

멀티태스크 러닝을 이용한 심층신경망 기반 기계 소리 이상 진단

최지훈, 최정우*

한국과학기술원, *한국과학기술원

wlgns2533@kaist.ac.kr, *khepera@kaist.ac.kr

Deep Learning based Machine Sound Anomaly Detection using Multi-Task Learning

Choi Ji Hoon, Choi Jung Woo*

KAIST, *KAIST

요약

소리에 기반하여 기계류의 이상 탐지를 수행하는 딥러닝 기법은, 기계류 작동 시에 발생할 수 있는 다양한 비정상 데이터를 사전에 획득하기 어려운 문제를 안고 있다. 때문에 정상 기계의 소리 데이터만을 가지고 사전 정의된 태스크를 수행하여 정상 데이터의 특징을 학습하는 자기 지도학습 방법이 널리 사용되고 있다. 본 논문은 기존 모델들이 특정한 단일 태스크로 자기 지도학습을 수행하면서 발생하는 한계를 극복하기 위하여 과거의 소리 데이터로 미래를 예측하는 복원 태스크와 서로 다른 기계 종류를 분류하는 분류 태스크를 동시에 수행하는 새로운 멀티태스크 모델을 제안한다. 제안하는 멀티태스크 모델은 WaveNet 기반의 특징 추출부를 소리 재현부와 분류부가 공유하는 구조를 가지고 있으며, 기계류 이상 탐지에 있어 기존의 단일 태스크 모델에 비해 더 높은 성능을 보여준다.

I. 서론

소리에 기반하여 기계류의 이상 유무를 탐지하는 딥러닝 기법은 산업의 여러 분야에서 사용되고 있다[1]. 이 과정에서 가장 큰 문제점은 기계의 정상 소리는 상시 측정이 가능해 데이터 확보가 쉽지만, 비정상 소리는 드물게 발생하고 그 발생 양상을 예측하기 어렵기 때문에 훈련 데이터를 수집하기 어렵다는 것이다. 이 때문에, 정상 소리 데이터만을 이용한 자기 지도학습 방법이 연구되고 있다.

기계 소리 이상 탐지를 위한 여러 모델 학습 방법 중에는, 과거의 정상 소리로 미래의 정상 소리를 예측 혹은 복원하는 자기 회귀 과정을 통해 정상 소리의 특징을 학습하는 복원 학습 방법과 기계 종류를 분류해 기계 간 소리의 특징을 학습하는 분류 학습 방법이 있다. 자기 회귀 태스크를 통해 훈련된 모델은 정상 기계 소리에 대해서는 학습한 기계의 특징을 이용해 잘 복원하게 되어 복원 오차가 작지만, 비정상 기계 소리에 대해서는 정상 기계 소리의 특징이 없으므로 복원을 잘하지 못해 큰 복원 오차를 나타내게 된다. 반면, 기계 종류 분류 태스크는 서로 다른 모델의 기계 데이터를 분류하도록 학습하며, 학습 과정에서 기계 종류별 정상 소리의 특징을 학습하게 된다. 학습된 모델은 정상 기계 소리에 대해서는 기계 종류를 예측할 확률이 높게 나오지만, 비정상 소리에 대해서는 분류 예측 확률이 낮게 나오게 된다.

본 논문에서는 기존에 존재하는 기계 소리 이상 탐지 모델의 성능을 향상하기 위해, 복원 학습 방법과 기계 id 분류 학습 방법을 동시에 적용한 새로운 멀티태스크 모델을 제안한다.

II. 본론

본 논문에서 제안하는 새로운 멀티태스크 모델은 입력 신호의 공통적인 특징을 추출하는 부분인 특징 추출 모델, 추출한 특징을 이용해 신호를 복원하게 되는 복원 모델, 특징을 이용해 기계 id를 분류하는 분류 모델 세 부분으로 나눌 수 있다. 모델의 전체 구조는 그림 1에 나타났다.

