

레이더 군집화를 위한 반복 K-means 클러스터링 알고리즘

박동현¹⁾ · 서동호²⁾ · 백지현²⁾ · 이원진²⁾ · 장동희^{*,1)}

¹⁾ 한국과학기술원 전기 및 전자공학부

²⁾ LIG넥스원(주) 전자전연구소

Repeated K-means Clustering Algorithm For Radar Sorting

Dong Hyun ParK¹⁾ · Dong-ho Seo²⁾ · Jee-hyeon Baek²⁾ · Won-jin Lee²⁾ · Dong Eui Chang^{*,1)}

¹⁾ School of Electrical Engineering, Korea Advanced Institute of Science and Technology, Korea

²⁾ Electronic Warfare R&D, LIG NEXI Co., Ltd., Korea

(Received 9 May 2023 / Revised 23 October 2023 / Accepted 22 November 2023)

Abstract

In modern electronic warfare, a number of radar emitters are in operation, causing radar receivers to receive high-density signal pulses that occur simultaneously. To analyze the radar signals more accurately and identify enemies, the sorting process of high-density radar signals is very important before analysis. Recently, machine learning algorithms, specifically K-means clustering, are the subject of research aimed at improving the accuracy of radar signal sorting. One of the challenges faced by these studies is that the clustering results can vary depending on how the initial points are selected and how many clusters number are set. This paper introduces a repeated K-means clustering algorithm that aims to accurately cluster all data by identifying and addressing false clusters in the radar sorting problem. To verify the performance of the proposed algorithm, experiments are conducted by applying it to simulated signals that are generated by a signal generator.

Key Words : Radar Clustering(레이더 군집화), Unsupervised Learning(비지도 학습), Electronic Warfare(전자전), K-means Clustering(K-평균 군집화)

1. 서론

전자전(electronic warfare)이란 각종 전자기기, 전자파, 전자통신 및 네트워크를 이용한 현대 전투 형태 중 하나이며 세부적으로 전자전 지원(electronic support),

전자공격(electronic attack) 그리고 전자보호(electronic protection)로 구분할 수 있다. 이때 전자전 지원 과정 중 레이더 위협 식별은 전투 중 복잡한 레이더 신호 환경에서 수신한 신호 펄스열을 분석하고 피아식별하는 과정이다. 수신기는 레이더 이미터에서 발생한 신호를 수신할 때, 그 신호의 주파수 특징 및 PRI(Pulse Repetition Interval) 특징 등을 분석하고 그 결과를 아군의 위협 식별 라이브러리와 매칭하여 위협을 식별

* Corresponding author, E-mail: dechang@kaist.ac.kr
Copyright © The Korea Institute of Military Science and Technology

한다^[1,2]. 실제 전자전 환경에서는 다중 위협이 존재하기 때문에 아군의 수신기에도 임의의 시간에서 다중 위협 신호를 수신할 수 있다. 고밀도의 레이더 펄스열은 그 분석이 쉽지 않고 정확도가 떨어지기 때문에 Fig. 1과 같이 펄스열 분석 단계 전 데이터 군집화 과정은 레이더 위협 식별 단계에서 매우 중요하다^[3].

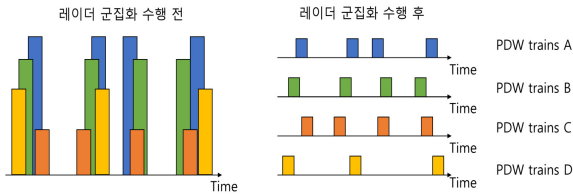


Fig. 1. Radar sorting

신호 수신기는 레이더를 수신할 때 그 레이더 데이터의 AOA(Angle Of Arrival), RF(Radio Frequency), PW(Pulse Width) 등의 특징정보를 PDW(Pulse Description Word)라는 벡터 형태로 전처리한다. 과거에는 이 PDW 정보를 가지고 엔지니어가 설정한 규칙 기반으로 펄스열 분석 단계 전 군집화 과정을 수행하였다. 규칙 기반의 군집화 과정으로는 순차 클러스터링 기법, 동시 클러스터링 기법, 연속 스캔 기법 등의 연구가 진행되어왔다^[4,7]. 순차 클러스터링 기법은 PDW의 각 특징들에 대해서 한 번에 한 특징씩 순서대로 고려하여 히스토그램 기반으로 군집화를 수행하는 방식이다. 개념이 쉽고 구현이 간단하지만 각 특징들의 상관성을 고려하지 못하는 단점이 존재한다. 이에 따라 각 특징들의 상관성을 반영하여 클러스터링을 수행하기 위해 동시 클러스터링 기법, 연속 스캔 기법 등과 같은 방법들이 연구되었지만 다차원 히스토그램이나 다차원 분리셀 등을 연산하기 위해 많은 메모리공간이 필요하다는 단점이 있다^[5,6]. 이러한 규칙 기반 군집화 기법은 적은 종류의 이미터들은 효율적으로 잘 군집화 하지만 전자전에서 사용되는 이미터 종류가 증가할수록 설계 복잡도가 증가하며 동시에 알고리즘의 효율성이 크게 떨어진다는 문제점이 존재한다^[8].

