

재생에너지 발전량 예측을 위한 LSTM-Autoencoder

박건하* · 김종찬*

LSTM-Auto-encoder for Renewable Energy Generation Forecast

Gun-Ha Park* · Jong-Chan Kim**

요약

본 논문은 재생에너지 발전량 데이터를 기반으로 발전량 예측값을 도출하기 위한 연구를 수행하였다. 발전량 측정을 위해 재생에너지의 발전량 데이터, 환경데이터와 같은 다변량 변수를 측정하였다. 측정값의 안정성과 신뢰성 확보를 위해 ARIMA(Auto Regressive Integrated Moving Average) 모델을 이용한 전처리를 수행하였다. 재생에너지의 발전량 예측을 위해 딥러닝 모형인 LSTM-Autoencoder 결합모델을 이용하여 발전량을 측정했고, 예측정확도를 높이기 위해 각 다변량 변수의 장기종속성 학습 방식을 이용하였다. 정제된 데이터를 활용한 환경변수들의 추세는 기존 데이터를 그대로 사용했을 때 보다 안정되었고, 상관관계 분석 결과를 반영하여 다변량 변수 중 상관성이 높은 변수만을 활용하여 재생에너지 발전량 예측 알고리즘을 개선하여 약 94%의 정확도가 측정되었다.

ABSTRACT

This paper conducted a study to derive the predicted value of power generation based on the renewable energy generation data. In order to measure the power generation, multivariate variables such as power generation data and environmental data of renewable energy were measured. In order to secure the stability and reliability of the measured values, preprocessing was performed using the ARIMA (Auto Regressive Integrated Moving Average) model. In order to predict the power generation of renewable energy, the power generation was measured using the LSTM-Autoencoder combination model, a deep learning model, and the long-term dependency learning method of each multivariate variable was used to increase the prediction accuracy. The trend of environmental variables using refined data was more stable when the existing data were used as it was, and the accuracy of about 94% was measured by improving the renewable energy generation prediction algorithm by using only the highly correlated variable among the multivariate variables by reflecting the results of the correlation analysis.

키워드

Renewable energy, Regression Model, ARIMA, Deep Learning, LSTM, Autoencoder
재생에너지, 회귀 모델, 아리마, 딥러닝, 엘에스티엠, 오토인코더

* 순천대학교 컴퓨터공학과 연구원
(pgh6151@naver.com)

** 교신저자 : 순천대학교 컴퓨터공학과
• 접수일 : 2024. 10. 30
• 수정완료일 : 2024. 11. 21
• 게재확정일 : 2024. 12. 12

• Received : Oct. 30, 2024, Revised : Nov. 21, 2024, Accepted : Dec. 12, 2024
• Corresponding Author : Jong-Chan Kim
Dept. Suncheon National University,
Email : seaghost@scnu.ac.kr

1. 서 론

재생에너지는 자연에서 지속적으로 공급되는 에너지 자원을 사용하여 전력을 생산하는 에너지를 의미한다. 재생에너지는 환경에 미치는 영향을 최소화하며 화석연료에 의존하는 기존의 에너지 생산 방식보다 지속 가능성이 커 미래 에너지 수급에 주요한 역할을 수행한다.

국내에서는 RE100(Renewable Energy 100%) 목표 달성을 위한 구체적인 제도가 도입됨에 따라 재생에너지 수요는 급증하고 있다. 기업들 또한 재생에너지의 전력구매계약(PPA:Power Purchase Agreement)을 통한 재생에너지 공급자 확보를 위해 움직이고 있다 [1]. 기업들의 전략적인 재생에너지 도입은 기업의 이익과 직결되며 성공적인 도입을 위해서 재생에너지의 전력 생산량은 정확하게 예측되고 관리되어야 한다.

재생에너지의 발전량은 일정하지 않은 경우가 대부분이며 재생에너지로 인한 전력이 너무 많은 경우 재생에너지는 출력제한을 받게 된다. 출력제한은 태양광, 풍력 등 재생에너지를 통해 생산된 공급이 일정 수준 이상으로 과도하게 증가할 때 전력 계통의 안정성을 유지하기 위해서 재생에너지 발전소에서 생산하는 전력을 제한하는 조치를 의미한다[2-4].

