

태양광 발전 이상감지를 위한 아웃라이어 추정 방법에 대한 연구

A study on the outlier data estimation method for anomaly detection of photovoltaic system

서종관*, 이태일*, 이휘성*, 박점배*

Jong Kwan Seo*, Tae Il Lee*, Whee Sung Lee*, Jeom Bae Park*

Abstract

Photovoltaic (PV) has both intermittent and uncertainty in nature, so it is difficult to accurately predict. Thus anomaly detection technology is important to diagnose real time PV generation. This paper identifies a correlation between various parameters and classifies the PV data applying k-nearest neighbor and dynamic time warping. Results for the two classifications showed that an outlier detection by a fault of some facilities, and a temporary power loss by partial shading and overall shading occurring during the short period. Based on 100kW plant data, machine learning analysis and test results verified actual outliers and candidates of outlier.

요약

태양광 발전은 특성상 간헐성과 불확실성이 항상 존재하기 때문에 정확한 예측은 어려우며, 실시간 발전량 진단을 위한 이상감지 기술이 중요하다. 본 논문에서는 다양한 파라미터의 상관관계를 도출하고 최근접 이웃 알고리즘을 적용하여 정상데이터와 비정상데이터를 분류한다. 두 분류의 결과는 발전 시스템의 결함에 의한 아웃라이어와 구름 등에 의해 단기간 동안 발생하는 부분 음영 및 전체 음영의 일시적인 전력손실을 보여준다. 100kW 발전소 데이터를 대상으로 머신러닝 분석을 수행하여 테스트 결과를 산출하였으며 실제 이상치와 이상치 후보지를 검증하였다.

Key words : Photovoltaic, Anomaly Detection, K-nearest neighbor, Classification, Dynamic Time Warping

1. 서론

태양광 발전은 온실가스 및 기후변화의 글로벌 에너지 문제에 대응하여 탈탄소화(Decarbonized), 탈중앙화(Decentralized), 디지털화(Digitized)로 대

표되는 에너지 전환의 핵심 재생에너지이며, 2025년에는 세계 태양광 발전용량은 128GW에 이를 것으로 전망된다[1]. 국내는 재생에너지 3020 이행과 신재생에너지의무사용 등 정부 주도적으로 다양한 보급 확대 정책을 시행하고 있으며, 풍력에너지와

* KEI Dept, I-ON communications

★ Corresponding author

E-mail : marcus@i-on.net, Tel : +82-2-3430-1200

※ Acknowledgment

This work was supported by the Korea Institute of Energy Technology Evaluation and Planning(KETEP) grant funded by the Korea government(MOTIE) (20182410104920)

Manuscript received Mar. 18, 2020; revised Jun. 17, 2020; accepted Jun. 22, 2020.

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

더불어 그리드패리티 시장으로 전환을 위한 주요 에너지원으로 여겨진다. 또한, 국내 태양광 용량은 작년엔 이어 2GW를 넘어설 것으로 예상되며, 에너지저장장치와 접목하여 지속적인 성장세를 보일 것으로 전망된다[2] [3].

이러한 태양광 발전은 본질적으로 기상정보 등 외부 요인에 민감하기 때문에 정확한 예측이 어렵고 간헐적인 출력변동성을 내포하고 있어 전력망 안정성에 큰 영향을 야기시키게 된다. 따라서 발전 데이터의 불확실성을 해소하기 위하여 실시간으로 발전설비들을 모니터링하고 발전량의 변화에 대응하는 이상감지 기술이 최근에 재조명 받고 있다[4], [5]. 일반적으로 발전량 이상감지는 시계열 또는 시퀀스 분석의 일종으로 데이터 마이닝의 일부로 급격한 값의 변동이나 평소와 다른 패턴을 찾는 것을 의미한다[6].

본 논문에서는 100kW 설비용량의 태양광 발전 데이터를 기반으로 이상치 및 결측치를 정의하고 정규화하여 데이터전처리를 시행한다. 그리고 발전 데이터와 일사량, 패널온도 등의 환경데이터를 결합하여 데이터 간 상관관계 도출하여 주요 피쳐 추출한다. 예측 발전량과 실제 발전량과의 잔차 수치별로 정상데이터와 비정상데이터를 그룹화하고 최근접 이웃 알고리즘을 적용하여 비정상데이터를 재분류하여 아웃라이어와 후보군을 선별한다. 마지막으로 partial-shading, overall-shading, 데이터 유사성 측정을 통해 아웃라이어를 판별한다.

