정한 것이다. 이 실험에서 표1과 같이 HDS의 F-측정과 던 지수 값이 제안하는 ACHC에 비하여 낮고 클러스터 내 분산이 높은 이유는 그림4(a)를 HDS로 수행하면 클러스터의 개수가 2개로 되기 때문이다. HDS는 처음에는 그림4(c)와 같이 클러스터가 2개로 나뉘지고 더 노드를 제거하면 그림4(e)와 같이 1개만 남는다. 이와 같이 HDS는 밀도가 상대적으로 낮거나 높은 부분은 나뉘지지 않는 경향이 있다. ACHC는 노드 페로몬이 로컬 영역의 상대 밀도를 측정하기 때문에 클러스터를 더 잘 분리되며 HDS는 클러스터 B와 C를 클러스터 A에 너무 가까이 있어서 클러스터 B와 C를 식별할 수 없다.

(표1) 그림4(a)의 데이터를 가지고 ACHC, HDS, K-means 알고리즘 성능을 측정 비교 분석한 결과

(Table 1) Evaluation measures of the performance from the proposed ACHC, HDS, and K-Means with the data set shown in fig. 4(a)

	F측정 (F Measure)	던지수 (Dunn Index)	클러스터내분산 (Intra Cluster Variance)
ACHC ($\alpha=1, \beta=0$)	0.948198	0.476222	23.391024
ACHC ($\alpha=1, \beta=1$)	0.939674	0.480762	23.528282
ACHC ($\alpha=0, \beta=1$)	0.942483	0.481347	23.352591
HDS	0.472604	0.245643	29.160772
K-means	0.937749	0.461142	22.972417

상대적인 밀도가 중요한 데이터에서는 ACHC 알고리즘이 HDS 알고리즘보다 우수하다.

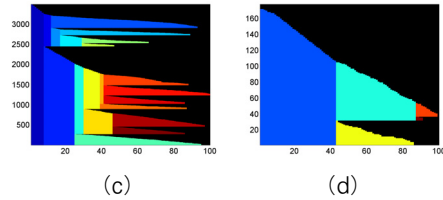
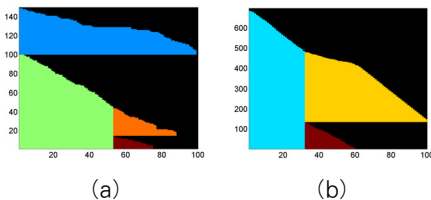
4.4 실제데이터집합(Real Data Set)

이 세션에서는 UCI 기계 학습 저장소 [19]에서 얻은 데이터 집합(Iris, Wisconsin, Wine, Digits)인 실제 데이터 집합에 대한 표2에 실험 결과를 제시한다.

(표 2) 실험 데이터 요약: n는 데이터의 포인트수, Ci는각 클러스터의 포인트수, d는 데이터 포인트의 차수, #C는 클러스터의 수, Type은 입력 데이터의 종류

(Table 2) Summarization of experimental data sets: N stands for number of data points, Ci for size of each cluster, D for dimension of data point, #C for number of cluster, and type for input data type

이름	N	C _i	D	#C	Type
Iris	150	3 x 50	4	3	Real
Wisconsin	699	458, 241	9	2	Integer
Digits	3498	363, 364, 364, 336, 364, 335, 336, 364, 336, 336	16	10	Integer
Wine	178	59, 71, 48	13	3	Real



(그림 5)(a) Iris, (b) Wisconsin, (c) Digits, (d) Wine 데이터를 Case 1 ($\alpha=1, \beta=0$)로 제한하는 ACHC를 수행하여 계층적 클러스터 결과를 시각화 (Figure 5) Hierarchy of clusters obtained from the ACHC of Case 1 ($\alpha=1$ and $\beta=0$) with (a) Iris (b) Wisconsin (c) Digits, and (d) Wine Data Sets.

