JKSCI

한국컴퓨터정보학회논문지
Journal of The Korea Society of Computer and Information
Vol. 25 No. 4, pp. 19-27, April 2020
https://doi.org/10.9708/jksci.2020.25.04.019

Study on Fault Detection of a Gas Pressure Regulator Based on Machine Learning Algorithms

Chan-Yang Seo*, Young-Joo Suh**, Dong-Ju Kim*

*Researcher, Postech Institute of Artificial Intelligence, POSTECH, Pohang, Korea
**Professor, Dept. of Computer Science and Engineering, POSTECH, Pohang, Korea
*Research professor, Postech Institute of Artificial Intelligence, POSTECH, Pohang, Korea

[Abstract]

In this paper, we propose a machine learning method for diagnosing the failure of a gas pressure regulator. Originally, when implementing a machine learning model for detecting abnormal operation of a facility, it is common to install sensors to collect data. However, failure of a gas pressure regulator can lead to fatal safety problems, so that installing an additional sensor on a gas pressure regulator is not simple. In this paper, we propose various machine learning approach for diagnosing the abnormal operation of a gas pressure regulator with only the flow rate and gas pressure data collected from a gas pressure regulator itself. Since the fault data of a gas pressure regulator is not enough, the model is trained in all classes by applying the over-sampling method. The classification model was implemented using Gradient boosting, 1D Convolutional Neural Networks, and LSTM algorithm, and gradient boosting model showed the best performance among classification models with 99.975% accuracy.

▸ **Key words**: Fault Detection, Gas Pressure Regulator, Gradient Boosting, Long Short-Term Memory, 1D Convolutional Neural Networks, Over-Sampling

[요 약]

본 논문에서는 정압기의 이상 상태 진단을 위한 기계학습 방법을 제안한다. 일반적으로 설비의 이상 상태 탐지를 위한 기계학습 모델 구현에는 관련 센서의 설치와 데이터 수집 과정이 동반되나, 정압기는 설비 특성상 안전문제에 매우 민감하여 추가적인 센서 설치가 매우 까다롭다. 이에 본 논문에서는 센서의 추가 설치 없이 정압기 설비에서 자체 수집되는 유량과 유압 데이터만을 가지고 정압기의 이상 상태를 조기에 판단하는 기계학습 모델을 제안한다. 본 논문에서는 정압기의 비정상데이터가 충분하지 않은 관계로, 모델 학습 시 오버 샘플링(Over-Sampling)을 적용하여 모델이 모든 클래스에 균형적으로 학습하도록 하였다. 또한, 그레이디언트 부스팅(Gradient Boosting), 1차원 합성곱 신경망(1D Convolutional Neural Networks), LSTM(Long Short-Term Memory) 등의 기계학습 알고리즘을 적용하여 정압기의 이상 상태를 판단하는 분류모델을 구현하였고, 실험 결과 그레이디언트 부스팅 알고리즘이 정확도 99.975%로 가장 성능이 우수함을 확인하였다.

▸ **주제어**: 이상 상태 진단, 정압기, 그레이디언트 부스팅, LSTM, 1차원 합성곱 신경망, 오버 샘플링

• First Author: Chan-Yang Seo, Corresponding Author: Dong-ju Kim
*Chan-Yang Seo (chan8616@postech.ac.kr), Postech Institute of Artificial Intelligence, POSTECH
**Young-Joo Suh (yjsuh@postech.ac.kr), Dept. of Computer Science and Engineering, POSTECH
*Dong-ju Kim (kkb0320@postech.ac.kr), Postech Institute of Artificial Intelligence, POSTECH
• Received: 2020. 02. 06, Revised: 2020. 04. 01, Accepted: 2020. 04. 03.

I. Introduction

정압기란 도시가스 압력을 사용처에 맞게 낮추는 감압 기능, 압력을 허용범위 내의 압력으로 유지하는 정압기능, 그리고 가스의 흐름이 없을 때는 밸브를 완전히 폐쇄하여 압력상승을 방지하는 폐쇄기능을 가진 기기로, 정압기용 압력조절기와 그 부속설비를 말한다. 정압기는 내부 부품의 마모 또는 손상, 자재 결함, 이물질 유입 등 다양한 원인으로 인해 고장 및 성능 저하가 일어날 수 있다. 정압기가 안정적인 도시가스 공급에 핵심 역할을 하는 설비인 만큼, 이상 상태를 조기에 진단하는 일은 매우 중요하다.

일반적으로 설비의 이상징후를 포착하기 위해서는 목적에 맞는 센서를 설치하여 데이터를 수집하는 과정이 수반된다. 하지만 정압기는 가스의 폭발 위험이 있는 환경에서 운영되는 설비로 센서를 부착하기 위해서는 방폭 인증 설비의 사용 등, 사전에 충족시켜야 하는 조건이 까다롭다. 이에, 본 논문에서는 추가 데이터 수집 장비를 설치하지 않고, 정압기에서 기본적으로 수집되는 유량과 유압 데이터만으로 정압기의 이상 상태 여부를 조기에 판단하는 기계학습 기법을 제안한다.

