

# 분산에너지 가상발전소의 정산금 향상을 위한 오차율 평가 기반 최대 이익 시계열 클러스터링 기법

조수빈<sup>†</sup>, 정영훈<sup>†</sup>, 조성령<sup>†</sup>, 서정일<sup>†</sup>, 천세진<sup>†</sup>

## MaxProfit Time Series Clustering Technique Based on Error Rate Evaluation to Improve Reward of Distributed Energy Virtual Power Plant

Subin Cho<sup>†</sup>, Yeonghun Jeong<sup>†</sup>, Seongryeong Jo<sup>†</sup>, Jeongil Seo<sup>†</sup>, Sejin Chun<sup>†</sup>

### ABSTRACT

Renewable energy is currently driving innovation in the energy industry. However, the widespread distribution of renewable energy resources across various regions and the temporal variability in production pose challenges for the operation and management of virtual power plant (VPP). To address this issue, a Dynamic Virtual Power Plant (DVPP) has been proposed for the dynamic configuration of VPP. In this paper, we propose the MaxProfit Time Series Clustering for Virtual Power Plant (MTPSC-VPP) to effectively aggregate virtual power plant. We perform time series clustering based on density using the normalized Mean Absolute Error (nMAE) between actual measurements and predictions. Subsequently, we compare the proposed algorithm with non-time series and existing algorithms through experimental evaluations. The results indicate that MTPSC-VPP yields approximately 93%p higher prediction revenue compared to non-time series methods and about 15%p higher compared to existing algorithms. However, it exceeds the calculation criteria of up to 30% of reward on specific dates, suggesting directions for achieving better performance in future research.

**Key words:** Renewable Energy, NMAE, Time Series, Clustering

### 1. 서 론

재생에너지 자원(Renewable Energy Sources, RES)은 세계적으로 에너지 산업을 혁신하고 있다. 재생에너지 '3020' 정책이 시행되면서 재생에너지를 이용한 발전 설비가 꾸준히 증가하고 있다. 그러면서, 소규모로 분산된 RES에 대한 관리가 더더욱 중

요해지고 있다. 분산된 형태의 재생가능 에너지 자원은 일반적으로 분산 에너지 자원(Distributed Energy Resources, DER)으로 지칭된다. 이러한 DER 들은 그리드 시스템 내에서 통합되어, 전력망의 안정성 및 신뢰성을 강화하는 데 기여하는 중추적인 역할을 수행한다. 이 통합 프로세스는 전력 그리드를 보다 유연하고 효율적인 제어가 가능하게 하여 에너지 공급

※ Corresponding Author: Jeongil Seo, Sejin Chun, Address: (49315) 37, Nakdong-daero 550, Saha-gu, Busan, Korea, TEL : +82-51-200-7796, +82-51-200-7923, FAX : +82-51-200-7783, E-mail : jeongilseo@dau.ac.kr, sjchun@dau.ac.kr

Receipt date : Dec. 12, 2023, Revision date : Mar. 15, 2024  
Approval date : Mar. 15, 2024

<sup>†</sup> Dept. of Computer Eng., Dong-A University  
(Subin Cho, E-mail : 2243372@donga.ac.kr)  
(Yeonghun Jeong, E-mail : jyhl1224@donga.ac.kr)  
(Seongryeong Jo, E-mail : 2243094@donga.ac.kr)

※ This research work was supported by the National Research Foundation of Korea(NRF) (No.RS-2023-00273349) and Institute of Information & communications Technology Planning & Evaluation(IITP) (No.2023-0-00076, National Program of Excellence in Software(Dong-A University)) grant funded by the Korea government (MSIT).

의 지속 가능성을 증진하고 있다. RES는 기상 조건의 변동성에 기인하여 발전량의 변동이 크고 예측이 어려운 특성을 보인다. 이러한 불확실성 때문에 DER의 발전량 예측에 있어 고도화된 예측 기술의 개발이 필요하다. 정확한 발전 예측은 전력 그리드의 효율적인 관리 및 안정적인 운영을 위한 필수적인 요소로 인식되고 있다. 예를 들어, 태양광 및 풍력 발전 시스템은 광범위한 지리적 분포와 함께, 일별 및 시간별로 발전량에 상당한 변동성을 보이는 경향이 있다. 이러한 변동성은 전력 그리드의 운영 및 관리 측면에서 복잡성을 증가시키며, 이에 대응하기 위한 세심하고 정교한 접근 방식을 요구한다.