출력제한의 영향을 최소화하고, 재생에너지의 사용을 확대하기 위해서는 변동성이 강한 재생에너지의 정확한 예측을 통해서 안정적이고 지속 가능한 전력 공급 체계를 구축하여야 한다.

본 논문에서는 시계열로 구성된 재생에너지 발전량 예측을 위해 LSTM과 Autoencoder 결합된 모델을 활용한다. LSTM-Autoencoder 결합모델을 통해서 차원 축소와 시간적 패턴학습을 동시에 수행하고, 발전량 예측을 수행하게 된다. Autoencoder는 입력 시계열 데이터의 차원을 저차원으로 압축하고 데이터의 특징을 추출한다. 이러한 Autoencoder는 시계열 데이터의 계절성, 과형 등의 특징을 활용하여 예측을 위한 학습을 수행한다.

II. 관련연구

2.1 시계열 데이터의 회귀분석 기반 예측

회귀분석은 두 개 이상의 변수 간의 관계를 모델링

하고 분석하는 기법이다. 시계열 데이터는 시간에 따라 증감이 보이는 추세, 특정 주기로 반복되는 패턴인 계절성, 이전 시간의 데이터가 현재 데이터에 영향을 미치는 자기 상관 등의 특성을 가지고 있다. 회귀분석은 주로 종속 변수와 하나 이상의 독립 변수 간의 관계를 이해하고 예측하기 위해 사용된다.

시계열 데이터를 활용한 회귀분석 방법은 단순 회귀(Simple Regression), 다중 회귀(Multiple Regression), ARIMA(AutoRegressive Integrated Moving Average) 등이 존재한다. 단순 회귀는 단순한 선형 회귀를 사용하여 시간과 관측값 사이 관계를 모델링하는 방식을 가지지만 시계열 데이터의 복잡한 특성을 반영하기에는 한계점이 존재한다[5-10].

다중 회귀 방식은 이러한 단순 회귀 방식을 개선하여 여러 개의 독립변수를 사용하여 현재 값을 예측하는 방식으로 시계열 데이터의 여러 기상 변수인 일사량, 온도, 습도 등의 환경변수와 태양광 발전량 간의 관계를 다중 선형 회귀로 표현할 수 있다.

2.2 LSTM을 활용한 예측 알고리즘

태양광 발전량 데이터는 시계열 데이터의 일종으로 시계열 데이터의 예측은 과거 시점으로부터 데이터의 분포를 학습한 것을 기반으로 미래 시점의 변수를 예측하는 것을 목표로 한다. 딥러닝은 복잡한 패턴 인식과 예측 문제를 해결하는 데 강력한 도구로 자리 잡고 있으며, 특히 시계열 데이터 예측 및 분석에 활용되는 딥러닝 알고리즘은 통계기반 예측 방식보다 높은 예측 정확도를 제공하였다. 시계열 데이터 예측에 딥러닝을 활용하면 시계열의 시간 의존성을 잘 모델링 할 수 있다.

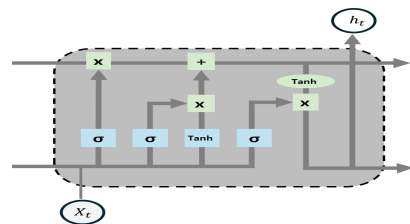


그림 1. 예측에 활용되는 LSTM 구조

Fig. 1 LSTM structure used for prediction

그림 1은 예측에 활용되는 기본적인 LSTM 구조를 나타낸다. LSTM(Long Short-Term Memory)은 순환

신경망(RNN)의 일종으로, RNN이 가지고 있는 장기 의존성 문제를 해결하기 위해 개발되었다[11-14]. RNN은 시계열 데이터의 시간순서 정보를 처리할 수 있지만, 장기 의존성을 학습하기에는 어려움을 가진다. LSTM은 이러한 문제를 해결하기 위해 ‘셀 상태(Cell state)’와 여러 게이트를 도입하여 이를 해결하고자 하였다.

LSTM 모델의 메모리 셀 구조를 보인다. LSTM 메모리 셀은 입력값과 이전 상태에 따라 값을 업데이트하고, 새로운 상태를 출력한다. 새로운 상태를 출력하기 이전 LSTM은 셀의 값을 얼마나 기억할지 결정하는 것이 가능한 게이트를 가지고 있어 필요한 정보만 기억하도록 제어할 수 있다.

