

인공 신경망 회귀 모델을 활용한 인버터 기반 태양광 발전량 예측 알고리즘

박건하* · 임수창* · 김종찬*

Inverter-Based Solar Power Prediction Algorithm Using Artificial Neural Network Regression Model

Gun-Ha Park* · Su-Chang Lim* · Jong-Chan Kim**

요약

본 논문은 전라남도에서 측정된 태양광 발전 데이터를 기반으로 발전량 예측값을 도출하기 위한 연구이다. 발전량 측정을 위해 인버터에서 직류, 교류, 환경데이터와 같은 다변량 변수를 측정하였고, 측정값의 안정성과 신뢰성 확보를 위한 전처리 작업을 수행하였다. 상관관계 분석은 부분자기상관함수(PACF: Partial Autocorrelation Function)을 활용하여 시계열 데이터에서 발전량과 상관성이 높은 데이터만을 예측을 위해 사용하였다. 태양광 발전량 예측을 위해 딥러닝 모델을 이용하여 발전량을 측정했고, 예측 정확도를 높이기 위해 각 다변량 변수의 상관관계 분석 결과를 이용하였다. 정제된 데이터를 활용한 학습은 기존 데이터를 그대로 사용했을 때 보다 안정되었고, 상관관계 분석 결과를 반영하여 다변량 변수 중 상관성이 높은 변수만을 활용하여 태양광 발전량 예측 알고리즘을 개선하였다.

ABSTRACT

This paper is a study to derive the predicted value of power generation based on the photovoltaic power generation data measured in Jeollanam-do, South Korea. Multivariate variables such as direct current, alternating current, and environmental data were measured in the inverter to measure the amount of power generation, and pre-processing was performed to ensure the stability and reliability of the measured values. Correlation analysis used only data with high correlation with power generation in time series data for prediction using partial autocorrelation function (PACF). Deep learning models were used to measure the amount of power generation to predict the amount of photovoltaic power generation, and the results of correlation analysis of each multivariate variable were used to increase the prediction accuracy. Learning using refined data was more stable than when existing data were used as it was, and the solar power generation prediction algorithm was improved by using only highly correlated variables among multivariate variables by reflecting the correlation analysis results.

키워드

ANN, ARIMA, Deep Learning, Inverter, Regression Model, Solar Power
태양광 발전, 인버터, 회귀 모델, 아리마 모델, 딥러닝, ANN 모델

* 순천대학교 컴퓨터공학과 연구원
(pgh6151@naver.com, suchangLim@sncnu.ac.kr)

** 교신저자 : 순천대학교 컴퓨터공학과

• 접수일 : 2024. 02. 03

• 수정완료일 : 2024. 03. 08

• 게재확정일 : 2024. 04. 12

• Received : Feb. 03, 2024, Revised : Mar. 08, 2024, Accepted : Feb. 12, 2024

• Corresponding Author : Jong-Chan Kim

Dept. Suncheon National University,

Email : seaghost@sncnu.ac.kr

I. 서 론

화석 연료 에너지는 오랜 기간 동안 주요 에너지원으로 사용되었다. 화석 연료를 통한 에너지는 연료를 태워 에너지를 조달하는 방식으로 에너지를 발생시키는데 이 과정에서 각종 오염물질, 초미세먼지, 오존 등이 배출된다. 화석 연료의 가열로 인한 화학 배출물들은 자연환경에 치명적이며 가용 가능한 화석 연료 또한 유한하다는 단점이 있다[1-2].

태양광 발전량 데이터는 시계열 데이터(Time Series Data)로 구분할 수 있으며, 발전량 데이터의 예측은 기상데이터, 일사량, 운량등 다변량 환경 변수에 의해 결정된다. 발전량의 측정을 위한 핵심 설비는 인버터와 태양광 모듈이 있다. 태양광 에너지를 전달하는 모듈은 태양광 에너지를 전기 에너지로 변환하는 인버터로 전달한다. 인버터는 태양광 모듈로부터 생산된 직류를 교류로 변환하는 장치로, 태양광 모듈과 인버터는 가동 시간이 흐름에 따라 고장, 열화 등을 발생시켜 성능 저하 및 발전량 감소를 초래한다[3]. 태양광 발전 설비의 열화 또는 고장은 태양광 발전소와 개인의 전력사용에 있어 치명적인 영향을 미칠 수 있어, 태양광 예측 시스템을 활용한 고장 및 설비 열화의 조기 발견을 위해서라도 예측 알고리즘의 정확도는 중요하다[4-6].