특히, 넓은 지역적 분포와 낮은 생산량을 가진 발전소들을 보완하기 위하여, 다양한 지역의 DER을 통합하는 가상발전소(Virtual Power Plant, VPP)의 개념이 등장하였다. VPP는 특정 중개 사업자가 개인 소유의 소규모 발전설비, 태양광, 풍력 등의 DER을 가상적으로 연결하여 운영하는 새로운 모델이다. 중개 사업자는 일일의 예측 생산량을 각 DER로부터 수집하여 전력 계통에 통합한다. 정산금 제도에 따라, 중개 사업자는 판매 이익을 얻고 이를 DER 소유자들에게 분배한다. 정산금 제도는 하루 전에 제출된 예측 생산량과 실제 생산량 간의 차이를 비교하고, 오차 범위 내에서 전력을 생산한 VPP에 정산금을 지급한다. 그러나, 다양한 DER의 생산량, 날씨 조건 및 계절 변화로 인해 중개 사업자는 적절한 VPP 구성과 효율적인 DER 집약 방법을 탐구해야 한다.

동적 가상발전소(Dynamic Virtual Power Plant, DVPP)[1]는 최신 연구에서 주목하는 에너지 관리 프레임워크로, 중개 사업자가 RES를 동적으로 통합하고 최적화된 운영을 할 수 있도록 설계되었다. 이는 고정된 매개변수에 의존하던 기존의 연구들과 대비되어, 변화하는 조건과 요구에 능동적으로 대응하는 동적인 접근을 요구한다. DVPP는 배치할 수 있는 RES와 그렇지 않은 RES의 조합을 통해 다양한 에너지 포트폴리오를 구성하며, 현재와 미래 시나리오 모두를 고려한 에너지 공급의 안정성과 최적화된 성능을 추구한다. 그럼에도 불구하고, DVPP를 자동으로 구성하는 시스템은 아직 제안되지 않았다는 한계가 있다.

본 연구는 정규화된 평균절대오차(Normalized Mean Absolute Error, nMAE)를 기반으로 한 시계

열 분석 모델을 통해 DER 발전량 예측의 불확실성을 DVPP가 반영할 방안을 제안한다. 이를 위해, 최대 이익을 실현할 수 있도록 각 DER의 데이터를 기반으로 가상발전소를 자동으로 집약하는 시계열 클러스터링 알고리즘(MaxProfit Time Series Clustering for Virtual Power Plant, MPTSC-VPP)을 제안하였다. 본 연구의 알고리즘 성능을 검증하기 위해, 기존의 클러스터링 방법들과 비교 실험을 실시하였다. 실험은 한 달 동안의 데이터로 모델을 훈련한 후, 이어지는 10일간의 데이터로 예측 정확도를 평가하는 방식으로 진행되었고, 정산금 지급액을 비교 분석함으로써 제안된 방법의 경제적 효율성을 측정하였다.

본 논문의 구성은 다음과 같다. 2장에서는 기존 클러스터링 알고리즘의 소개와 에너지 분야에서의 클러스터링 알고리즘 활용에 대해 소개한다. 이어지는 3장에서는 본 논문에서 제안한 MPTSC-VPP의 전체 framework와 제안 방법에 대한 내용을 다룬다. 4장에서는 MPTSC-VPP의 정산금 산정 성능 측정을 위한 실험과 검증에 대한 내용을 제시한다.

## 2. 관련 연구

### 2.1 클러스터링 기법

클러스터링 기법은 데이터를 서로 비슷한 특성을 가진 클러스터로 묶는 기법이다. 이는 데이터를 비슷한 속성이나 형태를 보이는 그룹으로 나누어 데이터를 이해하고 구조화한다. 클러스터링은 데이터 안에 숨겨진 구조를 찾아내거나 유사한 데이터를 동일한 집단으로 묶어서 데이터를 보다 이해하기 쉽게 한다. 주요 기법으로는 K-Means[2], Gaussian Mixture Model(GMM)[3], Density-Based Spatial Clustering of Applications with Noise(DBSCAN)[4] 등이 있다.

K-Means는 거리 기반의 알고리즘이다. 초기에 무작위로 선택한 중심을 조정하여 데이터를 K개의 클러스터로 그룹화한다. 각 데이터 포인트는 가장 가까운 중심에 할당되며, 클러스터 내의 데이터 포인트는 정확히 하나의 클러스터에 속한다. GMM은 확률적 모델로, 데이터를 여러 개의 Gaussian 분포로 모델링한다. 각 클러스터는 Gaussian 분포를 따르며, 데이터 포인트는 각 클러스터에 속할 확률을 가진다.