II. Preliminaries

1. Related works

정압기 결함 감지에 관한 연구는 Jian-ying JIAO et al.[1]의 연구와 Tian Shen et al.[2]의 연구, 그리고 Tsukasa Ishigaki et al.[3]의 연구가 있다. Jian-ying JIAO et al.의 연구[1]는 AE(Acoustic Emission) 센서를 사용하여 AE 신호데이터를 수집한 뒤, 웨이브릿 변환(Wavelet Packet Decomposition)을 통해 AE 신호데이터에서 추출한 특징으로 인공신경망(Artificial Neural Network)을 학습시켰다. 인공신공망의 출력은 정상 또는 결함, 둘 중 하나의 값을 갖는 이진 분류로 설계되었으며, 논문에 따르면 진단 정확도는 95% 이상을 보였다. 한편, Tian Shen et al.의 연구[2]는 정압기 출구 압력에 대해 비정상·비선형 데이터 분석에 적합한 CEEMDAN (Complete Ensemble Empirical Mode Decomposition with Adaptive Noise) 방식을 적용하여 특징을 추출한 후, 추출한 특징에 대해 힐버트 스펙트럼 분석(Hilbert Spectrum Analysis)을 적용하여 결함의 정도를 계산하고, 계산된 결함 정도에 클러스터링 기법을 적용하여 정압기의 상태를 진단한다. 그리고 정압기 상태 진단에 대한 Tsukasa Ishigaki et al.의 연구[3]는 정압기 내부의 고무 다이어프램(Rubber Diaphragm) 부위에 별도의 마이크를 설치하여 데이터를 수집하고, 이를 이용하여 정압기의 상태를 분류하고자 했다. 수집된 데이터는 정규화를 통해 주파수 스펙트럼으로 변화되어 쿨백-라이블러(KullBack-Leibler) 필터를 적용한 서포트 벡터 머신(Support Vector Machine)[4]에 입력된다. 서포트 벡터 머신은 정압기 상태를 정상 또는 낙후됨(Deteriorated)으로 분류하였으며, 연구에 의하면 분류 정확도는 92.41%를 보였다.

또한, 천연가스나 원유의 배관데이터를 분석하는 연구는 다양하게 진행되어왔는데, 자기장 누설(Magnetic Flux Leakage) 센서 데이터 세트[5][6]나 원거리 초음파 검사(Long Range Ultrasonic Testing) 데이터 세트[7][8]를 사용한 경우가 대부분이었으며, 음향데이터를 사용하여 배관의 누출을 탐지한 경우도 있었다[9]. 배관의 안전을 모니터링하는 연구 중에서 Wonhee Chung et al.의 연구[10]는 정압기에서 수집된 1차, 2차 압력데이터를 활용하여 배관압력을 예측하는 모델을 제안하였다는 점에서 선행연구로 포함시켰다. Wonhee Chung et al.의 연구는 랜덤 포레스트(Random Forest)[11], 서포트 벡터 머신, 그리고 LSTM 알고리즘[12][13]을 사용하여 실시간 배관압력 예측모델을 만들었으며, 모델 학습을 위해 정압기의 압력 데이터 외에도 시간변수, 가스의 수요량에 영향을 미치는 날씨, 유가, 출생률 등의 환경변수를 사용하였다.

III. The Proposed Scheme

본 논문에서 제안하고자 하는 정압기 상태 진단 시스템의 구조는 그림 1과 같다. 먼저, 전처리 작업을 통해 설정압력 변경으로 발생하는 데이터의 비일관성을 통일하고 이상치를 제거함으로써 수집된 데이터가 모델 학습에 적합하도록 만든다. 그 후, 그레이디언트 부스팅[14][15], 1차원 합성곱 신경망[16][17], LSTM 알고리즘[12][13]을 사용하여 정압기 상태 분류모델을 만들어 학습시킨다. 마지막으로 혼동행렬(Confusion Matrix)을 계산하여, 각 모델의 성능을 평가한다.

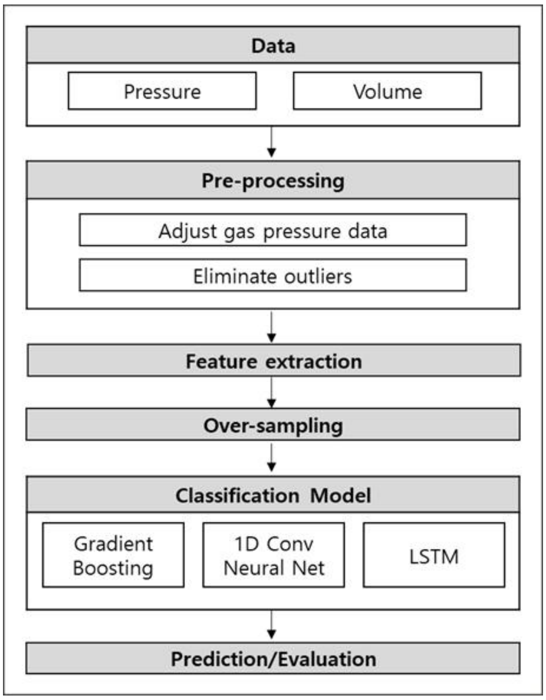

Fig. 1. System Architecture

1. Data Preprocessing and Data Analysis

본 논문에서는 2015년부터 2018년까지 가스공급기관에서 실제로 설비를 운용하면서 1분마다 샘플링하여 수집한 유압 및 유량 데이터를 학습 데이터로 사용하였다. 수집된 전체 데이터에서 정상 데이터는 약 172만 건, 비정상 데이터는 약 35만 건으로, 총 207만 건이다.

