

# 공간에서의 다중 균일 분포 군집의 데이터 포인트를 중심으로 하는 원의 누적합 시각화와 특성 분석

이진성

인천대학교 임베디드시스템공학과

awwdde22ljs@gmail.com

## Visualizing and Analysis of Cumulative Sum of Circular Centered on Data Points in Multi-Uniform Distribution Clusters in Space

Lee Jinseong

Dept. of Embedded Systems Engineering, Incheon National Univ.

### 요약

본 논문은 균일한 면적으로 분포된 군집에 대해 데이터 포인트를 중심으로 하는 원의 누적합의 시각화와 이를 분석한 내용을 제시한다. 원의 면적에 따라 보이는 누적값의 분포의 특징을 제시하였다. 누적값을 사용하여 군집에 대한 클러스터 헤드의 추정기법에 대해 다루었으며, 기존 클러스터링 방식들과의 차이에 대해 다루었다. 데이터 포인트 간의 거리를 사용하는 방식과 비교하여 누적합을 사용하는 방식에 의해 일반적인 상황에서의 사용 이점은 저조할 것으로 보인다. 이산상황에서의 컨볼루션, 병렬연산 등과 결합하였을 때 이점을 보일 것으로 전망하며, 이를 통해 이미지와 같은 이산환경에서의 데이터 마이닝과 이미지 프로세싱에 활용할 수 있을 것으로 보인다. 또한 통신 네트워크 분야에도 적용할 수 있을 것으로 기대한다.

### I. 서론

기존에 사용되고 있는 여러 클러스터링 알고리즘들은 다양한 방식으로 개발되어 왔다. K-means와 같은 방식의 경우 특정 데이터 포인트에 대해 다른 데이터 포인트의 거리를 사용하는 방식을 사용하며, 밀도기반 클러스터링은 특정 반경 내의 데이터 포인트의 수를 활용한다. 이와같이 대부분의 클러스터링 알고리즘들은 데이터 포인트 간의 거리를 통해 클러스터 헤드를 추출해내는 방식을 사용한다.

본 연구에서는 데이터가 분포한 다양한 상황 중 각 군집의 분포된 면적이 균일하고, 그 값을 알고있는 상황이라고 가정한다. 이때 군집들이 인접해 구분, 혹은 분할(Cluster Partitioning)해야하는 상황에 대해 모든 데이터 포인트들을 중심으로 하는 원 영역의 누적합을 구하고, 이들의 특징을 보이고자 한다.

### II. 본론

#### 2.1. 실험 방법

한 군집을 원 내에 분포시키기 위해 중심으로부터의 거리와 방위각에 대한 균일분포 난수를 생성하여 원 면적 내에 균일하게 분포시킨다. 이때 거리에 대해서는 균일분포에 대한 루트값을 사용하여 면적에 대한 균일하게 분포할 수 있게끔 하였다.

무작위로 분포된 군집을 만드는 과정에서 중심부터 외곽의 거리를 안다고 가정하였으며, 원형으로 배치하였으므로 군집의 경계를 이루는 원에 대한 반지름을 안다고 할 수 있다. 이 반지름 값을 동일하게 반지름으로 사용하는 원을 각 데이터 포인트를 중심으로 그리며, 모든 원에 대하여 해당되는 면적의 영역에 값을 1씩 누적한다. 이때 비교 분석을 위해 원의 반지름이 절반인 경우와 두 배인 경우에 대한 실험 및 분석을 동시에 진행하

였다.

본 연구에서는 1000 by 1000 해상도의 이미지에서 반지름이 100인 원 내에 무작위의 좌표 200개를 포함하는 군집 5개를 만들어 실험을 진행하였다.

#### 2.2. 실험 결과

먼저 거리와 방위각에 대한 균일분포 난수를 통해 반지름이 100인 원 내부에 200개의 균일분포된 데이터 포인트들을 생성하였다. 또한 동일한 방식으로 총 5개의 군집을 좌표평면 내에 위치시켰다.

