

차량 센서 데이터 조합을 통한 딥러닝 기반 차량 이상탐지

김송희, 김선혜, 윤병운*
동국대학교 산업시스템공학과

Deep Learning-Based Vehicle Anomaly Detection by Combining Vehicle Sensor Data

Songhee Kim, Sunhye Kim, Byungun Yoon*
Department of Industrial & Systems Engineering, Dongguk University

요약 4차산업혁명 시대에는 대량의 데이터를 학습하여 예측과 분류의 정확성을 향상시킬 수 있는 인공지능의 활용이 핵심적이다. 그러나, 기존 이상탐지를 위한 방법은 제한된 데이터를 다루는 전통적인 통계 방법에 의존하고 있어, 정확한 이상탐지가 어렵다. 그러므로, 본 연구는 인공지능 기반 이상탐지 방법을 제시하여 예측 정확도를 높이고, 새로운 데이터 패턴을 정의하는 것을 목적으로 한다. 특히, 자동차의 경우 공회전 기간의 센서 데이터가 이상 탐지에 활용될 수 있다는 관점에서 데이터를 수집하고 분석하였다. 이를 위해, 예측 모델에 입력되는 데이터의 적정 시간 길이를 결정하고, 공회전 기간 데이터와 전체 운행 데이터의 분석 결과를 비교하며, 다양한 센서 데이터 조합에 의한 최적 예측 방법을 도출하였다. 또한, 인공지능 방법으로 선택된 CNN의 예측 정확성을 검증하기 위해 LSTM 결과와 비교하였다. 분석 결과, 공회전 데이터를 이용하고, 공회전 기간보다 1.5배 많은 기간의 데이터를 이용하며 LSTM보다는 CNN을 활용하는 것이 더 좋은 예측결과를 보였다.

Abstract In the Industry 4.0 era, artificial intelligence has attracted considerable interest for learning mass data to improve the accuracy of forecasting and classification. On the other hand, the current method of detecting anomalies relies on traditional statistical methods for a limited amount of data, making it difficult to detect accurate anomalies. Therefore, this paper proposes an artificial intelligence-based anomaly detection methodology to improve the prediction accuracy and identify new data patterns. In particular, data were collected and analyzed from the point of view that sensor data collected at vehicle idle could be used to detect abnormalities. To this end, a sensor was designed to determine the appropriate time length of the data entered into the forecast model, compare the results of idling data with the overall driving data utilization, and make optimal predictions through a combination of various sensor data. In addition, the predictive accuracy of artificial intelligence techniques was presented by comparing Convolutional Neural Networks (CNN) and Long Short Term Memory (LSTM) as the predictive methodologies. According to the analysis, using idle data, using 1.5 times of the data for the idling periods, and using CNN over LSTM showed better prediction results.

Keywords : Anomaly detection, CNN, LSTM, Vehicle, Deep learning

본 논문은 2020년 대한민국 교육부와 한국연구재단의 인문사회분야 중견연구지원사업의 지원을 받아 수행된 연구임 (NRF-2020S1A5A2A01040428).

*Corresponding Author : Byungun Yoon(Dongguk University)

email: postman3@dongguk.edu

Received December 18, 2020

Accepted March 5, 2021

Revised January 26, 2021

Published March 31, 2021

1. 서론

최근 4차 산업혁명 시대가 도래하여, 대량의 정보를 학습하여 예측 및 분류의 정확도를 향상시키는 인공지능이 각광받으면서, 다양한 분야에서 이를 적용시키려는 움직임이 보이고 있다. 각 산업 분야에서 축적되어온 데이터를 활용하여 산업 내 다양한 시스템을 효율적으로 발전시키기 위해 인공지능과 같은 방법론들을 활용하고 있다. 최근에는, 이미지 분류, 언어 인식, 물체 인식, 이상 탐지와 같이 데이터의 특징을 추출하여 이를 토대로 공통점을 찾아줄 수 있는 방법 등이 개발되면서 적용 분야를 확장시키고 있다. 많은 연구자들과 기업들은 인공지능을 산업에 적용할 수 있는 분야들을 모색하고 있으며, 특히 자동차 산업에서는 자율주행 자동차, 자동 주차 시스템, 차량 이상 탐지 등에 주로 활용되고 있다.

이상 탐지는 이상한 것을 탐지하는 즉, 보편적이지 않은 것을 찾아내는 것으로 정의할 수 있다. 이상탐지는 비즈니스 분석, 컴퓨터 시스템 관리, 공정설비와 같은 분야에서 적용되고 있으며 최근에는 자동차 산업에서의 이상 탐지가 활발하게 연구되고 있다. 사용자가 차량의 고장을 진단하기 위한 방법으로 경험적인 방법 이외에 사용할 수 있는 방법이 제한적이므로 사용자가 이상을 탐지하기에 많은 시간과 노력이 필요하며 현실적으로 거의 불가능하다고 볼 수 있다. 따라서 이상을 탐지하기 위해서 전문가의 진단을 받는 것이 필요하며 전문가도 이상 탐지를 위한 규칙들을 확실히 하기 위해 차량의 내부를 직접 확인하는 것이 필수적이다. 뿐만 아니라, 차량의 이상을 탐지할 수 있는 특정한 패턴이 존재하지만, 이러한 모든 패턴을 전통적인 규칙 기반의 방법으로는 확인하기 어렵기 때문에, 이상 탐지를 위한 다양한 방법이 적용되어야 할 필요가 있다.

기존 연구들은 다양한 이상탐지 방법론을 제시하고 있지만, 전통적인 규칙기반(rule-based) 이상탐지는 이상을 탐지하는데 시간이 오래 걸리며, 실시간으로 이루어지기 힘들다는 한계점과 여러 개의 데이터를 한 번에 처리하기 어렵다는 한계점을 가지고 있다[1]. 뿐만 아니라, 클러스터링을 활용한 이상탐지에 대한 연구도 존재하지만, 최적화가 필수적이며 같은 분포를 가진 큰 데이터 셋이 필요하다는 한계점이 존재한다[2]. 이러한 한계점들을 극복하기 위해 기존 차량 이상탐지 분야에서 필요로 했던 인간이 발견하기 어려운 정보 내에 패턴 탐지와 분포가 상이한 데이터에서 이상의 징후 탐지를 인공지능 기술의 활용으로 식별하여 이상탐지를 가능하게 하고자 한다. 또

한, 하나의 센서를 사용하였던 규칙기반 이상탐지에서 탈피하여 다양한 데이터의 사용의 가능성을 보여주고자 한다. 더불어, 기존의 수동적 탐지나, 전조 증상을 활용한 방법들을 자동화시킴으로써 시스템의 유지 관리 및 효율성을 증대시키고자 하며, 기존 시스템에서 발견하기 어려웠던 이상 징후를 탐지하며, 수동적인 업데이트를 통해 지속적인 성능 개선이 가능할 것이다.