본 논문에서는 전라남도에서 측정된 태양광 발전량 데이터를 기반으로 태양광 발전소의 발전 효율을 최상으로 유지하기 위한 발전량 예측 알고리즘을 제안하며 신뢰성 높은 예측을 위해 통계적 회귀 기법 및 딥러닝 기법을 통합적으로 활용한다. 통계적인 방법은 통계적 회귀 모델인 ARIMA(Auto Regressive Integrated Moving Average) 모델을 활용하여 데이터에 대한 전처리를 수행하고, 발전량 예측 모델은 ANN(Artificial Neural Network) 모델을 사용한다. ANN 모델은 복잡한 패턴 및 비선형적인 관계를 학습하는 데 강점이 있고, 다변량 변수 간의 복잡한 상호작용을 효과적으로 모델링할 수 있다. 심층 신경망을 통해 발전량에 대한 정확하고 신뢰성 있는 예측 결과를 도출하며, 모델의 학습 및 성능 향상을 위한 최적화 기법도 함께 적용한다.

예측 정확도 향상을 위하여 다변량 변수 간의 상관관계를 분석하고 발전량 예측 모델의 특성을 더욱 강화하기 위해 부분자기상관함수(PACF: Partial

Autocorrelation Function)를 활용하여 상관관계 분석을 수행한다. PACF를 활용한 예측 모델은 시간적 패턴을 식별하여 특정 시간대의 발전량이 다음 시간대에 미치는 영향을 해석하여 예측 성능을 개선한다.

본 논문의 구성은 다음과 같다. 2장에서 시계열 데이터를 활용한 예측 알고리즘에 관련된 연구 내용을 기술했다. 3장에서는 제안한 알고리즘에 대해 서술하며 4장에서 전라남도에서 측정된 발전 데이터를 기반으로 발전량 예측 알고리즘을 평가한다. 마지막 5장에서는 결론을 맺으며 마무리한다.

II. 관련연구

2.1 시계열 데이터 예측 알고리즘

시계열 데이터에 대한 예측 알고리즘은 통계적인 방법[7-8]과 딥러닝을 사용한 방법으로 분류된다. 태양광 발전량 예측은 날씨, 계절, 지리, 일사시간, 패널의 방향 등 환경 변수에 의존하며 각 요소에 따라 불규칙하여 동일한 예측을 제공하지 못할 수 있다.

시계열 데이터에 대한 예측은 신경망 또는 회귀분석을 활용한 방식으로 제안되었다. 예측 모델의 한 종류인 ARIMA(Auto Regressive Integrated Moving Average) 모델[9]은 통계적인 특성을 가진 모델로 자기 회귀(AR:Auto Regressive) 및 이동 평균(MA:Moving Average)을 결합하는 방식을 사용한다. 모델의 차수는 ACF(Autocorrelation Function) 및 PACF(Partial Autocorrelation Function)를 사용하여 결정된다. 이를 통해서 시계열 데이터의 추세, 계절성 및 노이즈를 모델링하고 예측하는 데 사용된다. 통계적인 방법과 달리 RNN[10], LSTM[11], ANN[12-13]등의 연구에서는 다변량 변수로 구성된 시계열 데이터 처리에 인공신경망을 활용하였다. 다변량 변수와 같은 복잡한 패턴을 가진 데이터에서도 인공신경망을 활용한 예측은 강한 성과를 보였다.

III. 태양광 발전량 데이터를 활용한 발전량 예측 알고리즘 제안

3.1 태양광 발전량 데이터 수집

전라남도에서 측정된 태양광 발전량 데이터를 기반으로 예측 알고리즘을 구성하였으며 태양광 모듈과

연결된 데이터 수집장치를 통해 측정된다.

인버터는 태양광 에너지를 전기 에너지로 변환하는 핵심 장치로 실질적인 전력량을 측정한다. 태양광으로 파생되는 전력을 예측하기 위해서는 환경적인 요인이 고려되어야 하며, 환경적인 요인은 일사량, 운량, 날씨 등의 다변량 변수들의 상관관계 파악이 중요하다. 본 논문에서는 다변량 변수의 상관관계를 파악하기 위한 최초 단계로 태양광 발전량, 기상조건 등의 데이터를 수집하였다.

