

차량설비의 알람 유형 분류를 위한 오픈셋 인식 심층 신경망

김상훈¹ · 홍현욱² · 정의현² · 김성범^{1*}¹고려대학교 산업경영공학과 / ²현대자동차 빅데이터기획팀Open Set Recognition Deep Neural Network
for Classification of Alarm Types in Automobile FacilitySang Hoon Kim¹ · Hyun Wook Hong² · Ui Hyun Jeong² · Seoung Bum Kim¹¹Department of Industrial Management Engineering, Korea University²Big Data Planning Team, Hyundai Motor Group

It is important to predict critical alarms in the manufacturing process that can reduce the utilization rate of the production facility. In recent years, deep neural networks have been widely used for prediction of alarm types in the manufacturing processes. However, the existing deep neural network classifiers follow a closed-set assumption that all predictable categories should be learned in the training stage. For this reason, when an unknown alarm type comes into the classifier, it should be classified as one of the predefined alarm types. In the actual manufacturing processes, it is extremely difficult to collect all possible types of alarm data. Therefore, a model with open set recognition is required to identify unknown alarm types. In addition, because the alarm type data collected from production facilities occurs simultaneously, a multi-label classification model is necessary. In this study, we propose a multi-label open set recognition model combined with background data that can improve the ability to identify unknown alarm types. We demonstrated the usefulness and applicability of the proposed method by comparing it with existing open set classification methods using real process data obtained from an automobile industry.

Keywords: Automobile, Alarm Classification, Open-Set Recognition, Multi-Label Classification, Deep Neural Networks

1. 서론

산업용 알람 시스템은 공정 운영의 안전한 관리를 위해 필수적인 역할을 한다. 최근 하드웨어와 소프트웨어가 발전함에 따라 공정 진행 중에 다채널 센서로부터 거의 모든 프로세스 변수를 측정하고 수집하는 것이 가능해져 알람 프로세스를 실시간으로 모니터링할 수 있게 되었다(Bransby and Jenkinson, 1998). 하지만 공정 모니터링 및 제어 시스템의 장치 수가 지난 수십 년 동안 극적으로 증가함에 따라 각 개별 작업자가 평가하고 조치해야 하는 데이터의 양 또한 증가했다. 이는 공정의 전반적인 복잡성을 증가시켰을 뿐 아니라, 작업자에게 알람

시스템을 모니터링하고 프로세스를 운영하기 위해 더 높은 수준의 대응을 요구하게 되었다(Beebe *et al.*, 2013). 작업자가 겪는 혼란으로 인해 공정의 여러 문제가 발생하는 상황에서 공정 문제를 야기하는 중요 알람을 조기에 예측할 수 있다면, 작업자가 제 시간에 대응할 수 있기 때문에 안전과 비용적 측면에서 매우 효율적이다(Zhu *et al.*, 2016).

이상 상태를 조기에 감지하고 예측하기 위한 연구가 제조공정, 항공, 보안, 의료 등 많은 분야에서 진행되어 왔다. 이상치 탐지 모델은 데이터의 패턴을 학습하여 정상 또는 이상으로 분류한다. 많은 경우 정상 상태 데이터만이 가용하며, 이 경우 이상치 탐지 모델은 정상 데이터의 패턴을 학습하여 이 패턴

제16회 석사논문경진대회 수상논문.

* 연락저자 : 김성범 교수, 02841 서울특별시 성북구 안암로 145 고려대학교 산업경영공학과, Tel : 02-3290-3397, Fax : 02-3290-4550,

E-mail : sbkim1@korea.ac.kr

2021년 2월 1일 접수; 2021년 3월 22일 수정본 접수; 2021년 3월 24일 게재 확정.

에서 벗어난 데이터를 이상 데이터로 인식한다. 하지만 정상 데이터만으로 학습한 이상치 탐지 모델은 여러 클래스를 가지는 이상 데이터에 대해 정확한 분류를 수행할 수는 없다. 정상 상태 데이터뿐만 아니라 이상 상태 데이터도 수집하여 레이블을 정의할 수 있다면 각 상태별 데이터의 패턴을 파악할 수 있어 높은 정확도로 상태를 분류할 수 있다(Teng, 2010). 알람 시스템으로부터 수집된 설비 상태별 데이터의 경우 수집과 동시에 각 상태별 레이블을 정의할 수 있으므로 지도학습 기반의 이상치 탐지 모델을 적용할 수 있다.

심층 신경망의 비약적인 성능 향상으로 인해 심층 신경망 기반의 이상치 탐지 모델에 대한 연구들이 활발히 행해지고 있다. Zhu *et al.*(2016)는 이전 알람에 대한 프로세스 데이터로부터 각 알람 클래스별 발생 확률을 예측하기 위해 N-gram 확률 모델을 제안했다. N-gram 모델은 데이터의 순차적인 시계열 정보로부터 다음 항목을 예측하는 확률적 언어 모델이며 자연어 처리 분야에서 많이 사용되고 있다. 하지만 N-gram 모델은 시계열 프로세스 데이터의 장기적인 시계열 패턴을 고려하지 못하는 한계점이 존재한다. 이를 보완하기 위해 Cai *et al.*(2019)는 장기적인 시계열 패턴을 고려하는 recurrent neural networks(RNN) 구조의 long short term memory(LSTM) 네트워크를 알람 유형 분류에 적용하였다. 아울러 이미지 데이터에 많이 사용되는 convolutional neural network(CNN) 구조에서 1차원 convolution 연산을 사용한 채널 별 특징 추출을 통해 높은 알람 유형 분류 성능을 보였다(Ince *et al.*, 2016).