III. 재생에너지 발전량 예측을 위한 LSTM-Autoencoder

3.1 재생에너지 발전량 데이터 수집

태양광 발전은 빛 에너지를 전기 에너지로 변환하는 과정으로 주로 태양광 패널을 통해 이루어진다. 태양광 패널은 n형 반도체와 p형 반도체로 구성되어 광자를 흡수하고 전기장을 생성하여 전류를 형성하는 광전효과(Photovoltaic effect)에 의해 발전이 이루어진다. 일반적인 가정이나 산업용으로 사용되는 전기는 대부분 교류(Alternating Current)전기를 사용하지만, 태양광 패널에서 발생한 전기는 직류(Direct Current)이다. 이를 인버터를 통해 패널에서 발생한 직류전기를 활용할 수 있는 교류전기로 변환한다.

정확한 태양광 발전을 위해서는 환경변수에 대한 고려가 필요하다. 환경변수들은 태양광 패널에 도달하는 일사량을 결정하고, 패널의 효율성을 변화시키며, 최종적으로 발전량에 영향을 미친다. 태양광 발전 시스템의 예측을 위해서 태양광 발전량에 직접적인 영향을 미치는 변수들을 분류하고 통합하여 정확한 예측을 수행하여야 한다.

3.2 재생에너지 시계열 데이터 전처리

수집된 태양광 발전량 데이터는 전처리 이전의 데이터로 전처리를 거치지 않은 데이터는 모델의 성능에 큰 영향을 미치는 핵심 단계라고 할 수 있다. 특히

발전량 예측을 위해 모델을 학습시킬 때 결측치 또는 이상값을 모델이 학습하고 예측 결과를 내보낼 수 있기 때문에 데이터의 신뢰성과 안정성을 확보하기 위해서 거쳐야 하는 필수적인 과정이다.

본 논문에서는 태양광 발전량 데이터를 전처리하기 위해서 ARIMA 모델을 활용하여 전처리를 수행하여 각 데이터의 결측값과 이상값을 제거하고 데이터가 안정적인 시계열의 모습을 갖출 수 있도록 하였다.

수집된 태양광 발전량 데이터를 기반으로 ARIMA 모델은 이상값, 결측값을 발견하고, 발견된 데이터를 예측값으로 대체하여 시계열 데이터의 계절성 등을 유지할 수 있다. ARIMA 모델은 비교적 단순하고, 모델의 작동방식이 직관적이다. 시계열 데이터의 트렌드와 계절성을 포착하고 데이터의 구조적 특성을 명확하게 이해할 수 있다. 상대적으로 적은 양의 데이터로도 모델을 구축하여 전처리에 반영할 수 있어 데이터를 수정하고 안정적으로 만든다.

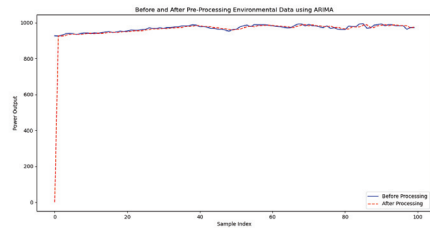


그림 2. 환경변수의 데이터 전처리
Fig. 2 Data preprocessing of environmental variables

그림 2는 수집된 환경 센서 데이터를 기반으로 데이터의 전처리가 수행된 모습을 보인다. 환경 센서 데이터를 ARIMA 모델을 활용하여 차분, 이상값과 결측치 제거 등의 전처리 작업을 수행하였다.

태양광 발전량 데이터는 시계열 데이터의 일종으로 데이터의 평균과 분산이 시간에 따라 변화하는 비정상성이 발생할 수 있다. ARIMA 모델은 이러한 비정상성을 제거하여 안정적인 패턴을 생성하기 위해 차분을 활용한다. 차분은 현재 시점의 데이터 값과 이전 시점 데이터를 비교하고 추세나 계절성을 제거하여 데이터가 일정한 분포를 따르도록 유도한다. 이러한 이점을 바탕으로 ARIMA 모델 기반 전처리를 수행하였다.

3.3 LSTM-Autoencoder 결합 모델

Autoencoder와 LSTM을 활용한 결합모델은 단일 모델로 활용되었을 때의 문제점들을 해결하고 날씨, 설비의 상태, 환경 변화 등 다양한 외부 요인에 영향을 받는 태양광 발전량 예측 알고리즘의 특성을 잘 파악하고 해결한다.