GMM은 클러스터의 모양이 임의적이며, 각 클러스터는 공분산 행렬을 통해 모델링 된다. DBSCAN 알고리즘은 데이터 포인트들의 공간적 밀도를 기반으로 클러스터를 형성한다. 이 알고리즘은 임의의 데이터 포인트를 시작점으로 선택하고, 설정된 반경(eps) 내에 위치한 최소 수(MinPts)의 이웃 데이터 포인트들을 기준으로 클러스터를 구성한다. 만약 eps 내에 MinPts 이상의 데이터 포인트들이 존재한다면, 해당 데이터 포인트는 '핵심 포인트'로 분류되고, 이를 중심으로 클러스터가 확장된다. 이 과정을 전체 데이터 집합에 대해 반복하며, 어떤 클러스터에도 속하지 않는 개별 데이터 포인트들은 이상치로 분류된다. DBSCAN의 장점은 클러스터의 형태가 임의적이며, 크기에 구애받지 않고 이상치에 강건하다는 점에서 다양한 데이터 세트에 적용 가능하다는 것이다.

nMAE는 예측값과 실측값 사이의 차이를 정량화하는 데 사용된다. 이 지표는 모든 개별 데이터 포인트에 대한 예측값과 실측값의 절대 오차를 계산하고, 이들의 평균값을 취하여 정규화함으로써 모델의 예측 성능을 평가한다.

## 2.2 에너지 분야에서 활용된 클러스터링 기법

본 연구는 VPP 형성을 위해 DER에서 데이터를 클러스터링하는 기법의 적용 사례에 주목하였다. A study on solar radiation prediction using medium-range weather forecasts[5]에서는 다중 선형 회귀, K-Nearest Neighbors(KNN), Random Forest, Support Vector Regression(SVR)과 같은 다양한 머신

러닝 방법론을 결합한 K-means 클러스터링 기법을 활용하여 각 클러스터에 대한 확률 밀도 함수를 계산하여 시간별 태양광 일사량을 추정하고, 이를 바탕으로 태양광 발전소의 일일 발전량을 예측하는 데 사용되었다. An Risk-based operation in microgrid using two-stage optimization approach considering uncertainty[6]에서는 분산 에너지 자원의 불확실성이 불가피하게 존재하며, 이러한 불확실성으로 인해 마이크로 그리드 운영자가 시스템을 운영자가 시스템을 운영하는 데 어려움을 겪는 문제에 대응하기 위한 방법을 제안했다. 추가적으로, A study on Grouping of Domestic Wind Power Generation Data through Clustering Methods[7]의 연구에서는 풍력 분야에 클러스터링 기법을 적용하여 다수의 풍력 발전소 단지를 15개의 클러스터로 분류하였고 이를 통해 풍속 및 풍향과 같은 변수들이 풍력 발전에 미치는 영향을 분석하였다. K-means를 이용한 수질센서 데이터의 클러스터링 모델[8]에서는 수질모니터링 시스템을 위하여 싱크노드에서 센싱 데이터들을 일정한 수의 클러스터로 K-means 기법을 사용하여 클러스터링 하였다.

## 3. 제안방법

본 장에서는 시계열 데이터 위의 확률 분포모델과 VPP를 위해 DER을 자동으로 집약하는 MPTSC-VPP 클러스터링 알고리즘을 제안한다.

Fig. 1에서 제시된 Framework는 MPTSC-VPP 알고리즘의 제안 방법을 보여준다. 예측 오차율은 중

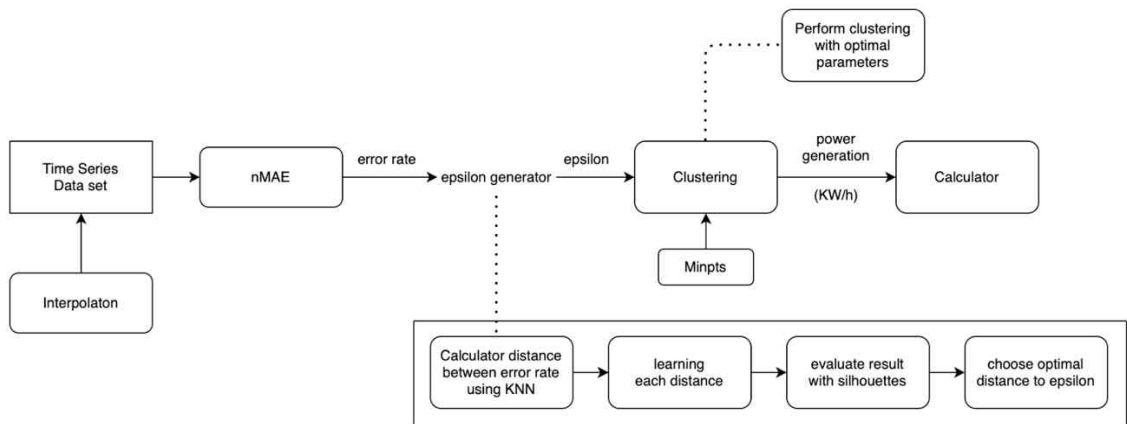


Fig. 1. Framework of proposed method.