기존 연구에서 설비의 이상을 탐지하는 모델은 주로 단일 레이블 분류를 목적으로 연구됐다. 그러나 단일 레이블 분류는 한 시간 단위의 알람 프로세스 데이터가 정상 상태이거나 하나의 알람이 발생한 상태만 대처할 수 있다는 한계점이 존재한다. 실제 공정에서는 설비에 이상이 생길 경우 여러 알람이 동시에 발생하는 경우가 많다. 따라서 설비에서 여러 알람이 동시에 발생하는 경우를 정확히 예측하기 위해서는 다중 레이블 분류가 가능한 모델이 필요하다. 또한 기존 심층 신경망 모델은 모든 예측 가능한 클래스가 학습된다는 닫힌 세계 가정을 따르기 때문에 학습 데이터에 없는 클래스의 데이터에 대해서도 높은 확률로 학습 클래스 중 하나로 분류하게 된다(Fei and Liu, 2016). 하지만 실제 공정에서는 수백 여 가지의 알람이 존재하기 때문에 모든 알람 클래스 데이터를 수집하여 학습에 이용한다는 것은 현실적으로 불가능하다. 따라서 모델이 학습하지 않은 클래스의 알람이 입력되었을 때 학습한 알람 클래스로 오분류 하지 않고, 모르는 클래스로 분류할 수 있는 모델이 필요하며, 이러한 모델을 오픈셋 분류 모델이라고 한다(Scheirer *et al.*, 2012).

본 연구에서는 실제 공정에서 동시다발적으로 발생하는 알람에 대처할 수 있는 다중 레이블 오픈셋 분류 모델 구축을 목표로 하고 있으며, 학습하지 않은 클래스의 보조 정보로 사용할 수 있는 백그라운드 데이터(background data)를 활용해서 오픈셋 분류 성능을 향상시키는 방법론을 제안하였다. 본 논문의 중요 기여점은 다음과 같이 요약할 수 있다.

- 백그라운드 데이터를 접목한 다중 레이블 오픈셋 분류 방법론을 제안하여 기존 다중 레이블 분류 방법론에 비해 우수한 분류 정확도를 보임을 입증하였다.
- 실제 차량설비 공정에서 수집된 알람 클래스별 다채널 시계열 데이터를 이용한 실험을 통해 제안 방법론의 현실 적용 가능성을 입증하였다.

본 논문은 다음과 같은 구조를 가진다. 제 2장에서는 오픈셋 분류 방법론을 다룬 연구들을 소개한다. 제 3장에서는 본 연구에서 제안하는 다중 레이블 오픈셋 분류 모델을 소개하고 백그라운드 데이터를 활용하여 오픈셋 분류 성능을 높인 제안 방법론을 설명한다. 제 4장에서는 실제 공정에서 수집된 알람 클래스별 다채널 시계열 데이터를 소개하고 실험을 통해 제안 방법론이 기존의 오픈셋 분류 방법론에 비해 우수함을 보인다. 마지막으로 제 5장에서는 본 논문의 결론을 제시한다.

2. 관련 연구

초기의 오픈셋 분류 연구는 각 클래스별 독립적인 분류 모델을 구축하는 방향으로 진행되었다. 이 방법은 각 클래스별 모델을 구축하여 입력 데이터가 해당 클래스에 속하는지 그렇지 않음을 판단하여 어떠한 클래스에도 속하지 않을 경우 모델이 학습하지 않은 클래스로 분류되도록 했다. Schölkopf *et al.*(2001)는 각 클래스에 대하여 독립적으로 one-class support vector machine(one-class SVM) 모델을 구축하여 개별 예측 확률을 산출했다. 그러나 각 클래스별 모델 구축 시 해당 클래스만을 학습 데이터로 사용했기 때문에 낮은 분류 성능을 보였다. Fei and Liu(2015)는 center-based similarity(CBS) 공간 학습 방법을 제안했다. 이 방법은 각 클래스의 중심을 계산하고 입력 데이터를 각 중심에 대한 유사도 벡터로 변환한 후, 변환된 유사도 벡터를 사용하여 이진 분류기를 구축하는 것이다. 구축된 이진 분류기는 각 클래스별 구 형태의 분류 결정 경계면을 생성하여 모델의 출력값이 분류 결정 경계면 밖에 있다면 해당 클래스에 속하지 않는 것으로 간주하고 동시에 어떠한 클래스에도 속하지 않을 경우 학습하지 않은 클래스로 판단했다. Scheirer *et al.*(2012)는 분류 모델이 분류 결정 경계면을 구축할 때 학습 클래스가 거의 없는 공간을 배제하도록 하여 SVM 분류 결정 경계면의 공간을 절반으로 축소하였고, 학습하지 않은 클래스가 학습 클래스로 오분류되지 않도록 유도하였다. 아울러 이어지는 연구에서는 확률적인 분류 임계값을 이용해 기존과 비슷한 접근 방식으로 오픈셋 분류 성능 향상을 꾀하였다(Scheirer *et al.*, 2014). 하지만 확률적인 분류 임계값을 사용하려면 사전적 지식이 필요하다는 한계가 있다. Dalvi *et al.*(2013)은 expectation-maximization(EM) 알고리즘을 기반으로 semi-supervised learning을 통해 오픈셋 분류 문제를 접근했으나 기존 방법론들에 비하여 좋은 성능을 보이지 못했다.

