

<https://doi.org/10.7236/IIBC.2021.21.4.57>
IIBC 2021-4-9

IoT 센서의 시계열 데이터 정확도 향상을 위한 인공지능 기반 분류 기법

Artificial Intelligence-based Classification Scheme to improve Time Series Data Accuracy of IoT Sensors

김진영*[†], 심이삭** , 윤성훈***

Jin-Young Kim*[†], Isaac Sim** , Sung-Hoon Yoon***

요 약 인공지능을 위한 병렬연산 능력이 향상됨에 따라 인공지능 적용 분야가 다양한 방향으로 확대되고 있다. 특히 방대한 데이터를 처리해야 하는 IoT센서의 데이터를 처리하기 위해 인공지능이 도입되고 있다. 하지만 시간에 따른 데이터의 중요도가 달라지는 IoT 시계열 데이터 특성상 기존의 인공지능 학습 기법을 그대로 적용하기에는 한계점이 있다. 본 과제에서는 IoT 센서 데이터를 효과적으로 처리하기 위해 시간가중치기반 및 사용자 상태값 기반 인공지능 처리 기법을 연구한다. 상기 기법을 통해 기존 인공지능 학습을 적용시키는 것 보다 높은 센서 정확도를 확보 할 수 있게 된다. 이에 더해, 해당 연구를 기반으로 다양한 분야에서 인공지능 학습을 적용하는 방안을 제시하고, 지속적인 연구를 통해 다양한 분야로의 확장을 기대할 수 있다.

Abstract As the parallel computing capability for artificial intelligence improves, the field of artificial intelligence technology is expanding in various industries. In particular, artificial intelligence is being introduced to process data generated from IoT sensors that have enormous data. However, the limitation exists when applying the AI techniques on IoT network because IoT has time series data, where the importance of data changes over time. In this paper, we propose time-weighted and user-state based artificial intelligence processing techniques to effectively process IoT sensor data. This technique aims to effectively classify IoT sensor data through a data pre-processing process that personalizes time series data and places a weight on the time series data before artificial intelligence learning and use status of personal data. Based on the research, it is possible to propose a method of applying artificial intelligence learning in various fields.

Key Words : IoT network, Artificial Intelligence, data post-process

* (주)스마트에버

** 광운대학교 전자융합공학과

*** (주)코젠

접수일자 2021년 4월 28일, 수정완료 2021년 7월 16일
게재확정일자 2021년 8월 6일

Received: 28 April, 2021 / Revised: 16 July, 2021 /

Accepted: 6 August, 2021

[†]Corresponding Author: jinyoung@kw.ac.kr
Smart Ever Co., Ltd.

I. 서론

최근 그래픽 카드를 기반으로 하는 병렬 연산 능력이 급증함에 따라 인공지능의 다양한 분야에서의 적용이 증가하고 있다^[1-4], 기존의 중앙 연산으로만 가능했던 빅데이터 및 IoT 데이터 처리 과정이 개인PC에서 충분히 수행 가능할 정도로 연산 능력이 증대하였고, 이에 비례하여 기하급수적으로 증가하는 데이터를 처리하기 위한 인공지능 기반 데이터 처리기술이 연구되고 있다^[5]. 현재 대부분의 인공지능 기능 처리기술은 학습 단계에서 학습의 난이도를 조정하는 기법을 연구하는 논문이 대부분이다. 하지만 이는 사용자가 정확히 파악할 수 없는 학습 결과물만을 가지고 학습 과정을 추론하여 경험기반의 학습 조정 (Empirical-based feedback)을 수행하기 때문에, 막대한 인적 자원이 소모된다^[6]. 또한, 짧은 시간간격으로 수집되는 IoT 데이터의 특성상 모든 데이터가 하나의 같은 학습방식으로 처리되면 인공지능 학습시 과도한 데이터에 의해 센서 정확도가 떨어지는 overfitting(과적응)이 발생하기 쉬워지게 된다.

위와 같은 학습 조정 및 overfitting 문제를 해결하기 위해 다양한 방법이 연구되고 있다. 대표적인 방법은 데이터의 전처리 기법이 있는데, 학습하고자 하는 데이터 중 거의 유사한 값을 가진 데이터를 선별하여 필터링하는 기법을 주로 사용하고 있다^[7]. 또한, 인공지능 학습 단계에서 계층 간의 weight transfer를 무작위로 선별하고, 학습데이터의 life cycle을 지정하여 일정 기간 이상 지난 시계열 데이터를 소거하는 방법도 연구되고 있다^[8]. 하지만 상기 방법들은 시계열데이터가 가지는 특적인 시간에 따른 데이터 중요도 변화를 고려하지 못하고, 오직 데이터값 또는 학습방법의 변화만으로 데이터를 전처리하고 있다.