이를 위해, 본 논문에서는 차량의 센서 데이터를 데이터 소스로 하여 딥러닝 기반 분류 모델을 구축하여 높은 성능의 이상 탐지 시스템을 만들고자 한다. 또한, 적용할 수 있는 대표적인 딥러닝 방법 중 CNN의 적용을 제시하고 이를 LSTM 적용 결과와 비교 분석하여 보다 정확한 이상 탐지 시스템을 구축하고자 한다. 본 논문에서는 센서 데이터를 데이터 셋으로 하여 딥러닝 기반 분류 모델을 구축하여 차량 이상을 탐지할 수 있는 시스템을 만드는 것을 목표로 하며, 딥러닝 기법들을 활용하여 보다 정확한 이상탐지 시스템을 구축하고자 한다. 또한 다양한 센서들의 활용 가능성을 확인하여 다양한 조건과 환경에 정확도 높은 이상 탐지가 가능한 시스템을 구축하고자 한다. 본 논문의 구성은 2장에서 본 연구의 배경이론, 3장에서는 활용한 데이터와 모델, 4장에서는 결과, 5장에서는 결과해석 및 시사점, 6장에서는 결론을 제시하는 형식으로 구성하였다.

2. 배경이론

2.1 CNN(Convolutional Neural Networks)

본 논문에서는 데이터의 특징을 파악하여 이상을 탐지하는 방법 중 하나로 CNN 알고리즘을 활용한다. CNN 알고리즘은 딥러닝(인공신경망) 알고리즘의 한 종류이며, 합성곱층(Convolution Layer)과 풀링층(Pooling Layer) 두 종류 층을 포함하는 순방향 신경망(Feedforward neural networks)이다[3]. 이는 전처리 과정을 추가한 다층 퍼셉트론의 한 종류로 2차원 데이터의 입력과 훈련에 용이하고 적은 매개변수만이 필요하다는 장점을 가지고 있다. CNN 알고리즘의 구조는 Fig. 1과 같이 합성곱층(Convolutional Layer)과 풀링층(Pooling Layer) 그리고 완전연결층(Fully-Connected Layer)으로 구성되어 있으며, 입력 층 이후에 합성곱층과 풀링층이 반복되며 마지막 첫번째 또는 두번째 층에서 완전연결층으로 진행하도록 구성되어있다. CNN 알고리즘은 특징을 직접 학습하기 때문에 특징을 수동으로 추출해야 할 필요

가 없다는 장점이 있으며 이미지와 영상 및 음성 인식과 같은 분야에서 좋은 성능을 보이고 있다.

CNN은 이미지 분류 기반으로 하여 확장한 모델을 사용하기도 한다. 센서 데이터를 활용한 CNN 모델에서는 CNN을 활용하여 회전 기계의 고장 진단을 위한 모델을 개발하였다. 이 모델에서는 입력층을, 가로축은 시간, 세로축은 각각 센서의 값으로 2차원으로 구성하였으며, 최종적으로 99.41%의 정확도를 확인하였다[4]. 또한 국내에서는 CNN 기반 소음 센서를 이용한 원동 구동장치 진단시스템을 제안하였으며 실제 적용될 현장에서 발생할 수 있는 소음의 예측 외의 소음환경에 유연하게 대처하기 위해 CNN을 적용하여 음성신호를 스펙트로그램(Spectrogram)으로 추출함으로써 이미지형태로 입력 데이터를 구성하였다. 본 모델을 통해 94.6%의 정확도를 확인하였다[5].

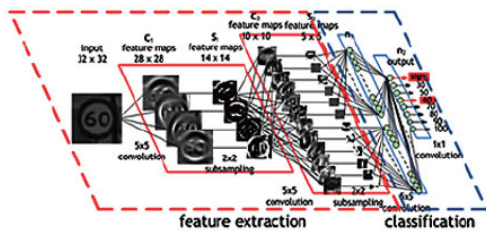


Fig. 1. Overview of CNN

2.2 LSTM(Long Short Term Memory)

LSTM 알고리즘은 RNN(Recurrent Neural Networks)의 확장 모델로, RNN 알고리즘은 순환 신경망의 이름으로 히든 노드가 방향을 가진 엣지로 연결되어 순환 구조를 이루는 인공신경망의 한 구조이다. 그러나, RNN 알고리즘은 장기 의존성(Long-Term Dependency) 문제를 가지고 있다. 다루는 데이터의 양이 클 경우에는 과거의 정보를 기억하는데 한계를 가지기 때문에 힘들다는 한계점을 가지고 있다. 따라서 이를 보완하고자 RNN의 주요 모델인 LSTM모델을 제안하였으며 이는 장기 의존성 문제를 해결하였다[6]. LSTM 모델의 구조는 Fig. 2와 같으며 총 4개의 게이트(gate)로 이루어져 있다. 학습하는 과정은 가장 먼저 망각 게이트(Forget Gate)에서 과거 정보를 버릴지 결정하게 되며 그 이후에 입력 게이트(Input Gate)로 저장된 정보들이 이동하며 게이트에서 현재 정보를 저장할 지 결정한다. 마지막으로 출력게이트(Output Gate)로 이동하여 어떤 값을 출력할지 결정되는 과정으로 학습이 이루어진다. LSTM은 이전의 데이터

를 기억하여 활용한다는 점에서 자연어 처리 및 시계열 데이터 처리에서 좋은 성능을 보이고 있다.