II. 본론

1. 태양광 발전 데이터 전처리

본 장에서는 100kW 설비용량의 태양광 발전소의 RTU(Remote Terminal Unit)을 통해 수집된 발전데이터와 환경데이터 간 상관관계를 도출하고 발전량 알고리즘 위한 피쳐 추출의 데이터 전처리를 기술한다. 대부분의 발전정보는 일정 주기로 RTU에서 수집되며, 본 논문에서는 1분 주기로 발전정보를 수집하였으며, '19.03.09~'19.06.21까지 90일간 총 138,388개의 데이터를 대상으로 정제, 통합, 축소, 변환, 정규화의 데이터 전처리를 수행하였다. 분석 수집 데이터의 출처에 따라 태양광 발전 데이터, 기상청 데이터, 환경센서 데이터로 분류하였으며, 그림 1은 전처리를 시행한 데이터 셋의 일부를

나타낸다.

- 태양광 발전 데이터 : 현재발전(인버터)전력, 누적발전량, 패널전력, 패널전류
- 기상청 데이터 : 기온, 강수량, 풍속, 풍향
- 환경센서 데이터 : 경사면일사량, 수평면일사량, 패널온도

Fig. 1. RTU dataset.

그림 1. RTU 데이터 셋

데이터 결측은 주요 피쳐 데이터에 해당하는 경사면일사량, 수평면일사량, 패널전력, 패널전류, 패널온도의 데이터가 누락된 경우와 발전데이터 날짜에 속하지 않는 환경센서와 기상데이터는 제외하였다. 이러한 누락 데이터는 대부분 인버터와 RTU의 순간적인 통신 이상으로 발생하며 데이터 신뢰성을 높이기 위해 발전시간(8시~21시) 구간 이외 데이터는 제외하였다.

발전데이터가 연속 10회 결측 시에는 해당 시각 전/후의 값을 활용하여 선형보간법의 방법으로 대체하였으며, 아래는 추정값 수식을 보여준다.

$$Y = Y_a + \frac{(X - X_a)(Y_b - Y_a)}{(X_b - X_a)} \tag{1}$$

표 1은 수집데이터의 결측치와 이상치의 데이터 전처리 유형을 나타낸다.

Table 1. Collection data preprocessing category.

표 1. 수집 데이터 전처리 항목

Case	data preprocessing category
1	- 경사면일사량, 수평면일사량, 패널전력, 패널전류, 패널온도 데이터가 NaN인 Case
2	- 경사면일사량, 수평면일사량, 패널온도 대비 패널전력이 작은 Case
3	- 경사면일사량, 수평면일사량, 패널온도 대비 패널전력이 큰 Case
4	- 시각 정보가 없는 Case
5	- 시각 정보가 중복되는 Case

2. 발전데이터 상관관계 및 분석

태양광 발전은 기상 정보에 의존적이기 때문에 일사량, 강수량, 패널온도 등의 정보가 발전량 이상 감지 판별을 위한 주요 데이터이다[7], [8]. 본 연구에서는 발전소 내 특정 패널에 장착된 환경센서를 통해 기상정보를 수집하였으며, 이는 외부 기상자료개방포털의 정보보다 높은 수준의 신뢰성을 가진다. 또한, 일반적으로 패널온도는 25℃에서 1℃의 온도가 상승함에 따라 약 0.5%의 효율이 감소하는 특성을 보인다[9].

그림 2는 데이터 전처리 과정을 통해 추출한 데이터의 피어슨 상관관계 히트맵을 나타내며, 패널 전력은 경사일사량 및 패널온도와 가장 높은 수치를 갖는 강한 양적 상관관계를 보이는 것을 확인하였다. 따라서 태양광 발전량은 일사량 및 패널온도가 주요 데이터이며, 해당 데이터 분석을 통해 발전량의 이상여부 확인이 가능하다.

