

# 클러스터링 알고리즘을 이용한 다수의 풍력 단지 발전량 시나리오 생성 방법

이륜경\*, 김지영\*\*, 신훈영\*\*

\*상명대학교 전기공학과

\*\*상명대학교 일반대학원 에너지그리드학과

e-mail: 201710839@sangmyung.kr

## A Scenario Generation Method for Multiple Wind Farms Using a Clustering Method

RyunGyeong Lee\*, Jiyoung Kim\*\*, Hunyoung Shin\*\*,

\*Dept.of Electrical Engineering, Sangmyung University

### 요약

에너지 전환 패러다임에 따라 전력계통에서 재생에너지 비중이 증가하고 있다. 재생에너지의 경우 출력에 있어 변동성과 불확실성이 존재하기 때문에 계통 연계 시 이를 정확하게 모델링 할 필요가 있다. 본 연구의 목적은 다수의 풍력 단지 출력들의 시나리오를 생성하는 알고리즘을 개발하는 데 있다. 다수의 풍력 단지 발전량 데이터를 머신러닝 기반의 K-means Clustering 알고리즘을 이용하여 발전 출력 패턴의 상관관계를 고려하여 군집화한 후 각 군집 별 풍력 단지들의 발전량 시나리오를 생성하였다. 국내 풍력 발전 데이터를 통해 제안하는 알고리즘으로 생성한 시나리오가 실제 발전량 데이터와 유사한 출력 패턴을 성공적으로 모의하는 것을 확인하였다.

## 1. 서론

전 세계적으로 화석 연료로 인한 환경 문제에 직면하면서, 온실가스 감축을 위해 기존 화석 연료에서 재생 가능 에너지로의 전환이 빠르게 확산 되고 있다. 이를 위해 각국에서는 재생에너지 목표량을 설정하여 재생에너지의 비중을 증가시키고 있다. 독일의 경우 총발전량 대비 재생에너지 발전량을 '50년까지 80%로 확대할 계획이다. 국내의 경우 '제8차 전력수급 계획'을 통해 '재생에너지 3020 이행 계획'에 따라 '17년 기준 1.2GW인 풍력 발전의 설비용량을 2030년까지 17.7GW로 확대할 계획이다[1-2]. 현재 국내의 풍력 발전은 '18년 기준 설비용량 1.4GW, '19년 기준 1.5 GW에 달하는 설비용량을 구축한 상태이다[3].

재생 가능한 에너지 자원의 보급이 증가함에 따라 기존의 전력계통과 연계 시에 더 많은 변동성과 불확실성을 갖게 될 것으로 예상된다. 특히 풍력 자원은 풍속, 풍향, 온도 등 다양한 요인에 영향을 받기 때문에 풍력 발전이 증가하게 되면 계통 운영에 어려움이 발생한다. 이에 따라 안정적인 전력계통과의 연계를 위해서는 정확한 재생에너지 모델링을 기반으로 운영 및 설비 투자 계획을 세울 필요가 있다.

본 연구에서는 다수의 풍력 단지 출력 시나리오를 생성하는 알고리즘을 제안한다. 본 알고리즘은 풍력 발전 단지들을 클러스터링 기법을 사용하여 유사한 출력 패턴을 기반으로 군집화한 후, 각 군집 별 풍력 단지들의 발전량 시나리오를 생성한다.

## 2. 본론

### 2.1 K-Means Clustering

클러스터링 기법은 머신러닝 기반의 비지도 학습(Unsupervised Learning) 중 하나로, 목적함수를 최소화하는 데 기반을 둔 클러스터링 중에서도 가장 널리 연구된 것이 K-means 클러스터링 기법이다 [4]. 주어진 데이터 집합을 K개의 클러스터로 그룹화할 수 있는 방법을 제공한다. 예를 들어  $X = \{x_1, x_1, \dots, x_i, \dots, x_n\}, x_i \in R^d$  라는  $n$ 차원의 데이터 집합과 K가 주어진다면, 클러스터링 알고리즘을 통해 K개의 군집으로 나눌 수 있다 [5]. 이때 군집마다 하나씩 K개의 중심점 (Centroid)을 갖게 된다.