점점 더 늘어나는 현대 전자전에서의 레이더 이미터 수를 대비하기 위해 최근에는 머신러닝 기법을 활용하여 레이더 군집화 문제를 해결하는 적용 연구가 활발히 진행되고 있다. 군집화 문제를 해결하는 대표적인 머신러닝 알고리즘으로는 K-means 클러스터링 기법이 있다^[9]. K-means 클러스터링 기법은 간단하고

연산 수행시간이 빠르다는 장점은 있지만, 초기 중심점 위치를 어떻게 설정하느냐에 따라 클러스터링 수렴 속도가 달라지고 때에 따라 과소 및 과대 클러스터링이 일어날 수 있다는 문제점을 가지고 있다. 또한 사전에 군집의 개수 K를 알고 있어야 정확하게 군집화된다는 단점을 가지고 있다. 레이더 군집화 문제에서 K-means 클러스터링을 적용할 때 초기 중심점 문제와 군집 개수 추정 문제를 해결하기 위해 다양한 데이터 과학 기법을 융합하는 연구가 활발히 진행 중이다^[8,10,11]. G. Qiang, et al.^[8]는 SVC(Support Vector Clustering)를 통해 데이터 일부를 가지고 대략적인 군집의 초기 중심점과 개수를 추정한 후 그 정보를 가지고 K-means 클러스터링을 수행하여 레이더 군집화를 수행하였다. M. Li, et al.^[10]과 X. Feng, et al.^[11]는 데이터 장(data field) 이론을 이용하여 초기 중심점과 개수를 추정한 후 K-means 클러스터링을 수행하여 레이더 군집화를 수행하였다. 이 밖에도 K-means++^[12], 팔꿈치 방법^[13], 실루엣 방법^[14] 등 초기 중심점 문제와 군집 개수 추정 문제를 해결하기 위해 다양한 연구들이 진행되고 있다. 이러한 연구들은 레이더 군집화 문제의 군집화 성능을 높이긴 하였지만, 여전히 모든 레이더 실험 데이터를 정확하게 군집화하는 결과는 보이지 못했다. 또한 데이터 과학 기법을 이용하였음에도 초기 중심점과 군집 개수를 잘못 추정할 경우에 대한 논의가 되고 있지 않다. 실제 전자전 환경을 고려한다면, K-means 클러스터링을 레이더 군집화에 사용할 때 적절하지 못한 초기 중심점이나, 오차가 있는 군집 개수에도 강건하게 동작하는 방법이 필요하다.

모든 레이더 데이터를 정확하게 군집화하면서 K-means 클러스터링 알고리즘의 문제점을 해결하기 위해 저자는 본 논문에서 반복 K-means 클러스터링 알고리즘을 제안한다. 본 논문의 구성은 다음과 같다. 2장에서는 K-means 클러스터링 결과 중 오답 클러스터를 검사하는 알고리즘을 소개하고 그것을 기반으로 수행되는 반복 K-means 클러스터링 알고리즘을 제안한다. 3장에서는 실험을 통해 2장에서 제안한 알고리즘이 실제로 잘 동작하는지 검증한다. 마지막 4장에서는 결론과 앞으로 진행할 연구에 대해 제시한다.

2. 알고리즘

2장에서는 레이더 군집화 문제에서 K-means 클러스

터링 알고리즘을 적용할 때 발생하는 두 종류의 오답 클러스터를 레이더 PDW 정보로 식별하는 방법을 제시한다. 또한 그렇게 찾아낸 오답 클러스터에 대해서 반복적으로 K-means 클러스터링을 수행하는 반복 K-means 클러스터링 알고리즘을 제안한다.