Bendale and Boulton(2016)이 제안한 OpenMax 방법론은 처음으로 심층 신경망에 오픈셋 분류를 적용한 것으로 극단치 이론(extreme-value theorem)에 기반하여 기존 분류 모델의 소프트맥스 함수를 개선해 학습하지 않은 클래스에 대한 확률값도 산출할 수 있도록 했다. 하지만 OpenMax는 하이퍼 파라미터에 따라 민감하게 반응하기 때문에 일반화 성능이 떨어지는 문제점을 지녔고, 소프트맥스 함수의 기본 구조를 차용함에 따라 모든 클래스의 예측 확률이 비슷하게 나올 경우 학습하지 않은 클래스로 인식하기 때문에 학습하지 않은 클래스가 학습 클래스와 유사할 경우에는 오픈셋 분류 성능이 떨어진다는 한계를 가진다. Shu *et al.*(2017)는 모델이 출력하는 각 클래스별 예측 확률의 임계값을 가우시안 분포 피팅을 통해 학습한 클래스 중 하나로 분류되도록 유도했다. Dhamija *et al.*(2018)은 모델의 판정 기준인 임계값을 조정하는 대신 백그라운드 데이터를 학습에 포함하여 모델이 출력하는 확률을 줄이는 방법을 모색했다. 학습 클래스에 속하지 않는 데이터로 구성된 백그라운드 데이터는 학습 단계에서 모르는 클래스의 데이터에 대한 보조 정보를 모델에 제공한다. 하지만 백그라운드 데이터를 어떤 데이터로 선정해야 하는지에 대한 명확한 기준을 제시하지 못했다. Ditra *et al.*(2020)는 생성 모델인 conditional generative adversarial networks(cGAN)의 고정된 클래스 정보에 제한된 분류 공간을 심층 메트릭 학습을 통해 개선하여 특징 공간에서 오픈셋 분류가 가능하도록 유도했다. 또한 구축한 생성 모델의 특징 공간에서 모델이 학습한 클래스 밖의 데이터 샘플을 생성하였고 이를 기존 데이터의 증강 기법에 활용하여 분류기의 성능을 증가시켰다. Sun *et al.*(2020)는 확률적 생성 모델인 variational auto encoder(VAE)와 adversarial auto encoder(AAE)를 개선한 오픈셋 분류 모델 conditional probabilistic generative models(CPGM)을 제안했다. CPGM은 VAE와 AAE의 구조를 결합하여 잠재 공간에서 오픈셋 분류 수행하였다. 결합된 오픈셋 분류기는 학습 과정에서 재구성 오류와 클래스별 학습 데이터의 조건부 가우시안 분포를 통해 클래스별 잠재 벡터가 균집을 이루도록 유도했다. 학습이 완료

된 CPGM에 학습하지 않은 클래스의 데이터가 입력되면 잠재 공간 상에서 모든 클래스별 균집 밖에 위치하게 되고 모르는 클래스의 데이터로 분류할 수 있다.

3. 제안 방법론

본 연구에서는 동시에 발생한 알람을 다중 레이블로 예측할 수 있고 모델이 학습하지 않은 클래스의 알람을 모르는 알람으로 분류할 수 있는 다중 레이블 오픈셋 분류 모델을 제안한다. <Figure 1>은 제안 모델의 구조를 요약한 그림이다. 특징 추출기와 시그모이드 출력 레이어를 결합한 구조는 모델이 각 클래스별 확률을 독립적으로 출력할 수 있게 하여 다중 레이블에 대한 예측을 가능하게 한다. 또한 제안 방법론은 두 단계의 오픈셋 분류 알고리즘을 포함한다. (1) 백그라운드 데이터를 학습 과정에 반영하여 학습하지 않은 클래스의 데이터가 입력되었을 때 각 클래스별 확률을 낮게 출력하도록 유도했다. (2) 다음으로 학습된 모델이 출력하는 클래스별 확률값에 높은 임계값을 적용하여 엄격한 기준으로 각 클래스 레이블을 예측하도록 했다.

3.1 Multi-Label Classification Network

다채널 설비 센서로부터 수집된 시계열 데이터는 일정 시간의 윈도우 단위로 나누어지고, 나누어진 한 윈도우 단위의 입력 데이터는 (윈도우 길이×채널 수)의 크기를 갖고 모델에 입력된다. 입력된 데이터는 CNN을 활용한 특징 추출기와 완전 연결 레이어(fully connected layer)를 통해 (1×클래스 수)의 크기를 가지는 특징 벡터로 요약된다. 특징 추출기는 시계열 데이터의 특징 추출기로 많이 활용되는 RNN 및 LSTM으로도 대체할 수 있다. 하지만 convolution 연산을 사용하는 CNN 특징 추출기 또한 다채널 시계열 데이터의 특징 추출에서 좋은 성능을 보이고(Ince *et al.*, 2016), 대부분의 오픈셋 분류 알고리즘이

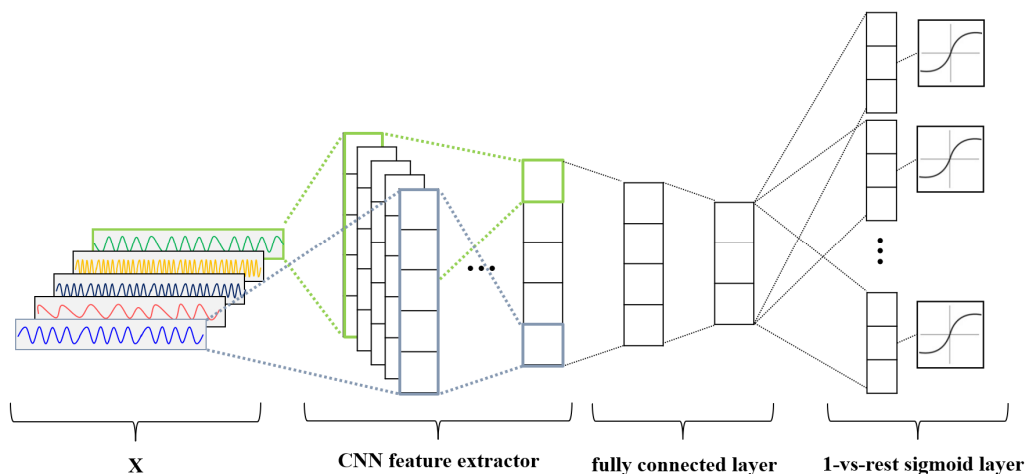


Figure 1. Deep Open Multi-label Classification Network

CNN을 기반으로 구성되어 있기 때문에 공정한 비교를 위하여 convolution 연산을 사용하는 CNN 특징추출기를 채택하였다.