LSTM 알고리즘은 이상 예측에서 자주 사용되고 있다. 시계열 데이터를 활용한 LSTM 이상예측 알고리즘을 제안한 Malhotra의 연구에서는 LSTM 알고리즘을 활용하여 EncDec-AD 알고리즘을 개발하여 이를 시계열 센서 데이터에 적용시켰다[7]. LSTM encoder는 입력된 시계열 데이터를 고정된 길이의 벡터로 변형하여 입력받고, LSTM decoder는 입력 받은 데이터 형태로 현재의 상태를 구성하여 학습이 진행된다. 또한 국내에서는 LSTM 기반 미세먼지 농도 예측 알고리즘을 제안하였으며 데이터를 일정한 범위로 가지고 있게 하면서 이동하는 방법인 슬라이딩 윈도우(sliding window) 방법을 기반으로 학습시켜 높은 정확도가 나오는 것이 확인하였다[8].

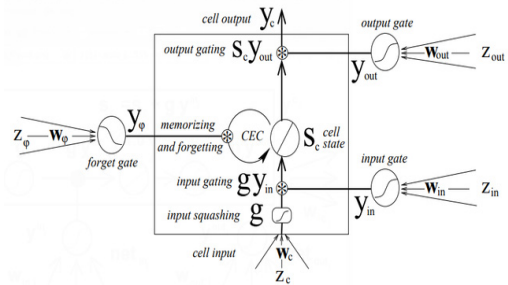


Fig. 2. Structure of LSTM

2.3 이상탐지(Anomaly Detection)

본 연구에서는 센서 데이터들을 기반 데이터로 하여 차량의 이상탐지를 하고자 한다. 이상탐지란, 예상된 행동에 부합하지 않는 특이한 패턴을 식별하기 위해 사용하는 방법이며 일반적으로 Novelty Detection, Outlier Detection이라고 불리기도 한다. 이상탐지는 크게 비정상 데이터의 사용여부와 정상과 비정상에 대한 라벨의 유무에 따라 지도 이상탐지(Supervised Anomaly Detection), 반지도 이상탐지(Semi-supervised Anomaly Detection), 비지도 이상탐지(Unsupervised Anomaly Detection)로 분류된다. 지도 이상탐지는 주어진 학습 데이터 셋에 정상 샘플과 비정상 샘플의 데이터와 라벨이 모두 존재하는 경우로, 다른 방법들에 비해 정확도가 높은 특징을 가진다. 반지도 이상탐지는 정상인 데이터만을 가지고 학습하는 방법으로 auto encoder[9]를 활용하거나 GAN을 활용하는 방법론[10]등 활발한 연구가 이루어지고 있지만, 지도 이상탐지 방법론과 비교했을 때 상대적으로 정확도가 떨어진다. 대표적인 반지도 학습이상탐지 모델

인 서포트 벡터머신(Support Vector Machine: SVM)은 특징 공간에서 정상 데이터를 둘러싸는 가장 작은 공간(구)을 찾고, 해당 경계 면을 기반으로 이상치를 탐지하는 방법으로 이상의 라벨(label)이 없을 때 이상탐지가 가능하도록 설계되었다. 데이터의 라벨이 부족할 때 이상 탐지가 가능하다는 장점이 있는 반면, 데이터의 패턴의 변화나 데이터 탐색 방법 및 설정 변수에 따라 정확도가 달라진다는 한계가 있다. 비지도 이상탐지는 데이터의 정상/비정상의 라벨이 없는 상태로 학습을 진행하는 방법으로, Auto encoder를 활용한 방법[11]이나, 주성분분석(Principal Component Analysis: PCA)을 사용하는 방법이 존재한다. 가장 대표적인 비지도 이상탐지 모델은 주성분분석 모델로, 데이터의 분포에 따라 분산을 최대로 하는 축을 찾아내는 방법이다. 데이터의 라벨이 불필요하다는 장점이 존재하지만, 데이터 분포가 비선형적이거나, 데이터의 성향이 벗어나는 경우 정확도가 낮다는 한계점이 있다. 또한, 비지도학습은 하이퍼 파라미터(Hyper Parameter)에 매우 민감하다는 한계점이 있다.

차량의 이상탐지는 다양한 분포를 가진 데이터인 다수의 운전자 데이터를 통해 이상탐지를 진행해야 하며, 차량 이상탐지는 이미 정의된 이상탐지를 위한 규칙들이 존재하기 때문에 높은 정확도를 위해 본 연구에서는 지도 이상탐지를 진행하였다. 또한 높은 정확도의 이상탐지의 가능성을 확인하기 위해 다양한 센서를 조합하여 학습도 함께 진행하였다.

이상탐지 방법은 대부분 시간의 흐름에 따라 달라지는 데이터인, 시계열 데이터를 사용하며, 컴퓨터분야[12], 의료분야[13], IoT분야[14] 등 다양한 분야에서 적극적으로 활용하고 있다. 최근 딥러닝을 활용한 이상탐지 및 이상예측을 활발하게 진행되고 있으며, 센서 데이터를 활용한 이상탐지로는 LSTM 알고리즘을 기반으로 이상을 예측하는 알고리즘을 LSTM-AD 알고리즘을 제안하였으며, 93% 이상의 정확도를 보였다[15]. 국내에서는 GAN 알고리즘(적대적 생성모델)을 활용하여 사내에서 발생하는 내부 위험을 탐지하는 알고리즘을 제안하였으며, 99%에 가까운 정확도를 탐지하였다[16].

3. 연구 프레임워크

본 연구는 Fig. 3과 같이 데이터베이스에 저장되어있는 센서 데이터를 사용하여 전처리를 진행한 후 시계열 데이터를 이미지화하는 과정을 수행하고, 딥러닝 방법론

을 거쳐 차량의 이상탐지를 감지한다. 본 장에서는 각 과정을 순차적으로 설명하고자 한다.