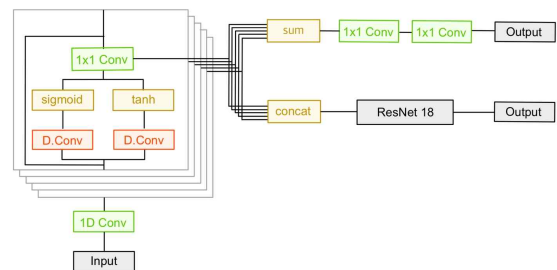


그림 1. 멀티태스크 모델의 구조

먼저 모델의 특징 추출부는 WaveNet[2]의 구조를 사용했다. WaveNet은 End-to-End로 음성 데이터를 생성하는 것을 목적으로 사용되는 생성 모델이지만, 이를 변형한 스펙트로그램 미래 예측 모델이 제안된 바 있다[3]. WaveNet의 특징인 causal convolution을 이용해 일정 시간의 스펙트로그램 데이터를 입력으로 받으면 다음 시간 프레임의 주파수 bin 값을 예측하게 된다. 본 연구에서는 WaveNet의 residual block 부분을 분리하여, 스펙트로그램의 복원 과정과 기계 소리 분류 과정에 사용될 공통 특징을 추출하는 추출부로 사용한다. 특징으로 뽑아낸 residual block의 출력은 각각 복원 모델과 분류 모델의 입력으로 사용되어 스펙트로그램 복원과 기계 소리 분류를 동시에 수행한다. 복원 모델은 1x1 convolution을 두 번 사용했고, 분류 모델은 18-layer ResNet[4]을 사용했다. 모델의 손실함수는 복원 모델의 손실함수와 분류 모델의 손실함수 합으로 나타났다.

$$L_{total} = L_{recon} + L_{class} \quad (1)$$

식 (1)에서의 L_{recon} 은 입력과 출력 간 스펙트로그램의 L_2 값을 사용했고, L_{class} 은 categorical cross entropy를 사용했다.

모델 학습을 위한 데이터 세트는 DCASE2020 Challenge Task2에서 제공한 기계 소리를 사용했다[5]. ToyCar, ToyConveyor, fan, slider, pump, valve 총 6개의 기계 종류가 있고, 각 기계 종류별로 다른 기계 번호를 나타내는 id로 세분되어있다. 10초의 기계 소리 음원 파일을 STFT를 통

해 스펙트로그램을 구했다. 이때 음원의 sampling rate는 16000, window 크기는 2048, hop length는 512, 주파수 bin 개수는 128개로 설정했다. 전 처리 과정으로 기계 종류마다 정규화를 진행한 후 모델의 입력으로 사용했다. 데이터의 다양성을 위해 데이터 증강 기법도 사용했다. 서로 다른 id를 갖는 같은 종류의 두 기계의 스펙트로그램을 일정 비율로 섞어 새로운 이미지를 생성한다. 이때 복원 모델은 새로운 이미지 자체를 복원하고, 분류 모델은 새로 생성된 이미지가 각 기계 id에서 얼마만큼의 비율로 섞였는지를 예측하게 된다. 데이터의 배치 크기는 64, 학습 방법은 Adam, learning rate는 10^{-4} 을 사용했고, 100 epoch까지 학습한 결과를 이용했다.

데이터의 평가 방법은 기계 종류에 따라 다른 방법을 사용한다. 전 주파수 영역에 대해 기계 소리가 관측되는 valve와 slider는 학습할 때처럼 입력과 출력 사이의 L_2 값을 사용한다. 나머지 기계들은 특정 주파수 대역에서 기계 소리가 관측되므로, 입력과 출력 이미지 사이의 시간 평균을 취해 주파수 벡터를 구한 뒤, Mahalanobis distance를 계산한다[3].

$$L_{\text{recon}}^{ij} = \sqrt{(x_{ij} - m_j)^T \Sigma^{-1} (x_{ij} - m_j)} \quad (2)$$

식 (2)에서의 x_{ij} 는 기계 종류 j 의 i 번째 데이터로부터 얻은 주파수 bin 벡터, m_j 는 학습시킨 기계 종류 j 의 데이터로부터 얻은 주파수 벡터들의 평균, Σ_j 는 기계 종류 j 의 데이터 세트로 구한 주파수 bin 간의 공분산 행렬이다. 분류 성능에 대한 평가는 학습할 때처럼 categorical cross entropy를 사용한다. 이상 탐지 분류를 위한 이상 점수는 복원에 대한 점수와 분류에 대한 점수의 Mahalanobis distance를 계산한다.

$$\text{Score} = \sqrt{(l_{ij} - d_j)^T S_j^{-1} (l_{ij} - d_j)} \quad (3)$$

식 (3)에서의 l_{ij} 는 기계 종류 j 의 i 번째 데이터로부터 얻은 손실함수 벡터, d_j 는 학습시킨 기계 종류 j 로부터 얻은 손실함수 벡터들의 평균, S_j 는 기계 종류 j 의 데이터 세트로 구한 손실함수 간의 공분산 행렬이다.

모델의 이상 탐지 성능 평가를 위한 지표로는 ROC-AUC와 pAUC를 사용했다[5]. 성능 비교를 위해, 자기 회귀 task만 수행하는 WaveNet 모델, 분류 task만 수행하는 ResNet 모델, 그리고 새로 제안하는 멀티태스크 모델과 DCASE 2020 Challenge에서 가장 좋은 성능을 보인 모델[6]의 ROC-AUC와 pAUC를 비교하였다.