유압은 가스사용자에게 공급되는 가스를 정압기 출구 또는 가스공급시설 끝부분의 배관에서 측정한 계기압력을 말한다. 본 논문에서 사용한 유압 데이터는 정압기의 출구측 압력을 기록하는 장치인 자가압력기록계를 통해 자동으로 수집된 정압기 후단압력과 가스공급시설의 끝부분의 계량기를 통해 측정된 계량기 후단압력이다. 유량은 수급지점에 설치된 거래용 계량기를 이용해 측정된 가스 공급량을 의미하며, 순시유량은 소비자의 수요에 따라 실시간으로 달라지는 유량을, 누적유량은 일정기간 동안 누적된 순시유량을 일컫는다. 본 논문에서는 이상 탐지를 위해 순시유량을 사용하였다.

1.1 Adjust Gas Pressure Data

정압기 운영 중 필요에 따라 설정 압력을 변경하는 경우가 있다. 설정 압력이 변경되면 정압기가 유지하고자 하는 압력의 기준이 달라지기 때문에, 설정 압력 변경 전의 수집 데이터와 변경 후 수집 데이터의 분포가 상이해진다. 상이한 분포를 갖는 데이터는 모델의 학습을 방해할 수 있기 때문에, 그림 2와 같이 기존에 수집된 압력 데이터를 설정 압력의 변경 폭만큼 조정해줌으로써 모델의 학습이 안정적으로 진행되도록 하였다.

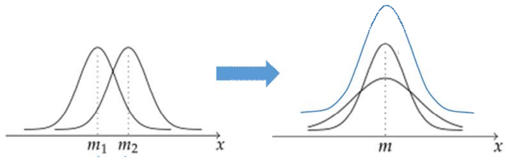

Fig. 2. Adjust Gas Pressure Data to Eliminate the Differences in Preference Value

1.2 Outliers Elimination

정압기의 유량 및 유압 데이터도 결국 정압기의 내부 센서에 의해 수집되는 데이터이기 때문에, 센서 오작동으로 인해 잘못된 값이 측정될 수 있다. 데이터 분석 결과, 대부분의 오측정된 데이터는 그림 3과 같이 정상 압력값보다 극단적으로 높거나 낮은 이상치의 형태로 나타나는 것을 확인하였다.

이상치는 모델의 학습을 저해하기 때문에, 학습 데이터에서 반드시 제거되어야 한다. 본 논문에서는 제트 스코어(Z-Score) [18]를 사용하여 이상치를 제거한 후, 모델 학습을 진행하였다.

Fig. 3. Outliers in gas pressure data

제트 스코어는 모집단이 정규분포를 따른다고 가정했을 때, 관측치가 표본평균에서 표준편차의 몇 배의 거리만큼 떨어져 있는지 나타내는 지표로 수식 1과 같다.

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

위 식을 계산한 결과의 절대값이 크면 클수록 분포의 양극단에 위치하며, 해당값이 설정한 임계값을 넘어서면 이상치로 판단하고 해당 값은 학습에서 제외시킨다.

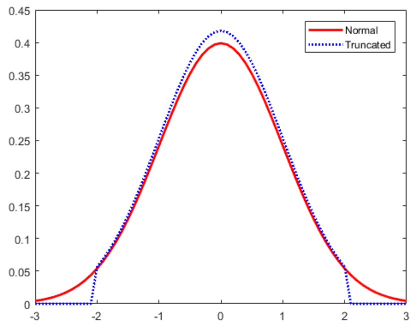

Fig. 4. Truncate outliers using z-score value

1.3 Over-sampling

정압기는 유량에 상관없이 일정한 압력 유지를 위한 방향으로 작동한다. 급격한 유량의 변화에는 어쩔 수 없이 유압이 영향을 받게 되지만, 이 상황에서도 정압기는 최대한 빨리 압력을 원래 설정된 값으로 되돌리는 방향으로 작동한다. 따라서, 유량의 변화에 따른 유압의 변화에 주목하면 정압기의 이상 상태를 진단할 수 있다. 이에 본 논문에서는 현재 유량값과 현재 유압값 그리고 과거 일정 구간 내의 유압값을 이용하여 정압기의 이상 상태를 진단하고자 하였다.

하지만 수집된 데이터에서 정상-비정상 클래스 간 데이터 수의 불균형으로 비정상 클래스에 대한 모델 학습이 제대로 이뤄지지 않았다. 비정상 클래스에 대해서도 모델이 학습하도록 만들기 위해, 본 논문에서는 오버 샘플링 기법의 하나인 SMOTE (Synthetic Minority Over-Sampling Technique)[19]를 사용하였다. SMOTE기법은 마이너리티 클래스(Minority Class)에 속한 데이터 중에서 가장 가까운 점들을 선으로 연결한 뒤, 연결선 위에 점을 찍어 새로운 데이터를 생성하고 생성된 데이터 역시 같은 클래스에 속한다고 여기는 방법이다.

SMOTE기법 적용 시 주의할 점은 일정 구간의 유압값을 특징으로 사용하는 경우, 과거의 유압값을 그대로 사용하는 것이 아니라 과거의 유압값과 현재 유압값의 차로 변형하여 사용해야 한다는 점이다. SMOTE 기법 적용 전 반드시 유압값의 차이를 특징으로 추출해야 하는 이유는 정압기의 이상 동작이 대부분 과거 유압과 현재 유압의 상대적인 차이의 크기로 나타나기 때문이다. SMOTE 기법의 처리 과정은 그림 5와 같으며, 여기서 빨간색(Red)는 정상 데이터, 보라색(Purple)은 비정상 데이터, 노란색(Yellow)은 생성된 데이터를 가정한다. 본래 정압기의 상태 진단에 쓰인 데이터는 훨씬 고차원의 벡터이지만, 설명의 편의를 위해 2차원 평면에 도식화하였다.