개 사업자가 예측한 발전량과 실제로 측정된 발전량 간의 오차율로 nMAE를 이용해 계산한다. 식(1)은 전력 시장 운영 규칙에 따른 nMAE 식이다.  $PG_a$ 는 실측 발전량이고  $PG_p$ 는 예측 발전량이고 PC는 발전 소용량을 의미한다.

$$nMAE = \frac{|PG_a - PG_p|}{PC} \times 100 \quad (1)$$

예측 정산금은 예측한 발전량과 실제로 측정된 발전량 간의 오차에 따라 어느 정도의 정산금을 얻을 수 있는지를 나타낸다. 예측 정산금을 높이기 위해 nMAE가 최소가 되도록 하는 것이 중요하다.

우리는 정산 금액을 높이기 위해 nMAE를 사용하여 최적의 오차율을 나타내는 epsilon 파라미터를 찾는다. eps를 통해 하이퍼 파라미터인 MinPts를 설정하여 클러스터링을 수행한다. 최적의 eps 값을 결정하기 위해 KNN을 이용하여 오차율 간의 거리를 계산한다. 각 거리를 학습한 후, 실루엣 계수를 통해 평가를 수행한다. 이를 통해 최적의 오차율 거리를 선택한다. 클러스터링 알고리즘을 통해 얻은 클러스터를 기반으로 발전량마다 정산금을 계산하였으며, 이때 단위는 kW/h이다.

본 연구는 DER에서 취득된 데이터의 정상성을 가정한다. 즉, 주어진 각 DER의 시간별 측정 데이터에 대하여 일일 발전량의 확률 분포를 시간대별로  $X_1, X_2, \dots, X_{24}$ 과 같이 정의하고, 식(2)와 같이 공통된 분포 함수  $F(\theta)$ 를 따른다고 할 수 있다:

$$X_i \sim F(\theta) \text{ for } i = 1, 2, \dots, 24 \quad (2)$$

$X_i$ 는  $i$ 번째 시간대에 대응하는 확률 변수를 나타내며,  $F(\theta)$ 는 해당 확률 변수들의 분포 함수를 의미한다.  $\theta$ 는 분포의 파라미터를 나타내며, 시계열 데이터 세트의 특성을 반영한다. 여기서  $\sim$ 은 “따른다(follows)”라는 의미로 사용된다.

본 연구에서는 RES의 발전량에 대한 확률 분포가 에너지 자원의 종류에 따라 서로 다른 패턴을 나타낸다는 것을 확인하였다. 예를 들어, 태양광 발전량은 주로 일사량에 의해 영향을 받으므로, 일출 후부터 일몰 전까지의 시간대에 높은 발전량을 보이지만, 일몰 후부터 일출 전까지는 상대적으로 발전량이 적다. 이와는 대조적으로, 풍력 발전은 주로 해당 시간대의 풍속에 의해 발전량이 결정되며, 이는 시간대별로 상이한 패턴을 나타낸다. 이러한 현상을 반영하여, 본

논문에서는 각 시간대의 발전량 데이터가 서로 독립적이라는 가정하에 모델을 구성한다. 이는 태양광 발전이나 풍력 발전과 같은 각기 다른 발전 방식에서 관찰되는 시간대별 발전량의 변화를 감안한 것으로, 각 시간대의 발전량이 다른 시간대의 발전량에 영향을 미치지 않음을 의미한다. 따라서, 각 시간대에 대한 발전량의 확률적 특성을 독립적으로 분석하고 예측하는 것이 가능하다.

### 3.1 최대 이익 시계열 클러스터링을 이용한 가상 발전소 집약

구성된 VPP가 최대 이익을 얻을 수 있도록 하기 위해 본 논문은 최대 이익 시계열 클러스터링 알고리즘인 MPTSC-VPP를 제안한다. 클러스터링 기법 중 하나인 DBSCAN을 확장하여 시계열 데이터를 밀도 기반으로 그룹화하고, 발전량의 예측 오차율을 줄이는 방향으로 학습하였다. 이를 통해 예측 정산금을 극대화하는 모델을 구성하였다.

Fig. 2는 본 논문에서 제안하는 MPTSC-VPP 알고리즘의 개요를 나타낸다. 알고리즘은 주어진 시계열 데이터 세트  $D$ 를 입력으로 받아, 최소한의 발전소 개수  $n_v$ 를 기준으로 가상발전소를 형성한다. 알고리즘은 먼저 모든 발전소를 ‘할당되지 않음(UNASSIGNED)’ 상태로 초기화한다. 이후, 각 발전소  $x$ 에 대하여,  $x$ 가 아직 클러스터에 할당되지 않았다면, SCAN 연산을 이용하여  $x$ 의 이웃하는 발전소들