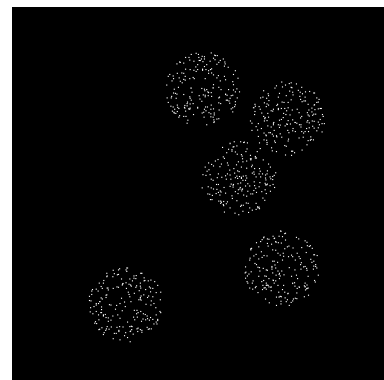


그림 1. 랜덤 데이터 분포  
Fig.1. Random Data Distribution

200개의 데이터 포인트를 갖는 5개의 군집을 랜덤으로 분포시켰다. 이때 군집의 영역이 겹쳐 특정 구역의 밀도가 높아지지 않는 랜덤값을 갖는 케

이스를 제외하였다.

1000개의 데이터 포인트에 대해 반지름이 100인 원에 대한 영역을 만들고, 해당 영역에 해당하는 픽셀들의 값을 1씩 누적하였다.

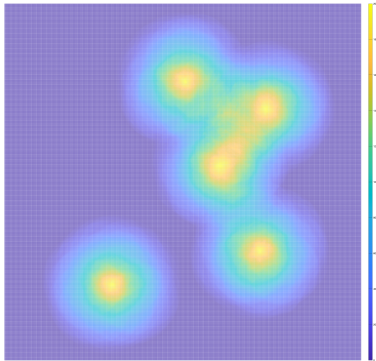


그림 2. 원의 누적합 결과의 시각화  
Fig.2. Visualization of Cumulative Sum Results of a Circle

데이터 포인트를 중심으로 하는 원에 해당되는 영역에 한번 이상 해당되어 누적합의 값이 0이 아닌 영역은 각 군집의 반지름의 두배에 해당하는 영역 내에 균일하게 분포하며, 원형을 띄게 된다.

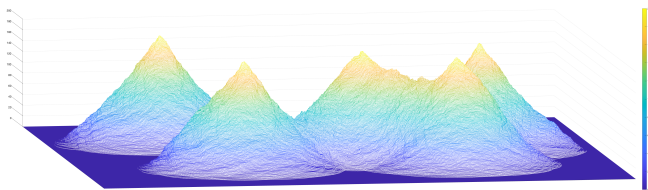


그림 3. 우측면 시점의 누적합 결과  
Fig.3. Cumulative Sum Results from a Right-Side Perspective

이때 누적값들은 군집의 중심에서 외곽으로 향할수록 일정하게 감소하는 형태를 띤다. 군집이 서로 인접한 경우 누적영역이 겹치면서 일정하게 감소하지 않으며, 각 군집의 중앙을 연결한 선 인근의 값이 다른 거리가 비슷한 좌표에 비해 높은 것을 볼 수 있다. 군집의 중심에서는 군집의 데이터 포인트의 수와 일치하거나 근사한 수치를 띄게 된다. 이때 누적값이 극대값인 경우의 좌표는 클러스터 헤드와 일치하거나 매우 인접함을 알 수 있다.

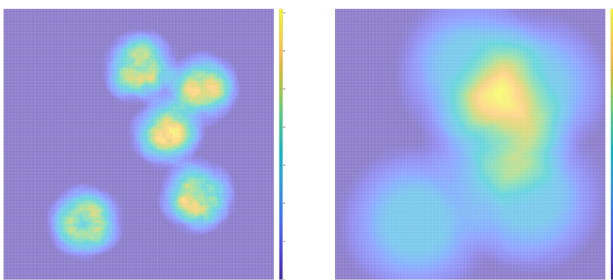


그림 4. 원의 면적이 군집 면적보다 작은 경우와 큰 경우의 누적값  
Fig.4. Cumulative Values for Cases Where the Area of a Circle is Smaller and Larger than the Cluster Area

누적원의 면적이 군집의 면적보다 작은 경우 누적값이 데이터 군집과 유사하게 분포되지만 중심으로부터 외곽으로 향할수록 지속적으로 감소하

는 것이 아닌 어느정도 평탄한 값을 볼 수 있으며, 데이터의 분포에 따라 군집의 중앙보다 어느정도 거리가 있는 위치에서 더 큰 값을 갖는 경우를 볼 수 있다.

