Combining Multiple Distributions based on Sub-Cluster AdaCos for Anomalous Sound Detection under Domain Shifted Conditions

Kevin Wilkinghoff

Fraunhofer FKIE

DCASE Workshop 2021, Barcelona

Introduction

- DCASE challenge task 2[1]:
  - detect anomalous data using normal training samples only
  - source domain: 1000 training samples per machine type/section
  - target domain: only 3 training samples per machine type/section
- two main approaches for ASD:

1. autoencoder:
   - uses reconstruction error as anomaly score
   - assumption: normal data can be reconstructed well, anomalous data cannot
2. discriminative embeddings:
   - estimate distribution of normal data
   - assumption: information to discriminate classes is sufficient to detect anomalous data

Sub-Cluster AdaCos Loss[2]

- AdaCos[3]: angular margin loss with adaptive scale parameter
  - no hyperparameters need to be tuned
- idea:
  - multiple mean values learned per class
  - use GMM to estimate distribution fo embeddings for each class
  - negative log-likelihood can be used as anomaly score
- yields state-of-the-art performance[2] on DCASE 2020 ASD dataset[4]

$$\hat{P}_{i,j} := \sum_{l\in\mathcal{M}(j)} \frac{\exp(\hat{s}\cdot\cos\theta_{i,l})}{\sum_{k=1}^{CS}\exp(\hat{s}\cdot\cos\theta_{i,k})}$$

Extracting Embeddings

- compute log-Mel spectrograms with 128 bins and standardize them
- train neural network with modified ResNet architecture to extract embeddings using two losses with equal weights:
  - classify among sections and machine types
  - classify among different attribute information
- only mixup is used for augmenting data

Modified ResNet Architecture

layer name	structure	output size
input	-	313×128
2D convolution	7×7, stride$=2$	$157\times 64\times 16$
residual block	$\begin{pmatrix}3\times 3\\3\times 3\end{pmatrix}\times 2$, stride$=1$	$78\times 31\times 16$
residual block	$\begin{pmatrix}3\times 3\\3\times 3\end{pmatrix}\times 2$, stride$=1$	$39\times 16\times 32$
residual block	$\begin{pmatrix}3\times 3\\3\times 3\end{pmatrix}\times 2$, stride$=1$	$20\times 8\times 64$
residual block	$\begin{pmatrix}3\times 3\\3\times 3\end{pmatrix}\times 2$, stride$=1$	$10\times 4\times 128$
max pooling	10×1, stride$=1$	4×128
flatten	-	512
dense (representation)	linear	128
sub-cluster AdaCos	-	42
sub-cluster AdaCos	-	199

Calculating Anomaly Scores

- source domain:
  - one GMM for each section
  - another GMM for each different attribute information

$$Z_{\text{source}}(x) := -\max_{k} \log P(x|s(x),k) - \max_{k} \max_{a\in a(s(x))} \log P(x|a,k)$$

  - for machine type ‚valve' also add GMM trained on temporal maxima of log-Mel spectrograms

$$\widetilde{Z}_{\text{source}}(x) := Z_{\text{source}}(x) - \max_{a\in a(s(x))} \log P_{t_{\max}}(t_{\max}(x)|a)$$

- target domain:
  - one GMM for each section (source domain)
  - another GMM with 3 components for target samples

$$Z_{\text{target}}(x) := -\max_{k} \log P(x|s(x),k) - \max_{k=1,2,3} \log P(x|X_{\text{target}}(s(x)),k)$$