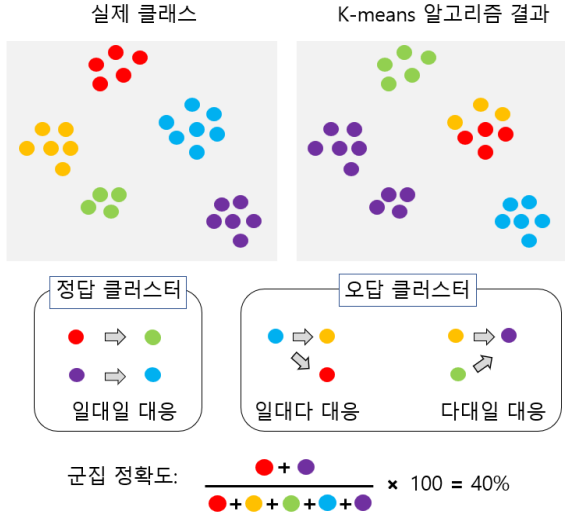


Fig. 2. K-means clustering result with K=5

Fig. 2는 5개의 클러스터를 가지고 있는 데이터에 대해서 K-means 클러스터링 기법을 적용했을 때 발생할 수 있는 결과의 예이다. 다섯 개의 클러스터 중 두 클러스터는 실제 클러스터와 일대일 대응되면서 정확한 군집을 이루는 것을 확인할 수 있다. 하지만 나머지 군집은 그 대응 관계가 각각 과소, 과대 클러스터링을 의미하는 일대다 대응, 다대일 대응 관계를 가지면서 정확하지 않은 군집을 이룬 모습을 확인할 수 있다. 마찬가지로 레이더 군집화 문제에서 K-means 클러스터링 알고리즘을 적용할 때도 이와 같은 오답 클러스터가 존재할 수 있다. 이때 그 결과 중 정답 클러스터와 오답 클러스터를 정확하게 구분할 수 있다면 오답 클러스터만 따로 K-means 클러스터링 기법을 다시 수행할 수 있고 이 과정을 반복하면 모든 클러스터를 정확하게 클러스터링할 수 있다.

2.1 오답 클러스터 검사알고리즘

2.1.1 다대일 대응 검사

레이더 PDW에는 AOA, RF, PW 등 신호의 다양한 정보들이 기록되어있다. 여기서 AOA는 레이더 신호

가 들어온 방향을 의미한다. 고정된 이미터에서 송신한 레이더 신호는 거의 일정한 AOA 가지기 때문에 레이더를 구분 짓는 군집화 문제에서 중요한 정보로 사용되고 있다. 이점을 이용하여 t 시점의 DAOA를 식 (1)과 같이 정의하고 군집화가 잘되었는지 아닌지에 대한 평가 도구로 사용할 수 있다.

$$DAOA_t = |AOA_t - AOA_{t-1}| \quad (1)$$

만약 K-means 클러스터링 알고리즘 수행 후 생성된 결과 클러스터 중 한 클러스터 데이터들의 DAOA가 거의 0 값을 일정하게 유지하지 않고 어떤 큰 값이 발생한다면 두 개 이상의 위치를 가진 레이더가 한 클러스터로 잘못 클러스터링 되었다는 것을 의미한다. Algorithm 1은 DAOA를 활용한 레이더 클러스터링 다대일 오답 클러스터 검사알고리즘이다. Algorithm 1의 4번째 줄과 같이 K-means 클러스터링을 수행한 결과의 각 클러스터에서 DAOA의 최댓값을 구한다. 이 값이 최댓값 임계치(ϵ_{n-1})보다 큰 경우에는 한 클러스터에 두 개 이상의 클러스터가 존재한다고 간주하고 해당 클러스터를 오답 클러스터로 지정한다.

Algorithm 1 n-to-1-check

Input: Radar data labeled K-means clustering result

- 1: **Set** *False_cluster*
- 2: **for** *i* **in** $1 \leq i \leq K$ **do**
- 3: *N* is the number of data of i_{th} cluster
- 4: $max_DAOA = \max_{t=2 \sim N} (|AOA(i, t) - AOA(i, t-1)|)$
- 5: **if** $max_DAOA > \epsilon_{n-1}$ **then**
- 6: *False_cluster.insert(i)*
- 7: **end if**
- 8: **end for**
- 9: **return** *False_cluster*

2.1.2 일대다 대응 검사

K-means 클러스터링을 수행한 결과 생성된 각 클러스터를 확인했을 때 어떤 두 군집이 매우 인접해 있다면 그 두 군집은 일대다 오답 클러스터로 간주한다. Algorithm 2는 레이더 클러스터링 일대다 오답 클러스터 검사알고리즘이다. 두 클러스터가 인접해 있는지를 확인하기 위해 $MinDist(i, j)$ 함수를 정의하고 이를 사용하여 일대다 대응 검사를 수행한다. $MinDist(i, j)$ 는 K-means 클러스터링을 수행한 결과 생성된 클러스터의 i 번째 클러스터와 j 번째 클러스터의 데이터들 사이에서 가장 짧은 거리를 의미한다. $MinDist(i, j)$

값이 임계치 (ϵ_{1-n})보다 작다면 i 번째 클러스터와 j 번째 클러스터는 하나의 클러스터로 간주하고 그 두 클러스터는 오답 클러스터로 지정한다.