현재 인공지능 관련 연구는 단말기기의 컴퓨팅 연산 성능이 상승함에 따라, IoT 센서 네트워크에 적용되는 인공지능 기술은 기존의 서버 중심의 연산에서 단말분산형 연산으로 확대되고 있다. 또한, 기존에 기계가 담당하기 힘들었던 모니터링 및 정밀제어 분야 또한 기계학습 및 신경망학습을 기반으로 사람 중심의 제어 시스템에서 탈피하여 인적자원 없이 자동화를 추진하고 있다^[9].

본 논문에서 제안하는 인공지능 기반 IoT 시계열 데이터 분류 기법은 기존의 데이터값 기반 분류 기법과 달리 데이터 뿐만 아니라 데이터에 연계되어 있는 각종 외부 요건을 정밀하게 비교분석한다. IoT 센서와 밀접한 연관이 있는 사용자 및 주변 환경을 데이터화하여, IoT

센서로 수집 할 수 없는 데이터를 추가하고 이를 기반으로 데이터를 분류하고자 한다.

본 논문에서 제안하는 인공지능 기반 IoT 시계열 데이터 분류 기법은 기존의 데이터값 기반 분류 기법과 비교하여 다음과 같은 차별성 및 기여도가 있다.

- 가. 기존의 데이터 전처리 및 분류기법에서 많이 차용되는 중복데이터 기반 필터링을 지양하고, 각 IoT 기기들이 가지는 인간의 생활패턴 기반 시계열 특성을 고려하여, 데이터 자체에 가중치를 부여하여 인공지능 학습시 이를 고려한 학습을 통한 데이터 향상을 제시하였다.
- 나. 현재 활발하게 연구되고 있는 IoT기술 및 스마트 홈 관련 기술에 개인 맞춤형 데이터를 testbed를 통해 가시화하여 연구하였다.
- 다. 데이터 분류 및 학습 능력 향상에 도움이 되는 외부요인을 조사하고, 이를 인공지능 신경망에 실제로 적용시켜 학습량은 줄이면서 정확도는 향상시킬 수 있는 방법을 제시하였다.

본 논문에서는 IoT센서의 시계열 데이터 수집을 위해 Arduino 및 센서를 통하여 다양한 비정형 데이터를 생성하고 이를 실시간을 수집하는 모델을 수립하였다.

본 논문의 2장에서는 본 논문과 관련된 IoT 데이터 관련 기술을 설명하고, 3장에서는 본 논문에서 제시하는 시스템 설계를 설명한다. 4장에서는 제시한 시스템 기반의 시뮬레이션 구현 및 시뮬레이션 결과를 진행하고, 5장에서는 결론을 설명하도록 한다.

II. 시스템 설계

1. IoT센서 단말에서의 데이터 분류

그림 1은 일반적으로 사용되는 분산센서 기반 IoT 네트워크와 본 논문에서 제시하는 IoT 센서의 시계열 데이터의 정확도 향상 기법이 적용된 시스템 모델을 비교하고 있다. 그림 1.(a)는 일반적인 IoT 시스템을 나타내었으며, IoT센서에서 획득한 정보를 그대로 사용하거나, 중복되는 데이터를 제거하고 학습을 진행하게 된다. 이와같은 방식으로 진행되면 시계열 데이터가 가지는 시간에 따른 데이터 가중치를 무시하고 학습을 수행하게 되며, 이는 연산량대비 학습 정확도가 떨어질 수 있다. 그

