

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

## 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.

### Abstract

본 논문은 균일한 면적으로 분포된 군집에 대해 데이터 포인트를 중심으로 하는 원의 누적합의 시각화와 이를 분석한내용을 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 같은이산환경에서의 데이터 마이닝과 이미지 프로세싱에 활용할 수 있을 것으로 보인다. 또한 통신 네트워크 분야에도 적용할 수 있을 것으로 기대한다.

### Introduction

#### 1. Existing Clustering Algorithms

K-means : 특정 데이터 포인트에 대해 다른 데이터 포인트의 거리를 사용하는 방식을 사용

BDSCAN(밀도기반 클러스터링) : 특정 반경 내의 데이터 포인트의 수를 활용

→ 대부분의 클러스터링 알고리즘들은 데이터 포인트 간의 거리를 통해 클러스터 헤드를 추출함

#### 2. Research Context

1. 각 군집의 분포된 면적이 균일
2. 군집의 면적을 알고있음
3. 일부 군집들이 인접해 구분, 혹은 분할(Cluster Partitioning)해야하는 상황

### Methods and Results

#### 1. Generation of Random Cluster Coordinates

$$A_1, A_2, \dots, A_{200} \sim U(0, 2\pi)$$

$$D_1, D_2, \dots, D_{200} \sim \sqrt{U(0, \text{radius of cluster})}$$

$$X = D \cos A + \text{random } X \text{ coordinate}_{\text{Cluster's Center}}$$

$$Y = D \sin A + \text{random } Y \text{ coordinate}_{\text{Cluster's Center}}$$

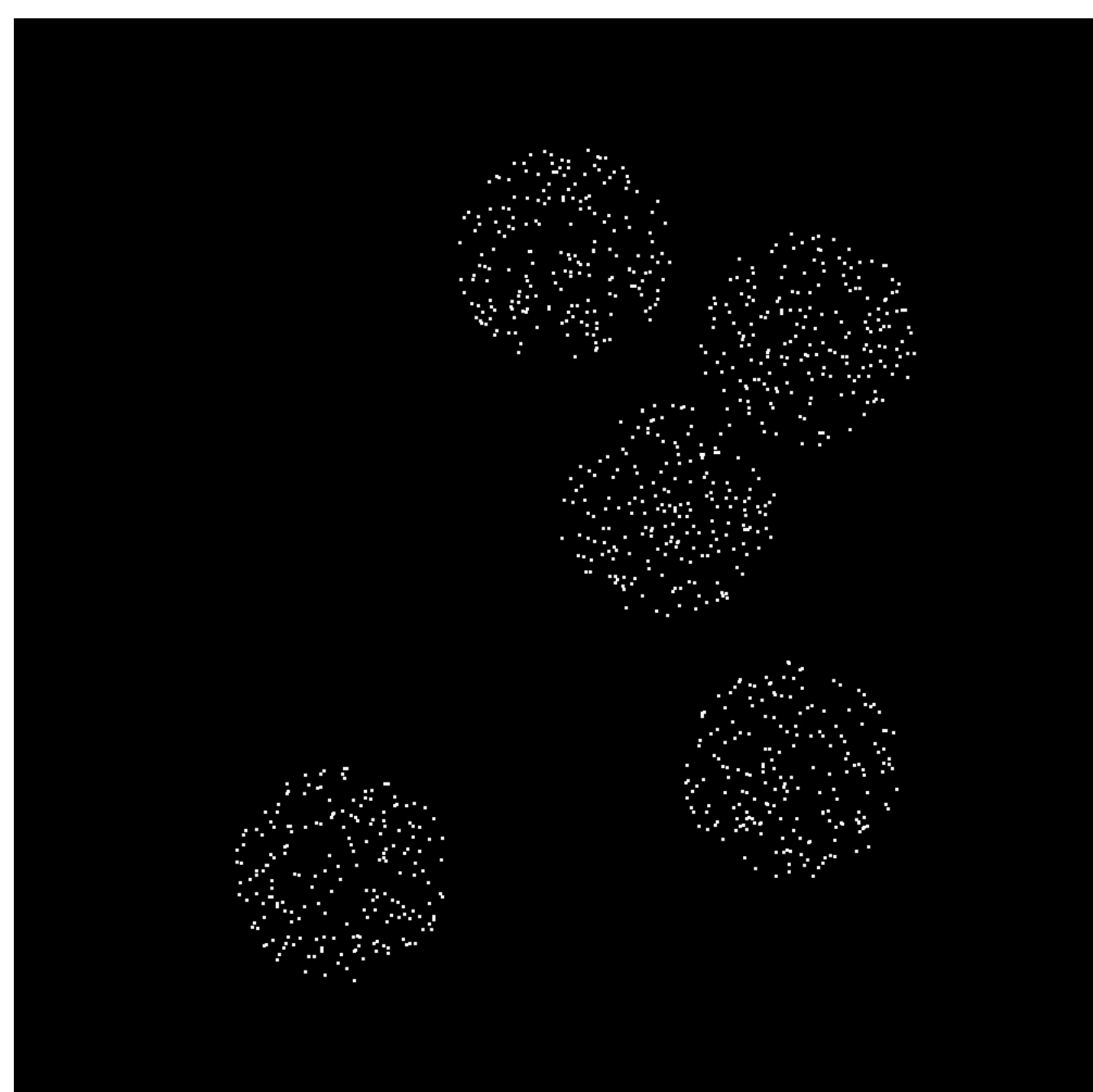


Fig.1. Random Data Distribution

#### 2. Cumulative Sum of a Circle

1. Radius of Cluster를 동일하게 반지름으로 사용하는 원을 각 데이터 포인트를 중심으로 생성
2. 모든 원에 대하여 해당되는 면적의 영역에 값을 1씩 누적
3. 반지름의 값이 Radius of Cluster의 절반, 두 배인 상황 분석