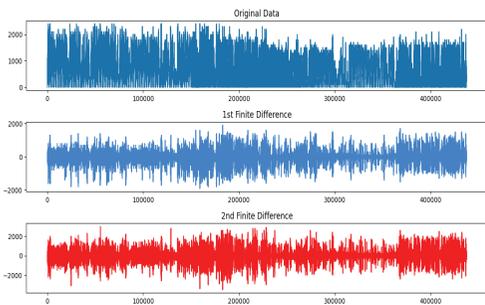
3.2 태양광 발전량 데이터 전처리

태양광 발전량 데이터에서의 전처리는 모델의 성능에 많은 영향을 미치는 핵심 단계라고 할 수 있다. 특히 발전량 예측 시 결측치 또는 아웃라이어 값은 예측 성능에 치명적인 영향을 미치며, 데이터의 안정성과 신뢰성을 확보하기 위해 필수적인 과정이다.

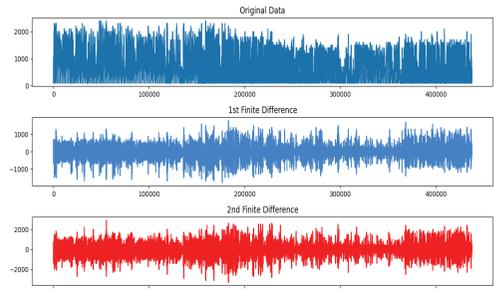
본 논문에서는 태양광 발전량 데이터는 시계열 데이터로 간주하였고, 시계열 데이터 전처리를 위한 통계적인 방법 중 좋은 결과를 보인 ARIMA 모델을 사용하여 전처리를 수행하였다.

ARIMA를 활용한 전처리 작업의 목표는 데이터의 안정성과 신뢰성을 확보하기 위한 작업으로 결측치와 이상치가 발생하지 않도록 하는 작업이 중요하다. 결측치를 제거하기 위해서는 시계열 데이터에서 결측치가 존재하는 부분을 탐색하고, 발견된 결측치 값을 ARIMA 모델을 활용한 예측값으로 대체한다.

결측치 제거 이후 태양광 발전량 데이터가 정상성을 띠지 않을 때를 대비하고, 이상값이 발생하는 경우를 최소화하기 위해 차분(Differencing) 법을 사용하며, 필요에 따라 2차까지 차분을 진행한다.



(a) DCP 1차, 2차 차분 결과



(b) ACP 1차, 2차 차분 결과

그림 1. 태양광 발전 모듈인 인버터에서 측정된 직류, 교류 발전량 데이터

Fig. 1 DC and AC power generation data measured by the solar power module inverter

그림 1은 태양광 발전 모듈인 인버터에서 측정된 직류, 교류 발전량 데이터를 기반으로 데이터 전처리를 위한 1, 2차 차분을 진행한 결과를 보인다. (a)의 첫 번째 행은 직류 발전량, 두 번째 행은 직류 발전량의 1차 차분, 세 번째는 2차 차분이며, (b)의 첫 번째 행은 교류 발전량, 두 번째는 교류 발전량의 1차 차분, 세 번째는 2차 차분 진행 결과를 보이며 전처리를 거치면서 데이터의 안정성을 찾아가고 있는 모습을 보인다.

태양광 발전량 예측을 위해서는 환경 변수들에 대한 상관관계 분석이 필수적이며 본 논문에서의 상관관계 분석은 전처리 된 데이터를 기반으로 PACF를 분석하는 과정이 필요하다. 예측하고자 하는 교류 발전량과 상관계수가 높은 값을 입력값으로 선정하였고, 상관 분석 결과 인버터의 데이터 중 교류 발전량과 직류 발전량이 높은 상관계수를 가짐을 보였다.

직류 발전량이 변환된 실측값이 교류 발전량이므로 이 둘의 관계는 인버터의 효율과 직접적인 연관성을 가진다고 볼 수 있다. 인버터의 설치 위치 및 기상 변수를 통제된 환경 센서 데이터에서 상관계수가 가장 높은 값은 일사량으로 측정되었으며 최종 결과로 일사량보다도 직류 발전량이 교류 발전량과 더 강한 상관관계를 가지고 높은 밀도와 높은 상관계수를 보였다.