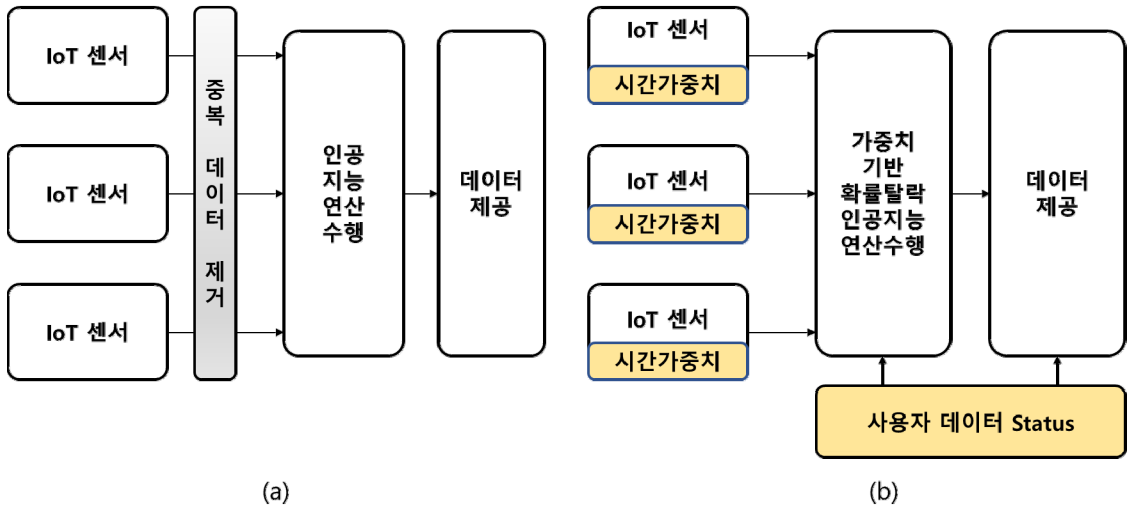


그림 1. 기존 IoT 네트워크의 인공지능 학습 및 본 논문에서 제안하는 시간가중치 및 사용자 상태 기반 인공지능 학습
 Fig. 1. Conventional IoT-AI scheme and proposed IoT-AI scheme based on time-weight and personal status.

림 1. (b)는 본 논문에서 제안하는 시간가중치 및 사용자 데이터 기반 IoT 센서 데이터 학습 시스템이다. 단말의 연산 능력이 향상됨에 따라 IoT 센서 단말에서도 센서정보를 전송하는 것 뿐만이 아닌 데이터에 대한 가공을 수행할 수 있게 되었고, 본 논문은 이에 영감을 받아 데이터의 전처리를 시스템이 아닌 센서에서 분담하는 방식을 제안하였다. IoT 센서는 시계열 데이터 전체를 전송하는 것이 아닌, 다음과 같은 임계점을 넘는 센서 데이터만 전송을 수행하게 된다:

$$Data_{IoT} = \sum_{i=1}^n d_i \cdot w_{ts} \cdot \delta_{det} \quad (1)$$

$$\delta_{det} = \begin{cases} 1, & d_{i-j} > \lambda_{det} \\ 0, & d_{i-j} < \lambda_{det} \end{cases} \quad (2)$$

위 식에서 $Data_{IoT}$ 는 총 전송되는 시계열 데이터의 합, d_i 는 시간 i 의 데이터, w_{ts} 는 시계열 데이터 내부의 시간 데이터를 기반으로 한 가중치, δ_{det} 는 시간 i 의 데이터와 시간 j 의 데이터의 차이를 기준으로 전송을 결정하는 임계값이다. 본 논문의 IoT센서는 데이터가 입력될 때마다 그 전값의 데이터와 차이를 비교하여 그 값이 정해진 임계점 λ_{det} 보다 높을 때, 시간에 대한 가중치를 곱하여 이를 전송하게 된다. 해당 과정을 통해 IoT 센서는 모든 데이터를 보내는 것이 아닌, 학습에 영향을 줄 수 있는 데이터만을 전송하게 되고, 인공지능 학습이 수행되기 전 데이터에 가중치를 부여함으로써 의미있는 데이

터에 대한 학습 성능을 향상시킬 수 있다.

2. 인공지능 신경층에서의 데이터 분류

그림 2는 본 논문에서 제시하는 데이터 탈락 층에 포함된 합성곱 신경망 층을 나타낸 도식이다. 입력 데이터로 식 (1-2)을 통해 전송된 데이터를 입력하며, 1차원 합성곱, Pooling을 거친 후 데이터 탈락이라는 층을 거친다. 데이터 탈락은 기존의 인공지능 학습의 drop out(생략)과 비슷하나 다른 의미를 가진다. Drop out은 층과 층 사이의 node 중 일부를 정해진 확률을 통해 무작위로 생략하고, 생략된 노드는 해당 연산에서는 계산되지 않는다. 본 논문에서 제시한 데이터 탈락은 기존의 Dropout과 비교하여 2가지 다른 점이 존재한다.