그림 5의 (a)는 전처리 직후의 데이터를 도식화한 것이고, 여기에 바로 SMOTE기법을 적용했을 시 발생가능한 문제를 그림으로 표현한 것이 그림 5의 (b)다. 그림 5의 (b)는 비정상데이터가 너무 적어 비정상데이터를 잇는 연결선이 정상구간을 지나가는 문제가 발생한 경우이다. 정상 구간을 지나는 연결선 위에 비정상데이터가 생성되면, 정상데이터와 비정상데이터 간의 구분이 모호해져 모델이 제대로 학습되기 어렵다.

따라서, 그림 5의 (c)처럼 먼저 유압의 차이로 특징을 추출하여 정상데이터와 비정상 데이터를 서로 분리하는 과정이 필요하다. 본 논문에서는 정압기로부터 수집된 정상데이터와 비정상데이터를 서로 분리한 후에 SMOTE 기법을 적용하여 모델 학습하였으며, 특징 추출 전보다 분류 정확도가 훨씬 높아짐을 확인할 수 있었다.

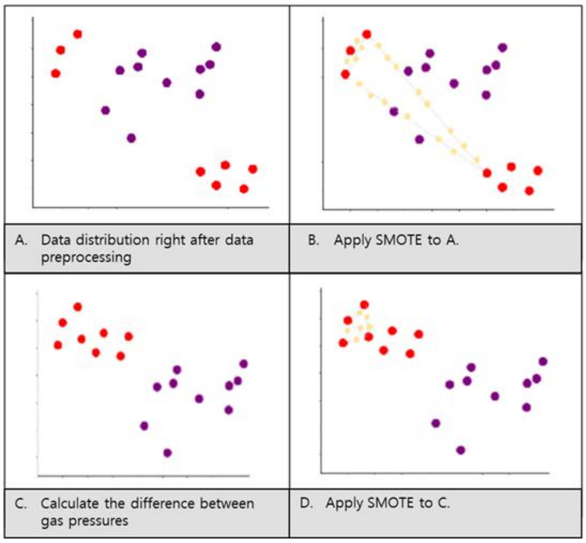

Fig. 5. Application of SMOTE
(Red: Abnormal, Purple: Normal, Yellow: Generated)

2. Machine Learning Algorithms for Fault Detection of a Gas Pressure Regulator

정압기의 이상 상태 진단 분류모델을 구현하기 위해, 본 논문에서는 그레이디언트 부스팅, 1차원 합성곱 신경망, LSTM 등의 세 가지 알고리즘을 사용하였다.

먼저, 그레이디언트 부스팅은 그림 6과 같이 약한 분류기 여러 개를 결합하여 강한 분류기를 만들어 내는 알고리즘으로, 새로운 분류기가 이전 단계의 분류기가 분류하지 못한 잔차(Residual Error)를 최소화하는 방향으로 학습하는 알고리즘이다.

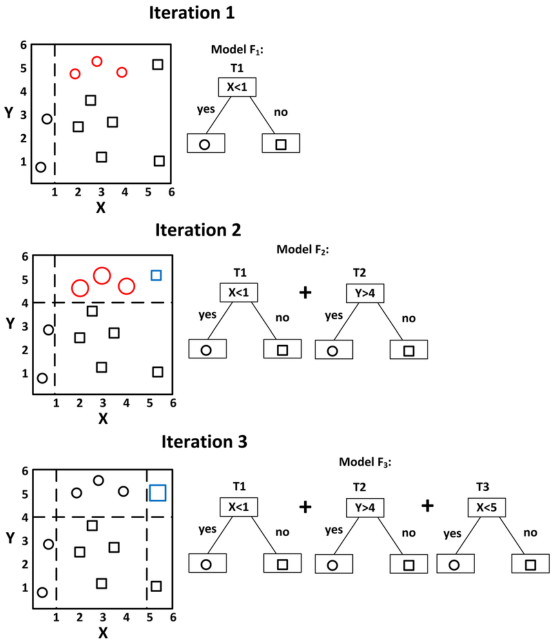

Fig. 6. Gradient Boosting

이를 수식으로 살펴보면 다음과 같다.

$$D = \{(x_i, y_i)\}(|D| = n, x_i \in \mathbb{R}^m, y_i \in \mathbb{R}) \qquad (2)$$

데이터 세트 D가 n개의 m차원 데이터로 구성되어 있다고 가정하면, 이에 대한 결과를 예측하기 위한 트리 앙상블 모델은 K개의 함수의 합으로 표현된다(단, f_k는 독립된 트리구조와 가중치를 가짐).

$$\hat{y_i} = \sum_{k=1}^{K} f_k(x_i) \qquad (3)$$

모델은 손실함수를 최소화하는 방향으로 학습하게 되며, 손실함수는 실제값과 예측값의 차이로 표현된다.

$$L = \sum_i l(y_i, \hat{y_i}) + Regularization\ Term \qquad (4)$$

트리 앙상블 모델은 유클리드 공간에서 최적화하는 것은 불가능하므로, 손실함수는 다음과 같이 i번째 데이터에 대한 t번째 함수까지의 예측 결과의 합과 실제값의 차이로 변형하여 학습에 사용한다.