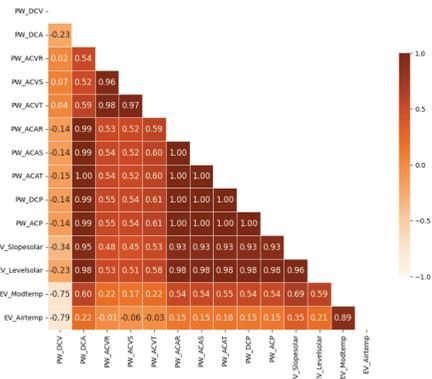


그림 2. 인버터에서 추출된 데이터의 상관계수 매트릭스와 상관관계 분석 결과
 Fig. 2 Correlation analysis results with correlation coefficient matrix of data extracted from inverter

그림 2는 태양광 발전 인버터에서 추출된 데이터를 기반으로 통계적인 방식으로 분석한 뒤 데이터 간의 상관관계를 분석한 결과이다. -1에서 1까지의 상관성 범위를 보여주며 DCP값은 직류와의 상관성이 높은 값을 가져 밀접한 관계를 가진다고 볼 수 있다. 교류 발전량이 높게 나온 경우 직류 발전량 또한 높기 때문에 상관계수는 높은 수치를 보인다. 상관관계 분석 결과를 바탕으로 상관성이 높은 변수를 활용하여 예측한다.

표 1. 태양광 발전량 데이터 수집을 위한 환경 데이터 스키마

Table 1. Environmental Data Schema for Solar Power Data Collection

Column	Data type	Description
EV_Date	date	DATE
EV_Time	time	TIME
EV_Slopesolar	float	Horizontal Solar Radiation
EV_LevelSolar	float	Vertical Solar Radiation

표 1은 태양광 발전량 데이터 수집을 위한 데이터 스키마이다. 데이터 스키마는 상관성 분석으로 높은 상관성을 보인 다변량 변수를 활용하여 설정하였다. EV_DATE와 EV_TIME은 각각 데이터가 수집된 시기를 의미하며, EV_Slopesolar, EV_LevelSolar는 태양광 복사열 데이터를 의미한다.

3.3 ANN 모델을 활용한 태양광 발전량 예측

ANN 모델은 대표적인 딥러닝 모델로 가장 많이

사용되는 모델 중의 하나인 모델이다. 기본적인 골격은 입력층, 은닉층, 출력층으로 이루어진 네트워크가 가중치를 학습해 나가며 최적의 예측치를 도출하는 모습을 가진다. 뉴런이라고 불리는 노드들이 서로가 유기적으로 연결된 구조를 가지며 각 노드에서 가중치를 곱하고 활성화 함수에 통과시켜 학습한다. ANN 모델의 관계는 은닉층의 뉴런 출력과 출력층의 입력 사이에 두 번째로 발생하게 되며 은닉층의 뉴런 개수 만큼 출력 레이어의 각 뉴런에 대한 입력이 주어지게 된다.

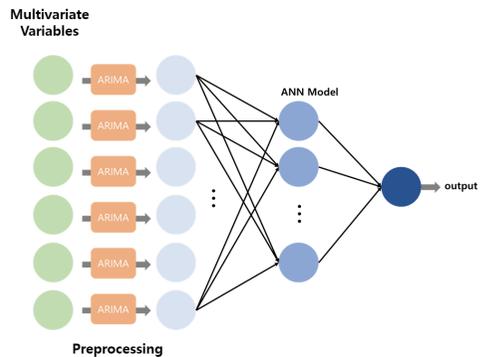


그림 3. 태양광 발전량 예측을 위한 파이프라인
 Fig. 3 Pipeline for Solar Power Prediction

그림 3은 태양광 발전량 예측을 위한 태양광 발전량 예측 모델의 전체 파이프라인을 나타낸다. 노드의 개수는 상관관계 분석을 활용하여 높은 상관성을 보인 6개의 노드를 활용한 예측을 활용하였고, 각 노드의 변수들은 ARIMA 모델을 통과하며 전처리 과정을 거치게 된다. 전처리 과정이 끝난 데이터들은 ANN 모델의 뉴런이 되어 가중치 값을 전달하고 학습하여 결과를 도출한다. ANN 모델은 입력층, 은닉층, 출력층 3가지 레이어로 구성이 되어있으며 각층에서 노드들은 가중치를 학습해나간다.