pvPower	pvCurrent	todayPower	temperature	irrModule	irrHorizontal	tempModule	tempAir	windSpeed
1.00000	0.99883	0.14496	0.09241	0.99541	0.99090	0.81257	0.54531	
0.99883	1.00000	0.14871	0.08123	0.99595	0.99146	0.81863	0.55213	
0.14496	0.14871	1.00000	0.04628	0.81196	0.81582	0.23270	0.22668	
0.09241	0.08123	0.04628	1.00000	0.09131	0.05477	-0.98147	-1.18965	
0.99541	0.99595	0.15106	0.09131	1.00000	0.99460	0.81992	0.55244	
0.99090	0.99146	0.13852	0.05477	0.99460	1.00000	0.83724	0.58010	
0.81257	0.81863	0.23370	-0.98147	0.81992	0.83724	1.00000	0.91027	
0.54531	0.55213	0.22668	-1.18965	0.55244	0.58010	0.91027	1.00000	

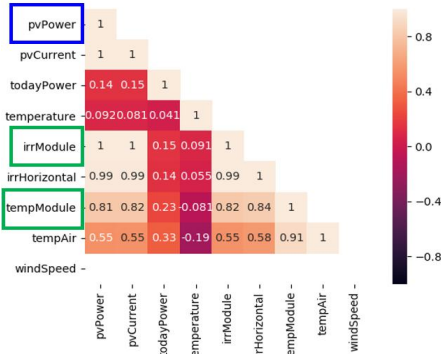


Fig. 2. Pearson correlation of PV data.
그림 2. 발전데이터 피어슨 상관관계

그림 3은 월별 일사량과 패널전력 관계를 나타내며, 일사량이 증가하면 패널전력이 증가하는 선형 증가의 추이를 보이며, 두 변수는 양의 직선형 상관관계임을 확인할 수 있다. 6월은 3월보다 일사량이 높고 긴 일조시간을 보이므로 패널전력은 양의 직선성 상관관계와 유사한 선형분포를 나타내었다. 또한, 일정 수치 이상의 일사량 대비 발전량이 현저하게 낮은 경우는 이상치 데이터로 구분하여 이상감지 현상이 발생한 것으로 판별 가능하며, 그림의 노란색 영역은 일사량 400 m²/W 이상인 구간에서 패널전력이 0으로 이상감지 후보군이 된다.

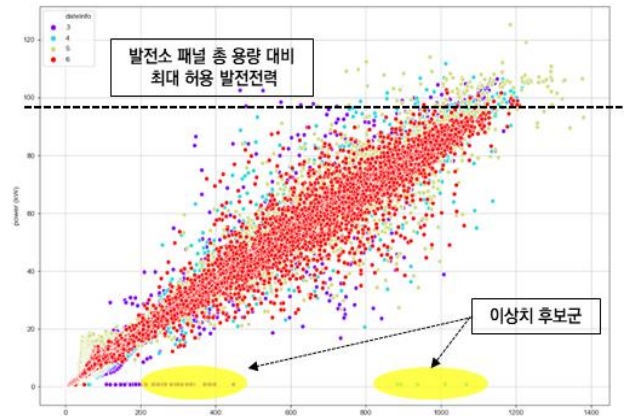


Fig. 3. Irradiance and panel power distribution plot monthly.
그림 3. 월별 일사량과 패널전력 분포

3. 발전량 이상감지 알고리즘 및 KNN 분류

발전 전력은 일정 수준 이상의 패널온도가 증가함에 따라 효율이 감소하고, 일사량에 정비례한 출력을 보이므로 식(2)와 같이 태양광 발전 알고리즘을 산정하였다. P는 패널전력, Q는 일사량, η_{T_{ref}}는 패널 변환효율, α는 온도계수, t는 표면온도를 나타낸다.

$$P = Q \cdot \eta_{T_{ref}} \cdot [1 - \alpha(t - 25)] \quad (2)$$

실제 발전량과 위 알고리즘에 의해 계산된 발전량을 비교하여 오차를 산정하고 오차 값의 수준에 따라 패널전력, 예측전력, 일사량 변화 추이를 산출할 수 있다. 그림 4는 10kW 초과 오차 구간에서의 월별 데이터 변화를 나타낸다.