$$L^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y_i}^{(t-1)} + f_t(x_i)) + Regularization\ Term \qquad (5)$$

위의 손실함수에서 알 수 있듯, 그레이디언트 부스팅 알고리즘은 손실함수를 최소화할 때까지 지속적으로 f_t 를 추가하기 때문에 적은 양의 데이터에도 잘 학습된다는 장점이 있다. 정압기는 가스 공급의 핵심 설비로, 정압기의 이상 동작은 거대한 비용문제와 심각한 안전문제를 초래할 수 있다. 따라서 가스공급업체는 정압기의 이상 동작을 미연에 방지하기 위해 여러 노력을 기울이므로, 정압기의 유량 및 유압 데이터는 정상데이터에 비해 비정상데이터의 양이 극히 적을 수밖에 없다. 이에 적은 양의 데이터에도 우수한 성능을 보이는 그레이디언트 부스팅 알고리즘을 활용하면 좋은 결과를 낼 수 있을 것으로 기대할 수 있다.

본 논문에서는 그레이디언트 부스팅 알고리즘의 입력 타임 스텝(Time Step)을 8로 설정하였다. 그 이유는 정압기의 특성상 가스 공급 시작 직후, 6분에서 8분 이내에는 유압이 급격하게 변동하더라도 정상으로 판별해야 하므로, 8분 이상의 시간 동안 유압의 변화를 관찰하여 이상 여부를 판단해야 하기 때문이다. 더불어, 그레이디언트 부스팅 알고리즘에 적용한 하이퍼파라미터(Hyperparameter)는 격자 찾기(Grid Search)를 통해 최적값을 선정하였으며, 선정 결과, 학습률은 0.16, 트리 최대 길이는 50, 트리 개수는 300개로 정하여, 실험에 사용하였다.

다음으로, 1차원 합성곱 신경망은 1차원의 고정 길이 데이터로부터 특징을 추출할 때 유용한 인공신경망이다. 그림 7과 같이 1차원 합성곱 필터는 한 방향으로 이동하면서 겹치는 영역의 내적값을 계산하여 특징을 추출하기 때문에 자연어, 신호데이터, 센서 데이터 등 시계열 데이터 분석에 적합하다. 정압기의 유량·유압 데이터 역시 시계열 데이터이므로, 데이터 특성상 1차원 합성곱 신경망이 적합할 것으로 기대할 수 있다.

본 논문에서는 1차원 합성곱 레이어에 완전 연결 레이어(Fully Connected Layers)를 추가하여 정압기의 상태를 정상 또는 비정상으로 분류하도록 설계하였다. 1차원 합성곱 신경망 네트워크 초반에는 노이즈의 영향을 최소화하기 위해 비교적 큰 크기의 필터(7x1)을 사용하였으며, 그 이후 레이어에서는 학습 파라미터 수를 줄이기 위해 상대적으로 작은 크기의 필터(3x1)를 여러 개 사용하여 충분히 넓은 수용 영역(Receptive Field)을 갖도록 구조를 디자인하였다. 네트워크 초반의 7x1 필터가 적용되기에 충분한 길이의 데이터를 입력으로 사용하기 위해 타임 스텝은 32로 적용하였다. 학습률은 0.0005로 설정하였으며, 1차원 합성곱 신경망의 예측 결과가 정상 또는 비정상 둘 중 하나로 산출되기 때문에 손실함수는 이진 교차 엔트로피(Binary Cross-Entropy)를 사용하였다.

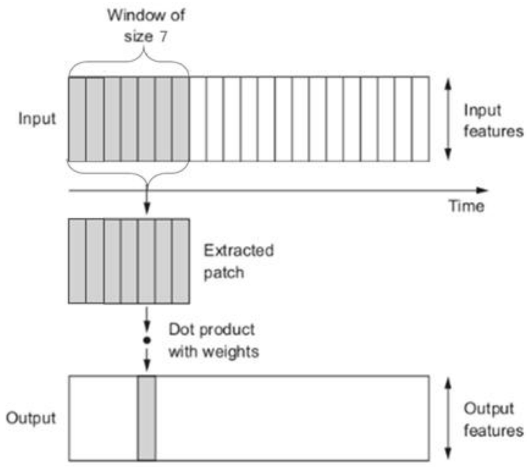

Fig. 7. 1D Convolution

LSTM은 장기적인 종속성을 학습할 수 있는 특수한 형태의 순환신경망(Recurrent Neural Network)이다. 기존 순환신경망 모델의 장기 기억을 잃어버리는 문제와 역전파 과정에서 발생하는 기울기 소실 문제를 해결하기 위해 명시적으로 설계된 네트워크로, 시계열 데이터 분석에 적합하다.

본 논문에서는 정압기의 정상 여부를 판단하기 위해 LSTM 셀에 완전 연결 레이어를 추가하였다. 그림 8의 마지막 완전 연결 레이어의 출력인 $\hat{y}$은 정압기의 이상 상태 여부를 예측한 결과로 0(정상) 또는 1(비정상)의 값을 갖는다. LSTM의 타임 스텝은 그레이디언트 부스팅과 마찬가지로 충분한 시간 동안 유압의 변화를 관찰하여 이상 여부를 판단하기 위해 8로 설정하였다. 학습률은 0.0005로 설정하였으며, 손실함수는 1차원 합성곱 신경망과 동일하게 이진 교차 엔트로피를 사용하였다.