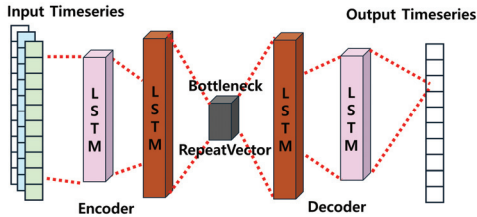


그림 3. LSTM-Autoencoder 결합모델 전체 구조
Fig. 3 LSTM-Autoencoder Combined Model Full Structure

그림 3은 LSTM과 Autoencoder를 활용한 모델을 활용한 데이터 처리 과정이다. LSTM 모델을 활용하여 Autoencoder의 encoder 부분을 재현하여 입력 데이터를 압축하여 함축적인 특징을 표현한다. Autoencoder 구조를 형성하여 압축된 특징 데이터를 활용해서 다시 LSTM 모델을 활용하여 데이터를 복원하면서 태양광 발전량 데이터를 예측하며 복원하는 역할의 LSTM 모델은 Autoencoder의 decoder를 의미한다.

태양광 발전량은 여러 제약에 따라 발전량이 변동될 수 있다. 본 논문에서의 데이터는 태양광 모듈의 열화에 대한 발전량 감소를 제외한 환경변수를 산정하였다. 태양광 발전량에 직접적인 영향을 주는 환경 데이터는 태양광 모듈에서 측정되는 직류, 교류발전량 데이터와 모듈의 온도, 외부온도, 기온 등을 활용한다.

본 논문에서는 Autoencoder 구조를 활용한 재생에너지의 발전량 예측에 대해 연구한다. 특히, 시계열 데이터 예측에 강력한 성능을 보이는 LSTM(Long Short-Term Memory)과 Autoencoder의 결합모델을 활용하여 태양광 발전량을 예측하는 방법론을 제시한다. LSTM은 시퀀스 데이터를 처리하고 장기 의존성을 학습하는 데 매우 유용한 구조로, 발전량 예측에

있어서 시간적 상관관계를 효과적으로 학습할 수 있다. Autoencoder는 입력 데이터를 압축하고 주요 특징만을 추출하여 학습하는 데 사용되며, 이를 통해 데이터의 노이즈를 제거하고 중요한 패턴을 학습할 수 있다.

제한된 결합모델의 아키텍처는 Autoencoder의 구조를 기반으로 하여, Encoder와 Decoder로 구성된다. Encoder 부분에서는 두 개의 LSTM 레이어를 사용하여 입력 데이터의 중요한 특징만을 추출하고 압축한다. 이는 복잡한 시계열 데이터를 더 간결하게 표현함으로써, 모델이 중요한 정보에 집중할 수 있도록 도와준다. 이후 Decoder 부분에서는 다시 두 개의 LSTM 레이어를 사용하여 데이터를 복원하는 과정을 거친다. 이 복원 과정에서 데이터의 계절성(seasonality)과 추세(trend)를 재구성하며, 이를 통해 태양광 발전량을 예측하게 됩니다. 이렇게 설계된 결합 모델은 태양광 발전량 예측에서 높은 정확도를 달성할 수 있고, 환경 변수의 복잡한 패턴을 효과적으로 반영할 수 있는 강점이 있다.

IV. 성능평가

4.1 정량적 지표

학습모델은 LSTM과 Autoencoder의 결합모델을 활용하여 예측을 진행하였다. 기본적인 아키텍처 구조는 시계열 데이터를 입력값으로 가져와 LSTM을 활용하여 데이터의 특징만을 압축하여 인코딩한다. 이때 사용되는 LSTM Layer는 두 개의 LSTM Layer를 사용한다.

첫 번째 LSTM 인코딩 Layer에는 모든 타임 스텝의 출력을 전달하고 두 번째 LSTM Layer는 인코딩된 압축 벡터를 출력한다. 이후 decoder 파트에서의 LSTM Layer에 들어가기 전 마지막 encoder의 출력을 반복하고 반복된 입력을 받아 모든 타임 스텝의 출력을 반환한다.