첫 번째로, Dropout은 노드마다 생략될 확률이 동일하지만, 데이터 탈락은 노드마다 생략될 확률은 다음과 같다:

$$0 \leq (\text{탈락확률}) = k \cdot w_{ts} \leq 1 \quad (3)$$

여기서 k 는 각 노드별 가지고 있는 계수이고, w_{ts} 는 노드가 가진 가중치가 아닌 중재목 1에서 수행된 IoT 센서의 가중치를 뜻한다. 이를 계산하기 위해 사용자 데이터에 k 와 w_{ts} 를 저장하고 이를 상태값(status value)으로 활용하게 된다. 이 값은 인공지능 연산에 의해 변경되지 않는 상수값으로 활용이 된다.

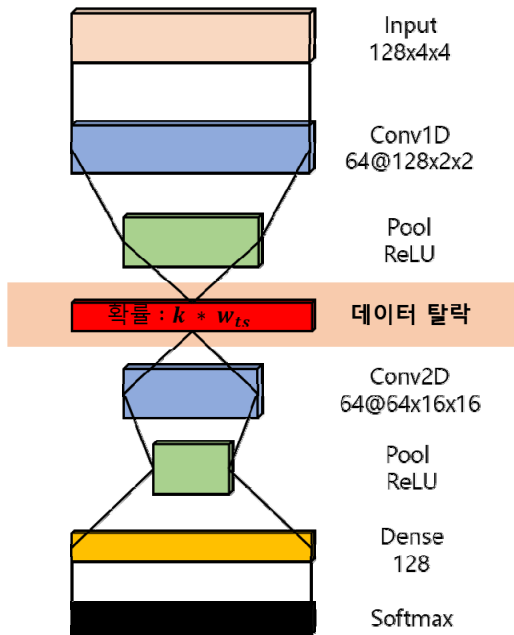


그림 2. 본 논문에서 제시하는 데이터 탈락 층이 추가된 합성곱 신경망 기반 데이터 분류 인공지능 학습 모델.
Fig. 2. CNN classification artificial intelligence model with data leave-out layer.

두 번째로, Dropout는 생략된 노드는 해당 연산에서 연산이 진행되지 않지만, 데이터 탈락은 탈락된 노드는 이전의 weight 및 bias값을 가지고 연산을 수행하게 된다.

따라서, 본 논문에서 제시하는 인공지능 학습모델에 추가된 데이터 탈락 과정은 데이터 탈락을 결정하는 연산이 단발성으로 추가되고, 그 외의 과정은 기존의 CNN(Convolutional Neural Network)기반 분류 과정과 동일한 과정을 거쳐 데이터를 분류하게 된다.

3. 사용자 데이터 상태값으로의 활용

중제목 2에서 언급한 사용자데이터 기반의 상태값은 사용자별로 필요한 값 k를 저장하여 활용하게 된다. 이 뿐만 아니라, 사용자에게 맞는 데이터를 상태값으로 저장하고, 이를 학습에 활용하여 모델링을 수행한다. 모델링 데이터는 IoT네트워크를 이용하는 사용자마다 맞춤형으로 모델링하게 되며, 이를 구현하기 위해 python tensorflow의 keras 라이브러리를 활용하여 코드를 구현하였다.

4. Keras를 활용한 인공지능 모델 구현

표 1은 본 과제의 CNN 모델 레이어 구성 관련 부분을 발췌한 코드이다. keras에서 제공하는 model 함수를 이용하여 1차원 convolution 및 pooling을 수행한 뒤, 개인 상태값 w_p를 로드하여 이를 이용한 데이터 탈락을 진행하였다. 해당 값을 다시 2번째 모델을 생성한 후 이후 과정을 수행하였다.