그림3에서 $\alpha=1, \beta=0$ 을 사용하여 ACHC 알고리즘에서 클러스터의 계층구조를 얻는다. 결과화면은 관련 데이터 세트의 데이터 구조를 쉽게 발견할 수 있다.

표3은 ACHC(Ant Colony Hierarchical Clustering)와 HDS(Hierarchical Density Shaving) 알고리즘에 대한 성능을 비교한 것이고 K-Means 그 성능의 우수성을 보이기 위해 ACHC 알고리즘에서 찾은 클러스터의 개수를 사용하여 비교한다. ACHC는 노드 페로몬이 확률적으로 결정되기 때문에 ACHC를 50회 실험하고 각 결과에 대해 평균을 계산하였다. 그 결과 표3에서 보는 것과 같이클러스터의 수를 찾았고 ACHC 알고리즘이 HDS 보다 결과가 우수하다는 것을 확인할 수 있다. 이 실험에서 보듯이 식(5)에서 노드 페로몬을 사용하는 ACHC의 경우 성능이 우수하다는 것을 확인할 수 있다.

(표 3) Iris, Wisconsin, Digits, Wine 데이터를 제한하는 ACHC 알고리즘과 HDS, K-means와 성능을 비교 분석 (Table 3) Evaluation measures of the performance from the proposed ACHC, HDS, and K-Means with the data sets of IRIS, WISCONSIN, DIGITS, and WINE.

이름	ACHC ($\alpha=1, \beta=0$)	ACHC ($\alpha=1, \beta=1$)	ACHC ($\alpha=0, \beta=1$)	HDS	K-means	
Iris	F Measure	0.8468	0.8353	0.8379	0.8905	0.8609
	Dunn Index	0.5028	0.4555	0.4709	0.7090	0.6133
	Intra Cluster Variance	0.6868	0.6735	0.6780	0.7514	0.7605
	# of Clusters	3.8	3.98	3.94	3	3

이름		ACHC ($\alpha=1, \beta=0$)	ACHC ($\alpha=1, \beta=1$)	ACHC ($\alpha=0, \beta=1$)	HDS	K-means
Wisconsin	F Measure	0.9673	0.9620	0.9617	0.8766	0.9192
	Dunn Index	0.4658	0.5121	0.5095	0.5929	0.5901
	Intra Cluster Variance	5.3804	5.4390	5.4931	5.7632	5.3877
	# of Clusters	2.28	2.12	2.1	2	2
Digits	F Measure	0.7980	0.8224	0.8316	0.7996	0.6888
	Dunn Index	0.2326	0.2287	0.2067	0.2388	0.3083
	Intra Cluster Variance	64.8927	63.5327	63.1956	63.0883	65.7765
	# of Clusters	11.04	12.08	12.54	12	10
Wine	F Measure	0.8756	0.9111	0.9114	0.7446	0.9495
	Dunn Index	0.4006	0.4056	0.4062	0.3765	0.3788
	Intra Cluster Variance	2.7272	2.7024	2.7024	3.0925	2.6933
	# of Clusters	2.98	3	3	2	3

5. 결 론

본 논문은 새로운 개미 기반의 계층적 클러스터링 알고리즘인 ACHC(Ant Colony Hierarchical Clustering) 알고리즘을 제안한다. 제안하는 ACHC 알고리즘은 그래프에서 노드로 들어오는 에지 페로몬의 합인 노드 페로몬의 개념을 기반으로 한다. 노드 페로몬의 양은 노드 밀도 관점에서 그 주변의 이웃과 경쟁하며 개미 수신 빈도를 반영한다. 따라서 노드 페로몬은 그래프에서 개미 확률적 이동을 통해 상대 밀도를 측정할 수 있다. 이 상대 밀도를 기반으로 계층적 클러스터링 하는 ACHC 알고리즘은 절대 밀도 기반의 HDS(Hierarchical Density Shaving) 알고리즘의 결과보다 더 좋은 성능을 보여준다.