두 번째 decoder의 LSTM Layer에서는 첫 번째 Layer의 출력을 전달받아 복원된 시퀀스를 출력한다. 모델에 입력과 출력의 데이터 사이즈는 동일하게 반환된다. 여기서 LSTM은 장기기억 Layer를 통해서 모델을 복원하는 방법을 학습하게 되며 학습하는 과

정에서 예측되는 시계열을 배출할 수 있게 된다.

LSTM-Autoencoder 결합 예측 모델을 평가하기 위해서는 Mean Absolute Error(MAE), Root Mean Square Error(RMSE), Mean Absolute Percentage Error(MAPE)등을 사용하여 정확도를 산출한다.

4.2 실험 결과

본 논문에서는 LSTM-AE 결합모델을 구축하였고 시계열 데이터 예측 성능평가를 진행하였다.

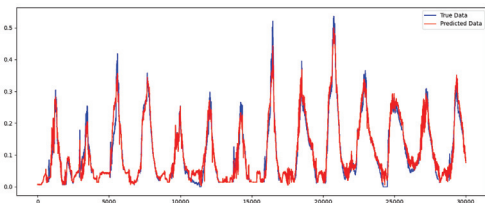


그림 4. 재생에너지 중 태양광 발전량 데이터를 활용한 예측 결과 그래프

Fig. 4 Graph of prediction results using solar power generation data among renewable energy

그림 4는 본 논문에서의 예측 모델이 도출한 예측 발전량과 실제 값의 오차를 계산해 각 시점의 데이터를 비교하였고, 실제 값과 추세와 계절성은 유사하게 예측하여 정확도는 약 94%로 측정되었다.

V. 결론 및 향후연구

최근 인공지능(AI) 기술의 눈부신 발전은 다양한 산업 분야에서 혁신을 가져오고 있고, 특히 예측 분야에서는 그 영향력이 두드러진다. 딥러닝 알고리즘은 대량의 데이터를 학습하여 복잡한 패턴을 이해하고 미래를 예측하는 능력을 갖추고 있고, 이는 기존의 통계기반 예측 기법이나 회귀분석보다 훨씬 더 높은 정확도를 제공하고 있다. 딥러닝 모델들은 특히 시계열 데이터와 이미지 데이터와 같은 복잡한 구조의 데이터를 처리하고 분석하는 데 뛰어난 성능을 발휘하고 있고, 예측과 분류 등 다양한 문제 해결에 있어 큰 강점으로 작용하고 있다.

재생에너지원 중 하나인 태양광 발전은 환경을 오염

시키지 않으며 지속 가능한 에너지원으로 주목받고 있다. 태양광 발전은 지구상의 태양 복사 에너지를 직접 변환하여 전력을 생성하는 방식으로, 환경에 미치는 영향이 적고 온실가스 배출량을 줄이는 데 기여한다.

태양광 에너지의 발전량은 기상 및 환경 조건에 매우 민감하게 반응하며, 그 변동성이 매우 크며 변동성은 태양광 발전을 대규모 에너지원으로 활용하는 데 있어 주요한 도전 과제 중 하나로 자리 잡고 있다. 특히, 에너지 밀도가 낮다는 특성으로 인해 대규모 전력 생산을 위해서는 넓은 면적에 다수의 태양광 모듈이 설치되어야 한다.

본 연구는 LSTM-Autoencoder 결합 모델을 통해 태양광 발전량 예측의 정확도를 높이는 동시에, 재생에너지의 안정적 운영을 위한 기반 기술을 제공하는 것을 목표로 한다. 이를 통해, 재생에너지의 변동성으로 인한 전력망의 불안정을 최소화하고, 보다 효율적인 에너지 자원 관리가 가능하도록 기여할 수 있을 것이다.

References

- [1] D. Kim, C. Park, J. Park & J. Roh, "An Analytical Approach for Transmission Use-of-System Charges for a Renewable Energy Power Purchase Agreement," *The Transactions of the Korean Institute of Electrical Engineers*, vol. 72, no. 7, 2023, pp. 801-808. <https://doi.org/10.5370/KIEE.2023.72.7.801>.
- [2] O. Kwon, S. Kim & S. Song, "Status and Cases of Wind Power Curtailment of Overseas Grid Operator with a High Penetration of Wind Power Generation," *Proceedings of the Korean Institute of Electrical Engineers Conference*, 2018., pp. 906-907. <https://doi.org/10.1016/j.rser.2016.06.082>.
- [3] C. Lee, Y. Kim & S. Kim, "Minimization of Renewable Energy Output Limits and Reduction Methods," *Journal of the Korean Institute of Illuminating and Electrical Installation Engineers*, vol. 33, no. 12, 2019, pp. 41-48. <http://dx.doi.org/10.5207/JIEIE.2019.33.12.041>.
- [4] Y. Yang, S. Bremner, C. Menictas & M. Kay, "Battery energy storage system size determination in renewable energy systems: A review,"