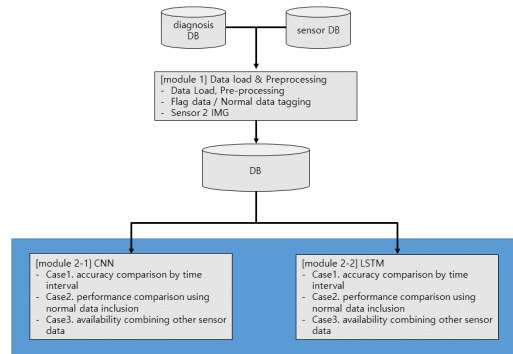


Fig. 3. Overall Framework

3.1 데이터 로드 및 전처리

차량의 센서 데이터는 차량 주행 데이터로 다양한 운전자가 차량을 운전한 데이터이며 기존의 규칙 기반 방법(Rule based method)으로 사용했던 형태로 구축되어있는 데이터베이스를 사용하였다. 본 연구에서 활용한 이상 진단을 위한 센서는 'LPG 연료 레일 압력'을 측정하는 센서이며, 기존의 규칙 기반 방법에서는 공회전 시위 센서에서 측정된 값이 214이상으로 5분 이상 지속되었을 때가 이상으로 정의된다. 사용한 센서 데이터는 8개의 차량을 통해 2018년 7월 20일 21시 20분 12초부터 2018년 11월 30일 16시 54분 05초까지 측정된 데이터로 수집 간격은 1초 간격으로 총 8826116건의 센서 데이터가 저장되어 있다. 사용한 진단 데이터는 실무에서 사용하는 규칙 기반의 이상탐지를 통해 작성된 데이터베이스이며, 총 5개의 카테고리(LPG 연료 레일 압력 이상, 연소분사 시간 및 실린더 이상, 산소센서전압S1(좌측) 이상, 산소센서전압(우측)이상, 배터리전압 이상)의 이상으로 구성되어 있다. 진단 기간은 2019년 1월 16일 10시 02분 58초부터 2019년 1월 16일 10시 14분 27까지로 총 진단된 데이터는 12816건이다. 수집된 데이터를 본 연구에 적용하기 위해서 데이터에 이상에 대한 라벨을 작성하는 전처리 과정이 필요하다. 이는 실무에서 직접 작성한 규칙들을 기반으로 생성되었다. 라벨이 생성된 이후에는 각 케이스에 맞는 데이터로 전처리가 진행된다. 기존의 규칙기반 이상탐지에서 사용했던 공회전 데이터만을 사용하기 위해 공회전 데이터만을 추출하는 과정이 포함된다. 연구에서 공회전은 냉각 수온이 70℃ 이상인 시점, 차량 속도가 0 km/h 이후 30초 이상 지난 시점,

차량 속도가 0km/h보다 커지기 6초 이전의 시점으로 정의하였다. 또한, 전처리 과정에서 센서 데이터를 딥러닝 분야에서 높은 성능을 보이고 있는 이미지 분류에 적용할 수 있도록 센서 데이터를 이미지화하는 Sensor2IMG 과정을 수행하였다. 이는 정확도 높은 딥러닝 기반 이상 탐지를 위한 것뿐만 아니라, 사람이 확인할 수 없는 새로운 특징이나, 센서의 조합 등과 같은 새로운 패턴을 정의가 가능할 수 있도록 진행하였다.

3.2 딥러닝 기반 이상탐지 수행

센서 데이터를 이미지화한 데이터를 기반으로 딥러닝 방법론인 CNN 방법론을 사용하여 이상탐지를 수행하고, 기존에 이상탐지에 활용되는 LSTM 방법론의 결과와 비교한다. 본 연구에서는 딥러닝 방법론을 활용한 높은 성능 차량의 이상탐지를 목적으로 하고 있다. 따라서 본 연구에서는 다양한 케이스들을 설계하여 비교 분석하고자 하였으며 최근 이상탐지 및 이미지 분석에서 높은 성능을 보이고 있는 CNN과 LSTM과 같은 딥러닝 모델을 사용하여 이상탐지를 진행한다. 딥러닝 기반 이상탐지 수행의 결과를 비교하기 위해 CNN과 LSTM 알고리즘 모두 3가지 모듈에 대한 분석을 수행하였으며, 각 분석 모듈은 이미지 데이터 변환시켜 진행하였다. 시간 축의 크기 따른 정확도 비교, 공회전 데이터만이 아닌 전체 주행 데이터 사용에 따른 정확도 비교, 이상을 탐지하는 하나의 센서만이 아닌 다른 센서를 함께 사용했을 때에 따른 정확도 비교를 위해 설계되었다.

첫번째 분석 모듈로, Sensor2IMG시 시간 축의 크기에 따른 정확도 비교에서는 시계열인 센서 데이터를 이미지로 변환할 때, 시간 간격을 어떻게 설정하느냐에 따라 이미지에서 담을 수 있는 데이터의 양이 달라지므로 정확도에 영향을 미칠 수 있다고 판단하여 시간을 실무에서 고장을 탐지하는 규칙에서 시간기준의 1배, 1.2배, 1.5배를 수행하였다. 정의된 규칙 기반으로 Fig. 4와 같이 x축을 시간으로 두고, y축은 센서에서 감지한 값을 나

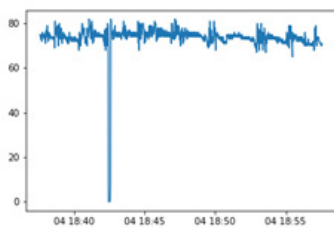


Fig. 4. Sensor2IMG Result

타내는 꺾은선 그래프를 그려 이미지화를 진행하였다. 본 연구에선 이 과정을 Sensor2IMG 과정이라고 명명하였으며, 모든 데이터를 입력을 위한 데이터로 변형하였다.

두번째 분석 모듈에서는, 공회전 데이터만이 아닌 전체 주행 데이터 사용에 따른 정확도 비교를 수행하였으며, 실무에서 사용하는 규칙이 공회전시에만 센서의 데이터를 확인하였다. 이는 전체 데이터를 사용할 수 없어 자동화하는데 어려웠기 때문에 딥러닝 방법론을 사용하였을 때 전체 주행데이터를 사용할 수 있을지에 대한 여부를 확인하였다. 세번째 분석 모듈은 이상을 탐지하는 하나의 센서만이 아닌 다른 센서를 함께 사용했을 때에 따른 정확도 비교에서는 이상탐지를 직접적으로 보여주는 센서만이 아닌 이와 연관이 될 가능성이 있는 센서들을 전문가가 선정하여, 이상탐지하는 센서와 함께 사용하여 정확도를 높일 수 있을지에 대한 여부를 확인하기 위해 수행하였다.