Algorithm 2 1-to-n.check

Input: Radar data labeled K -means clustering result

```

1: Set False_cluster
2: for  $i$  in  $1 \leq i \leq K$  do
3:   for  $j$  in  $1 \leq j \leq K$  do
4:     if  $i \geq j$  then
5:       continue
6:     else
7:       if  $MinDist(i, j) < \epsilon_{1-n}$  then
8:         False_cluster.insert(i, j)
9:       end if
10:    end if
11:  end for
12: end for
13: return False_cluster.unique()

```

2.2 반복 K-means 클러스터링 알고리즘

Algorithm 3은 본 논문에서 제안하는 반복 K-means 클러스터링 알고리즘에 대한 설명을 나타낸다. 2.1장의 오답 클러스터 검사알고리즘들로 K-means 클러스터링을 수행한 결과 중 정답 클러스터와 오답 클러스터를 구분할 수 있다. 찾아낸 오답 클러스터들에 대하여 반복적으로 K-means 클러스터링을 수행한다. 오답 클러스터에 대해서 다시 K-means 클러스터링을 수행할 때는 새로운 임의의 초기 중심점에서 수행하기 때문에 이전에서 찾지 못했던 정답 클러스터를 찾아낼 수 있고 이 과정을 통해 오답 클러스터를 줄여 나갈 수 있다. 이 과정을 오답 클러스터가 없어질 때까지 반복한다면 전체 데이터셋에 대하여 정확한 클러스터링 결과를 확인할 수 있다. 한편 반복 K-means 클러스터링 알고리즘에 설정한 K 값과 실제 군집 개수 사이에 오차가 존재한다면 정답 클러스터를 대부분 걸러낸 후에 데이터가 남아 있음에도 불구하고 정답 클러스터가 발생하지 않는 순간이 온다. 이때 남은 데이터들은 규칙 기반의 고전 군집화 방법이 적용된다. Algorithm 3의 20번째 줄 *rule_based()* 함수는 규칙 기반 군집화를 의미한다. 반복 K-means 클러스터링으로 많은 수의 정답 군집들이 이미 형성된 후이기 때문에 남은 데이터의 군집화해야 할 군집의 개수가 많지 않다. 따라서 남은 데이터들은 규칙 기반 군집화 방법만으로도 충분히 정확한 군집이 이루어진다.

Algorithm 3 Repeated K-means clustering

Input:

X : Radar data
 K : input K value
patience : patience number
iter : iteration number for K -means clustering

```

1:  $iter = 0$ 
2:  $flag = 0$ 
3: while  $iter = 0$  or  $False\_cluster \neq \emptyset$  do
4:    $X_{labeled} = K\_means(X, K, iter)$ 
5:    $False\_cluster\_1 = n\text{-to-}1\text{-check}(X_{labeled})$ 
6:    $False\_cluster\_2 = 1\text{-to-}n\text{-check}(X_{labeled})$ 
7:    $False\_cluster = False\_cluster\_1 \cup False\_cluster\_2$ 
8:    $True\_cluster = False\_cluster^c$ 
9:   if  $len(True\_cluster) == 0$  then
10:     $flag = flag + 1$ 
11:    if  $flag == patience$  then
12:      break
13:    end if
14:  else
15:     $flag = 0$ 
16:  end if
17:   $K = K - len(True\_cluster)$ 
18:   $iter = iter + 1$ 
19: end while
20: rule_based(False_cluster)

```

3. 실험 결과

3.1 데이터셋

본 논문에서 제안하는 반복 레이더 K-means 클러스터링 알고리즘의 성능을 확인하기 위해 92종의 서로 다른 레이더 데이터를 신호 모의 발생기를 이용하여 생성하였다. 각 클러스터별로 약 900에서 1,200개 사이의 PDW 셋을 생성하여 총 99,421개의 데이터를 실험에 사용하였다. 모의 데이터는 실제 현대 전자전에서 발생할 수 있는 주파수 패턴(agile, fixed, hopping, positive, negative, sine, triangular), PRI 패턴(stable, dwell & switch, stagger, negative, positive, triangular wobble), scan(steady, circular) 패턴을 섞어 제작하였다. 생성된 PDW는 AOA, RF, PW를 포함하여 총 14가지의 레이더 특징 벡터로 이루어져 있다.

