Ensemble Of Complementary Anomaly Detectors Under Domain Shifted Conditions

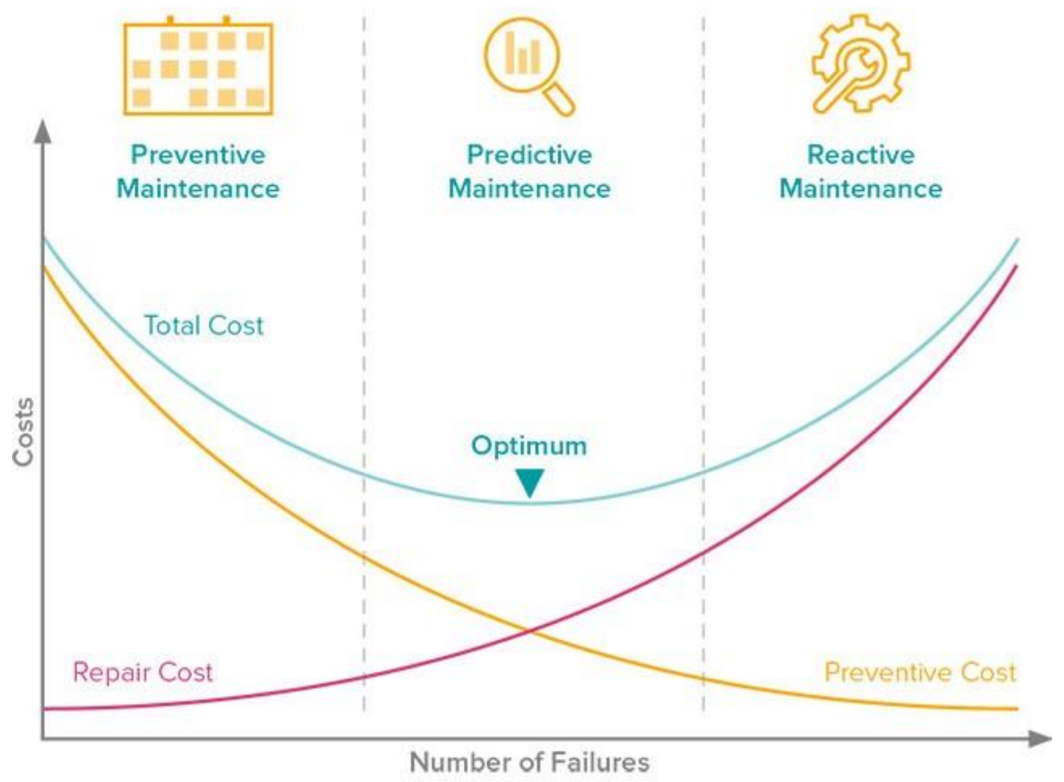
Jose A. Lopez, Georg Stemmer, Paulo Lopez-Meyer, Pradyumna S. Singh, Juan A. del Hoyo Ontiveros, Hector A. Cordourier
Intel Corporation

Abstract

- We participated in the annual DCASE machine learning competition for acoustic anomaly detection.
  - This year, the DCASE Challenge featured domain shifts by changing the acoustic conditions between the source and target domains.
- We found that using existing domain adaptation approaches were insufficient to handle domain shifts in general.
- Our ensemble of three complementary models obtained first place in this year's contest.

Motivation

- The benefits of predictive maintenance in factories are well understood and include reduced downtime, longer machine use, reduced product waste, etc.
- Existing approaches are expensive and generally require physical contact to the machine.
  - Audio anomaly detection can address these two pain points.
- The audio machine learning community is aware of these benefits as evidenced by the inclusion of audio anomaly tasks in the DCASE challenges since 2020.

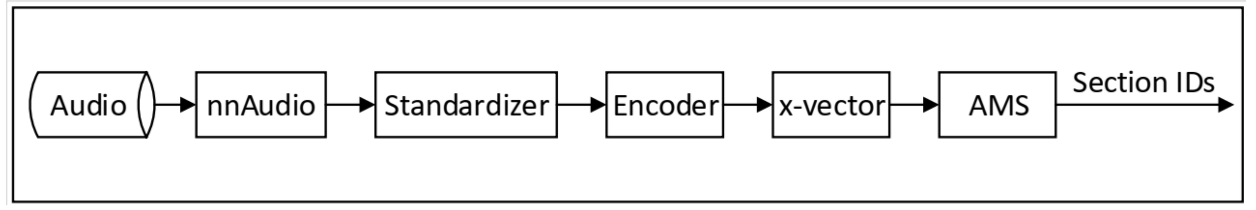


source: https://www.wipotec-ocs.com/us/dynamic-weighing-systems-cep/predictive-maintenance