데이터 집합의 계층적 분석은 노드 페로몬 양에 따라 노드의 반복적인 제거에 의해 수행된다. 제안하는 방법은 기존의 방법과 비교하여 데이터 집합의 계층 구조 분석에 매우 효과적임을 보여준다. 앞으로는 대용량의 데이터를 처리할 수 있는 알고리즘을 연구할 것이다.

참 고 문 헌(Reference)

- [1] Rui Xu and Donald Wunsch II, "Survey of Clustering Algorithms," IEEE Transactions On Neural Networks, Vol. 16, No. 3, pp. 645-678, May 2005.
- [2] Yan Yang and M. S. Kamel, "An aggregated clustering approach using multi-ant colonies algorithms," Pattern Recognition Vol. 39, Issue 7, pp. 1278-1289, July 2006.
- [3] Deneubourg, J.-L., Goss, S., Franks, N., Sendova-Franks, A., Detrain, C., and Chretien, L., "The dynamics of collective sorting: Robot-like ants and ant-like robots," in Proceedings of the First International Conference on Simulation of Adaptive Behavior: From Animals to Animats 1, 1991, pp. 356-365, Cambridge, MA: MIT Press.
- [4] Lumer, E. and Faieta, B., "Diversity and adaptation in populations of clustering ants," in Proceedings of the Third International Conference on Simulation of Adaptive Behavior: From Animals to Animats 3, pp. 501 - 508, Cambridge, MA: MIT Press.
- [5] Handl, J. and Meyer, B., "Improved ant-based clustering and sorting in a document retrieval interface," in Proceedings of the Seventh International Conference on Parallel Problem Solving from Nature, 2002, Vol. 2439 of Lecture Notes in Computer Science, pp.913-923. Berlin, Germany: Springer-Verlag.
- [6] Handl, J., Knowles, J., and Dorigo M., "Ant-based clustering and topographic mapping," Artificial Life, Vol. 12, No. 1, pp. 35 - 62, 2006.
- [7] Shang Liu, Zhi-Tong Dou, Fei Li, Ya-Lou Huang, "A New Ant Colony Clustering Algorithm Based on DBSCAN," in Machine Learning and Cybernetics, Proc. of 3th International Conf., 2004, vol. 3, pp 1491 - 1496.
- [8] N. Monmarché, M. Slimane, G. Venturini, "AntClass: discovery of clusters in numeric data by a hybridization of an ant colony with the K-means algorithm," Internal Report No 213, France, Jan. 1999.
- [9] Ling Chen, Li Tu, Hong-Jian Chen, "Data Clustering by Ant Colony on a Digraph," in Machine Learning and Cybernetics, Proc. of 4th International Conf., Guangzhou Aug. 18 - 21, 2005, vol. 3, pp 1686 - 1692.