다루고자 하는 데이터의 차원이 크면 클수록 차원의 저주 때문에 연산시간이 늘어나고 복잡도가 증가하여 정확도가 떨어질 수 있다. 뿐만 아니라 기존의 많은 규칙 기반의 레이더 군집화 과정에서 레이더의 PDW 정보 중 AOA, RF, PW를 사용하여 군집화 수행하였기 때문에^[4,7] 본 논문의 실험에서는 그 세 가지 특징만 사용하였다. 실험에 사용한 데이터의 클러스터

별 AOA, RF, PW의 분포는 Fig. 3과 같다. PDW의 AOA와 RF의 경우는 식 (2)와 같이 min-max 정규화를 수행하고 PW의 경우 데이터 분포를 고려하여 식 (3)과 같이 log-min-max 정규화를 수행하여 0에서 1 사이의 값으로 재구성하였다.

$$x_{normalized} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (2)$$

$$x_{log_normalized} = \frac{\log(x) - \log(x_{min})}{\log(x_{max}) - \log(x_{min})} \quad (3)$$

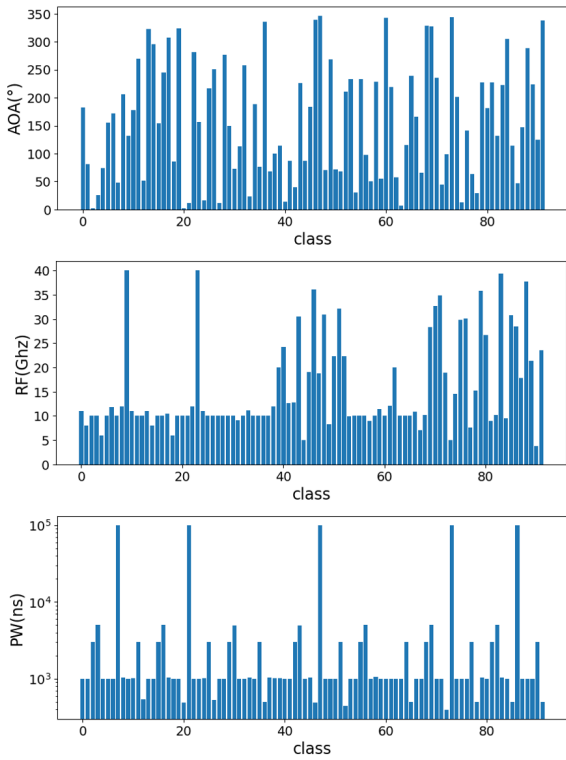


Fig. 3. PDW data distribution about feature

3.2 평가지표

클러스터링 성능을 비교하기 위한 평가지표는 군집 정확도로 한다. Fig. 2와 같이 클러스터링 수행 결과를 실제 군집 클러스터와 비교하여 완벽하게 1:1 대응 되는 클러스터를 정답 클러스터로, 그렇지 않은 클러스터는 오답 클러스터로 정하고 전체 클러스터 대비 정답 클러스터 비율을 군집 정확도로 정의한다. 예

를 들어 Fig. 2의 경우는 전체 5개의 클러스터 중 정답 클러스터가 2개이므로 군집 정확도는 40 %로 계산된다.

3.3 규칙 기반 클러스터링 결과

3.3장에서 본 논문의 실험 데이터에 서론에서 소개한 규칙 기반 클러스터링 중 순차 히스토그램 기법을 적용해본다. 순차 히스토그램에 필요한 규칙을 정하고 클러스터링을 수행하였다. 이때 군집화해야 할 클러스터 수에 따라 규칙 기반 클러스터링의 성능이 얼마나 차이 나는지 확인하기 위해 같은 규칙으로 데이터의 클러스터 개수를 반씩 줄여나가며 반복 실험하였다. Table 1은 규칙 기반 클러스터링 수행 결과이다. 전체 데이터셋에서 12개의 클러스터만 가지고 순차 히스토그램 기법을 적용했을 때는 100 % 정확하게 군집화를 하였다. 하지만 군집화 해야 할 클러스터의 개수가 늘어나면 늘어날수록 순차 히스토그램 기법의 군집화 성능이 떨어지는 것을 확인할 수 있다.