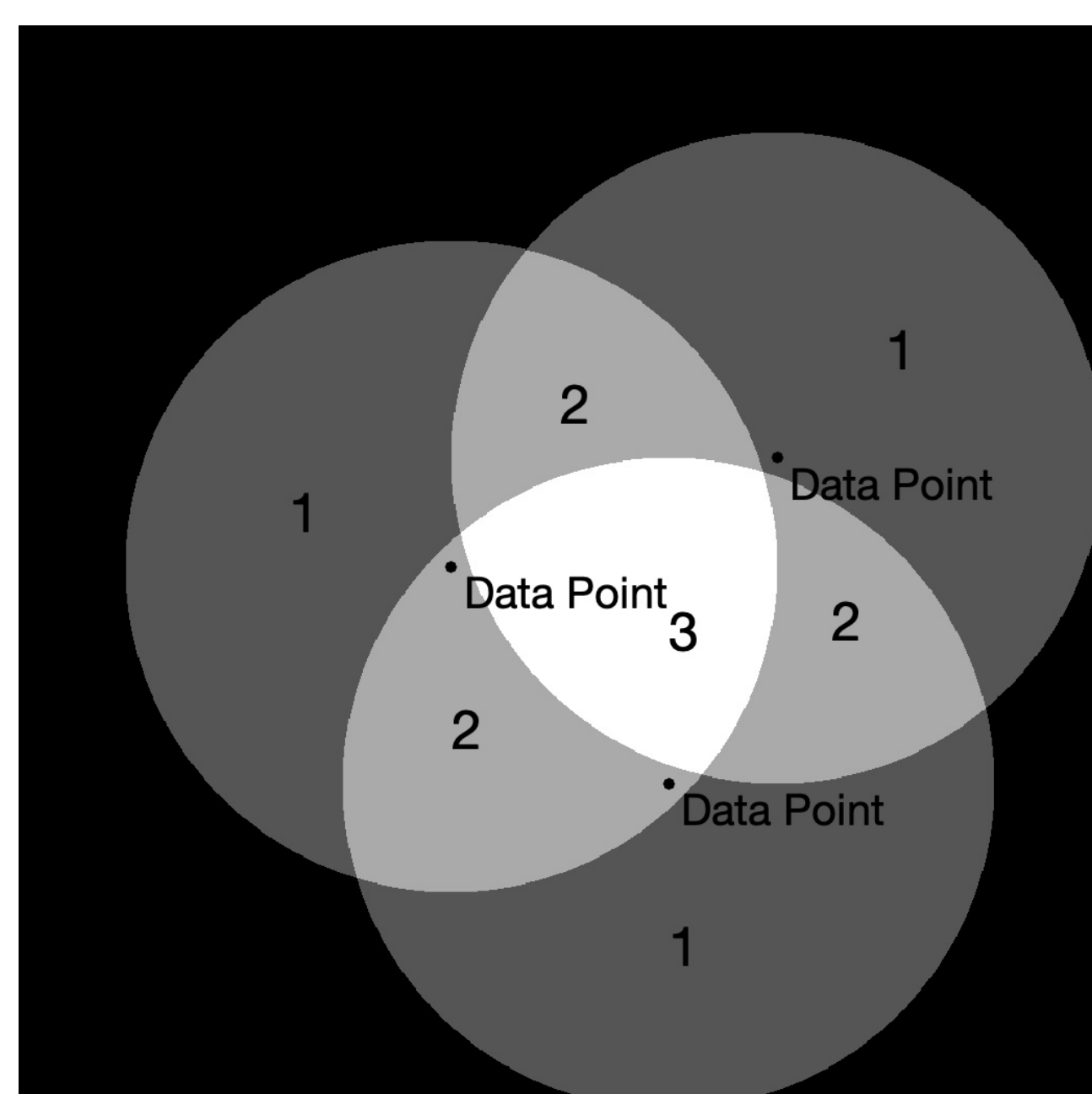


Fig.2. Cumulative Value of a Circle Example

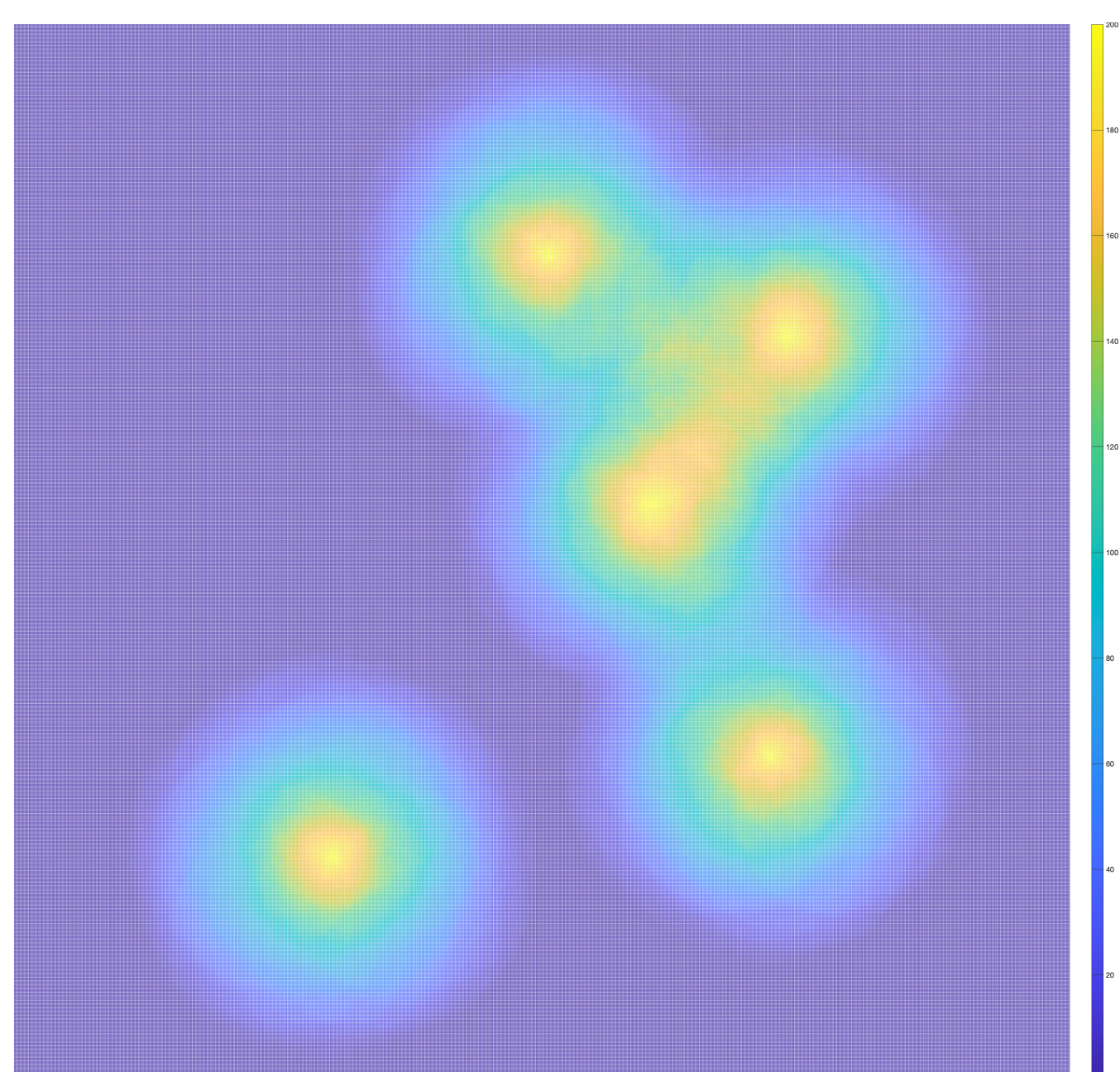


Fig.3. Visualization of Cumulative Sum Results of a Circle

1. 값이 0이 아닌 영역은 각 군집의 반지름의 두배에 해당하는 영역 내에 균일하게 분포함
2. 1의 반지름의 두배에 해당하는 영역은 원형을 띄게 됨