---

#### Algorithm 1: MPTSC-VPP

---

**Input:** dataset  $\theta$   
           neighbor plant difference  $D$ ,  
           minimum # of plant to form vpp  $n_v$   
**Output:** Cluster  $c$

```

1  $k = 0$  // number of plants
2 for each  $x \in \theta$  do
3    $y_x \leftarrow \text{UNASSIGNED}$ 
4 end
5 for each  $x \in \theta$  do
6   if  $c_x = \text{UNASSIGNED}$  then
7      $\theta_x = \text{SCAN}(p, D)$  // Search neighbors of  $x$ 
8     if  $|\theta_x| \geq n_v$  then
9        $k \leftarrow k + 1$ 
10       $c_x \leftarrow k$ 
11      for each  $z \in \theta_x$  do
12         $c_z \leftarrow k$ 
13         $\theta_x = \text{SCAN}(p, D)$ 
14        if  $|\theta_x| \geq n_v$  then
15           $\theta_x \leftarrow \theta_x \cup \theta_z$ 
```

---

16 **return**  $c$

---

Fig. 2. MPTSC-VPP Algorithm.

Table 1. Power generation dataset of 169 solar power plant nationwide from march 1, 2023 to march 31, 2023.

Date	Class	Power Plant Code	Capacity (kW)	0:00	1:00	...	22:00	23:00	Sum
2023-03-01	Actual	G1053	249.48	0	0	...	0	0	640.9
2023-03-01	Predict	G1053	249.48	0.95	0.97	...	0.89	0.98	739.26
...	...	...	...	...	...	...	...	...	...
2023-03-31	Actual	B1739	98.825	0	0	...	17.44	0	572.96
2023-03-31	Predict	B1739	98.825	0.28	0.28	...	16.83	0.28	567.76

Algorithm 1: SCAN

**Input:** Plant  $p$   
 neighbor plant difference  $D$ ,  
**Output:** neighbor  $\theta_p$   
 1 **for each**  $z \in X$  **do**  
 2   **if**  $diff(p, z) \leq D$  **then**  
 3      $\theta_p \leftarrow \theta_p \cup \theta_z$   
 4 **return**  $\theta_p$

Fig. 3. SCAN Algorithm.

의 집합  $\theta_x$ 를 찾는다. SCAN 연산은 Fig. 3에 정의된 바와 같이, 데이터 포인트  $\theta$ 의 주변에 있는 이웃 데이터 포인트  $\theta_p$ 를 탐색하는 함수이다. 만약  $\theta_x$ 의 원소 개수가 미리 설정된 최소 발전소 개수  $n_v$  이상이면, 발전소  $x$ 를 핵심 데이터 포인트로 분류하고 새로운 클러스터  $C_x$ 를 형성한다. 그리고  $\theta_x$ 에 속한 각 데이터 포인트에 대하여, 다시 SCAN 함수를 적용하여 해당 데이터 포인트가 핵심 데이터 포인트인 경우, 발전소  $x$ 와 동일한 클러스터에 속하도록 한다. 이러한 과정을 모든 데이터에 대해 반복하여, 발전소 간의 연관성을 기반으로 VPP를 구성한다. 이 알고리즘은 클러스터 형성 과정에서 발전소의 최적 집약을 도모하여, 최대의 이익을 발생시킬 수 있는 VPP 구성을 목표로 한다.

## 4. 실험

### 4.1 실험 환경

본 논문에서는 모델의 학습을 위하여 Table 1의 전국 169개 태양광 발전소의 실측 발전량과 예측 발전량을 2023년 3월 1일부터 2023년 3월 31일까지 24시간 단위로 작성한 데이터를 사용한다. 결측값 확인과 데이터 전처리를 거쳐 모델이 데이터 패턴을 학습하고 예측할 수 있도록 하였다. 모델을 학습하는 데

Table 2. Hyperparameter for experiments.

Hyperparameters	Values
eps	2.2
MinPts	4
A Number of Cluster	6

Table 2와 같은 하이퍼 파라미터를 사용한다. 이를 위해 KNN을 사용하여 최적의 eps를 탐색하고 2차원 데이터에 대한 MinPts를 4로 설정하였다. 2차원보다 많은 변수를 가진 데이터 세트의 경우에는 MinPts를  $2 \times \text{dim}$ 으로 설정한다. 또한 클러스터의 개수는 6개로 설정하여 실험을 진행한다. 검증 데이터는 10일 치로 구성되어 있다.