Table 1. The result of rule-based clustering

군집 개수	군집 정확도(%)
92	51.08
46	60.87
23	91.30
12	100

3.4 K-means 클러스터링 결과

다음으로 레이더 군집화 문제에서 K-means 클러스터링 알고리즘이 잘 적용되는지 실험을 수행해본다. 본 논문에서 수행한 K-means 클러스터링 알고리즘의 초기 중심점 포인트는 임의로 지정하기 때문에 그 결과값이 수행할 때마다 다르다. 따라서 동일한 실험을 5번 수행하여 그 결과를 분석한다. 또한 실제 무기체계에서 부정확한 K 추정값을 고려하기 위해 K값 설정을 실제 군집 개수에 ±5, ±10으로 설정하여 실험을 수행해본다. Table 2는 전체 실험 데이터에 대해서 본래의 K-means 클러스터링 적용하여 레이더 군집화를 수행한 실험 결과를 나타낸다. 본래의 K-means 클러스터링을 수행한 결과 초기 중심점 문제로 인해 모든 실험에서 좋지 않은 결과를 확인할 수 있다.

Table 2. The result of K-means clustering

회차	군집 정확도(%)				
	K=82	K=87	K=92	K=97	K=102
1	48.91	46.73	45.65	42.39	39.13
2	41.30	50.00	40.21	44.56	48.91
3	46.73	53.26	50.00	48.91	48.91
4	39.13	46.73	52.17	45.65	44.56
5	45.65	38.04	48.91	47.82	46.73
평균	44.34	46.95	47.39	45.86	45.65

3.5 반복 K-means 클러스터링 실험 결과

3.5장에서는 2.1장에서 제안한 일대다 검사알고리즘과 다대일 검사알고리즘이 잘 적용되어 클러스터링 결과 중 정답 클러스터와 오답 클러스터가 잘 구분되었는지를 실험한다. 또한 2.2장에서 제안한 반복 클러스터링 알고리즘으로 모든 데이터에 대해 정확하게 클러스터링이 되었는지 확인한다.

Fig. 4는 3.4장의 실험 수행 중에 발생한 오답 클러스터의 결과에서 다대일 오답 클러스터의 예시와 일대다 오답 클러스터의 예시를 각각 시각화한 것이다. 다대일 검사알고리즘 수행 결과 모든 다대일 오답 클러스터들에 대해서 DAOA의 이상을 확인할 수 있었다. Fig. 5의 왼쪽 그림은 Fig. 4 (a)의 다대일 오답 클러스터의 예시에 대한 DAOA의 그래프이다. 두 종류의 레이더가 한 클러스터에 묶여있기 때문에 DAOA가 시간에 따라 진동하는 그래프를 확인할 수 있다. 이 DAOA 값들의 최댓값이 설정한 임계값 (ϵ_{n-1})을 넘기

때문에 오답 클러스터로 잘 구분되었다. 반면 Fig. 5의 오른쪽 그림은 정상 클러스터의 DAOA 그래프이다. 고정된 하나의 레이더로부터 들어온 신호만 클러스터에 존재하기 때문에 DAOA 값이 거의 0에 근접하고 설정한 임계값을 넘지 않는 것을 확인할 수 있다.

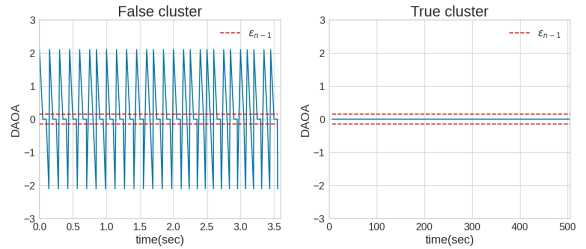
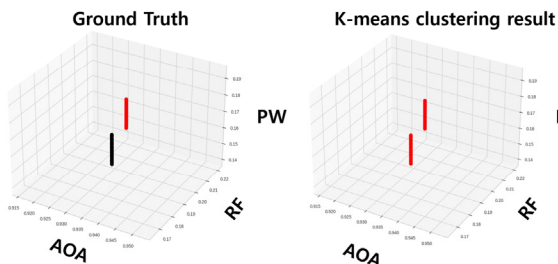


Fig. 5. DAOA graph of false cluster(left) and true cluster(right)

또한 Fig. 4 (b)의 일대다 오답 클러스터의 그래프를 확인해보면 서로 인접해 있는 일대다 오답 클러스터들은 실제로 하나의 클러스터였음을 확인할 수 있다. 종합적으로 3.4장의 실험 중에 발생한 결과 클러스터들에 대해서 제안한 오답 클러스터 검사알고리즘을 수행해본 결과 모두 정확하게 정답 클러스터와 오답 클러스터를 구분하는 것을 확인할 수 있었다.