표 1. 데이터 탈락층이 추가된 인공지능 프로세스 코드
Table 1. AI learning code with data leave-out

```
#데이터 모델 생성
model = models.Sequential()

#Conv1d 및 pooling 수행
model.add(layers.Conv1D(128,(2, 2), activation='relu',
input_shape=(28, 28, 1)))
model.add(layers.MaxPooling2D((4, 4)))

#데이터 모델 가중치 추출
X, _, _, _ = train.test.split(X, y, size, state)

#데이터 탈락 과정
w_p = pd.read.csv('/users/user1.csv')
X = np.array(input.iloc[:,1:])
for i in range(X)
    if (X[i]!="data") < w_p[i][k] * w_p[i][weight]
        X[i]=undefined;

#2번째 데이터 모델 생성
#Conv2d 및 Pooling, dense, softmax 수행
model_2 = models.Sequential()
model.add(layers.Conv2D(64, (3, 3), activation='relu'))
model.add(layers.MaxPooling2D((2, 2)))
model.add(layers.Conv2D(64, (3, 3), activation='relu'))

#최종 데이터 도출
history_dict = history.history
accuracy = history_dict['accuracy'] # 정확도
fig, ax = plt.subplots(1, 2)
```

III. 시뮬레이션 결과

시뮬레이션은 표 2의 환경에서 수행되었다.

그림 3은 표 2의 시뮬레이션 환경을 기반으로 수행한 CNN 모델의 트레이닝-테스팅 데이터 손실율(에러율)을 나타낸 그래프이다. 적색-청색은 기존의 CNN 코드를 적용한 결과이고, 녹색-황색은 본 과제에서 제시한 데이터 탈락층이 추가된 CNN 모델 레이어다. 기존의 CNN 방식 대비 약 8~11%의 손실을 완화가 있음을 확인하였다.

표 2. 본 논문 시뮬레이션의 환경변수 설정
 Table 2. Parameter settings used in proposed simulations.

| 환경 | 변수값 |
|----------------------------------|---|
| OS | 리눅스 우분투 19.10 |
| 프레임워크 | TensorFlow 1.15.2 |
| 사용 언어 | Python 3.6 |
| 학습율 | 0.15 |
| 활성화함수 | ReLU |
| 시계열데이터 획득주기 | 60seconds |
| 시계열데이터 전송 임계값 λ_{det} | $0.01 * \text{abs}[-30,.]$ (이전 30개 데이터 값 평균의 1% 이상의 변동) |
| 시계열가중치 w_{ts} | $0 < w_{ts} < 0.5$ @ 08-17, |
| | $0 < w_{ts} < 1$ @ 17-22, |
| | $0 < w_{ts} < 0.1$ @ 22-08 (시간기준 최대값 설정) |

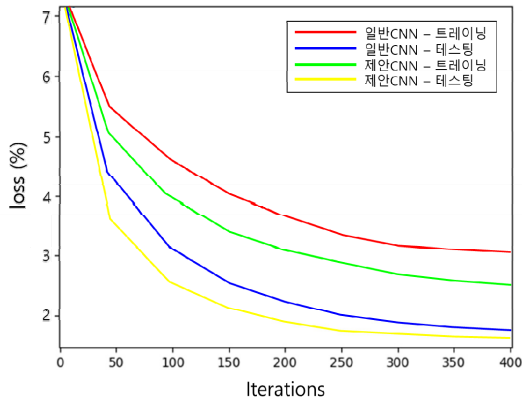


그림 3. 학습 횟수 대비 센서 트레이닝-테스팅 데이터 손실을 그래프.
 Fig. 3. training-testing data loss versus learning iteration for the proposed and conventional CNN scheme

IV. 결 론

본 논문에서는 IoT 시계열 데이터를 효과적으로 학습하기 위해 다양한 기법을 연구하였다. 시계열 데이터 학습 정확도 향상을 위해 단말 연산성을 이용한 IoT 센서의 시간가중치 전처리, 가중치 기반 데이터 탈락 인공 지능 연산, 사용자상태값을 통한 학습 모델 개인화를 제시하였다. 본 연구를 통해 기존 단순한 데이터 처리를 넘어 데이터 특성에 맞는 인공지능 학습 전략 수립을 위한 토대를 마련하였다.

향후 IoT 네트워크 및 빅데이터 연구에 맞춤형 인공 지능을 적극적으로 활용할 수 있는 방안을 제시하였다.