실험 지표로는 식(1)의 nMAE와 예측 정산금을 선정하였다. nMAE는 편차를 정규화하여 미래의 발전량을 비교할 수 있고, 값이 작을수록 예측이 정확함을 나타낸다. 이는 발전소들을 서로 다른 그룹으로 분류하였을 때, 오차를 패턴이 비슷한 발전소들의 특성을 이해하고 비교할 수 있게 한다. 또한, 예측 정산금을 비교하는데, 이 예측 정산금은 예측 오차율이 낮을수록 높아진다. 예측 정산금이 높다는 것은 해당 VPP가 안정적으로 전력을 공급했다는 지표이며, 각각의 DER이 많은 수익을 얻을 수 있음을 나타낸다. 예측 정산금을 계산하기 위해 식(3)을 사용하였다. 이때,  $P_G$ 는 실측 발전량의 평균이고, nMAE는 식(1)을 나타낸다. M은 한 달의 일수를 나타내며, h는 하루 측정 시간이다. 이때 단위는 원(\)이다.

$$\begin{cases} P_G \times M \times h \times 4 & (nMAE < 6\%) \\ P_G \times M \times h \times 3 & (6\% < nMAE < 8\%) \end{cases} \quad (3)$$

각각의 모델을 학습시키고 비교 실험함에 있어서 Table 3과 같은 환경을 구성하였다.

Table 3. Environment for experiments.

Requirements	Version
Python	Ver3.10
Scikitlearn	Ver2.2.0
Timeserieslearn	Ver0.6.2

## 4.2 실험 결과

### 4.2.1 시계열과 비시계열 모델 간의 성능 비교

Fig. 4는 시계열을 고려하지 않은 발전소별 평균 오차율 그래프이다. x축은 발전소, y축은 발전소별 오차율이다. 각각의 클러스터는 비슷한 오차율을 보이고 있으며, cluster 0에 낮은 오차율을 보이는 발전소들이 모인다. Fig. 5는 시계열이 고려되지 않았을 때 각 클러스터에 밀집된 분포를 보여준다. Cluster 0에는 99개의 가장 많은 발전소가 밀집됐고, cluster 1과 cluster 3에 가장 적은 발전소가 속한다.

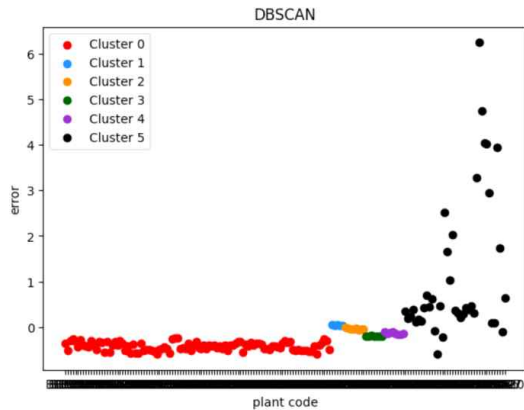


Fig. 4. Result of clustering as the average error of power plant.

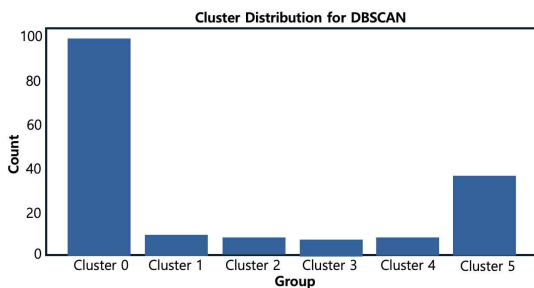


Fig. 5. Distribution of power plant based on non-time series VPP formation.

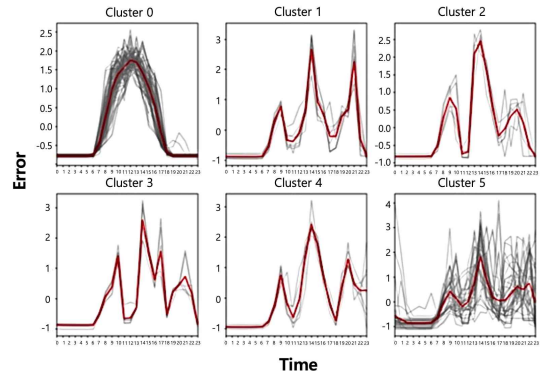


Fig. 6. Result of clustering as the time series error of power plant.

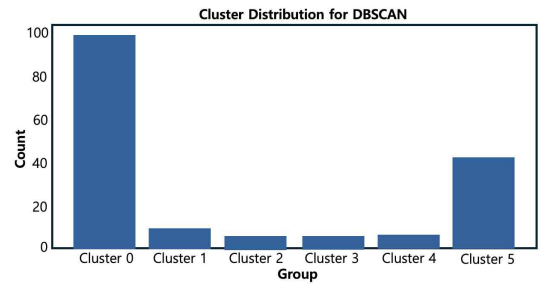


Fig. 7. Distribution of power plant based on time series VPP formation.