검증된 오답 클러스터 검사알고리즘을 사용하여 실험 데이터에 대하여 본 논문에서 제안한 반복 K-means 클러스터링을 수행한다. 3.4장의 실험과 같은 이유로 K=92와 그 값을 기준으로 ± 5 , ± 10 한 경우를 각각 5번씩 수행하였다. Table 3은 반복 K-means 클러스터링

(a) False cluster result : n to 1



(b) False cluster result : 1 to n

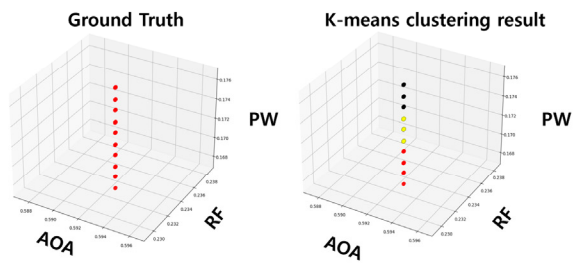


Fig. 4. False cluster result example, The left side of each picture represents the actual clusters, and the right side represents the incorrectly clustered results. In the 'n to 1'(a) false cluster example, two actual clusters were incorrectly clustered into one cluster. In the '1 to n'(b) false cluster example, one actual cluster was incorrectly clustered into three clusters

알고리즘 수행 과정의 3가지 예를 나타낸 것이다. 가운데 결과는 K=92로 설정하여 정확한 군집 개수를 알고 있는 경우의 실험 결과 중 하나이다. K-means 클러스터링 수행 결과에서 오답 클러스터 검사알고리즘으로 정확하게 정답 클러스터와 오답 클러스터를 구분 지었고 오답 클러스터에 대해 같은 과정을 반복하면서 최종적으로 92개의 클러스터를 모두 정확하게 군집화하는 데 성공하였다. Table 3의 K=87과 K=97의 결과는 오차를 고려한 K값의 실험 결과 중 하나이다. K=97로 잘못된 군집 개수를 설정한 경우, 반복적인 K-means 클러스터링 수행으로 대부분의 정답 클러스터를 군집화하는 데 성공하였지만 5번째 반복 이후부터는 K-means 클러스터링이 더 이상 정상 동작하지 못하는 것을 확인할 수 있다. 하지만 남은 데이터들에 대하여 클러스터링해야 할 군집의 개수는 3개로 고전적인 규칙 기반의 군집화 방법으로도 충분히 군집화 수행이 될 만큼 대폭 줄어들었다. 실제로 3.3장의 같은 규칙을 가지고 남은 3개의 클러스터에 대하여 군집화를 수행하였을 때 100 % 군집화 정확도를 확인하였다. 즉 K=97로 알고리즘을 수행했음에도 불구하고 최종적으로 92종의 클러스터를 모두 정확하게 군집화하는 데 성공하였다. 마지막으로 K=87로 설정한 경우, K=97의 실험 결과와 같이 K-means 클러스터링이 어느 정도 반복이 진행되면 K-means 클러스터링이 더 이상 정상 동작 하지 못하는 순간이 온다. 이때도 마찬가지로 남은 데이터에 대해서 규칙 기반 군집화 방법을 수행할 수 있다. 다만 알고리즘 특성상 실제 군집 개수보다 적은 개수의 K가 설정이 되어있다면

반복 후 남은 데이터의 군집화해야 할 군집의 개수가 상대적으로 많이 남게 되어 규칙 기반 군집화 기법이 쉽게 군집화를 수행하지 못할 수 있다.

Table 4는 전체 실험에 관한 결과이다. 설정한 K 값이 실제 군집 개수보다 같거나 클 때는 각각 5번의 실험 결과 모두 92종의 클러스터를 완벽하게 군집화하는 데 성공하였다. 설정한 K가 실제 군집 개수보다 적을 때는 반복 K-means 클러스터링 결과 남은 오답 클러스터가 비교적 많기 때문에 간혹 규칙 기반 클러스터링 단계에서 몇몇 클러스터를 클러스터링 하는 데 실패하는 경우가 있었다. 하지만 이 역시도 전체 데이터에서 규칙 기반 군집화 기법이나 본래의 K-means 클러스터링 기법을 적용한 결과보다 좋은 성능이고 마지막의 규칙 기반 군집화 방법을 더 좋은 규칙 기반 군집화 방법으로 변경하거나 더 최적화된 규칙으로 개선한다면 성능 향상을 기대할 수 있다.