요약된 ($1 \times$ 클래스 수)의 크기를 갖는 특징 벡터는 모델의 마지막 출력 레이어인 시그모이드 출력 레이어로 입력된다. 시그모이드 출력 레이어는 학습 데이터에 존재하는 클래스 수만큼 시그모이드 함수를 포함하며 이는 특징 벡터를 ($1 \times$ 클래스 수)의 크기를 갖는 각 클래스별 예측 확률로 변환한다.

손실함수(loss function)는 식 (1)과 같이 학습 데이터에 대한 모든 클래스별 시그모이드 함수가 출력하는 확률값의 로그 손실 총합으로 정의하였다. 식 (1)에서 y_i 는 i 번째 클래스의 정답 값을 나타내고, 시그모이드 출력 확률값 $f(s_i)$ 는 i 번째 클래스의 예측 확률을 나타낸다. 모델이 해당 손실함수를 최소화하는 방향으로 학습이 진행되면, i 번째 클래스 레이블 정답이 $y_i = 1$ 인 학습데이터가 입력될 때는 i 번째 시그모이드 출력 확률값 $f(s_i)$ 는 1에 가까워지도록 학습되고, $y_i = 0$ 인 학습데이터가 입력되면 $f(s_i)$ 는 0에 가까워지도록 학습된다.

$$Loss = - \sum_{i=1}^C \{y_i \log(f(s_i)) + (1 - y_i) \log(1 - f(s_i))\} \quad (1)$$

모델 학습이 완료되면 새로운 데이터 x_i 가 입력되었을 때 모델은 각 클래스별 시그모이드 출력 확률 $f(s_i)$ 를 산출하고 이 산출값이 임계값 t_i 보다 크면 i 번째 클래스 예측 레이블은 1이 되고, 반대로 임계값보다 작으면 i 번째 클래스 예측 레이블은 0이 된다(식 (2)). 만일 모든 클래스에서 시그모이드 출력 확률이 임계값보다 작으면 예측 레이블은 모든 클래스에 대해 0이 되고, 이는 어떠한 클래스에도 속하지 않는 모르는 클래스로 분류된다. 또한 여러 클래스에서 예측 레이블이 1이 될 수도 있으므로 다중 레이블 예측으로의 확장이 가능하다.

$$\hat{y}_i = \begin{cases} \hat{y}_i = 1, & \text{if } f(s_i) > t_i \\ \hat{y}_i = 0, & \text{if } f(s_i) < t_i \end{cases} \quad (2)$$

3.2 Background Data

모델이 학습하지 않은 클래스의 데이터를 모르는 클래스로 분류하기 위해서는 해당 클래스가 아닌 데이터에 대해서는 낮은 예측 확률을 출력할 수 있어야 한다. 특히, 어떠한 학습 클래스에도 속하지 않는 모르는 데이터의 경우 모든 학습 클래스에서 낮은 예측 확률을 출력할 수 있어야 한다. 학습하지 않은 클래스 데이터의 모든 기존 클래스 확률을 낮추기 위한 방법으로 학습 단계에서 백그라운드 데이터를 활용할 수 있다 (Dhamija et al., 2018). 백그라운드 데이터를 활용한 방법론은 모델이 학습하고자 하는 클래스 외의 데이터를 백그라운드 데이터로 활용하여 해당 데이터들이 모든 학습 클래스에서 낮은 예측 확률을 출력하도록 모델을 학습시킨다. 이 때 백그라운드 데이터는 모델이 학습하지 않은 클래스의 데이터에 대한 가이드 라인으로 작용하여 추후 학습된 모델에 모르는 클래스

의 데이터가 입력되었을 때 모든 학습 클래스에서 낮은 예측 확률을 출력하도록 유도한다.

기존 백그라운드 데이터 활용 방법론은 백그라운드 데이터에 대한 손실함수를 새롭게 정의하여 이를 기존 손실함수에 결합하는 방식을 채택한다. 이를 통해 모델은 학습 클래스의 데이터에 대해서는 해당 클래스로 잘 분류할 수 있도록 하고, 학습 클래스에 속하지 않는 백그라운드 데이터에 대해서는 어떠한 클래스로도 분류되지 않도록 학습한다. 하지만 기존 방법론은 단일 레이블 분류 상황을 다루고 있기 때문에 소프트맥스 함수를 통해 각 클래스별 예측 확률값을 산출한다. 소프트맥스 함수가 출력하는 클래스별 예측 확률은 정의에 따라 모든 클래스 확률합이 1이 되어야 한다(식 (3)). 즉, 학습 클래스에 속하지 않는 백그라운드 데이터가 어떠한 클래스로도 분류되지 않게 하려면 모델이 출력하는 클래스별 예측 확률값이 모든 클래스에 대해 동일해야 한다. 하지만 모든 클래스에 대해 동일한 예측 확률을 출력하게끔 모델을 학습시키는 것은 특정 클래스에서 높은 확률을 출력하게 하는 것보다 상대적으로 쉬운 학습에 속한다. 따라서 손실함수를 각각 정의하여 결합한 모델의 경우 학습과정에서 상대적으로 쉬운 작업인 백그라운드 데이터에 대한 손실함수 부분이 빠르게 수렴하여 분류 모델의 본래 목적인 학습 클래스를 분류하는 작업이 잘 이루어지지 않는다는 한계가 있다.

$$f(s_i) = \frac{e^{s_i}}{\sum_{j=1}^C e^{s_j}}, \quad (3)$$

제안 방법론은 다중 레이블 예측을 위해 각 클래스별 이진 분류를 수행하기 때문에 손실함수를 새롭게 정의하지 않고도 백그라운드 데이터를 모델 학습에 쉽게 반영할 수 있다. 제안 방법론의 손실함수를 최소화하는 목적식은 시그모이드 출력 확률이 학습 데이터의 레이블에 수렴하도록 한다. 즉, 모델은 학습 데이터의 레이블에 따라 출력하는 각 클래스별 예측 확률값이 결정된다. 따라서 손실함수를 재정의할 필요없이 백그라운드 데이터의 모든 정답 레이블을 0으로 입력하여 모델을 학습시키면 백그라운드 데이터에 대한 모든 클래스 예측 확률값이 0에 가까워지고 학습 클래스 중 어디에도 속하지 않는 모르는 클래스의 예로 학습될 수 있다. 또한 학습이 완료된 모델은 해당 클래스에 속하지 않는 데이터에 대해 0에 가까운 낮은 확률을 출력하기 때문에 모르는 클래스의 데이터를 단일 레이블 문제보다 용이하게 분류할 수 있다.