본 연구에서 CNN 모델은 보편적으로 이미지 분류에서 높은 정확도를 보였던 CNN 알고리즘을 사용하였으며, 모델의 도식화는 Fig.5 와 같으며, 구조는 3개의 레이어로 구성되어 있으며, 세부 파라미터는 Table 1과 같다. LSTM모델은 가장 기본 LSTM 알고리즘을 사용하였으며, 구조는 Fig. 6과 같다.



Fig. 5. Structure of CNN model

Table 1. Structure of CNN model

	Layer 1	Layer2	Layer3
Filters	16	64	64
Kernel size	(3,3)	(3,3)	(3,3)
Activation function	ReLU	ReLU	ReLU
Dropout	0.25	0.25	0.25
Pooling	Max (2,2)	Max (2,2)	Max (2,2)

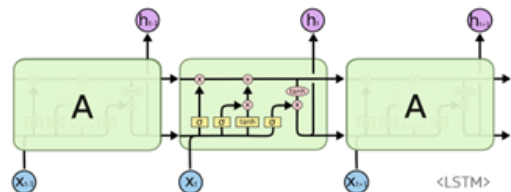


Fig. 6. Structure of LSTM model

4. 분석 결과

4.1 CNN 기반 이상탐지 결과

본 연구에서는 이미지 생성 모델에서 높은 성능을 보이고 있는 LeNet모델을 사용하여 지도 학습을 진행하였고, 모델 생성을 위해 100개의 epoch를 활용하였으며, batch size는 32로 학습을 진행하였다. 모델의 성능을 향상시키기 위해 3500개의 epoch를 사용하였고, batch size로는 10으로 학습시켰으며, Early stopping을 적용하였다. 활용한 하이퍼 파라미터(Hyper Parameter)는 CNN의 이미지 분류에서 가장 높은 성능을 보인 하이퍼 파라미터로 설정하였으며, 본 연구에서도 위와 같은 파라미터로 사용하였을 때 가장 높은 성능을 보였다. 학습의 검증을 위해 전체 데이터를 training set: test set = 7:3으로 나누어 학습과 테스트를 진행하였다. 전통적인 이상탐지는 분류(classification)모델과 같은 방법으로 성능을 계산하였다. 분류 모델의 성능 계산 방법으로는 정확도(Accuracy), 재현율(Recall), 정밀도(Precision), F1 점수 등을 주로 사용하고 있다. 재현율은 실제 참인 것들 중 예측으로 참이 되는 비율이며 이는 실제 참을 거짓으로 잘못 판단하면 위험이 큰 문제들일 때 주로 사용한다. 정밀도는 예측한 참인 것들 중 실제 참인 것의 비율이며 거짓을 판별하는 것이 중요한 경우에 자주 사용된다. 정확도는 전체 데이터 중 올바르게 예측한 비율을 구하는 방법으로 분류 모델이 얼마나 정확하게 판단하는지를 중점적으로 파악하며, F1 점수는 재현율과 정밀도의 조화 평균으로 작성된 지표이며, 이는 거짓을 더 잘 판별하는 방법으로 거짓을 얼마나 정확하게 판단하는지를 파악한다. 본 연구는 자동차 이상 탐지이기 때문에 고장인 경우 고장이 아니라고 판단하는 것이 가장 큰 문제이므로, 참을 정확하게 판단하는 것이 중요하다. 따라서 참을 잘 판단하는 지표인 정확도를 사용하였다.

4.1.1 데이터 변환 시 시간 축의 크기에 따른 정확도 비교

센서 데이터를 이미지화하여 삽입하기 위해 x축(시간 축)의 크기를 통일시켜야 모델에 데이터를 입력하기 용이하고, 높은 정확도를 얻을 수 있다. 이 과정을 Sensor2IMG 과정으로 명명하였으며 x축의 크기(시간)를 이상을 감지하는 시간 기준의 1배, 1.2배, 1.5배를 각각 수행하여 Fig. 7과 같이 이미지화를 진행하였다. 결과는 Table 2에서 제시한 바와 같이 x축의 크기를 1.5배

를 했을 때 가장 높은 정확도를 보였다. 따라서 이후 모든 분석모델에 대한 그래프는 시간 축의 1.5배를 한 결과를 사용하였다.

Table 2. Result of CNN case 1

	1 time	1.2 times	1.5 times
Accuracy	0.4776119	0.6865716	0.8059701

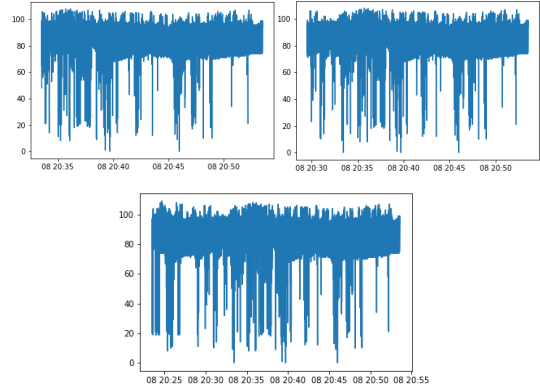


Fig. 7. Input Data of CNN case 1

4.1.2 전체 주행 데이터 사용에 따른 정확도 비교

기존 이상 감지를 위해 전통적인 방식인 공회전 데이터만을 활용하여 이상을 탐지하였으나, 자동적인 방법에서 활용하는데 한계점이 있을 수 있다고 판단하였다. 따라서 전체 주행 데이터를 사용할 수 있을지 확인하고자 하였다. 분석 결과 Table 3과 같이 공회전 데이터를 활용하였을 때 더 높은 정확도가 나타났다.

Table 3. Result of CNN case 2

	Idling data	All data
Accuracy	0.805970149	0.40299

4.1.3 다른 센서를 함께 사용했을 때에 따른 정확도 비교

사용한 하나의 센서가 아닌 다른 센서와의 조합을 통해 정확도를 증가시킬 수 있을 것으로 예상하였다. 조합에 사용된 데이터는 본 연구에서 사용한 LPG 연료 레일 압력 센서와 관련 가능성이 높은 센서 9개(엔진 회전수, 흡기압센서, 냉각수온센서, 산소 센서 전압 1(좌측에 위치), 산소 센서 전압 2(우측에 위치), 연소 분사시간, 에어컨 컴프레서, 타이머, ETC모터 듀티)를 선택하였으며 각

센서와 LPG 연료 레일 압력 센서를 함께 데이터로 활용하여 분석을 진행하였다. Table 4에서 제시한 바와 같이, 하나의 센서만을 사용하였을 때는 80.59%정도의 정확도를 나타냈지만 보다 다양한 센서의 조합으로 사용했을 때는 모두 85%이상의 정확도가 도출되었음을 확인하였다. 이 중에서는 산소센서전압의 센서와 함께 사용하였을 때 97.06%로 가장 높은 정확도를 나타내었다.