Table 4. The result of repeated K-means clustering

회차	군집 정확도(%)				
	k=82	k=87	k=92	k=97	k=102
1	100	100	100	100	100
2	98.91	97.82	100	100	100
3	100	100	100	100	100
4	97.82	98.91	100	100	100
5	98.91	100	100	100	100
평균	99.13	99.35	100	100	100

Table 3. The example of K-means clustering process

K = 87					K = 92					K = 97				
반복 횟수	K	(T)	K-(T)	(F)	반복 횟수	K	(T)	K-(T)	(F)	반복 횟수	K	(T)	K-(T)	(F)
1	87	32	55	60	1	92	42	50	50	1	97	43	54	49
2	55	22	33	38	2	50	22	28	28	2	54	23	31	26
3	33	22	11	16	3	28	12	16	16	3	31	19	12	7
4	11	5	6	11	4	16	7	9	9	4	12	3	9	4
5	6	2	4	9	5	9	6	3	3	5	9	1	8	3
6	4	0	4	9	6	3	3	0	0	6	8	0	8	3
7	4	0	4	9	7					7	8	0	8	3
8	4	0	4	9	8					8	8	0	8	3

K: number of clusters, (T): number of true clusters among clustering results (F) number of actual false clusters

4. 결론

본 논문에서는 레이더 군집화 문제에서 K-means 클러스터링 알고리즘을 적용할 때 그 결과 중 정답 클러스터와 오답 클러스터를 구분하는 검사 방법을 제시하고 그 방법을 통한 반복 K-means 클러스터링 알고리즘을 제안한다. 다양한 실험 조건에서 수행해본 결과 고전적인 규칙 기반의 클러스터링 기법과 본래의 K-means 클러스터링보다 높은 군집 정확도를 확인하였다. 본 논문의 연구에서 더 발전하여, 향후 전시 상황 노이즈나 신호 간의 중첩 등의 상황이 첨가된 데이터에서 어떻게 군집화 문제를 해결할지에 관한 연구도 이루어질 것이다.

5. 사 사

이 연구는 LIG NEX1 산학협력과제(G01220631) 지원으로 연구되었음.

References

- [1] Y. Kim, K. Yu and J. Han, "Research for radar signal classification model using deep learning technique," Journal of the Korea Institute of Military Science and Technology, Vol. 22, No. 2, pp. 170-178, 2019.
- [2] H. E. Hassan, "A new algorithm for radar emitter recognition," Proceedings of the 3rd International Symposium on Image and Signal Processing and Analysis, pp. 1097-1102, 2003.
- [3] M. G. S. Ahmed and B. Tang, "Sorting radar signal from symmetry clustering perspective," Journal of Systems Engineering and Electronics, Vol. 28, No. 4, pp. 690-696, 2017.
- [4] J. W. Han, S. K. Ha and L. D. Woen, "An adaptive clustering scheme for ES," Conference on Information and Control Systems, pp. 366-368, 2006.
- [5] J.-T. Lee, Y.-K. Ju, G.-T. Kim, and J.-N. Jeon, "A Clustering Technique of Radar Signals using 4-Dimensional Features," Journal of the Institute of Electronics and Information Engineers, Vol. 51, No. 10, pp. 137-144, 2014.
- [6] D.-W. Lee, J.-W. Han, and W.-D. Lee, "A Kernel Density Signal Grouping Based on Radar Frequency Distribution," Journal of the Institute of Electronics Engineers of Korea, Vol. 48, No. 6, pp. 124-132, 2011.
- [7] D.-W. Lee, and J.-W. Han, "Adaptive radar pulses clustering based on density cluster window," International Technical Conference on Circuits Systems, Computers and Communications, pp. 1377-1380, 2008.
- [8] G. Qiang, X. Zhang and Z. Li, "SVC & k-means and type-entropy based de-interleaving/recognition system of radar pulses," 2006 IEEE International Conference on Information Acquisition. IEEE, pp. 742-747, 2006.
- [9] A. K. Jain, "Data clustering: 50 years beyond K-means," Pattern Recognition Letters, Vol. 31, No. 8, pp. 651-666, 2010.
- [10] M. Li, M. He, J. Han, and Y. Tang, "A new clustering and sorting algorithm for radar emitter signals," Journal of Physics: Conference Series, Vol. 1617, No. 1, pp. 012009, 2020.
- [11] X. Feng, X. H and Y. Liu, "Radar signal sorting algorithm of K-means clustering based on data field," 2017 3rd IEEE International Conference on Computer and Communications, pp. 2262-2266, 2017.
- [12] D. Arthur and S. Vassilvitskii, "K-means++ the advantages of careful seeding," Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 1027-1035, 2007.
- [13] P. Bholowalia and A. Kumar, "EBK-means: A clustering technique based on elbow method and k-means in WSN," International Journal of Computer Applications, Vol. 105, No. 9, 2014.
- [14] F. Wang, H.-H. Framcp-Penya, J. D. Kelleher, J. P. and R. Ross, "An analysis of the application of simplified silhouette to the evaluation of k-means clustering validity," Machine Learning and Data Mining in Pattern Recognition: 13th International Conference, pp. 291-305, 2017.