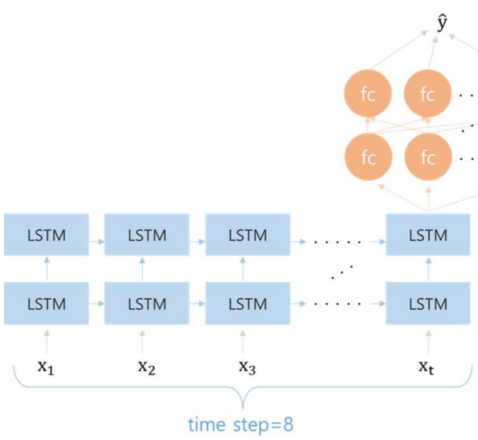

Fig. 8. LSTM

IV. Experiment and Result

본 논문에서는 그레이디언트 부스팅, 1차원 합성곱 신경망, LSTM 알고리즘을 사용하여 정압기 상태 분류모델을 구현하고 그 성능을 테스트하였다. 실험 환경은 다음과 같다.

Table 1. Test Environment

Category	Used
HW(GPU)	Nvidia Tesla V100 16GB
SW	python
	Keras
	XGboost

테스트 데이터는 2019년도 1월부터 9월까지 수집된 정압기의 유량 데이터와 유압데이터로, 입력 데이터 구성 시 알고리즘의 개별 특성을 고려하여 각기 다른 timestep을 적용하였다. 따라서, 성능 평가에 사용된 데이터 건수는 알고리즘마다 다르며, 알고리즘별 테스트에 사용된 데이터 건수는 표 2와 같다.

Table 2. The number of Data used for testing each algorithm

Algorithm	The number of Data	
	Normal	Abnormal
Gradient Boosting	269,151	21,543
1D ConvNet	269,095	21,543
LSTM	269,108	21,543

분류모델의 정확도는 모델의 정상·비정상 데이터의 분류 결과로 혼동행렬을 구성한 후 수식 6에 의해 계산한다.

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \qquad (6)$$

Table 3. Confusion Matrix

		Predicted	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN

정압기의 유량 및 유압 데이터에 대하여 모델 구현 후 혼동행렬을 통해 정확도를 평가한 결과는 표 4와 같다.

Table 4. Accuracy of each algorithm

Algorithm	Accuracy
LSTM	98.071%
1D CNN	99.930%
Gradient Boosting	99.975%

데이터의 특성상 테스트 데이터에서 정상데이터가 차지하는 비중이 비정상데이터에 비해 훨씬 크므로, 전반적으로 모든 알고리즘의 정확도가 매우 높게 나타난다. 따라서, 정압기 상태 진단 문제에 대한 알고리즘의 적합성 여부를 판단할 때에는 단순히 정확도만 고려해서는 안 되며, 틀린 예측(혼동행렬의 FN과 FP에 해당)을 최소로 하는 알고리즘을 본 연구의 적합 알고리즘으로 채택해야 한다.

그림 9부터 11은 각각 LSTM, 1차원 합성곱 신경망, 그레이디언트 부스팅 알고리즘의 성능 평가 결과를 혼동행렬로 나타낸 결과이다. LSTM의 오예측건수는 FP 1,828건, FN 1,763건 총 3,591건으로 가장 많았으며, 그 다음은 1차원 합성곱 신경망이 FP 148건, FN 54건으로 총 202건의 오예측을 하였다. 반면에, 그레이디언트 부스팅의 오예측건수는 FP 22건, FN 49건, 총합 71건으로 가장 적었다. 따라서, 본 연구에서는 그레이디언트 부스팅이 정압기 상태 진단에 가장 적합한 알고리즘이라고 판단하였다.

상기 실험 결과가 의미하는 바는 다음과 같이 추론할 수 있다. 보편적으로 딥러닝은 최적의 답을 구할 수 없는 복잡한 문제에 대하여 전문가 수준의 지식을 요구하는 피처 엔지니어링(Feature Engineering) 없이도 데이터 기반의 학습을 통해 우수한 성능을 보이는 것이 사실이다. 이는 1차원 합성곱 신경망이나 LSTM의 정확도가 98%이상으로 측정되었다는 점에서도 확인할 수 있다. 하지만, 본 논문에서 해결하고자 했던 정압기의 이상 상태 진단이라는 문제는 비정상데이터가 정상데이터보다 훨씬 적기 때문에, 단순히 정확도가 충분히 높다는 사실만으로 정압기의 이상 상태를 잘 진단했다고 말하기 어렵다. 따라서, 정압기의 이상 상태 발생을 최대한 놓치지 않고 진단하기 위해서는 잔차가 최소화될 때까지 탐욕스러운 방식으로 모델을 확장할 필요성이 있다. 이러한 문제 상황에서, 그레이디언트 부스팅 알고리즘이 가장 적합하다는 결론은 필연적이었다고 할 수 있다.