Table 4. Result of CNN case 3

	Engine RPM	Manifold pressure sensor	Cooling temperature sensor
Accuracy	0.8955	0.9522	0.9552
	Oxygen sensor voltage 1	Oxygen sensor voltage 2	Combustion injection time
Accuracy	0.9706	0.9701	0.9145
	air conditioner compressor	Timer	ETC motor duty
Accuracy	0.9552	0.8657	0.91045

4.2 LSTM 기반 이상탐지 결과

CNN 기반 이상탐지와 비교하기 위하여, LSTM 기반 이상탐지를 CNN과 동일한 3가지 분석모듈을 활용하여 진행하였다. 정확도도 CNN과 마찬가지로 전체 데이터를 training set: test set = 7:3으로 나누어 training과 test를 진행하였으며 정확도는 training으로 학습한 모델에 test 데이터를 입력하였을 때 정답을 맞히는 비율을 정확도로 정의하였다.

4.2.1 데이터 변환시 시간 축의 크기에 따른 정확도 비교

센서 데이터를 이미지화하여 삽입하기 위해 x축(시간 축)의 크기를 통일시켜야 모델에 데이터를 입력하기 용이하고, 높은 정확도를 얻을 수 있다. x축의 크기(시간)를 이상을 감지하는 시간 기준의 1배, 1.2배, 1.5배를 각각 수행하여 이미지화하였다. 결과는 Table 5에 정리하였으며, x축의 크기를 1.2배를 했을 때 가장 높은 정확도를 보였다. 따라서 이후 모든 case에 대한 그래프는 시간 축의 1.2배를 한 결과를 사용하였다.

Table 5. Result of LSTM case 1

	1 time	1.2 times	1.5 times
Accuracy	0.65	0.73	0.70

4.2.2 전체 주행 데이터 사용에 따른 정확도 비교

CNN에서와 마찬가지로 기존에는 이상 탐지를 위해 데이터 전처리를 활용하여 공회전 데이터만을 활용하여 이상을 탐지하였다. 따라서 전체 주행 데이터를 활용하는 것이 데이터 분석 시 더 유용할 수 있다고 판단하여 전체 주행 데이터의 활용 가능성을 확인하였다. 분석 결과 전체 데이터를 활용하였을 때 더 높은 결과가 도출되었다.

Table 6. Result of LSTM case 2

	Idling data	All data
Accuracy	0.73	0.79

4.2.3 다른 센서를 함께 사용했을 때에 따른 정확도 비교

Table 7에 제시한 바와 같이, LSTM 알고리즘에서는 센서의 조합을 사용하였을 때와 하나의 센서를 사용하였을 때의 정확도인 73%와 차이가 거의 존재하지 않았으며, 산소센서전압2를 사용하였을 때에는 92%의 높은 정확도가 도출됨을 확인하였다. 이는 LSTM 모델에서는 다양한 센서의 조합을 사용하는 것이 정확도를 높일 수는 있지만, 유의미하지는 않았다.

Table 7. Result of LSTM case 3

	Engine RPM	Manifold pressure sensor	Cooling temperature sensor
Accuracy	0.74	0.73	0.71
	Oxygen sensor voltage 1	Oxygen sensor voltage 2	Combustion injection time
Accuracy	0.71	0.92	0.66
	air conditioner compressor	Timer	ETC motor duty
Accuracy	0.73	0.73	0.65

5. 결과 해석 및 시사점

5.1 학습 방법의 비교를 통한 검증

본 연구에서는 같은 데이터 입력을 활용하여 CNN 적용결과를 LSTM 적용결과와 비교함으로써, CNN 기반 이상탐지 방법을 검증하고자 하였다. 전체적으로 이미지 분류에서 좋은 성능을 보였던 CNN은 80% 이상의 정확도를 보이는 반면 LSTM은 73%정도로 CNN이 보다 더 높은 결과를 도출했다. 이는 CNN 모델의 강점인 데이터

의 특징을 뽑아내는 데 높은 정확도를 보인다는 점에 따라 나온 결과라고 해석된다. 입력 데이터를 이미지로 변환하여 사용하였으므로 CNN에서 더 높은 결과가 도출되었다. LSTM 모델은 시계열 데이터를 활용하였을 때 높은 결과를 도출한다는 모델의 특징이 있다. 따라서 본 연구에서 사용한 데이터를 LSTM에 적합하도록 입력 데이터를 조정한다면 높은 정확도를 도출할 수 있을 것으로 예상된다.

5.2 사용 데이터 별 비교

본 연구에서는 분석 모듈 별로 정확도를 비교하였다. CNN과 LSTM 활용하였을 때 공회전 이외의 일반 데이터를 활용하는데 정확도 차이가 확연히 존재하였다. CNN을 활용하였을 때에는 공회전 데이터만을 활용했을 때 더 높은 정확도가 도출되었다. 이는 일반 데이터를 사용하게 되면 이상을 탐지하기 위한 데이터 이외의 불순한 데이터가 함께 사용되는데 이 때, 불순한 데이터를 걸러내기 위한 데이터가 충분하지 않았기 때문에 낮은 정확도가 도출되었다고 판단되었으며, 일반 주행 데이터를 활용하기 위해서는 불순 데이터를 무시하기 위해 더 많은 데이터를 활용해야 할 것으로 예상된다. 반면 LSTM을 활용하였을 때는 일반 데이터를 활용했을 때 더 높은 정확도가 도출되었다. 이는 LSTM 모델의 특징이 반영된 결과이다. LSTM은 이전의 데이터를 기억하여 다음 학습에 반영하는 특징을 가지고 있다. 따라서 일반 데이터로 학습시키게 되어도 불순 데이터를 무시하기 위한 데이터가 충분히 확보가 되었다고 할 수 있다. 따라서 일반 데이터를 활용하여 실시간으로 이상 탐지를 위해서는 CNN알고리즘보다는 LSTM을 활용하는 것이 효과적이라는 결과를 도출할 수 있었다.