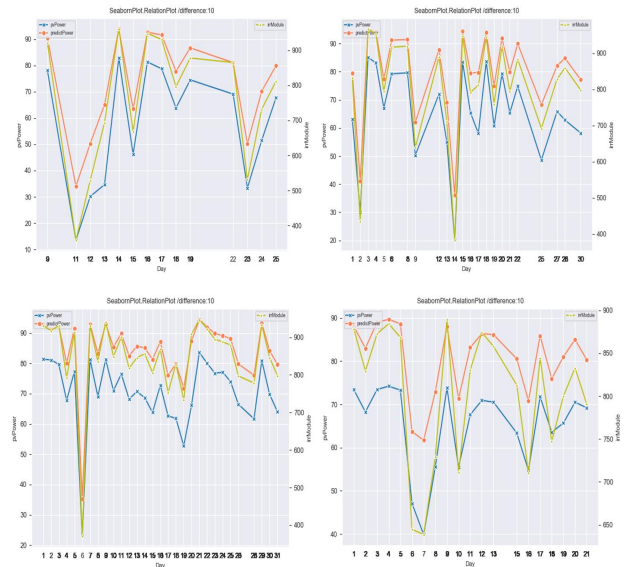


Fig. 4. Variation of module power, predict power, irradiance.
그림 4. 패널전력, 예측전력, 일사량 변화 추이

패널전력 변동 추이일 사량이 변화하는 시점에 따라 대체적으로 일치한 형태의 그래프 유형을 보인다. 일사량과 예측전력이 증가하지만 패널전력이 감소하는 경우는 총 4회(5월 16, 17일/6월 17일, 18일)이며, 10분 이내의 단기간 동안 발생한 것으로 보아서는 외부 환경에 의한 일시적인 발전 손실로 추정 가능하다. 그림 6은 5월과 6월 기간 동안 10kW 초과 구간의 패널전력, 일사량, 예측전력 데이터 변화 추이 그래프이며, 빨간색 부분은 일사량과 패널전력이 반비례한 일시를 나타낸다.

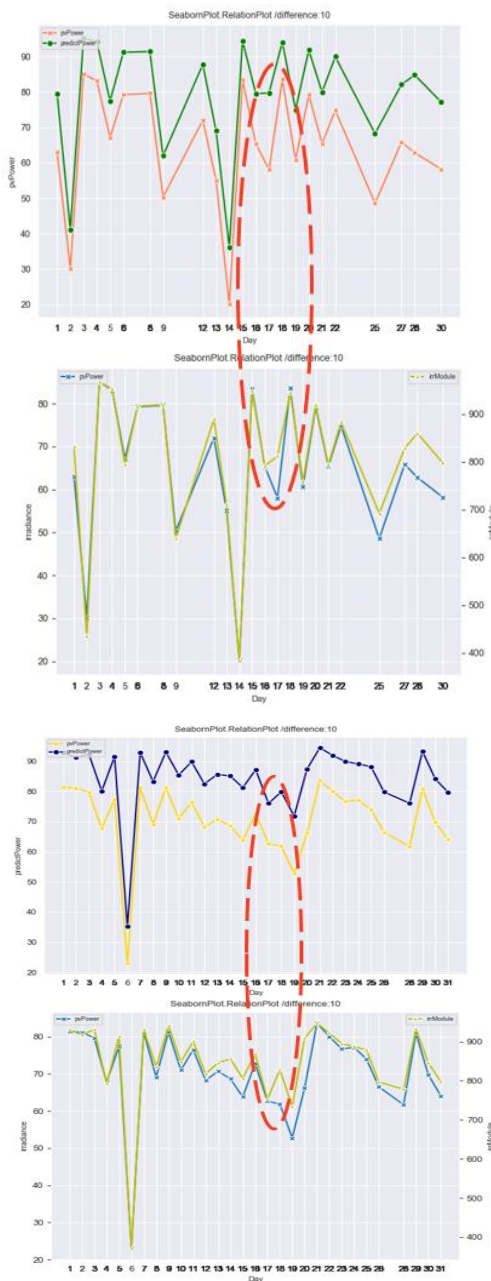


Fig. 5. Data variation transition in May and June. 그림 5. 5월과 6월 데이터 변화 추이

발전 오차값을 10kW 초과로 산정한 근거는 100kW 발전 설비용량의 10%는 패널수명, 인버터 변환효율 등의 손실을 고려한 수치이며, 발전소의 설비 특성에 따라 상이하지만 30% 이상 손실이 발생하는 경우도 있다. 그림 6은 10kW로 구분한 오차 구간 별 통계 정보를 보여준다. 10kW 이하 구간은 앞서 설명한대로 태양광 발전의 설비효율 손실이며, 20~30kW 구간이 주요 이상감지 구간으로 볼 수 있다. 또한, 40kW 초과 구간은 57개로 오차가 거의 발생하지 않았으며 구름 등에 의한 일시적인 발전저하로 판단가능하다.