	WaveNet	ResNet	Ours	Giri[6]
ToyCar	0.82/0.65	0.88/0.87	0.90/0.86	0.95/0.90
Toy-Conveyor	0.67/0.54	0.55/0.55	0.70/0.57	0.81/0.66
fan	0.91/0.78	0.70/0.70	0.92/0.86	0.82/0.79
pump	0.90/0.78	0.77/0.77	0.89/ 0.83	0.87/0.80
slider	0.94/0.82	0.73/0.72	0.99/0.96	0.97/0.90
valve	0.78/0.57	0.87/0.87	0.97/0.92	0.97/0.91

표 1. 기계 종류별 이상 진단 결과 (ROC-AUC/pAUC)

표 1에서 볼 수 있듯이 6종의 기계 중 pump를 제외한 모든 기계에 대해 이상 진단 성능이 향상됨을 알 수 있다. 특히 valve에 대해서는 ROC-AUC 값이 0.1이나 증가한다. 표 2에서의 기계 id 별 결과를 비교해보면 자기 회귀 모델이나 분류 모델 중 한쪽의 성능이 좋지 않더라도, 멀티태스크 모델은 기계 이상 진단 성능이 높은 것을 알 수 있다.

III. 결론

본 논문에서는 기계 소리 이상 진단을 위해 활용되고 있는 여러 딥러닝 학습 방법 중 자기 회귀 방법과 분류 방법을 동시에 학습하는 새로운 모델을 제안했다. 이상 탐지 성능 평가 지표인 ROC-AUC를 이용해 모델의 성능을 평가했을 때 하나의 작업을 수행하는 단일 모델에 비해 멀티태스크 모델이 좋

은 성능을 보이는 것을 확인하였다.

기계 종류 (개수)		WaveNet	ResNet	Ours
ToyCar (7000 / 2459)	id_01	0.798	0.656	0.774
	id_02	0.857	0.881	0.857
	id_03	0.704	0.969	0.980
	id_04	0.923	0.999	0.995
ToyConveyor (6000 / 3509)	id_01	0.696	0.607	0.732
	id_02	0.642	0.510	0.674
	id_03	0.673	0.532	0.701
fan (6251 / 1875)	id_00	0.800	0.502	0.788
	id_02	0.924	0.801	0.967
	id_04	0.938	0.566	0.940
	id_06	0.962	0.919	0.998
pump (5766 / 856)	id_00	0.903	0.783	0.935
	id_02	0.881	0.570	0.707
	id_04	0.913	0.996	0.991
	id_06	0.906	0.735	0.923
slider (5174 / 1290)	id_00	0.999	0.945	0.997
	id_02	0.862	0.987	0.991
	id_04	0.982	0.500	0.994
	id_06	0.916	0.500	0.986
valve (5822 / 879)	id_00	0.635	0.950	1.00
	id_02	0.878	0.529	0.898
	id_04	0.885	1.00	0.983
	id_06	0.713	1.00	1.00

표 2. 기계별 데이터 개수(train / test)와 기계 id 별 ROC-AUC 결과

ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. NRF-2020M2C9A1062710)

참 고 문 헌

- [1] Yumi, O., Yoshifumi, O., Takafumi, K., Soichiro, T., and Osamu, H. "Anomaly detection of motors with feature emphasis using only normal sounds," Proc. IEEE ICASSP pp. 2800 - 2804, 2013.
- [2] Oord, A. V. D., Dieleman, S., Zen H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., and Kavukcuoglu, K. "Wavenet: A generative model for raw audio," 2016, (<https://arxiv.org/abs/1609.03499>).
- [3] Daniluk, P., Gozdziwski, M., Kapka, S., and Kosmider, M. "Ensemble of auto-encoder based and wavenet like systems for unsupervised anomaly detection," Challenge on Detection and Classification of Acoustic Scenes and Events, 2020.
- [4] He, K., Zhang, X., Ren, S., and Sun, J. "Deep residual learning for image recognition." Proc. IEEE conference on computer vision and pattern recognition, 2016.
- [5] Koizumi, Y., Kawaguchi, Y., Imoto, K., Nakamura, T., Nikaido, Y., Tanabe, R., ... and Harada, N. "Description and discussion on DCASE2020 challenge task2: Unsupervised anomalous sound detection for machine condition monitoring," 2020, (<https://arxiv.org/pdf/2006.05822.pdf>)
- [6] Giri, R., Tanneti, S. V., Cheng, F., Helwani, K., Isik, U., Krishnaswamy, A. "Unsupervised Anomalous Sound Detection Using Self-Supervised Classification and Group Masked Autoencoder for Density Estimation" Challenge on Detection and Classification of Acoustic Scenes and Events, 2020.