3.3 Threshold

각 클래스별 시그모이드 출력 확률에 대한 임계값 조정을 통해 오픈셋 분류 성능을 한 단계 더 향상시킬 수 있다. 일반적으로 각 클래스별 시그모이드 출력 확률에 대한 임계값은 0.5를 사용한다. 즉, 시그모이드 출력 확률이 0.5보다 크면 해당 클래스로, 그렇지 않으면 해당 클래스가 아닌 것으로 판정한다.

그러나 제안 방법론의 손실함수를 최소화하는 방향으로 학습된 모델은 학습데이터에 대해 해당하는 클래스는 시그모이드 출력 확률이 1에 수렴하는 값을 출력하고, 해당하지 않는 클래스는 0에 수렴하는 값을 출력하게 되므로 모델은 학습 데이터에 대해 0 또는 1에 가까운 양극단의 확률만을 출력하게 된다. 즉, 양극단이 아닌 중간 영역은 모델이 학습하지 않은 클래스의 데이터가 출력하는 확률이 잠재적으로 존재할 수 있는 공간이 된다. 따라서 임계값 0.5를 사용한 분류 결정 경계는 단순히 분류 공간을 양분하기 때문에 중간 영역에 존재하는 모델이 학습하지 않은 클래스의 데이터를 학습 클래스로 오분류할 가능성이 높아진다(Scheirer *et al.*, 2012). 이는 학습 데이터의 클래스별 출력 확률의 분포를 반영하여 임계값을 증가시킴으로써 개선할 수 있다(Shu *et al.*, 2017). 각 클래스별 정답 레이블 $y_i = 1$ 인 학습 데이터의 시그모이드 출력 확률 $f(s_i)$ 가 평균이 1인 가우시안 분포의 절반을 따른다고 가정하고, 해당 가우시안 분포의 표준 편차(σ_i)를 사용하여 각 클래스별 임계값을 $\max(0.5, 1 - 3\sigma_i)$ 로 대체하면 학습된 클래스의 데이터에 기반하여 더욱 엄격한 분류 결정 경계를 가질 수 있다. 학습한 클래스의 정보를 반영한 엄격한 분류 결정 경계는 각 클래스의 학습 데이터 분포로부터 가까운 데이터에 대해서만 해당 클래스로 분류할 수 있게 된다.

4. 실험 결과

4.1 데이터 소개

본 연구에서 사용한 데이터는 2020년 4월 1일부터 2020년 5월 31일까지 자동차 제조 업체 A사의 엔진 공장 헤드 가공 공정 총 12개의 설비에서 수집된 42채널의 알람 데이터이다. 알람 데이터에는 발생한 알람 클래스의 정보와 수집 시점이 기록되어 있는데 이를 토대로 수집 시점 3일 전까지의 기간을 해당 알람 구간으로 정의했다. 정상 구간은 최대한 알람의 전조 증상을 배제하기 위해 알람 발생 간격이 3주 이상인 구간에서 추출하였다.

실험에서는 정상 클래스와 총 6종의 알람 클래스를 학습 클래스로 사용했다. 학습 클래스에 활용한 6종의 알람 클래스는 헤드 가공공정에서 발생할 수 있는 총 89종의 알람 클래스 중 설비 가동율을 저하시키는 중요한 알람 유형이다. 다음으로 백그라운드 데이터는 학습 클래스와 유사하지만 다른 클래스로는 구분될 수 있는 클래스의 데이터를 활용하는 것이 바람직하기 때문에(Lee *et al.*, 2017) 본 연구에서는 헤드 가공공정의 설비와는 다른 작업을 수행하는 블락 가공공정의 설비 데이터 중 3종의 알람 클래스 데이터를 활용했다. 마지막으로 테스트 과정에서 오픈셋 분류 성능의 평가를 위해 사용될 모델이 모르는 클래스의 데이터는 헤드 가공공정에서 발생할 수 있는 3종의 중요하지 않은 알람 클래스를 활용했다.

4.2 데이터 전처리

먼저 데이터 수집 간격이 고르지 못한 다채널 시계열 데이터를 1초 단위의 평균으로 요약하여 수집 간격을 통일해 주었다. 다음으로 한 관측치 단위를 60초 간격의 윈도우로 설정하여 전체 데이터가 각각(윈도우 크기×채널 수)의 크기를 가지도록 전처리를 진행했다. 데이터는 단일 클래스 7종과 다중 클래스 2종으로 구성되어 있다. 단일 클래스에는 정상 클래스와 단독으로 발생한 6종의 알람 클래스가 포함되어 있으며 다중 클래스에는 알람 클래스 1과 알람 클래스 2가 동시에 발생한 케이스와 알람 클래스 1과 알람 클래스 5가 동시에 발생한 케이스가 포함되어 있다. 모델의 학습과 평가를 위해 전체 데이터를 학습 데이터(training data) 8,797개, 테스트 데이터(testing data) 1,830개로 분리하였고 이를 <Table 1>에 요약하였다. 제안 방법론에 사용될 백그라운드 데이터는 총 2,549개로 레이블 값을 모든 학습 클래스에 대해 0으로 입력하여 학습에 포함시켰고, 모델이 모르는 클래스의 데이터(unknown data)를 2,995개로 구성하여 오픈셋 분류 성능 평가에 사용하였다(<Table 2> 참조).