CNN과 LSTM을 각각 LPG 연료 레일 압력 센서와 다른 센서의 조합으로 학습하였을 때는 두 모델 모두 LPG 연료 레일 압력 센서만을 사용하는 것보다는 다른 센서들과의 조합을 하여 사용하는 것이 높은 정확도가 도출되었다. 이는 하나의 센서만으로 이상을 탐지하는 것 보다는 다른 센서를 함께 확인하는 것이 이상탐지를 위한 효율적인 방법이라는 것을 확인하였다. 즉, 하나의 센서만으로 빠른 시간 내 차량의 이상을 확인하는데 한계가 있으며 이러한 한계를 다른 센서들이 보완할 수 있음을 확인하였다. 이는 규칙기반방식에서 나아가 인공지능 방식으로 새로운 패턴이 추출되었으며, 더 방대한 데이터를 입력으로 진행된다면, 새로운 패턴 중 하나로 정의할 수 있을 것으로 예상된다.

6. 결론

본 연구는 센서 데이터를 기반으로 딥러닝의 방법론 별, 센서 데이터의 다양한 변형을 통해 효과적인 이상탐지 시스템을 만들기 위한 실험을 진행하였으며 다음과 같은 기여점이 있다. 첫째, 시간 간격 별 이상 탐지 정확도 차이가 존재함에 따른 모델의 입력에 적합한 데이터 변형 방법을 확인하였다. 원래의 이상탐지 시스템은 시간 기반이 아닌 규칙 기반 방법론을 주로 이용하여 기존의 방법에서 시간 블록 방법으로 딥러닝 모델의 입력에 적합하게 변형하면서 배제된 데이터가 존재한다. 따라서 본 연구에서는 시간 블록을 만드는 간격에 따른 실험을 진행하였다. CNN 알고리즘은 규칙 기반 방법에서 고장을 모니터링 하는 시간의 1.5배를 하였을 때 높은 성능을 보였으며, LSTM 알고리즘은 1.2배 하였을 때 높은 성능을 보였으므로 시간 축을 이상 탐지 기준 조건보다 더 큰 시간을 기준으로 작성하였을 때 좋은 성능을 보이는 것을 확인하였다. 따라서 딥러닝과 같은 자동화 방법을 활용하였을 때, 적합한 데이터로 변형하기 위한 방법에 대한 시도는 의미를 가진다. 둘째, 전체 주행데이터를 사용하기 위해서는 LSTM과 같이 시계열 데이터 처리에 용이한 알고리즘을 사용하는 것이 의미가 있음을 확인하였다. CNN 알고리즘을 사용하였을 때 전체 주행 데이터보다 공회전 데이터만을 사용하였을 때 높은 정확도를 보인 반면에, LSTM 알고리즘은 전체 데이터를 사용하였을 때 더 높은 정확도를 보였으므로, 전체 데이터 활용 시에는 시계열 데이터에 적합한 모델을 활용하는 것이 의미가 있음을 확인하였다. 따라서 향후에는 자동적인 이상탐지를 위해 LSTM모델을 활용하는 연구가 진행될 수 있을 것이다. 셋째, 하나의 센서가 아닌 다양한 센서의 조합에 따른 정확도는 대체적으로 하나의 센서가 아닌 하나 이상의 센서를 사용하였을 때 높은 정확도를 보였으며, 이는 차량의 고장이 하나의 센서에서만 나타나지 않는다는 것을 의미하며, 높은 정확도의 이상탐지를 위해서는 다양한 센서들을 확인하는 것이 유의미하다는 결과를 얻었다. 또한 이를 통해 향후에는 새로운 센서 조합의 이상탐지 규칙을 뽑아낼 수 있을 것이다. 넷째, 딥러닝 알고리즘을 비교하였을 때 데이터의 특징을 뽑아내는 것에 특화된 CNN 알고리즘이 시계열 데이터에 특화된 LSTM 알고리즘을 사용하는 것 보다 유의미한 차이가 있음을 확인하였다. 마지막으로, 딥러닝 기반 이상탐지 모델을 사용하여 기존의 이상탐지 모델의 한계점들을 보완할 수 있었다. 기존의 SVM이나 PCA 모델 등의 한계점이었던 다양

한 데이터 분포를 가진 이상탐지가 어렵다는 점과 다양한 특징을 가진 데이터에서의 분석이 어렵다는 점을 해결하기 위해 다양한 특징을 반영할 수 있는 최신의 모델인 CNN과 LSTM과 같은 딥러닝 모델을 사용하였다. 다양한 센서들의 조합들을 사용한 이상탐지를 통해 다양한 특징을 가진 데이터의 가능하게 하였으며, 실제 주행 데이터를 사용하여 다양한 데이터 분포를 가진 경우의 이상탐지가 가능함을 확인하였다. 더불어, 기존의 규칙기반 이상탐지의 수동적 탐지와 전조 증상을 자동화의 가능성을 보였다. 또한 향후에 모델의 수동적인 유지보수를 진행한다면 높은 성능의 이상탐지가 가능할 수 있음을 보였다.

본 연구의 다양한 기여점에도 한계점이 존재한다. 우선, 본 연구에서 활용한 CNN 모델은 일반적으로 99% 이상의 정확도를 보일 수 있는 모델이다. 하지만 이보다 낮은 수준의 정확도를 보이므로 아직 완벽하게 적용하는 데는 한계가 있으므로 기존의 전통적인 방법과 함께 활용하는 수준에서 적용할 필요가 있다. 또한, 이후에 다양한 데이터 변형 기법들을 활용하여 데이터를 풍부하게 하여 이상탐지 모델을 학습시키게 된다면 본 연구보다 더 높은 정확도의 모델을 구축할 수 있을 것으로 예상된다. 마지막으로, 전체적으로 높은 이상탐지 정확도에도 불구하고 일반 주행 데이터를 활용하는 데 안정화가 필요하다. 이는 이상탐지를 위한 충분한 데이터의 부재와 기존의 규칙기반 방법에서 시간 블록 방법으로 변형하면서 도출되는 문제점으로, 데이터 샘플링을 통해 더 많은 데이터로 변형하거나 딥러닝 기반 이상 탐지를 위한 새로운 고장에 대한 기준이 설정된다면 좀 더 높은 정확도의 실시간 이상 탐지가 가능할 것으로 예상된다.