누적원의 면적이 군집의 면적보다 큰 경우에는 인접한 군집을 모두 포함하는 데이터들에 대한 중앙이 극대와 유사함을 볼 수 있었으며, 인근의 값들에 대해서도 각 군집의 중심을 연결한 다각형에 대한 중선 다각형의 변에 집중되는 것을 볼 수 있다.

원의 누적합의 시각화 이미지를 통한 클러스터 헤드의 추정과 다른 클러스터링 방식과의 차이점을 찾을 수 있다. 기존의 클러스터링 방식들은 각 데이터 간의 거리에 의존하여 클러스터링을 진행하였다. DBSCAN과 같이 특정 영역 내에 포함된 데이터 포인트를 활용한다는 점까지 유사하다고 할 수도 있다. 하지만 본 실험과정에서 보인 바에 따라 기존의 거리를 통해 데이터 포인트 간의 관계를 보이는 것이 아닌 군집의 반지름에 해당하는 거리를 반지름으로 하는 원의 영역을 누적하였으며, 누적값에 대한 극대를 통해 클러스터 헤드를 추정한다는 차이점이 존재한다.

### III. 결론

본 논문에서는 2차원 공간에 분포된 데이터 포인트에 대해 해당 데이터 포인트를 중심으로 하는 원의 누적합의 시각화와 원의 크기와 군집의 크기에 따른 누적값의 변화, 누적값을 사용한 클러스터 헤드의 추정과 기존 방식들과의 차이에 대해 기술하였다. 군집면적과 근접한 면적의 원을 사용하였을 때 극대를 통해 클러스터 헤드를 결정하기 용이하였다. 면적이 상대적으로 작거나 넓은 경우 클러스터 헤드를 결정하기에는 부적절하거나 무관한 값이 도출되는 것을 보였다.

누적합의 극대값을 사용하는 방식에서 기존 클러스터링 방식들의 공통적인 특징과 상이한 특징이 있음을 보였다. 하지만 각 좌표만을 사용하는 기존의 방식과 달리 해당이 없는 좌표도 연산에 사용하므로, 기본적인 연산시간에 대한 이점은 없는 것으로 보인다. 때문에 시간의 제약을 받는 상황에서 적용하기에는 어려울 것으로 보인다.

픽셀단위 연산으로 실험을 진행함을 통해 이미지의 중복 다중 객체를 분리하는 등의 상황에서 컨볼루션과 병렬연산 등에 결합하였을 때 이점을 가질 수 있을 것으로 전망한다. 본 연구의 결과를 바탕으로 향후 연구에서는 유사한 다양한 데이터 마이닝과 이산환경에서의 적용 등을 탐색하고자 하며, 통신 네트워크에서의 통신 거리에 따른 클러스터링 최적화와 같은 분야에도 적용할 수 있을 것으로 기대한다.

### ACKNOWLEDGMENT

본 논문은 농촌진흥청 공동연구사업(과제번호 : RS-2023-00232847)의 지원에 의해 이루어진 것임

### 참 고 문 헌

[1] Hyesoo Shin et al., "Deep k-means node clustering based on graph neural networks", Journal of KIISE, vol. 50, no. 12, pp.1153-1162, Dec. 2023

[2] Seok-Hwan Choi et al., "An adversarial attack type classification method using linear discriminant analysis and K-means algorithm", Journal of The Korea Institute of Information Security & Cryptology, vol. 31, No. 6, pp.1215-1225, Dec. 2021