- Renewable and Sustainable Energy Reviews, vol. 91, 2018, pp. 109-125.
<https://doi.org/10.1016/j.rser.2018.03.047>.
- [5] W. Kim, K. Kim, J. Lee & H. Lee, "Linear, Non-Linear and Time Series Model Verification for Forecasting the Change of Grade Increase Applicants-Focusing on Gyeonggi Region," Kukkiwon Taekwondo Research, vol. 8, no. 1, 2017, pp. 169-190
<https://doi.org/10.24881/tjk.2017.8.1.169>.
- [6] S. Kim, Y. Lee, R. Kim, S. Seo & G. Ryu, "An Incremental Regression Model for Time Series Data Prediction," Proceedings of the 25th Spring Conference of the Korea Information Processing Society, 2006, vol. 13, no.1, pp. 23-26.
- [7] R. Shumway & D.Stoffer, "ARIMA models," Time series analysis and its applications, pp. 75-163, 2017.
https://doi.org/10.1007/978-3-319-52452-8_3.
- [8] K. Kalpakis, D. Gada & V. Puttagunta, "Distance Measures for Effective Clustering of ARIMA Time-Series," Proceedings 2001 IEEE international conference on data mining, IEEE, 2001, pp. 273-280
<https://doi.org/10.1109/ICDM.2001.989529>.
- [9] D. Lee, A. Jung, J. Kim, C. Kim, H. Kim & Y. Lee, "Solar Power Generation Forecast Model Using Seasonal ARIMA," Journal of the Korean Solar Energy Society, , 2019, vol. 39, no. 3, pp. 59-66.
<https://doi.org/10.7836/kses.2019.39.3.059>.
- [10] H. Jeong, S. Hong, J. Jeon, S. Lim, J. Kim, H. Park & C. Park, "A Research of Prediction of Photovoltaic Power using SARIMA Model," Journal of Korea Multimedia Society, vol. 25, no. 1, 2022, pp. 82-91.
<https://doi.org/10.9717/kmms.2022.25.1.082>.
- [11] L. Wang, M. Mao, J. Xie, Z. Liao & H. Zhang, "Accurate solar PV power prediction interval method based on frequency-domain decomposition and LSTM model," Energy, vol. 262, 2023, pp. 1-18.
<https://doi.org/10.1016/j.energy.2022.125592>.
- [12] F. Altche & A. Fortelle, "An LSTM Network for Highway Trajectory Prediction," 2017 IEEE 20th international conference on intelligent transportation systems (ITSC), IEEE, 2017.
<https://doi.org/10.1109/ITSC.2017.8317913>.
- [13] A. Djaafari, A. Ibrahim, N. Bailek, K. Bouchouicha, M. Hassan, Alban Kuriqi, Nadhir Al-Ansari & El-Sayed M. El-Kenawy, "Hourly predictions of direct normal irradiation using an innovative hybrid LSTM model for concentrating solar power projects in hyper-arid regions," Energy Reports, vol. 8, pp. 15548-15562, 2022,
<https://doi.org/10.1016/j.egy.2022.10.402>.
- [14] L. Ren, J. Dong, X. Wang, Z. Meng, L. Zhao & M. Deen, "A Data-Driven Auto-CNN-LSTM Prediction Model for Lithium-Ion Battery Remaining Useful Life," IEEE Transactions on Industrial Informatics, vol. 17, no. 5, 2020, pp. 3478-3487.
<https://doi.org/10.1109/TII.2020.3008223>.

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

김종찬(Jong-Chan Kim)



2000년 순천대학교 컴퓨터공학과 졸업(이학사)

2002년 순천대학교 대학원 컴퓨터공학과 졸업(이학석사)

2007년 순천대학교 대학원 컴퓨터공학과 졸업(이학박사)

2013년 서울대학교 자동차 시스템 연구소 선임연구원

2021년 ~ 현재 순천대학교 컴퓨터공학과 조교수

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