

# 안정적인 태양광 발전소 운영을 위한 오토 인코더 기반 태양광 발전량 이상 탐지에 관한 연구

김보우, 황준화, 최정훈 서동준\*  
경북대학교

\*corresponding author

{kbw5913, jhhwang1907, cjh34544, \*dongjunsuh}@knu.ac.kr

## A Study on Anomaly detection in Solar PV Generation Based on Auto Encoder for stable Solar PV Power Plant Operation

Bowoo Kim, Junhwa Hwang, Jeonghoon Choi, Dongjun Suh\*  
Kyungpook National University

### 요 약

본 연구는 안정적이고 효율적인 태양광 발전소의 운영을 위하여 비지도 학습을 활용한 오토 인코더 기반 태양광 발전량 이상 탐지 모델을 제안한다. 제안된 모델은 정상 데이터와 비정상 데이터의 불균형 문제를 해결하고자 비지도 학습 방법을 채택하였으며, 별도의 부착형 센서 데이터가 아닌 인버터 단에서 제공되는 발전량 데이터를 활용하였다. Kaggle 에서 제공되는 인도의 태양광 발전소 발전량에 대해 실험을 진행하였으며, F1-score 를 통한 성능 측정 결과 0.9997 이라는 매우 우수한 성능을 확인하였다.

### I. 서론

기후변화 문제의 해결법 탐색과 탄소 배출의 노력의 일환으로 기존의 화력 발전에 주로 사용되는 화석 연료를 대체하기 위해 전 세계적으로 많은 신재생 에너지가 활용되고 있다. 태양광 발전은 다른 신재생 에너지 발전과 대비하여 환경 및 경제적 관점에서 매우 우수하다. 태양광 패널의 재료 및 설치 비용의 감소를 통해 전 세계적으로 대규모 태양광 발전소가 건설되고 있으며, 개별 건물의 옥상이나 벽면에 수많은 태양광 패널이 설치되고 있다. 하지만, 태양광 발전 패널의 보급률 대비 고장이나 시스템 오작동으로 인한 부분들에 대한 관리가 미흡한 실정이다. 따라서, 효율적이고 안정적인 유지보수와 운영을 위하여 태양광 발전 설비 이상의 초기 진단 및 예방이 필수적이다 [1].

태양광 발전을 포함한 기존의 다양한 산업분야에서 사용되는 이상 탐지의 경우 시스템의 고장 진단을 위하여 다양한 센서 정보를 사용한다 [2]. 하지만, 센서 정보를 사용할 경우 고장과 시스템 오류를 진단하기 위해 고장 사례 별로 데이터를 분류 및 라벨링과 같은 전처리를 진행한 후 수집해야 되므로 대용량의 데이터를 보유해야 한다. 또한, 대다수의 가정용 태양광 패널 발전 및 태양광 발전소는 비용 및 유지 관리의 문제로 인하여 발전 설비 내에 다수의 센서를 부착하지 않고 운영하는 경우가 다수 존재한다. 따라서, 본 연구에서는 기존에 설치된 태양광 패널과 태양광 발전소에 대해 비정상 데이터 취득 문제와 센서 부착 문제를 해결하고자 비지도 학습을 활용한 오토 인코더 (Auto encoder) 기반 태양광 발전량 이상 탐지 모델을 제안한다.

### II. 본론

기계학습 기법을 적용한 산업 현장에서의 이상 탐지 문제는 정상 데이터와 결함 데이터 수의 불균형이 가장 큰 문제로 작용한다. 예를 들어 제조업의 경우, 정상 데이터와 결함 데이터의 비율이 매우 극단적으로 정상 데이터가 약 1 억 개가 수집되는 동안 결함 데이터는 약 100 개 정도 수집된다. 이는 데이터의 비율뿐만 아니라, 비정상 데이터를 수집하는 시간 또한 매우 오랜 기간이 소요된다. 또한, 비정상 데이터를 개별적으로 고장 분류에 따른 라벨링과 같은 전처리를 별도로 진행한 후 데이터를 수집해야 한다. 따라서, 본 연구는 앞서 언급한 문제를 해결하기 위하여 인버터 단에서 수집되는 태양광 발전량 데이터 이상 탐지를 위하여 비지도 학습 기반의 오토 인코더 태양광 발전량 이상 탐지 모델을 제안한다.

제안된 모델에 사용된 데이터는 데이터 예측 모델 및 분석 대회를 주최하는 Kaggle 에서 제공되는 공개 데이터를 사용하였다 [3]. 이는 인도에 위치한 태양광 발전소의 15 분 단위로 측정된 34 일간의 데이터로써, 측정 일시, 발전소 ID, source key, DC 발전량, AC 발전량, 일일 생산량, 총 생산량, 외기 온도, 모듈 온도, 일사량을 제공한다. 전체 실험 기간은 2020 년 5 월 15 일부터 6 월 17 일이며 그림 1 을 통해 해당 기간의 AC 발전량과 일사량을 나타내었다. 그림 1 에서 파란 선으로 표시된 AC 발전량은 6 월 7 일을 나타내는 두번째 봉우리와 6 월 9 일의 봉우리에서 이상 데이터가 존재하는 것을 확인할 수 있다. 해당 기간의 데이터는 총 7 개의 타임 스텝에 해당하는 AC 발전량 데이터가 0 으로 기록되어 이상 데이터로 판단 후 실험을 진행하였다.