Fig. 6은 시계열 클러스터링을 통해 구성된 클러스터의 시간별 오차율을 보여준다. x축은 시간대, y축은 오차율이다. 각각의 클러스터가 시간대별로 확실한 오차율 차이를 보인다. Fig. 7은 시계열 클러스터링으로 구성된 각 클러스터에 밀집된 분포이다. Fig. 5와 마찬가지로 cluster 0에 가장 많은 발전소가 포함되었다. 하지만 시계열을 고려하지 않았을 때와 달리 cluster 2와 cluster 3에 가장 적은 발전소가 포함되었다.

Table 4는 각 발전소의 오차율을 비시계열 방식, K-Means, GMM과 제안 방법인 MPTSC-VPP를 이용해 클러스터링하여 클러스터별 정산금을 구한 결과이다. 모든 클러스터에 대해 제안 방법의 정산금이 증가한 것으로 확인되었다. 제안 방법으로 클러스터링을 수행한 결과, 가장 높은 예측 정산금을 보이는 클러스터는 cluster 0으로 4,728,894원이었다. 반대로 가장 낮은 예측 정산금을 보이는 클러스터인 cluster 2는 예측형 집합 자원의 기준 발전량인 20MW를 넘지 않아 정산금을 산정받지 못한다. 이를 통해 기준

Table 4. Results of reward of Non-time series and Time series by Cluster.

Cluster	Non-Time Series reward(\)	Ours(\)	K-MEANS(\)	Gaussian Mixture(\)
Cluster 0	618,338	4,728,894	2,806,366	2,309,040
Cluster 1	19,427	1,313,630	364,294	686,797
Cluster 2	32,749	0	0	0
Cluster 3	19,359	1,754,023	3,667,487	2,006,527
Cluster 4	84,308	395,223	1,367,500	1,677,725
Cluster 5	137,592	3,774,707	2,665,638	2,985,771
Total Reward	911,773	11,966,477	10,871,285	9,665,860

발전량 미만의 발전소를 필터링하는 효과를 얻을 수 있음이 확인되었다. 이를 제외한 나머지 클러스터에서는 기존의 알고리즘보다 약 15%p 더 높은 예측 정산금을 산출하면서 더 높은 수익성을 보장한다.

#### 4.2.2 벤치마크 데이터 기반 정확성 평가

연구 결과를 검증하기 위해, 2023년 4월 1일부터 2023년 4월 10일까지의 실측 발전량 데이터를 이용하였다. 해당 검증 데이터를 통해 제안 방법이 실제 문제 해결에 적합한지에 대하여 클러스터링 결과의 품질과 신뢰성을 확인하고자 한다.

Table 5는 MPTSC-VPP 알고리즘을 사용하여 발전소들을 모은 각 VPP별 검증 결과를 나타낸다. 도메인 전문가의 성능 검증으로 10일의 검증 결과 중 7일이 오차율 8% 이내의 일별 nMAE와 그룹별 nMAE 모두 정산금 산정기준에 포함된다. 이로써 비슷한 측정 오류 형태를 보인 DER 들을 통합하여 VPP를 클

러스터링한 제안 방법이 우수함을 증명하였다. 또한 cluster 0에 99개의 발전소가 밀집되었으나, 2.8%의 가장 적은 오차율을 보이면서 문제가 없음을 확인하였다. 다만, 4월 4일부터 4월 6일까지의 오차율은 정산금 산정 기준에서 벗어났으므로 중개 사업자는 해당일의 경우에 발전량 예측의 정확도를 높일 수 있는 알고리즘을 사용해야 한다.

## 5. 결 론

본 논문에서는 VPP의 예측 정산금을 높이기 위하여 재생에너지 발전량 데이터들의 변동성과 시계열 특성을 고려한 클러스터링 알고리즘을 제안한다. 제안한 방법은 산업통상자원부에서 제공하는 재생에너지 발전량을 이용하여 파라미터 데이터를 수집한 후, 시계열 데이터 특성의 고려한 MPTSC-VPP 알고리즘으로 클러스터링한다. MPTSC-VPP를 통하

Table 5. Result of validation with 10 days' data (nMAE).

	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Daily nMAE
Capacity	88,271	24,950	3,262	17,982	5,965	57,370	
2023-04-01	2.8%	7.3%	2.1%	7.6%	1.6%	2.0%	3.9%
2023-04-02	0.9%	7.0%	2.3%	6.5%	2.1%	1.9%	3.5%
2023-04-03	1.6%	4.2%	2.3%	8.9%	2.1%	2.7%	3.6%
2023-04-04	7.9%	11.7%	11.5%	11.0%	5.8%	8.2%	9.3%
2023-04-05	0.0%	24.6%	0.0%	33.9%	9.8%	1.7%	11.7%
2023-04-06	8.0%	23.6%	4.5%	17.7%	24.1%	6.7%	14.1%
2023-04-07	1.9%	11.5%	6.4%	17.5%	6.5%	3.8%	7.9%
2023-04-08	2.3%	4.6%	0.6%	4.9%	4.3%	3.3%	3.3%
2023-04-09	2.3%	3.0%	2.2%	5.8%	4.9%	3.0%	3.5%
2023-04-10	0.5%	4.4%	1.7%	5.7%	2.1%	2.5%	2.8%
Group nMAE	2.8%	10.2%	3.7%	11.9%	6.3%	3.6%	6.4%