References

- [1] M., Heinrich, A., Gözl, T., Arul, S. Katzenbeisser, "Rule-based Anomaly Detection for Railway Signalling Networks", arXiv preprint arXiv:2008.05241, 2020.
- [2] H. Sarmadi, A. Karamodin, "A novel anomaly detection method based on adaptive Mahalanobis-squared distance and one-class kNN rule for structural health monitoring under environmental effects.", *Mechanical Systems and Signal Processing*, Vol. 140, 106495, 2020. DOI: <https://doi.org/10.1016/j.ymssp.2019.106495>
- [3] S. Lawrence, C.L. Giles, A.C. Tsoi, A.D. Back, "Face recognition: A convolutional neural-network approach." *IEEE transactions on neural networks*, Vol. 8, No. 1, pp. 98-113, 1997. DOI: <https://doi.org/10.1109/72.554195>
- [4] M. Xia, T. Li, L. Xu, L. Liu, C.W. de Silva, "Fault Diagnosis for Rotating Machinery Using Multiple Sensors and Convolutional Neural Networks", *IEEE/ASME Transactions on Mechatronics*, Vol. 23, No. 1, pp. 101-110, 2018. DOI: <https://doi.org/10.1109/TMECH.2017.2728371>
- [5] S.H. Lee, J.S. Kim, B.B. Shin, "CNN-Based Noise System for Motorized Driving Unit Fault Causes Classification System", *Proceedings of the Korean Computer Information Conference*, Vol. 26, No. 1, pp. 7-8, 2018.
- [6] F.A. Gers, J. Schmidhuber, "Recurrent nets that time and count", In *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks, Neural Computing: New Challenges and Perspectives for the New Millennium*, Vol. 3, pp. 189-194, 2000. DOI: <https://doi.org/10.1109/IJCNN.2000.861302>
- [7] P. Malhotra, L. Vig, G. Shroff, P. Agarwal "Long short term memory networks for anomaly detection in time series", In *Proceedings, Presses universitaires de Louvain*, p.89, 2015.
- [8] S.G. Kim, T.I. Oh, "Real-time PM10 Concentration Prediction LSTM Model Based on IoT Streaming Sensor Data" *Korea Journal of the Korean Society of Science and Technology*, Vol. 19, No. 11, pp. 310-318, 2018. DOI: <https://doi.org/10.5762/KAIS.2018.19.11.310>
- [9] B. Zong, Q. Song, M.R. Min, W. Cheng, C. Lumezanu, D. Cho, H. Chen, "Deep autoencoding gaussian mixture model for unsupervised anomaly detection", In *Proceeding of International Conference on Learning Representations*, 2018.
- [10] D. Li, D. Chen, J. Goh, S.K. Ng, "Anomaly detection with generative adversarial networks for multivariate time series", arXiv preprint arXiv:1809.04758, 2018.
- [11] C. Baur, B. Wiestler, S. Albarqouni, N. Navab, "Deep autoencoding models for unsupervised anomaly segmentation in brain MR images", In *Proceeding of International MICCAI Brainlesion Workshop*, pp. 161-169, 2018. DOI: https://doi.org/10.1007/978-3-030-11723-8_16
- [12] ,M.R. Moore, J.M. Vann. (2019, January). Anomaly detection of cyber physical network data using 2D images. In *Proceeding of 2019 IEEE International Conference on Consumer Electronics (ICCE)* (pp. 1-5), 2019. DOI: <https://doi.org/10.1109/ICCE.2019.8662084>
- [13] Q. Wei, Y. Ren, R. Hou, B. Shi, J.Y. Lo, L. Carin, "Anomaly detection for medical images based on a one-class classification", In *Proceeding of Medical Imaging 2018: Computer-Aided Diagnosis*, 2018. DOI: <https://doi.org/10.1117/12.2293408>
- [14] M. Hasan, M.M. Islam, M.I.I. Zarif, M.M.A. Hashem, "Attack and anomaly detection in IoT sensors in IoT

sites using machine learning approaches”, Internet of Things, 7, 100059, 2019.
DOI: <https://doi.org/10.1016/j.iot.2019.100059>

- [15] P. Malhotra, L. Vig, G. Shroff, P. Agarwal, “Long short term memory networks for anomaly detection in time series”, In Proceedings of Presses universitaires de Louvain, Vol. 89, pp. 89-94, 2015.
- [16] N.Y. Choi, W.H. Kim, “Detecting user behavior anomalies using Generative Adversarial Networks”, Intelligence Information Research, 25(3), 43-62, 2019.

김 송 희(Songhee Kim)

[준회원]



- 2020년 2월 : 동국대학교 산업시스템공학과(공학 학사)
- 2020년 2월 ~ 현재 : 동국대학교 산업시스템공학과 석사과정

<관심분야>

기술 인텔리전스, 자연어 처리, 딥러닝

윤 병 윤(Byungun Yoon)

[정회원]



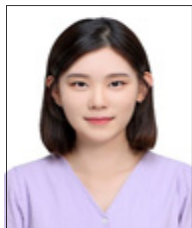
- 1998년 2월 : 서울대학교 산업공학과 (공학학사)
- 2000년 2월 : 서울대학교 산업공학과 (공학석사)
- 2005년 2월 : 서울대학교 산업공학과 (공학박사)
- 2007년 3월 ~ 현재 : 동국대학교 산업시스템공학과 교수

<관심분야>

기술경영, 특허분석, 텍스트마이닝, 인공지능

김 선 혜(Sunhye Kim)

[정회원]



- 2018년 8월 : 동국대학교 산업시스템공학과 (공학사)
- 2020년 2월 : 동국대학교 일반대학원 산업시스템공학과 (공학석사)
- 2020년 3월 ~ 현재 : 동국대학교 일반대학원 산업시스템공학과 (박사과정)

<관심분야>

기술경영, 특허분석, 데이터 어널리틱스, 텍스트마이닝