Table 1. Summary of the Training and Testing Data

Class	Training data	Testing data	Total
Normal	1,804	395	2,199
Alarm 1	1,158	223	1,381
Alarm 2	443	78	521
Alarm 3	1,420	302	1,722
Alarm 4	658	129	787
Alarm 5	793	172	965
Alarm 6	267	55	322
Alarm 1+Alarm 2	1,329	279	1,608
Alarm 1+Alarm 5	925	197	1,122
Total(Known)	8,797	1,830	10,627

Table 2. Summary of the Background and Unseen Data

Class	Training data	Testing data
Background 1	868	-
Background 2	1,018	-
Background 3	663	-
Unknown 1	-	729
Unknown 2	-	824
Unknown 3	-	1,442
Total	2,549	2,995

4.3 실험 설계 및 모델 학습

먼저 제안 방법론과 오픈셋 알고리즘을 반영하지 않은 multi-label classification(MLC) 모델을 비교하여 오픈셋 알고리즘의 효과를 검증하였다. 다음으로 다중 레이블 분류에 적용할 수 있는 최신 오픈셋 분류 알고리즘을 적용한 모델과 비교하여 제안 방법론의 우수성을 살펴보았다. 비교한 오픈셋 분류 모델은

시그모이드 출력 확률값의 분포를 임계값에 반영한 deep open classification(DOC) 기반의 다중 레이블 오픈셋 분류 모델(Shu *et al.*, 2017)과 극단치 이론(extreme value theorem; evt)에 근거하여 임계값을 설정하는 다중 레이블 오픈셋 분류 모델(Jang and Kim, 2020), 백그라운드 데이터에 대한 손실함수를 재정의하여 학습에 포함시킨 다중 레이블 분류 모델(Dhamija *et al.*, 2018)로 구성된다.

본 실험에서는 공정한 비교를 위해 모든 비교 방법론에서 제안 방법론과 동일한 모델 구조를 적용하였다. 실험에 사용된 CNN 특징추출기는 반복 실험 과정에서 실험 데이터에 대하여 성능이 높게 나온 3-dense block으로 이루어진 densely connected convolutional network(DenseNet)을 사용하였다. Optimizer는 Adam optimizer를 사용하였으며, 학습율은 초기값 0.01에서 10번째 epoch 이후의 학습부터 이전 epoch에서의 학습율의 90%로 점차적으로 줄어듦을 설정하였다. 학습 데이터는 훈련:검증=8:2로 분할하였고, 분할된 데이터셋의 batch size는 훈련 데이터, 검증 데이터 모두 128로 설정하였다. 모델은 훈련 데이터만을 사용하여 학습을 진행했으며 100번의 epoch 중 검증 데이터에 대한 손실함수 값이 가장 작은 epoch에서의 모델을 최적 모델로 선정하였다. 실험에 대한 최종 결과는 훈련, 검증 데이터의 분할 시드를 변경해가며 총 20회 반복실험을 통해 산출하였다.

4.4 평가 지표

모델의 분류 성능을 평가하기 위해 다중 레이블 분류 상황에서 널리 쓰이고 있는 Jaccard similarity score를 사용했다. 식 (4)에서와 같이 Jaccard similarity score는 예측된 레이블(Z_i)과 실제 레이블(Y_i)을 모두 포함하는 레이블 대비 예측된 레이블과 실제 레이블이 일치하는 레이블의 비율을 나타낸다.

$$Jaccard\ Similarity\ Score = \frac{1}{n} \sum_{i=1}^n \frac{|Y_i \cap Z_i|}{|Y_i \cup Z_i|}, \quad (4)$$

또한 실험한 모든 모델은 각 클래스별 독립적인 이진 분류를 수행하므로 클래스별 정오행렬을 구축할 수 있다. 클래스별 정

오행렬에서 참 양성, 거짓 양성, 거짓 음성, 참 음성을 산출할 수 있다. 참 양성은 해당 클래스의 데이터를 해당 클래스로 정분류한 것으로, 거짓 양성은 해당 클래스가 아닌 데이터를 해당 클래스로 오분류한 것으로 정의된다. 또한 거짓 음성은 해당 학습 클래스의 데이터를 해당 클래스가 아닌 데이터로 오분류한 것으로, 참 음성은 실제 해당 클래스가 아닌 데이터를 모델이 정분류한 것으로 정의된다. 참 양성, 거짓 양성, 거짓 음성, 참 음성으로 나타나는 전체 분류 결과 중 올바르게 분류된 참 양성, 참 음성의 비율을 계산하면 모델의 분류 성능을 평가하는 지표인 정확도(accuracy)를 계산할 수 있다. 하지만 각 클래스별 독립적인 이진 분류를 수행할 경우 정확도는 다소 과대 평가될 수 있기 때문에 본 연구에서는 정밀도(precision)와 재현율(recall)의 조화평균인 F_1 score를 이용하여 성능평가를 수행하였다. 정밀도는 참 양성과 거짓 양성을 합한 값 중에 참 양성의 비율로 모델이 해당 클래스로 분류한 데이터 중에서 실제 해당 클래스의 데이터 비율을 나타내며, 재현율은 참 양성과 거짓 음성을 합한 값 중에 참 양성의 비율로 계산된다.