IV. 성능평가

본 논문에서는 ARIMA 모델을 활용한 전처리를 통해 데이터의 안정성을 보장하고, 딥러닝 모델을 활용하여 예측 정확도를 높이는 방법으로 태양광 발전량을 예측했다. 예측 모델의 성능을 평가하기 위해 예

측값과 실제 값 간의 차이를 제공하고 평균을 구하는 평균 제곱 오차(MSE: Mean Squared Error)와 절댓값의 평균을 구하는 평균 절대 오차(MAE: Mean Squared Error) 그리고 실제 값 대비 예측값의 오차를 백분율로 나타낸 평균 백분율 오차(MAPE: Mean Absolute Percentage Error)를 활용하여 시계열 데이터의 예측 성능을 평가한다. MSE와 MAE 값은 성능 차이 표현을 위해 ARIMA를 활용한 전처리 전후 값을 비교하여 모델의 성능을 비교한다.

학습 진행시 실측값을 기반으로 학습 과정에서는 Validation Loss 값이 일정하지 못한 모습을 보였으며 일정하지 않은 Validation Loss는 부정확한 예측으로 이어진다.

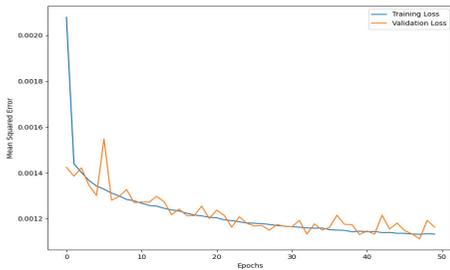


그림 4. 전처리 후 데이터를 활용한 학습 결과
Fig. 4 Learning results using data after preprocessing

그림 4는 ARIMA를 통한 전처리 과정을 거친 후 학습을 진행한 결과를 보인다. 전처리 이전의 Validation Loss 값과 비교해 보았을 때 추세가 안정되었으며, Training Loss와 추세가 유사해지는 모습을 볼 수 있다.

표 2. 태양광 발전량 데이터의 정량적 평가
Table 2. Quantitative evaluation of solar power data

Column	Before	After
ACP_MSE	16903.33(w)	9238.06(w)
ACP_MAE	69.25(w)	52.71(w)
DCP_MSE	16933.24(w)	10457.67(w)
DCP_MAE	63.29(w)	52.71(w)

표 2는 ARIMA를 통한 전처리 과정을 거친 전후 예측값에 대한 평균 제곱 오차와 평균 절대 오차 값을 보인다. 전처리 이후 오차값이 감소하였고 오차값 감소는 예측 정확도 상승으로 이어진다.

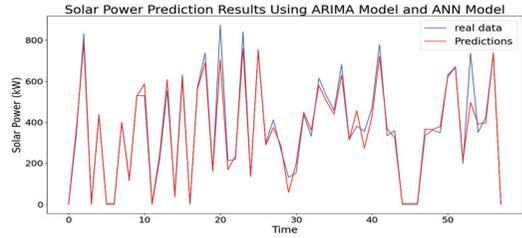


그림 5. 태양광 발전량 예측 결과
Fig. 5 Solar Power Prediction Results

그림 5는 발전량 데이터의 예측 결과의 일부분들 시각화한 자료이다. 실제 값과 유사한 추세와 높은 예측 결과를 보였으며 MAPE 값을 활용한 예측 정확도는 92%를 보였다.

V. 결론 및 향후연구

본 논문에서는 전라남도에서 측정된 태양광 발전량 데이터를 기반으로 인공신경망 회귀 알고리즘을 활용하여 발전량을 예측하는 방법에 관해 연구하였다. 태양광 발전량 데이터를 추출하기 위해서 데이터 수집 장치에서 측정 가능한 다변량 변수를 추출하였고, 데이터의 안정성을 얻기 위하여 ARIMA를 활용한 데이터 전처리 작업을 진행하였다. 이후 태양과 발전량 예측을 위해 딥러닝 모델인 ANN 모델을 활용하여 데이터에 대한 예측을 진행하였다. 데이터를 정제하여 학습 결과의 Validation Loss 값을 안정시켰으며, 평균 절대 백분율 오차를 활용한 정확도는 92%로 평가되었다. 향후 연구를 통해 발전량 예측 정확도를 더욱 올리기 위해 기후 데이터를 통합한 앙상블 모델을 활용한 연구 및 실시간 예측 시스템으로 실시간 발전량과 고장 예측 모니터링에 관한 연구를 진행한다.