여 클러스터링 된 각각의 클러스터들은 비시계열 기반의 클러스터링 알고리즘과 기존 알고리즘보다 더 높은 예측 정산금을 산출하면서 높은 수익성을 보장하였다. 제안한 방법의 검증에 위한 2개월 검증 데이터 중 10일 치 발전량 데이터를 예측한 결과 10일 중 7일이 산정 기준에 포함되므로, 약 70%의 정확도를 보여주었다. 하지만 특정 날짜에서는 최대 30%의 정산금 산정 기준을 넘었다. 따라서 향후 연구 과제는 정산금 산정 기준에서 벗어난 특정 날짜에 관해서 더 정확한 발전량을 예측할 수 있는 연구가 필요하다.

## REFERENCE

- [1] B. Marinescu, O. Gomis-Bellmunt, F. Dörfler, H. Schulte, and L. Sigrist, "Dynamic Virtual Power Plant: A New Concept for Grid Integration of Renewable Energy Sources," *IEEE Access*, Vol. 10, pp. 104980-104995, 2022.
- [2] J. MacQueen, "Some Methods for Classification and Analysis of Multivariate Observations," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281-298, 1967.
- [3] K. Pearson, "Mathematical Contributions to the Theory of Evolution—On a Form of Spurious Correlation Which May Arise When Indices are Used in the Measurement of Organs," *Proceedings of the Royal Society of London*, pp. 489-498, 1897.
- [4] E. Schubert, J. Sander, M. Ester, HP. Kriegel, and X. Xu, "DBSCAN Revisited: Why and How You Should (Still) Use DBSCAN," *Association for Computing Machinery Trans.*, Vol. 42, pp. 1-21, 2017.
- [5] S.J. Park, H.J. Kim, and S.H Kim "A Study on Solar Radiation Prediction Using Medium-Range Weather Forecasts," *Korean Journal of Applied Statistics*, Vol. 36, No. 1, pp. 49-62, 2023.
- [6] H.S. Ryu, *Risk-based Operation in Microgrid using Two-Stage Optimization Approach Considering Uncertainty*, Master in the Department of Energy Systems Engineering in the Graduate School of Chung-Ang University, 2021.
- [7] R.Y. Lee, J. Hur, J.G. Lee, S.Y. Jung, and H.Y. Shin, "A Study on Grouping of Domestic Wind Power Generation Data through Clustering Methods," *Proceeding of Summer Conference of KIEE*, pp. 481-482, 2020.
- [8] D.H. Kwon and S.S. Cho, "A K-Means Clustering Model on a Water Quality Monitoring System," *Korean Information Processing Society Annual Conference of KIPS*, pp. 1666-1669, 2010.





조 수 빈

2022년~현재 동아대학교 컴퓨터  
공학과 학사과정  
관심분야: 영상 처리, 딥러닝, 컴  
퓨터비전



서 정 일

1994년 경북대학교 전자공학과  
(공학사)  
1996년 경북대학교 대학원 전자  
공학과(공학석사)  
2005년 경북대학교 대학원 전자  
공학과(공학박사)

1998년~2000년 LG반도체 주임연구원  
2010년~2011년 영국 Southampton University, ISVR  
방문연구원  
2000년~2023년 한국전자통신연구원 실감미디어연구실장  
2023년~현재 동아대학교 컴퓨터AI공학부 부교수  
관심분야: 멀티미디어 부호화, 컴퓨터비전, 실감 영상 및  
음향, 멀티미디어 표준화



정 영 훈

2020년~현재 동아대학교 컴퓨터  
공학과 학사과정  
관심분야: 영상 처리, 딥러닝, 컴  
퓨터비전



천 세 진

2009년 8월 서울시립대 컴퓨터공  
학부(공학사)  
2018년 8월 연세대 컴퓨터공학과  
(공학박사)  
2021년 3월~현재 동아대학교 컴  
퓨터AI공학부 조교수

2018년 10월~2020년 8월 미국 표준기술연구원(NIST)  
ITL 근무  
관심분야: 지식그래프 표현 및 추론, 온톨로지 기반 모델  
링, 스트림 데이터 처리



조 성 령

2022년~현재 동아대학교 컴퓨터  
공학과 학사과정  
관심분야: 인공지능, 데이터 과학