index	count	min	max	sum	mean
(0, 10]	40469	0.10000	10.00000	127955.40000	3.16181
(10, 20]	9992	10.10000	20.00000	121245.70000	13.35011
(20, 30]	431	20.10000	30.00000	10187.80000	23.63759
(30, 40]	139	30.10000	40.00000	4772.50000	34.33453
(40, 50]	38	40.10000	49.70000	1666.70000	43.86053
(50, 60]	12	50.10000	59.40000	642.90000	53.57500
(60, 70]	2	60.20000	61.50000	121.70000	60.85000
(70, 80]	0	nan	nan	0.00000	nan
(80, 90]	2	86.50000	88.20000	174.70000	87.35000
(90, 100]	3	92.80000	95.00000	282.80000	94.26667

Fig. 6. Count of power difference. 그림 6. 발전량 오차 구간별 통계 정보

20~30kW 구간에 해당하는 데이터를 이상데이터 후보군으로 분류하고, 실제 발전량 데이터를 대상으로 K-Nearest Neighbor 모델을 적용하여 정상데이터와 이상데이터 분류하여 이상데이터 후보군과 일치여부를 확인하였다. train data와 test data는 3:1 비율이며, k-fold 교차검증을 적용하여 오버피팅을 방지하였다. 그림 7은 최적 k값 산정 그래프로 k=3 이후부터 일정함을 보여준다.

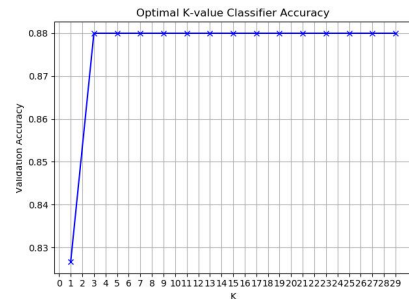


Fig. 7. Detection of optimal K-value plot. 그림 7. 최적의 K값 선정 그래프

그림 8은 훈련데이터, 테스트데이터, 모델의 정확성 결과이며, 제시한 이상감지 알고리즘은 88% 정도의 확률로 이상데이터를 분류할 수 있다.


```
KNeighborsClassifier : train score:0.94286 / test score:0.88000
KNeighborsClassifier : model accuracy:0.88000 / cross validation:0.92000
Mean Absolute Error: 0.12
Mean Squared Error: 0.12
Root Mean Squared Error: 0.34641016151377546
```

Fig. 8. Verification of data classification based KNN.
그림 8. KNN 기반 정상데이터 및 이상데이터 검증

dateinfo	DTW
2019-05-17 18:00:00	16.66667
2019-03-12 08:20:00	12.89003
2019-03-11 15:30:00	11.62130
2019-03-17 10:00:00	11.49058
2019-04-22 11:00:00	11.46237
2019-06-02 10:00:00	10.61360
2019-03-11 16:00:00	9.34014
2019-05-08 10:00:00	8.07187
2019-06-17 11:00:00	7.88979

Fig. 9. DTW distance result.
그림 9. DTW 유사도 결과

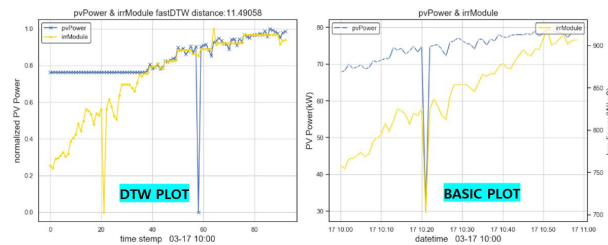


Fig. 10. DTW distance result plot.
그림 10. DTW 거리 측정 결과 그래프

최종적으로 이상데이터를 대상으로 경사면일사량과 패널전력의 동적시간왜곡(DTW, Dynamic Time Warping) 거리 알고리즘을 적용하여 패널전력과 일사량 간 유사도 측정을 통해 이상데이터의 아웃라이어를 판별하였다.