Costs vs. Number of Failures

Architectures

- Our submission included three models and an ensemble of the three.
  - The first two models are variants of the model we submitted in last year's contest, which was composed of a 2D CNN encoder and an x-vector classifier.
  - The third model is probabilistic and is itself an ensemble of a collection of normalizing flow (NF) based conditional density estimators. We call this model NF-CDEE.
  - Due to the different input modalities and learning approaches, we call these models complementary. Additionally, the performance on the various machine data categories was complementary.



XVector1D High-level Architecture

XVector1D

- Converts audio to spectrograms using the nnAudio library for PyTorch
- Standardizes the spectrogram with either a batch norm or AutoDIAL layer
- The encoder uses 1D convolutions.
- An x-vector type classifier processes the embeddings and uses an added margin softmax (AMS) top layer.

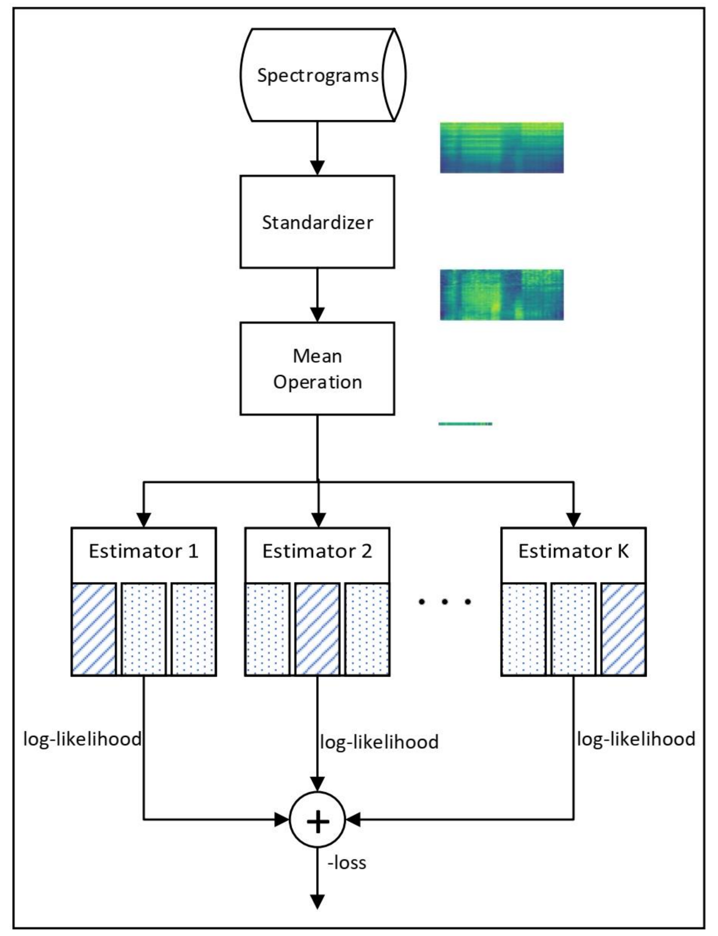
WaveNet-Xvector

- Replaces the encoder with a WaveNet block and operates on the audio signal directly, without computing spectrograms.

NF-CDEE

- NF-CDEE began as an attempt to fit the density of Mel spectrogram bins, but due to numerical difficulties evolved into density estimation of a subset of Mel bins conditioned on the remaining bins.
- For simplicity, the subsets of Mel bins are taken to be adjacent 32-bin segments without overlap, but this is not required.
- The sum of the negative log-likelihood outputs of the estimators are summed to produce the training loss and anomaly scores.
- In order to increase the stability of training the ensemble of density estimators, the normalized spectrogram samples are averaged along the time axis.
- This model was implemented with the Pyro probabilistic programming library.

Architectures



NF-CDEE Inference

Ensemble

- To obtain near optimal weights to use for combining the scores of the three models, we searched exhaustively over a grid of convex combinations.

Results

The results included here are for the ensemble system. Please see our workshop paper for individual model performance.