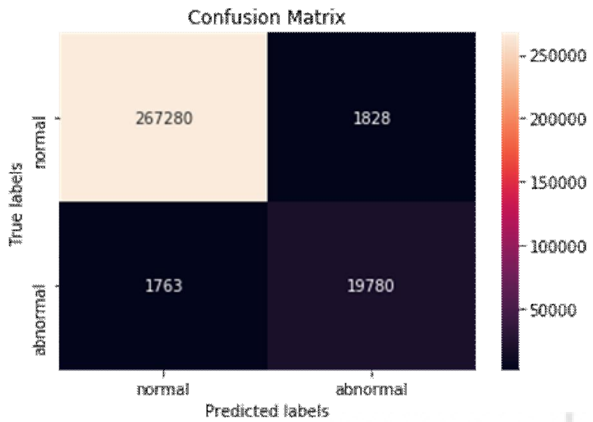

Fig. 9. Confusion Matrix(LSTM)

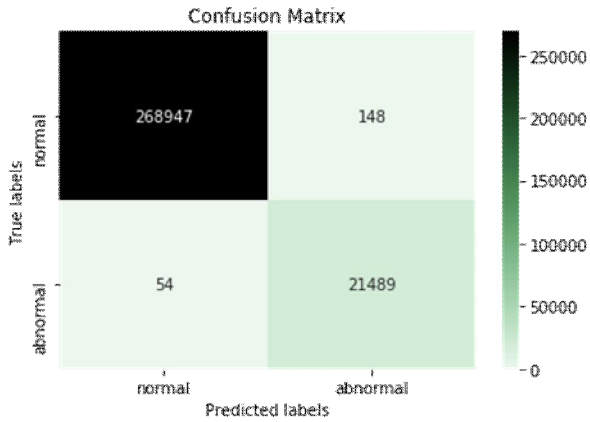

Fig. 10. Confusion Matrix(1D ConvNet)

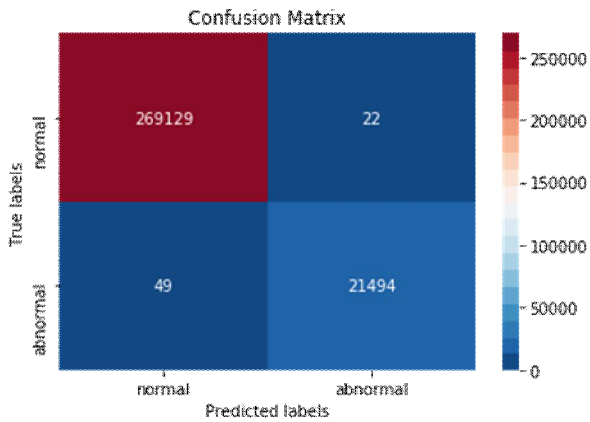

Fig. 11. Confusion Matrix(Gradient Boosting)

V. Conclusions

본 논문은 정압기에서 수집된 유량 및 유압 데이터를 활용하여 정압기의 이상 상태를 진단하는 분류모델을 제안하였다. 모델 학습에 사용된 데이터는 1분 단위로 수집된 정압기 후단압력과 계량기 후단압력 그리고 순시유량으로 구성되어 있다. 해당 데이터는 먼저 평균 조정 및 이상치 제거 등의 데이터 전처리 과정을 거친다. 전처리 완료 후 모델 학습 시에는 정상·비정상 클래스의 양적 불균형을 해소하기 위해 오버 샘플링을 수행하였다. 학습을 마친 모델의 성능은 혼동행렬을 사용하여 평가하였으며, 실험 결과, 구현 알고리즘 중 그레이디언트 부스팅이 테스트 데이터에 대하여 정확도 99.975%로 가장 좋은 성능을 보였다.

본 논문에서 제안한 정압기의 상태 분류 기술은 정압기 이상 상태의 조기 발견과 효율적이고 정확한 정비 스케쥴링이 가능하게 될 것으로 기대한다. 이는 치명적인 안전사고를 방지하고, 정압기의 이상 동작으로 전체 도시가스 공

급이 중단되는 사태를 막아 비용 손실 최소화에 기여할 수 있을 것으로 사료된다.

ACKNOWLEDGEMENT

This work was supported by Basic Research Project(No.2016R1D1A1B01014776) funded by the government(Ministry of Education) in 2016, and 2019 R&D support project based on science and technology according to the regional demand funded by the Ministry of Science and ICT(CN19100GB001).

REFERENCES

[1] Jiao, J. Y., Wei, Y. P., Li, H. L., Liu, Y., Jiang, F., Song, M. X., & Tan, S. L. “A gas regulator fault detecting method based on acoustic emission technology.” IEEE In 2017 Symposium on Piezoelectricity, Acoustic Waves, and Device Applications(SPAWDA), pp. 91-94, 2017, October. doi:10.1109/spawda.2017.8340295

[2] Tian, S., Bian, X., Tang, Z., Yang, K., & Li, L. “Fault Diagnosis of Gas Pressure Regulators Based on CEEMDAN and Feature Clustering”, IEEE Access, Vol. 7, pp. 132492-132502, 2019. doi:10.1109/access.2019.2941497

[3] Ishigaki, T., Higuchi, T., & Watanabe, K. “Spectrum Classification for Early Fault Diagnosis of the LP Gas Pressure Regulator Based on the Kullback-Leibler Kernel.” 2006 16th IEEE Signal Processing Society Workshop on Machine Learning for Signal Processing, Arlington, VA, pp. 453-458, 2006, doi:10.1109/mlsp.2006.275593

[4] Heo G. “Context Dependent Fusion with Support Vector Machines.” Journal of The Korea Society of Computer and Information, Vol. 18, No. 7, pp. 37-45, 2013. doi:10.9708/jksci.2013.18.7.037

[5] Layouni, M., Hamdi, M. S., & Tahar, S. “Detection and sizing of metal-loss defects in oil and gas pipelines using pattern-adapted wavelets and machine learning.” Applied Soft Computing, Vol. 52, pp. 247-261, 2017. doi:10.1016/j.asoc.2016.10.040