4.5 실험 결과

<Table 3>은 본 연구에서 제안하는 방법론과 비교 방법론의 전체 테스트 데이터에 대한 Jaccard similarity score와 F_1 score를 나타낸다. 성능 평가 측도는 20번 반복 실험 결과의 평균값을 사용하였으며 평균값과 함께 표준편차를 괄호 안에 함께 표기하였다. 실험 결과 학습 클래스(known classes)의 분류 성능은 오픈셋 분류 알고리즘을 적용하지 않은 MLC 모델에서 가장 높게 나왔지만 다른 모델과 유의미한 차이는 보이지 않았다. 모델이 모르는 클래스(unknown classes)의 분류 성능은 오픈셋 분류 알고리즘을 적용하지 않은 MLC 모델에서 Jaccard similarity score 기준 0.1209로 가장 낮게 나왔으며 이는 대부분의 모르는 클래스의 데이터를 모델이 학습한 클래스 중 하나로 오분류한 것으로 볼 수 있다. 반면 제안 방법론은 Jaccard similarity score 기준 0.8397로 다른 비교 방법론에 비해 독보적으로 높은 성능으로 모델이 학습하지 않은 클래스의 데이터를 모르는 데이터로

Table 3. Average Classification Results for Testing Datasets over 20 Repeated Experiments. Standard Deviations are Indicated in Parentheses

Algorithm	Jaccard Similarity Score			F_1 Score		
	Known classes	Unknown classes	Overall classes	Known classes	Unknown classes	Overall classes
MLC(t = 0.5)	0.9813 (0.0141)	0.1209 (0.0232)	0.5940 (0.0182)	0.9852 (0.0124)	0.1157 (0.0251)	0.5698 (0.0207)
EVT-based MLC	0.9514 (0.0258)	0.4401 (0.0173)	0.7450 (0.0242)	0.9534 (0.0242)	0.4288 (0.0167)	0.7324 (0.0175)
DOC	0.9604 (0.0235)	0.7429 (0.0244)	0.8625 (0.0238)	0.9632 (0.0157)	0.7389 (0.0146)	0.8597 (0.0155)
background	0.9598 (0.0152)	0.6691 (0.0243)	0.7794 (0.0211)	0.9601 (0.0178)	0.6705 (0.0155)	0.7803 (0.0170)
Proposed method	0.9645 (0.0114)	0.8397 (0.0218)	0.8870 (0.0187)	0.9685 (0.0151)	0.8401 (0.0148)	0.8881 (0.0149)

Table 4. Average Classification Results for Multi-label Classes over 20 Repeated Experiments. Standard Deviations are Indicated in Parentheses

Class	Algorithm				
	MLC(t = 0.5)	EVT-based MLC	DOC	Background	Proposed method
Alarm 1+Alarm 2	0.9655 (0.0124)	0.9950 (0.0257)	0.9526 (0.0222)	0.9838 (0.0237)	0.9785 (0.0199)
Alarm 1+Alarm 5	0.9918 (0.0155)	0.9575 (0.0155)	0.9828 (0.0099)	0.9690 (0.0132)	0.9949 (0.0118)

올바르게 분류했다. 특히 DOC 모델과 제안 방법론은 동일한 모델 구조와 손실 함수를 가지지만 백그라운드 데이터의 활용 여부에 기인해 성능의 차이가 발생했다. 또한, 백그라운드 데이터에 대한 손실 함수를 재정의하여 학습한 비교 모델(background)에 비해 제안 방법론이 학습 클래스 데이터와 모르는 클래스 데이터 모두에서 우수한 성능을 보였다. 이는 본 연구에서 제안하고 있는 백그라운드 데이터 사용이 분류 성능을 높이는 데 큰 기여를 했다고 볼 수 있다.

<Table 4>는 각 모델별 2종의 다중 레이블 클래스에 대한 20번 반복 실험 결과의 평균 Jaccard similarity score와 표준편차를 나타낸다. 알람 클래스 1과 알람 클래스 2가 동시에 발생한 279개의 테스트 데이터에 대해서는 오픈셋 분류 알고리즘을 적용하지 않은 MLC 모델에서 Jaccard similarity score가 0.9838로 가장 높게 나타났고, 알람 클래스 1과 알람 클래스 5가 동시에 발생한 197개의 테스트 데이터에 대해서는 제안 방법론에서 0.9949로 가장 높게 나왔다. 하지만 제안 방법론과 비교 방법론들 모두 유의미한 차이 없이 높은 정확도로 다중 레이블 분류를 수행함을 확인할 수 있었다.

다음으로 백그라운드 데이터와 모르는 클래스 데이터의 유사도에 따른 성능의 변화를 확인하기 위해 모르는 클래스 데이터의 최대값과 최소값을 범위로 가지는 균등 분포로부터 임의 추출한 데이터를 입력하여 제안 방법론의 성능을 확인하였다. 균등분포로부터 임의 추출한 데이터에 대해서도 총 20번의 반복 실험을 하였고, 그 결과 Jaccard similarity score의 평균이 0.7950, 표준편차는 0.0325로 나타났다. 이는 임의로 생성된 모르는 패턴의 데이터가 입력되어도 제안 방법론은 모르는 클래스의 데이터로 잘 분류할 수 있음을 뜻한다. 즉, 백그라운드 데이터와 모르는 클래스 데이터가 유사하지 않은 경우에도 제안 방법론은 이를 학습한 클래스 중 하나로 오분류하지 않음을 알 수 있었다.