- [10] Gunjan Gupta, Alexander Liu, Joydeep Ghosh, "Hierarchical Density Shaving: A clustering and visualization framework for large biological datasets," in Proc. 6th IEEE International Conf. on Data, Washington, 2006, pp 89 - 93.
- [11] Ramos, V. and Merelo, J., "Self-organized stigmergic document maps: Environment as a mechanism for context learning," in Proceedings of the First Spanish Conference on Evolutionary and Bio-Inspired Algorithms, Spain, 2002, pp. 284-293.
- [12] Kuntz, P and Snyers, D., "Emergent colonization and graph partitioning," in Proceedings of the Third International Conference on Simulation of Adaptive Behavior: From Animals to Animats 3, 1994, pp. 494-500, Cambridge, MA: MIT Press.
- [13] Kuntz, P and Snyers, D., "New results on an ant-based heuristic for highlighting the organization of large graphs," in Proceedings of the Congress of Evolutionary Computation, 1999, pp. 1451-1458, IEEE Press.
- [14] Gupta G, Liu A, Ghosh J., "Automated hierarchical density shaving: a robust automated clustering and visualization framework for large biological data sets." IEEE/ACM Trans Comput Biol Bioinform. 2010 Apr-Jun;7(2):223-237
- [15] KiYoung Lee, Dae-Won Kim, K. H. Lee, and D. Lee, "Density-Induced Support Vector Data Description," IEEE Transactions on Neural Networks, Vol. 18, No. 1, pp. 284-289, January 2007.
- [16] Marco Dorigo and Thomas Stutzle, Ant Colony Optimization, MIT Press, 2004.
- [17] Soumen Chakrabarti, mining the Web: Discovering Knowledge from Hypertext Data, Morgan Kaufmann Publishers, 2003.
- [18] Richard O. Duda, Peter E. Hart, and David G. Stork, Pattern Classification, John Wiley & Sons, 2001.
- [19] Blake, C. and Merz, C. UCI repository of machine learning databases, Technical report, Department of Information and Computer Sciences, University of California, Irvine.
- [20] M. Halkidi, M. Vazirgiannis, and I. Batistakis, "Quality scheme assessment in the clustering process," in Proceedings of the Fourth European Conference on Principles of Data Mining and Knowledge Discovery, Vol. 1910 of Lecture Notes in Computer Science, pp. 265 - 267, 2000.
- [21] J. Han and M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, 2006.
- [22] Debajit Sensarma, Koushik Majumder, "A Novel Hierarchical Ant based QoS aware Intelligent Routing Scheme for MANETS", International Journal of Computer Networks & Communications (IJCNC) Vol.5, No.6, November 2013
- [23] Hanene Azzag, Gilles Venturini, Antoine Oliver and Christiane Guinot, "A hierarchical ant based clustering algorithm and its use in three real-world applications", European Journal of Operational Research, Vol. 179, Issue 3, 16 June 2007, pp. 906-922
- [24] J. Chircop and C. D Buckingham, "A Multiple Pheromone Ant Clustering Algorithm", Proceedings of NICSO 2013, to be published in Studies in Computational Intelligence, Springer, 2013
- [25] Jan Chircop, Christopher D. Buckingham, "The Multiple Pheromone Ant Clustering Algorithm and its application to real world domains", Proceedings of the 2013 Federated Conference on Computer Science and Information Systems pp. 27 - 34
- [26] Kumar V and Balasubramanie P, "Ant Colony Optimization Using Hierarchical Clustering in Mobile Ad Hoc Networks", European Journal of Scientific Research, 2011, Vol. 61 Issue 4, p549-560.
- [27] O.A. Mohamed Jafar and R. Sivakumar, "Ant-based Clustering Algorithms: A Brief Survey" ,International Journal of Computer Theory and Engineering, Vol. 2, No. 5, October, 2010, 1793-8201
- [28] Wafa'a Omar, Amr Badr and Abd El-Fattah Hegazy, "Hybrid Ant-Based Clustering Algorithm With Cluster Analysis Techniques", Journal of Computer Science 9 (6): 780-793, 2013

● 저 자 소 개 ●



강 문 수 (Mun-Su Kang)

2005년 한국항공대학교 컴퓨터공학과 졸업(학사)

2007년 한국항공대학교 대학원 컴퓨터공학과 졸업(석사)

2014년 한국항공대학교 대학원 컴퓨터공학과 졸업(박사)

2014년~현재 티켓몬스터 기술정보랩

관심분야 : 웹데이터마이닝, 클러스터링, 빅데이터 etc.

E-mail : mskang@kau.ac.kr



최 영 식 (Young-Sik Choi)

1985년 연세대학교 전자공학과 졸업(학사)

1987년 연세대학교 대학원 전자공학과 졸업(석사)

1996년 미국 Missouri주 Columbia대학교 대학원 Computer Engineering 졸업(박사)

2000년~2002년 KT(주) 멀티미디어연구소 연구실장

2002년~현재 한국항공대학 항공전자정보통신공학부 교수

관심분야 : 퍼지, 웹데이터마이닝, 영상처리, 이미지검색, etc.

E-mail : choimail@kau.ac.kr