References

- [1] Erbo Shang, Xiaohua Liu, Hailong Wang, Yangfeng Rong, Yuerong Liu, "Research on the Application of Artificial Intelligence and Distributed Parallel Computing in Archives Classification", 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chengdu, China, pp. 1267-1271, 2019.
DOI: <https://doi.org/10.1109/IAEAC47372.2019.8997992>.
- [2] Pill-Won Park, "Text-CNN Based Intent Classification Method for Automatic Input of Intent Sentences in Chatbot", The Journal of Korean Institute of Information Technology, Vol. 18, No. 1, pp. 19-25, 2020.
DOI: <https://doi.org/10.14801/jkiit.2020.18.1.19>.
- [3] Jong-Bok Kim, Dong-Hwa Jang, Yang Kayoung, Kyeong-Seok Kwon, Jung-Kon Kim, Joon-Whoan Lee, "Recognition of Dog's Front Face Using Deep Learning and Machine Learning", Journal of the Korea Academia-Industrial Cooperation Society, Vol. 21, No. 12 pp. 1-9, 2020.
DOI: <https://doi.org/10.5762/KAIS.2020.21.12.1>
- [4] Ji-su Moon, Min-ji Kim, Young-kyu Lim, Ki-sok Kong, "Conversion Program of Music Score Chord using OpenCV and Deep Learning", The Journal of The Institute of Internet Broadcasting and Communication (IIBC), Vol. 21, No. 1, pp. 69-77, 2021.
DOI: <https://doi.org/10.7236/JIIBC.2021.21.1.69>.
- [5] Jie Yang, Yue Gao, Ying Ding, Yingyi Sun, Yang Meng, Wei Zhang, "Deep Learning Aided System Design Method for Intelligent Reimbursement Robot", in IEEE Access, Vol. 7, pp. 96232-96239, 2019.
DOI: <https://doi.org/10.1109/ACCESS.2019.2927499>.
- [6] Kazuo Okamura, Seiji Yamada, "Empirical Evaluations of Framework for Adaptive Trust Calibration in Human-AI Cooperation", IEEE Access, Vol. 8, pp. 220335-220351, 2020.
DOI: <https://doi.org/10.1109/ACCESS.2020.3042556>.
- [7] Daoqu Geng, Chengyun Zhang, Cengjing Xia, Xue Xia, Qilin Liu, Xinshuai Fu, "Big Data-Based Improved Data Acquisition and Storage System for Designing Industrial Data Platform", IEEE Access, Vol. 7, pp. 44574-44582, 2019.
DOI: <https://doi.org/10.1109/ACCESS.2019.2909060>.
- [8] Waldemar Hummer, Vinod Muthusamy, Thomas Rausch, Parijat Dube, Kaoutar El Maghraoui, Anupama Murthi, Punleuk Oum, "ModelOps: Cloud-Based Lifecycle Management for Reliable and Trusted AI", 2019 IEEE International Conference on Cloud Engineering (IC2E), Prague, Czech Republic, pp. 113-120, 2019.
DOI: <https://doi.org/10.1109/IC2E.2019.00025>.
- [9] Hui-Eun Lee, Jin-ku Choi, "Implementation of Smart Ventilation Control System using IoT and Machine

Learning”, The Journal of The Institute of Internet, Broadcasting and Communication, Vol. 20, No. 2, pp. 283-287, 2020.

DOI: <https://doi.org/10.7236/JIIBC.2020.20.2.283>.

저 자 소 개

김 진 영(Jin-Young Kim)



- 1998년 2월 : 서울대학교 전자공학과 공학박사
- 2001년 2월 : SK텔레콤 네트워크 연구소 책임연구원
- 2001년 3월 ~ 현재 : 광운대학교 전자융합공학과 교수

- 2019년 9월 ~ 현재 : ㈜스마트에버 연구소장
- 관심분야 : 차세대이동통신, 가시광통신, 전력선통신, 인공지능

심 이 삭(Isaac Sim)



- 2016년 2월 : 광운대학교 전자융합공학과 졸업
- 2016년 3월 ~ 현재 : 광운대학교 전자융합공학과 석박통합과정
- 관심분야 : 전력선통신, 무선 에너지 하비스팅, 인공지능

윤 성 훈(Sung-Hoon Yoon)



- 2000년 8월 : 서울과학기술대학교 재료공학과 졸업
- 2003년 2월 : 한양대학교 재료공학과 공학석사
- 2006년 1월 ~ 2010년 9월 : ㈜태영건설
- 2010년 9월 ~ 2017년 12월 : ㈜대우건설

- 2018년 1월 ~ 현재 : ㈜코젠
- 2018년 3월 ~ 현재 : 상명대학교 에너지그리드학과 박사과정
- 관심분야 : 디지털통신, 무선 측위 시스템, 전력전통신, 인공지능

※ 본 연구는 2020년도 중소벤처기업부의 중소기업기술혁신개발사업 지원에 의한 연구임 [S2855466].