Development Data

		ToyCar	ToyTrain	fan	gearbox	pump	slider	valve
ensemble	h-mean AUC	0.8745	0.7756	0.8122	0.8613	0.7958	0.8287	0.9032
	h-mean pAUC	0.7837	0.7048	0.8025	0.7635	0.6790	0.6925	0.7724
	AUC rank	1	4	1	2	2	2	2
	pAUC rank	1	1	1	2	1	3	2

		ToyCar	ToyTrain	fan	gearbox	pump	slider	valve
baseline	h-mean AUC	0.6249	0.6171	0.6324	0.6597	0.6192	0.6674	0.5341
	h-mean pAUC	0.5236	0.5381	0.5338	0.5276	0.5441	0.5594	0.5054
	AUC rank	19	21	15	21	21	20	23
	pAUC rank	24	22	22	23	23	24	22

Evaluation Data

		ToyCar	ToyTrain	fan	gearbox	pump	slider	valve
ensemble	h-mean AUC	0.7527	0.6915	0.6101	0.6307	0.8676	0.8318	0.6536
	h-mean pAUC	0.5971	0.5991	0.6079	0.6156	0.8155	0.7412	0.6015
	AUC rank	1	3	18	12	1	4	2
	pAUC rank	6	3	13	2	1	7	1

		ToyCar	ToyTrain	fan	gearbox	pump	slider	valve
baseline	h-mean AUC	0.6593	0.6851	0.6068	0.6549	0.5830	0.5722	0.5187
	h-mean pAUC	0.5232	0.5756	0.5050	0.5686	0.5098	0.5141	0.5007
	AUC rank	8	4	19	6	19	24	20
	pAUC rank	23	4	26	5	25	26	24

Conclusions & Future Work

- It seems that existing domain adaptation approaches, often developed with other sensor modalities in mind, are not as successful for audio.
- Ensemble methods remain a good option for improving performance, even under domain shifted conditions.
- Due to the unsupervised nature of NF-CDEE, and its ensembled architecture, we plan to explore and develop this model further.
  - e.g., one question worth exploring is the role of the mean operation used to stabilize training.

References

Please see our workshop paper for a complete list of references.

Y. Kawaguchi, K. Imoto, Y. Koizumi, N. Harada, D. Niizumi, K. Dohi, R. Tanabe, H. Purohit, and T. Endo, "Description and discussion on dcase 2021 challenge task 2: Unsupervised anomalous sound detection for machine condition monitoring under domain shifted conditions," arXiv:2106.04492, 2021.

G. Wilson and D. J. Cook, "A survey of unsupervised deep domain adaptation," 2020.

F. M. Carlucci, L. Porzi, B. Caputo, E. Ricci, and S. R. Bulo, "Autodial: Automatic domain alignment layers," 2017.

D. Snyder, D. Garcia-Romero, G. Sell, D. Povey, and S. Khudanpur, "X-vectors: Robust dnn embeddings for speaker recognition," in 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018, pp. 5329–5333.

F. Wang, J. Cheng, W. Liu, and H. Liu, "Additive margin softmax for face verification," IEEE Signal Processing Letters, vol. 25, no. 7, p. 926–930, Jul 2018.

A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, "Wavenet: A generative model for raw audio," 2016.

K. W. Cheuk, H. Anderson, K. Agres, and D. Herremans, "nnaudio: An on-the-fly gpu audio to spectrogram conversion toolbox using 1d convolution neural networks," 2020.

E. Bingham, J. P. Chen, M. Jankowiak, F. Obermeyer, N. Pradhan, T. Karaletsos, R. Singh, P. A. Szerlip, P. Horsfall, and N. D. Goodman, "Pyro: Deep universal probabilistic programming," J. Mach. Learn. Res., vol. 20, pp. 28:1–28:6, 2019.

Yuma Koizumi, Yohei Kawaguchi, Keisuke Imoto, Toshiki Nakamura, Yuki Nikaido, Ryo Tanabe, Harsh Purohit, Kaori Suefusa, Takashi Endo, Masahiro Yasuda, and Noboru Harada, "Description and discussion on DCASE2020 challenge task2: Unsupervised anomalous sound detection for machine condition monitoring," arXiv: 2006.05822, June 2020, pp. 1–4.

J. A. Lopez, H. Lu, P. Lopez-Meyer, L. Nachman, G. Stemmer, and J. Huang, "A speaker recognition approach to anomaly detection," in Proceedings of the Detection and Classification of Acoustic Scenes and Events 2020 Workshop (DCASE2020), Tokyo, Japan, November 2020, pp. 96–99.