### 2.2 시나리오 생성

전체적인 시나리오 생성을 알고리즘은 아래 [그림 1]과 같다. 클러스터링을 통해 다수의 풍력 발전 단지를  $j$ 개의 클러

스터로 군집화 한 다음, 각 클러스터의 대푯값을 산출한다. 이때  $t$ 시간대의  $j$ 번째 클러스터의 대푯값을  $\mu_j(t)$ 라 하고,  $j$ 번째 클러스터에 속한  $i$ 번째 발전단지의 출력값을  $x_{ij}(t)$ 라 한다. 여기서 대푯값인  $\mu_j(t)$ 와 출력값인  $x_{ij}(t)$ 의 차이를 다음의 식 (1)과 같이  $\epsilon_{ij}(t)$ 로 정의할 수 있다.

$$\epsilon_{ij}(t) = x_{ij}(t) - \mu_j(t) \quad (1)$$

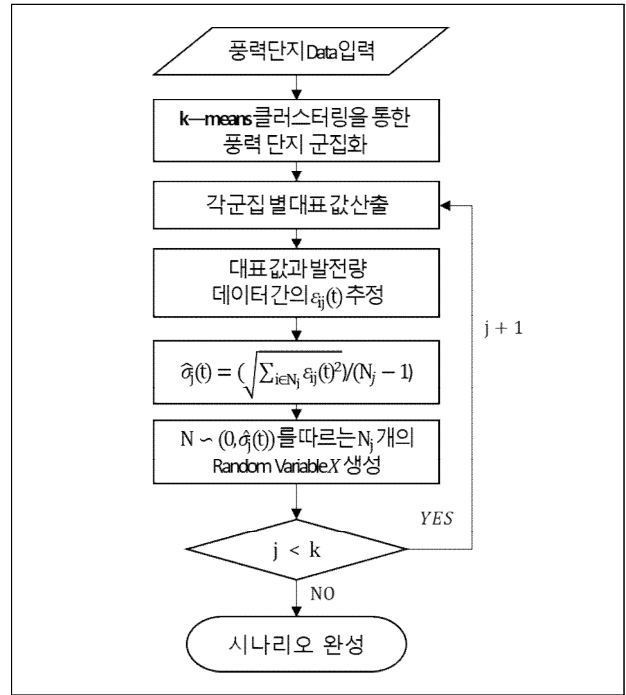
클러스터 내 풍력 단지들의 시나리오는 대푯값에 랜덤노이즈를 더하여 생성하게 된다. 랜덤노이즈는 가우시안 분포로 가정하고 이때의 표준편차  $\hat{\sigma}_j(t)$ 를 정하기 위한 식은 다음 식 (2)와 같다.

$$\hat{\sigma}_j(t) = \frac{\sqrt{\sum_{i \in N_j} \epsilon_{ij}(t)^2}}{N_j - 1} \quad (2)$$

여기서  $N_j$ 는 클러스터  $j$ 에 속한 발전단지의 개수를 뜻한다.

$$X \sim N(0, \hat{\sigma}_j(t)) \quad (3)$$

식 (3)의 분포를 따르는 랜덤 변수를  $N_j$ 개만큼 생성하여 대푯값인  $\mu_j(t)$ 에 더하여 최종적으로 시나리오를 생성하게 된다. 각  $j$ 개의 클러스터 내에서 이 과정은 독립적으로 이루어진다.



[그림 1] 시나리오 생성 알고리즘

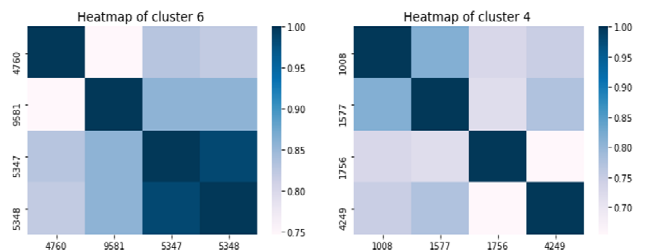
### 2.3 시뮬레이션 결과

시뮬레이션에서는 '18년 국내 전국 풍력 단지 82개의 발전량 데이터를 사용하였다. 클러스터링 기법을 사용하는 데 필요한 클러스터 개수 (K)는 임의로 6이라 지정하였다. 또한, 각각의 풍력 단지의 설비용량이 다르기 때문에 평균을 0, 분산을 1로 지정하여 같은 스케일의 값으로 변환해주는 표준화의 과정을 거쳤다. 클러스터링 과정을 수행한 결과는 다음 [표 1]과 같다.