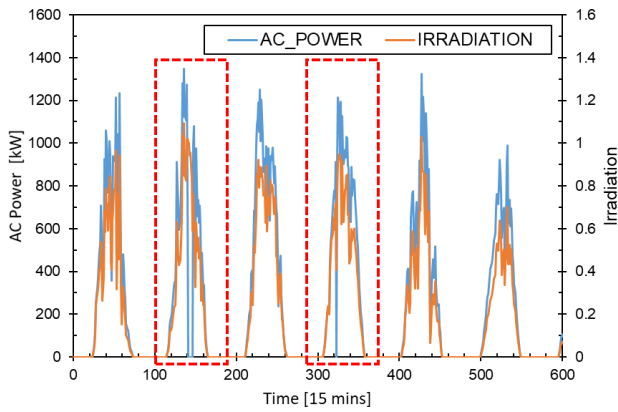


Figure 1 The AC power generation and irradiation data from 6 to 7 June, 2020.

		Ground Truth	
		True	False
Prediction results	True	True Positive (TP)	False Positive (FP)
	False	False Negative (FN)	True Negative (TN)

Figure 2 The factor of f1-score.

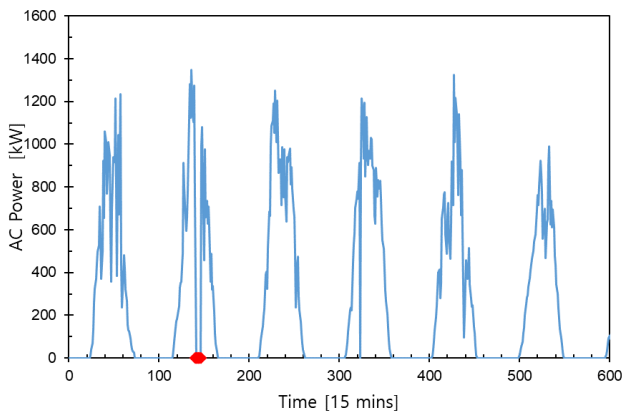


Figure 3 The Anomaly detection results of the proposed model.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{f1-score} = \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

제안된 모델은 오토 인코더 기법을 적용하였다. 오토 인코더는 입력 데이터의 가장 주요한 특징을 학습함으로써, 노이즈와 같은 정상 범주에서 벗어난 데이터를 제외한 주요 특징에 대해서만 학습을 진행한다 [4]. 만일, 학습된 모델에 이상 치가 입력될 경우, 이를 노이즈로 판단하여 해당 노이즈를 제외한 정상 데이터와 유사한 결과를 출력하여 입력과 출력의 차이가 극명하게 표현된다.

제안된 모델의 성능을 평가하기 위하여 f1-score를 통하여 측정하였다. F1-score는 정상 데이터와 비정상 데이터의 비율이 균일하지 않을 때 주로 사용되며 식 1~3

으로 표현된다. F1-score를 계산하기 위한 인자는 그림 2로 나타내었다.

실험 결과는 그림 3으로 나타내었으며, 검출한 이상 데이터는 붉은 점으로 표시하였다. 총 7개의 이상 데이터 중 6월 7일에 해당하는 6개의 데이터는 전부 이상 데이터로 검출하였지만, 6월 9일의 데이터는 검출하지 못하여 f1-score는 0.9997을 기록하였다. 이는 매우 높은 성능의 정확도임에는 분명하나, 전체 데이터 중 비정상 데이터의 비율이 매우 희소하므로 균등한 데이터의 분포가 이루어진다면 더욱 신뢰성 있는 결과를 도출할 수 있을 것이다.

### III. 결론

본 논문은 센서가 부착되지 않은 태양광 패널과 태양광 발전소에 대해 인버터 단에서 수집되는 태양광 발전량 데이터를 통해 이상 탐지를 진행하였다. 제안된 모델은 라벨링 되지 않은 이상 발전량 데이터를 검출하기 위하여 비지도 학습을 적용하였으며, 효과적인 이상 데이터 검출을 위하여 오토 인코더를 기반으로 설계하였다. 인도에 위치한 태양광 발전소에 대해 실험을 진행한 결과 f1-score: 0.9997이라는 매우 우수한 성능의 결과를 도출하였다.

향후 연구에서는 전체 실험 기간의 확장과 더불어 이상 데이터 증강을 통해 더욱 신뢰성 있는 이상 탐지 모델 개발 연구를 수행할 예정이다.

### ACKNOWLEDGMENT

본 연구는 정부(과학기술정보통신부, 교육부)의 재원으로 한국연구재단의 지원을 받아 수행되었음 (과제번호: 2021R1A5A8033165, 2021R1I1A3049503).

### 참고 문헌

- [1] A. Livera, et al, "Recent advances in failure diagnosis techniques based on performance data analysis for grid-connected photovoltaic systems", Renewable Energy, 133, pp. 126-143, Apr, 2019.
- [2] C. Tang, et al, "GRU-Based Interpretable Multivariate Time Series Anomaly Detection in Industrial Control System", Computers & Security, Jan, 2023.
- [3] Kaggle, <https://www.kaggle.com/datasets/anikannal/solar-power-generation-data>.
- [4] S. Zhai, et al, "Deep Structured Energy Based Models for Anomaly Detection", Proc. Int. Conf. Learn. Represent. (ICLR), vol. 48, pp. 1100-1109, 2016.