[6] Mohamed, A., Hamdi, M. S., & Tahar, S. “A machine learning approach for big data in oil and gas pipelines.” IEEE, In 2015 3rd International Conference on Future Internet of Things and Cloud, pp. 585-590, 2015, August. doi:10.1109/ficloud.2015.54

[7] Akram, N. A., Isa, D., Rajkumar, R., & Lee, L. H. “Active incremental Support Vector Machine for oil and gas pipeline defects prediction system using long range ultrasonic transducers.” Ultrasonics, Vol. 54, No. 6, pp. 1534-1544, 2014. doi:10.1016/j.ultras.2014.03.017

[8] Lee, L. H., Rajkumar, R., Lo, L. H., Wan, C. H., & Isa, D. “Oil and gas pipeline failure prediction system using long range ultrasonic transducers and Euclidean-Support Vector Machines classification approach.” Expert Systems with Applications, Vol. 40, No. 6, pp. 1925-1934, 2013. doi:10.1016/j.eswa.2012.10.006

[9] Yeo, D., Bae, J. H., & Lee, J. C. “Unsupervised Learning-Based Pipe Leak Detection using Deep Auto-Encoder.” Journal of the Korea Society of Computer and Information, Vol. 24, No. 9, pp. 21-27, 2019. doi:10.9708/JKSCI.2019.24.09.021

[10] Chung, W. H., Park, G., Gu, Y. H., Kim, S., & Yoo, S. J. “City Gas Pipeline Pressure Prediction Model.” Journal of Society for e-Business Studies, Vol. 23, No. 2, 2019. doi:10.7838/JSEBS.2018.23.2.033

[11] Kim, J., Kim, H., Jang, K., Lee, J., and Moon, Y. “Object Classification Method Using Dynamic Random Forests and Genetic Optimization.” Journal of the Korea Society of Computer and Information, Vol. 21, No. 5, pp. 79-89, 2016. doi:10.9708/jksci.2016.21.5.079

[12] Hochreiter, S., & Schmidhuber, J. “Long short-term memory.” Neural computation, Vol. 9, No. 8, pp. 1735-1780, 1997. doi:10.1162/neco.1997.9.8.1735

[13] Wang, G., Shin, S. Y., & Lee, W. J. “A Text Sentiment Classification Method Based on LSTM-CNN.” Journal of The Korea Society of Computer and Information, Vol. 24, No. 12, pp. 1-7, 2019. doi:10.1049/cje.2018.11.004

[14] Chen, T., & Guestrin, C. “Xgboost: A scalable tree boosting system.” In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785-794, 2016. doi:10.1145/2939672.2939785

[15] Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu TY. “Lightgbm: A highly efficient gradient boosting decision tree.” In Advances in neural information processing systems, pp. 3146-3154, 2017. doi:10.1109/iccse.2019.8845529

[16] Ince, T., Kiranyaz, S., Eren, L., Askar, M., & Gabbouj, M. “Real-time motor fault detection by 1-D convolutional neural networks.” IEEE Transactions on Industrial Electronics, Vol. 63, No. 11, pp. 7067-7075, 2016. doi:10.1109/tie.2016.2582729

[17] Mukhopadhyay, R., Panigrahy, P. S., Misra, G., & Chattopadhyay, P. “Quasi 1D CNN-based Fault Diagnosis of Induction Motor Drives.” IEEE In 2018 5th International Conference on Electric Power and Energy Conversion Systems (EPECS), pp. 1-5, 2018, April. doi:10.1109/epecs.2018.8443552

[18] Jain, A., Nandakumar, K., & Ross, A. “Score normalization in multimodal biometric systems.” Pattern recognition, Vol. 38, No. 12, pp. 2270-2285, 2005. doi:10.1016/j.patcog.2005.01.012

[19] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. "SMOTE: synthetic minority over-sampling technique." Journal of artificial intelligence research, Vol. 16, pp. 321-357, 2002. doi:10.1613/jair.953

Authors

Chan-Yang Seo received the B.S. degrees in International Trade from Inha University, Korea, in 2015. Ms. Seo joined the Researcher of the Research Department of Postech Information Research Laboratories, POSTECH, Pohang, Korea, in 2019. She is currently a Researcher of the Research Department of POSTECH Institute of Artificial Intelligence. She is interested in Computer Vision, and Natural Language Processing.

Young-Joo Suh received the B.S., M.S. degrees in Electronics Engineering from Hanyang University, Korea, in 1985, 1987, respectively and Ph.D. degree in Electrical and Computer Engineering from Georgia Institute of Technology, Georgia in 1996. Dr. Suh joined the faculty of the Department of Computer Science and Engineering at POSTECH, Pohang, Korea, in 1998. He is currently a Professor in the Department of Computer Science and Engineering, POSTECH. He is interested in wireless LAN MAC protocol, mobility management, 5G mobile networks, and indoor positioning systems.

Dong-Ju Kim received the B.S., M.S. degrees in Radio Wave Engineering from Chungbuk National University, Korea, in 1998, 2000, respectively and Ph.D. degree in Electric, Electronic and Computer Engineering from Seongkyungwan National University, Korea, in 2010. Dr. Kim joined the Senior Researcher of the Research Department of Postech Information Research Laboratories, POSTECH, Pohang, Korea, in 2016. He is currently a Research Processor of the Research Department of Postech Institute of Artificial Intelligence. He is interested in Face Recognition, HCI, Computer Vision, Artificial Intelligence.