[표 1] 풍력 발전단지의 클러스터링 결과

클러스터	발전단지 개수	클러스터	발전단지 개수
1	18	4	15
2	12	5	14
3	15	6	8

6개의 클러스터 내에 82개의 발전단지가 적절하게 그룹화된 것을 확인할 수 있다. [그림 2]는 6개의 클러스터에서 임의로 2개의 클러스터를 추출한 다음, 각 클러스터 내에 분류된 발전단지들의 상관관계를 히트맵으로 나타낸 것이다.

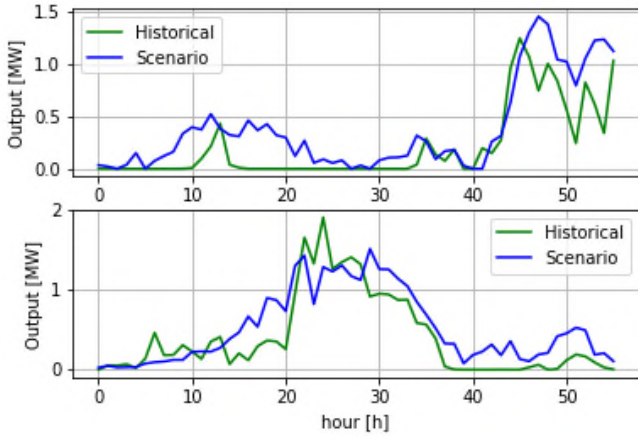


[그림 2] 임의의 클러스터 내 발전단지의 상관관계 히트맵

참고문헌

[1] 산업통상자원부, “제8차 전력수급기본계획”, 2017  
 [2] 산업통상자원부, “재생에너지 3020 이행계획(안)”, 12, 2017.  
 [3] KPX, “Annual Generation Capacity Report”, 2015~2019  
 [4] T.Kanungo et al., “An Efficient k-Means Clustering Algorithm: Analysis and Implementation”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.24 pp.881-892, August, 2002  
 [5] R.C. Duan et al., “Data Mining & Pattern Recognition of Voltage Sag Based on K-means Clustering Algorithm”, 2015 IEEE Power & Energy Society General Meeting

우측의 수치들은 상관관계를 나타내는 지표이며, 각 클러스터 내의 풍력단지들 간에 높은 상관성을 갖는 것을 확인하였다. [그림 2]를 통해 발전 패턴이 유사한 풍력단지들끼리 그룹화된 것을 확인할 수 있었다.



[그림 3] 발전단지의 발전량 데이터와 시나리오의 비교

[그림 3]에서는 82개의 풍력 발전단지들 중 임의로 2개의 발전단지를 추출하여 원래의 발전량 데이터와 본 논문에서 제시하는 알고리즘을 통해 생성한 시나리오의 값을 비교해보았다. 두 개의 발전단지 모두 생성된 시나리오와 실제 발전량 출력 데이터와 비슷한 패턴을 가지고 있는 것을 확인할 수 있었다.

### 3. 결론

본 논문에서는 풍력 발전 출력 데이터를 가지고 클러스터링을 통해 군집화하여 각 군집 간의 상관관계를 고려하여 시나리오를 생성하는 알고리즘을 수행하였다. 알고리즘 수행 후, 상관관계가 높은 발전단지들 별로 그룹화된 것을 확인하였다. 또한 실제 발전단지의 출력 데이터와 유사한 패턴으로 시나리오가 생성된 것을 확인할 수 있었다. 클러스터링 기반 샘플링 시나리오 생성 알고리즘은 계통 안정화를 위한 분석 시 재생에너지를 확률적으로 모델링 할 때 활용 가능할 것이다. 더 나아가 전력계통 분석 및 설비계획에도 사용 가능할 것 이라고 생각된다.

#### 감사의 글

본 연구는 2016년도 산업통상자원부의 재원으로 한국에너지기술평가원(KETEP)의 에너지인력양성사업으로 지원받아 수행한 인력양성 성과입니다. (No. 20164030300230)