5. 결 론

설비 가동율을 저하시키는 중요 알람을 분류하여 조기에 감지하는 일은 중요하다. 하지만 실시간으로 수많은 양의 알람 데이터가 발생하는 상황에서 작업자가 이를 모두 관리하기 어렵기 때문에 이를 보완할 수 있는 인공지능 모델은 꼭 필요한 실정이다. 기존의 분류 모델은 수집된 데이터에 존재하는 클래스에 대

해서만 분류할 수 있도록 학습되었다. 하지만 실제 공정에서는 각 설비별로 수십 가지의 알람 클래스가 정의되어 있고 알람 클래스마다 그 발생 시점이 특정되지 않을 뿐만 아니라 발생되지 않는 경우도 많기 때문에 발생 가능한 모든 클래스의 알람 데이터를 수집하는 것은 불가능에 가깝다. 따라서 한정적으로 확보된 데이터를 이용하여 모델을 구축해야 하는데 이는 데이터 수집 단계에 존재하지 않은 클래스의 데이터가 입력되었을 때 기존 클래스 중 하나로 강제 분류하게 되는 오류를 발생시킨다. 본 연구에서는 백그라운드 데이터를 접목한 다중 레이블 오픈셋 분류 방법론을 제안하여 기존 오픈셋 분류 방법론에 비해 우수한 분류 성능을 보임을 입증하였고, 알람이 동시에 발생하는 실제 공정 데이터를 통한 실험으로 높은 현실 적용 가능성을 입증하였다. 제안 방법론을 활용하면 실제 공정에서 동시다발적으로 발생하는 알람을 다중 레이블을 통해 예측할 수 있으며, 현재까지 수집된 데이터만으로 분류 모델을 학습하여도 데이터 수집 단계에 존재하지 않았던 다른 클래스의 데이터를 학습한 클래스 중 하나로 오분류하지 않을 수 있기 때문에 다양한 알람 유형이 존재하는 실제 공정 상황에서 유연하게 대처할 수 있을 것으로 기대한다.

참고문헌

- Beebe, D., Ferrer, S., and Logerot, D. (2013), The Connection of Peak Alarm Rates to Plant Incidents and what you can do to Minimize, *Process Safety Progress*, **32**(1), 72-77.
- Bendale, A. and Boulton, T. E. (2016), Towards Open Set Deep Networks, *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1563-1572.
- Bransby, M. L. and Jenkinson, J. (1998), *The Management of Alarm Systems*, Sudbury : HSE Books, 193-195.
- Cai, S., Palazoglu, A., Zhang, L., and Hu, J. (2019), Process Alarm Prediction Using Deep Learning and Word Embedding Methods, *ISA Transactions*, **85**, 274-283.
- Dalvi, B., Cohen, W. W., and Callan, J. (2013), Exploratory Learning, *In Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, Springer, Berlin, Heidelberg, 128-143.
- Dhamija, A. R., Günther, M., and Boulton, T. (2018), Reducing Network Agnostophobia, *In Advances in Neural Information Processing Systems*, 9157-9168.
- Ditria, L., Meyer, B. J., and Drummond, T. (2020), OpenGAN : Open Set Generative Adversarial Networks, *In Proceedings of the Asian Conference on Computer Vision*.

- Fei, G. and Liu, B. (2015), Social Media Text Classification Under Negative Covariate Shift, *In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 2347-2356.
- Fei, G. and Liu, B. (2016), Breaking the Closed World Assumption in Text Classification, *In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies*, 506-514.
- Ince, T., Kiranyaz, S., Eren, L., Askar, M., and Gabbouj, M. (2016), Real-Time Motor Fault Detection by 1-D Convolutional Neural Networks, *IEEE Transactions on Industrial Electronics*, **63**(11), 7067-7075.
- Jang, J. and Kim, C. O. (2020), One-vs-Rest Network-based Deep Probability Model for Open Set Recognition, *arXiv preprint arXiv:2004.08067*.
- Lee, K., Lee, H., Lee, K., and Shin, J. (2017), Training Confidence-Calibrated Classifiers for Detecting Out-of-Distribution Samples, *arXiv preprint arXiv:1711.09325*.
- Scheirer, W. J., de Rezende Rocha, A., Sapkota, A., and Boulton, T. E. (2012), Toward Open Set Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **35**(7), 1757-1772.
- Scheirer, W. J., Jain, L. P., and Boulton, T. E. (2014), Probability Models for Open Set Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **36**(11), 2317-2324.
- Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., and Williamson, R. C. (2001), Estimating the Support of a High-Dimensional Distribution, *Neural Computation*, **13**(7), 1443-1471.
- Shu, L., Xu, H., and Liu, B. (2017), DOC : Deep Open Classification of Text Documents, *In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2911-2916.
- Sun, X., Zhang, C., Lin, G., and Ling, K. V. (2020), Open Set Recognition with Conditional Probabilistic Generative Models, *arXiv preprint arXiv:2008.05129*.
- Teng, M. (2010), Anomaly Detection on Time Series, *In 2010 IEEE International Conference on Progress in Informatics and Compu-*

ting, IEEE, 603-608.

- Zhu, J., Wang, C., Li, C., Gao, X., and Zhao, J. (2016), Dynamic Alarm Prediction for Critical Alarms Using a Probabilistic Model, *Chinese Journal of Chemical Engineering*, **24**(7), 881-885.

저자소개

김상훈 : 고려대학교 산업경영공학과에서 2019년 학사 학위를 취득하고 고려대학교 산업경영공학과에서 석·박사 통합과정에 재학 중이다. 연구분야는 Machine Learning Algorithms, Predictive Analytics for Sequential Patterns, Open-Set Recognition Algorithms이다.

홍현욱 : 현대자동차CDO 산하 빅데이터 기획팀에서 재직 중이며, 연구 분야는 People Analytics(HR Network), Statistical Modeling & Visualization, NLP, 설비 예지 보전 분야이다.

정의현 : 현대자동차CDO 산하 빅데이터 기획팀에서 재직 중이며, 연구 분야는 Computer Vision, Data Engineering & Mining, 설비 예지 보전 분야이다.

김성범 : 한양대학교 산업공학과에서 1999년 학사를 취득하고 2001년과 2005년 미국 Georgia Institute of Technology에서 산업공학 석사학위, 박사학위를 취득하였다. 미국 텍사스 주립대학교 교수를 역임하고 2009년부터 고려대학교 산업경영공학부 교수로 재직하고 있다. 연구 분야는 인공지능, 머신러닝이다.