그림 9는 DTW를 적용한 거리 결과이며, 그림 10은 DTW가 가장 큰 시점의 일사량과 패널전력 추이를 비교한 그래프를 나타낸다. DTW가 클수록 데이터의 유사성이 작다는 것을 의미하기 때문에 서로 다른 패턴 형태를 보여주며 이는 두 데이터는 직선형 선형관계가 성립하지 않는 것으로 판단 가능하다. 또한, 평균 오차보다 큰 오차가 발생한 것으로 확인하였다.

III. 결론

본 논문에서는 태양광 발전 이상감지를 위해 데이터 전처리를 통해 주요 피처를 추출하고 발전 오차를 그룹화하여 이상치와 이상치 후보군을 구분하였다. 또한 KNN 모델을 적용하여 이상데이터 분류하였으며, 88%의 정확성을 보였다. 또한, 이상

데이터를 대상으로 DTW 유사성을 적용하여 최종적으로 아웃라이어를 판별하였다.

향후 KNN 모델 이외 랜덤포레스트, 서포트벡터머신, 앙상블 등의 머신러닝 모델을 적용하여 아웃라이어의 분류 정확도를 향상시키는 연구를 추진하고자 한다.

References

- [1] “Global Market Outlook for solar power 2019-2023,” *The solar power europe*, 2019.
- [2] “Renewable energy 3020 plan (RE3020),” *MOTIE of Korea*, 2017.
- [3] “2019 Quarterly Report,” The Export-import Bank of Korea, Vol.9, 2019.
- [4] Hye-won Ahn, Yong-sung Cho, “Study on Certification Policy for Stabilization of Photovoltaic Industry - A Case Study of PV Power Plant O&M,” *Journal of Climate Change Research*, Vol.6 No.2, 2015.
- [5] Tae-hoon Jeon, Young-Ju Lee, Myeong-Hwan Kim, Jeong-Cheol Lee, Jun-Ho Park, Gwon-Jong Yu, “Factors and solutions of decline of power generation efficiency during O&M of PV systems,” *Journal of Korean Solar Energy Society*, Vol.5, No.1, 2015.
- [6] Jeong-won Kim, “A Solar Power Prediction Scheme Based on Machine Learning Algorithm from Weather Forecasts,” *Journal of KIIT* Vol.17 No.9, 2019.
- [7] Kyung-Eun Park, Gi-Hwan Kang, Hyun-Il Kim, Gwon-Jong Yu, Jun-Tae Kim, “Study on variation of electrical characteristics of vertical PV module according to the change of irradiance and temperature,” *Journal of the Korean Solar Energy Society autumn conference*, 2008.
- [8] Hee-Jung Youn, Seong-Wan Jang, Eun-Heui Lee, “Temperature Monitoring of Vegetation Models for the Extensive Green Roof,” *Journal of the Korea Institute of Ecological Architecture and Environment*, Vol.13, No.5, 2013.
DOI: 10.12813/kieae.2013.13.5.089
- [9] W. Durish, B. Bitnar, Jean-C Mayor, H. Kiess,

K. Lam, J. Close, "Efficiency model for photovoltaic modeules and demonstration of its application to energy yield estimation," *Solar Energy Materials and Solar Cells*, Vol.91, pp.79-84, 2007.

DOI: 10.1016/j.solmat.2006.05.011

BIOGRAPHY

Jong-kwan Seo (Member)



2007 : BS degree in Information Communicating Engineering, Inha University.
2012 : MS degree in Electonic Engineering, Kwangwoon University.
2008~2018 : Senior Researcher, Korea Electrotechnology Research Institute(KERI)

2019~Present : Senior Researcher, I-ON Communications.

Tae-il Lee (Member)



2017 : BS degree in Energy Information Technology, Gachon University.
2019 : MS degree in Electrical Engineering, Gachon University.
2019~Present : Project Assistant, I-ON Communications

Whee-sung Lee (Member)



2010 : BS degree in Management Information Systems, Konkuk University.
2014~Present : Team Leader of Energy Business Team, I-ON Communications.

Jeom-bae Park (Member)



2009 : BS degree in Computer Science, Seoul Digital University.
2016 : MS degree in IT Convergence, Soongsil University.
2020 : PhD under course in IT Convergence, Soongsil University.
2007~Present : Team Leader of Energy ICT, I-ON Communications.