References

- [1] G. Bang, "Problems in Hydrogen Economy and the Need for the Development of New Energy Source," *Journal of the Korean Psychiatric Association*, vol. 13, no. 1, pp. 73-80, 2009.
- [2] J. Song, Y. Jeong, and S. Lee, "Analysis of

prediction model for solar power generatio," *Journal of digital convergence*, vol. 12, no. 3, pp. 243-248, 2014.

[3] J. Hong, J. Park, and Y. Kim, "Fault Prediction of Photovoltaic Monitoring System based on Power Generation Prediction Model," *Journal of Platform Technology*, vol. 6, no. 2, pp. 19-25, 2018.

[4] L. Gutierrez, J. Patino, and E. G. Eduardo, "A Comparison of the Performance of Supervised Learning Algorithms for Solar Power Prediction," *Energies*, vol. 14, no. 15, pp. 4424, 2021.

[5] M. Han, J. Woo, and J. Lee, "Power Generation Change According to Angle Control of Solar Power Plant Panel," *Journal of the KIECS*, vol. 14, no. 4, pp. 685-692, 2019.

[6] S. Jung, J. Koh, S. Lee, "Recurrent Neural Network based Prediction System of Agricultural Photovoltaic Power Generation," *Journal of the KIECS*, vol. 17, no. 5, pp. 825-832, 2022.

[7] H. Long, Z. Zhang, and Y. Su, "Analysis of daily solar power prediction with data-driven approaches," *Applied Energy*, vol. 126, pp. 29-37, 2014.

[8] V. Prema and K. U. Rao, "Development of statistical time series models for solar power prediction," *Renewable energy*, vol. 83, pp. 100-109, 2015.

[9] J. Kim, "Deep Learning Model based on Imaging Time Series Data for Demand Forecasting," Doctoral dissertation, *Hanyang University*, 2023.

[10] J. Zhang and K. F. Man, "Time series prediction using RNN in multi-dimension embedding phase space," *SMC'98 Conference Proceedings 1998 IEEE International Conference on Systems*, vol. 2, no. 98CH36218, pp. 1868-1873, 1998.

[11] B. Kim, S. Jung, M. Kim, J. Kim, H. Lee, and S. Kim, "Solar Power Generation Forecasting based on LSTM considering Weather Conditions," *Journal of Korean Institute of Intelligent Systems*, vol. 30, no. 1, pp. 7-12, 2020.

[12] D. O'Leary and J. Kubby, "Feature Selection and ANN Solar Power Prediction," *Journal of*

Renewable Energy, vol. 2017, pp. 1-7, 2017.

[13] E. Izgi, A. Öztopal, B. Yerli, M. K. Kaymak, and A. D. Şahin, "Short - mid-term solar power prediction by using artificial neural networks," *Solar Energy*, vol. 86, no.2, pp. 725-733, 2012.

저자 소개

박건하(Gun-Ha Park)



2023년 순천대학교 컴퓨터공학과 졸업(공학사)

2023년 ~ 현재 순천대학교 대학원 컴퓨터공학과 재학(공학석사)

※ 관심분야 : 인공지능, 컴퓨터비전, 객체 추적, 데이터 분석 및 예측

임수창(Su-Chang Lim)



2015년 순천대학교 컴퓨터공학과 졸업(공학사)

2017년 순천대학교 대학원 컴퓨터공학과 졸업(공학석사)

2022년 순천대학교 대학원 컴퓨터공학과 졸업(공학박사)

2020년 ~ 현재 ㈜티이에프 기업부설연구소 연구소장

※ 관심분야 : 인공지능, 컴퓨터비전, 딥러닝, 객체 추적, 시계열 데이터 분석 및 예측

김종찬(Jong-Chan Kim)



2000년 순천대학교 컴퓨터과학과 졸업(이학사)

2002년 순천대학교 대학원 컴퓨터과학과 졸업(이학석사)

2007년 순천대학교 대학원 컴퓨터과학과 졸업(이학박사)

2013년 서울대학교 자동화 시스템 연구소 선임연구원

2021년 ~ 현재 순천대학교 컴퓨터공학과 조교수

※ 관심분야 : 영상처리, HCI, 콘텐츠, 컴퓨터그래픽스, 기계학습, 객체추적, 컴퓨터비전, 데이터 분석 및 예측