### Discussion

#### 1. Perspective Shift

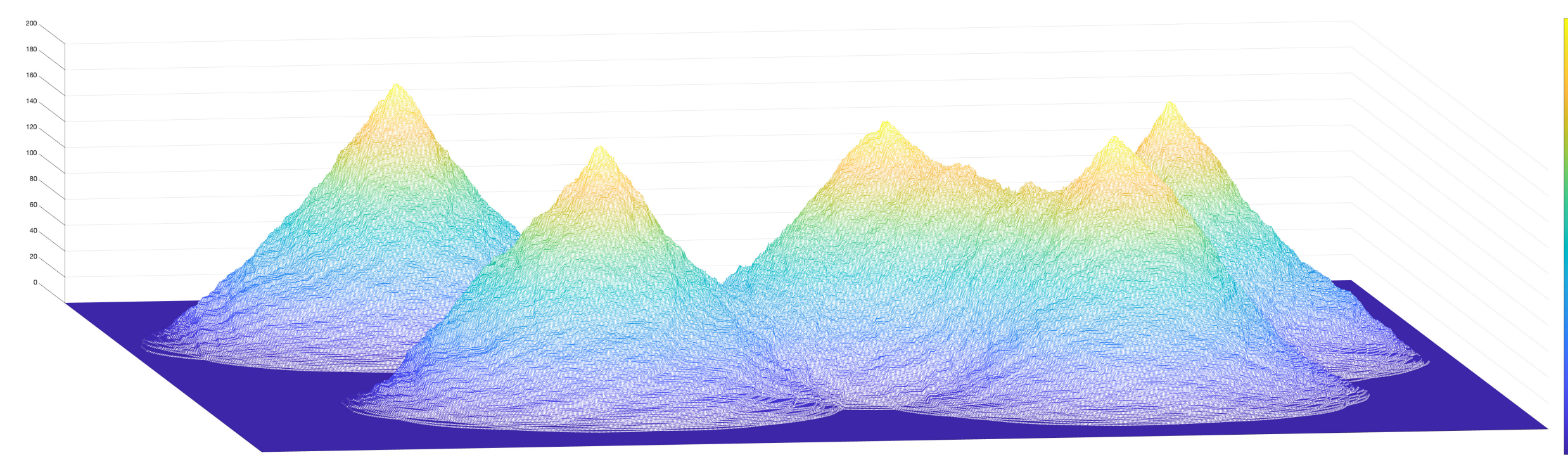


Fig.4. Cumulative Sum Results from a Right-Side Perspective

- 누적값들이 군집의 중심에서 외곽으로 향할수록 일정하게 감소함
- 군집이 서로 인접한 경우 누적영역이 겹치면서 일정하게 감소하지 않음
- 각 군집의 중앙을 연결한 선 인근의 값이 다른 거리가 비슷한 좌표에 비해 높음
- 군집의 중심에서는 군집의 데이터 포인트의 수와 일치하거나 근사한 수치를 가짐
- 누적값이 극대값인 경우의 좌표는 클러스터 헤드와 일치하거나 매우 인접함

#### 2. Circle Size Adjustment

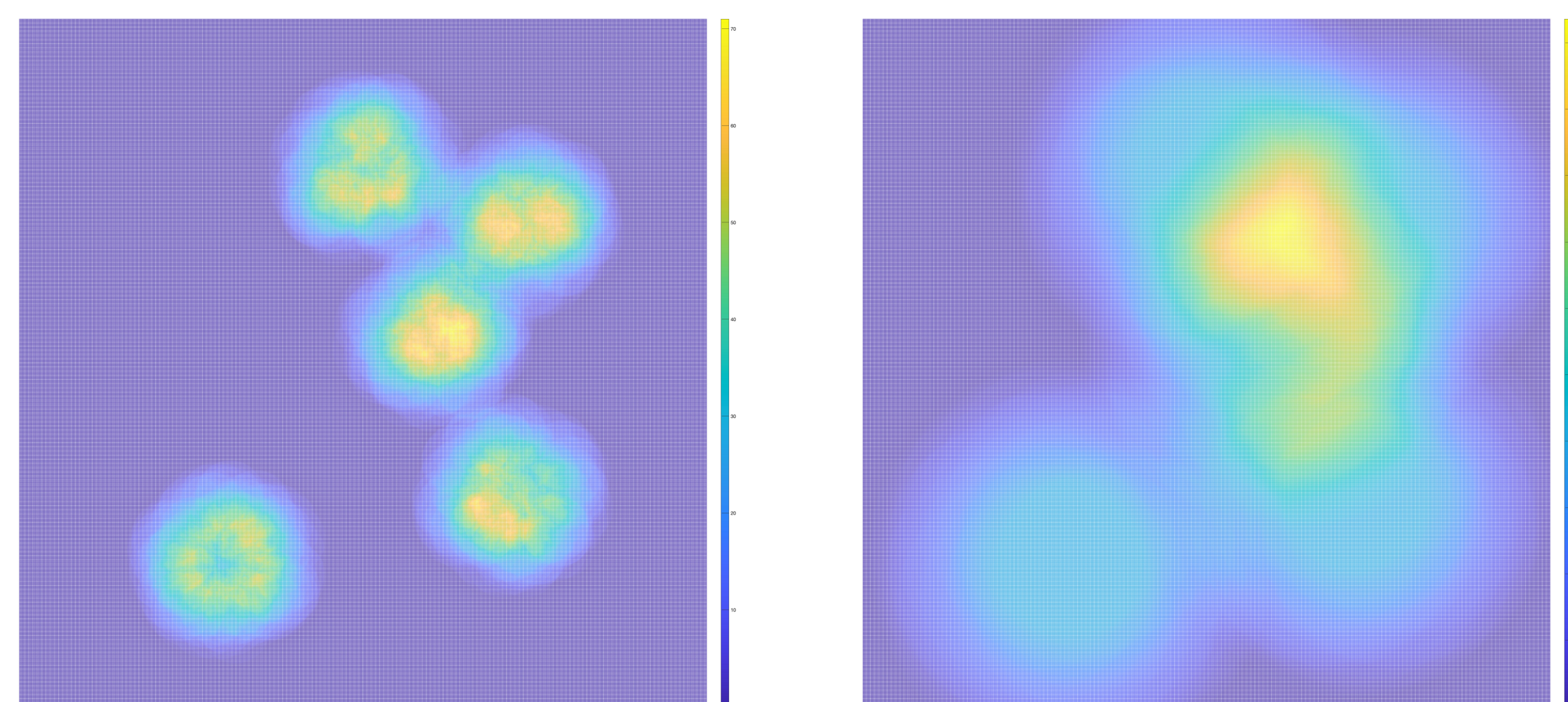


Fig.5. Cumulative Values for Cases Where the Area of a Circle is Smaller and Larger than the Cluster Area

1. 누적원의 면적이 군집의 면적보다 작은 경우
  - 누적값이 데이터 군집과 유사하게 분포
  - 중심으로부터 외곽으로 향할수록 지속적으로 감소하는 것이 아닌 어느정도 평탄함
  - 군집의 중앙보다 어느정도 거리가 있는 위치에서 더 큰 값을 갖는 케이스 발생
2. 누적원의 면적이 군집의 면적보다 큰 경우
  - 인접한 군집을 모두 포함하는 데이터들에 대한 중앙이 극대와 유사함
  - 인근의 값들에 대해 각 군집의 중심을 연결한 다각형에 대한 중선 다각형의 변에 집중 됨

#### 3. Comparison

- 기존의 클러스터링 방식과 달리 거리를 통해 데이터 포인트 간의 관계를 사용하지 않음
- DBSCAN과 비교하여 특정 영역 내에 포함된 데이터 포인트를 활용한다는 점은 유사함
- 누적값에 대한 극대를 통해 클러스터 헤드를 추정할 수 있음

### Conclusion

#### 1. Review

- 데이터 포인트를 중심으로 하는 원의 누적합의 시각화와 원의 크기에 따른 누적값의 변화, 클러스터 헤드의 추정과 기존 방식들과의 차이를 기술함

#### 2. Analysis

- 군집면적과 근접한 면적의 원을 사용하였을 때 극대를 통해 클러스터 헤드를 결정하기 용이함
- 면적이 상이한 경우 경우 클러스터 헤드를 결정하기에는 부적절하거나 무관한 값이 도출됨

#### 3. Limitations

- 데이터 포인트 외의 좌표를 사용, 연산시간에 대한 이점이 없어 보임, 시간 제한 상황에는 부적합

#### 4. Potential Applications

- Convolution과 Parallel Processing 등에 결합하였을 때 이점을 가질 수 있을 것으로 보임
- 다양한 데이터 마이닝과 이산환경, 이미지의 객체 구분 등에 응용할 수 있을 것으로 보임
- 무선 환경의 Centralized Communication에서 Range Limitation에 응용할 수 있을 것으로 보임

#### ACKNOWLEDGMENT

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

#### REFERENCES

- [1] Hyesoo Shin et al., "Deep k-means node clustering based on graph neural networks", Journal of KIIE, 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