Comparison of Different Distributions on Development Set

dataset split		distribution $P(.\|s,k)$		$P(.\|a,k)$		$P(.\|X_{\text{target}}(s),k)$		proposed ensemble	
machine type	domain	AUC	pAUC	AUC	pAUC	AUC	pAUC	AUC	pAUC
ToyCar	source	82.49%	65.82%	86.22%	69.39%	-	-	85.04%	69.25%
ToyCar	target	65.02%	58.45%	-	-	76.02%	64.13%	73.21%	64.55%
ToyTrain	source	90.54%	65.60%	91.55%	66.11%	-	-	91.30%	66.01%
ToyTrain	target	54.03%	53.21%	-	-	68.13%	57.29%	66.81%	56.14%
fan	source	81.05%	73.34%	80.55%	72.12%	-	-	80.98%	73.20%
fan	target	67.52%	65.71%	-	-	69.17%	60.20%	69.30%	62.20%
gearbox	source	75.20%	58.63%	75.66%	55.65%	-	-	75.88%	56.46%
gearbox	target	81.02%	57.93%	-	-	75.68%	55.34%	76.64%	55.59%
pump	source	83.63%	70.60%	83.83%	70.83%	-	-	83.70%	70.66%
pump	target	69.90%	60.11%	-	-	70.16%	55.51%	70.52%	56.81%
slide rail	source	91.10%	76.23%	90.97%	75.59%	-	-	91.11%	76.20%
slide rail	target	65.90%	56.17%	-	-	71.63%	54.99%	71.56%	54.94%
valve	source	83.85%	73.84%	89.34%	68.89%	-	-	89.33%	70.77%
valve	target	71.66%	61.64%	-	-	73.31%	59.16%	74.26%	60.00%
all	source	83.67%	68.66%	85.09%	67.81%	-	-	85.00%	68.38%
all	target	67.00%	58.80%	-	-	71.90%	57.93%	71.63%	58.41%

Comparison of Different Losses on Development Set

dataset split		sub-cluster AdaCos losses for sections		sections and file endings		proposed ensemble	
machine type	domain	AUC	pAUC	AUC	pAUC	AUC	pAUC
ToyCar	source	74.49%	60.67%	88.58%	71.58%	85.04%	69.25%
ToyCar	target	67.44%	61.55%	75.53%	65.84%	73.21%	64.55%
ToyTrain	source	86.88%	62.00%	91.15%	67.03%	91.30%	66.01%
ToyTrain	target	63.73%	54.18%	68.10%	56.43%	66.81%	56.14%
fan	source	81.66%	73.16%	80.22%	72.60%	80.98%	73.20%
fan	target	69.11%	62.55%	69.00%	61.16%	69.30%	62.20%
gearbox	source	74.97%	57.04%	75.34%	56.16%	75.88%	56.46%
gearbox	target	78.53%	59.53%	72.02%	51.90%	76.64%	55.59%
pump	source	83.44%	69.75%	83.75%	71.35%	83.70%	70.66%
pump	target	68.77%	56.31%	71.30%	56.34%	70.52%	56.81%
slide rail	source	90.44%	74.91%	90.99%	76.43%	91.11%	76.20%
slide rail	target	68.84%	54.44%	73.13%	55.41%	71.56%	54.94%
valve	source	88.90%	72.11%	89.51%	68.95%	89.33%	70.77%
valve	target	75.41%	61.12%	72.36%	56.55%	74.26%	60.00%
all	source	82.54%	66.44%	85.26%	68.58%	85.00%	68.38%
all	target	69.96%	58.34%	71.56%	57.37%	71.63%	58.41%

Comparison to Baseline Systems on Evaluation Set

dataset split		baseline autoencoder		MobileNetV2		proposed system	
machine type	domain	AUC	pAUC	AUC	pAUC	AUC	pAUC
ToyCar	source	76.33%	51.26%	34.32%	53.49%	67.07%	63.05%
ToyCar	target	58.02%	53.42%	56.62%	58.89%	72.83%	63.77%
ToyTrain	source	69.89%	55.49%	47.30%	52.49%	70.87%	56.19%
ToyTrain	target	67.18%	59.78%	39.27%	48.75%	48.38%	52.39%
fan	source	66.58%	51.36%	70.88%	57.76%	89.07%	69.85%
fan	target	55.74%	49.68%	59.96%	58.53%	88.89%	70.55%
gearbox	source	67.81%	55.71%	53.16%	53.47%	61.19%	50.97%
gearbox	target	63.32%	58.06%	49.27%	49.83%	54.68%	49.40%
pump	source	62.75%	51.18%	67.12%	60.77%	70.89%	65.52%
pump	target	54.43%	50.79%	68.85%	59.79%	79.20%	67.81%
slide rail	source	64.13%	50.91%	73.06%	60.47%	88.06%	64.38%
slide rail	target	51.65%	51.92%	72.78%	60.94%	85.66%	69.69%
valve	source	51.56%	50.89%	54.71%	53.03%	73.19%	55.97%
valve	target	52.19%	49.27%	51.64%	50.10%	54.90%	51.47%
all	source	64.76%	52.32%	53.82%	55.73%	73.13%	60.21%
all	target	57.03%	53.01%	54.80%	54.80%	65.76%	59.47%
all	both	56.38%		54.77%		64.20%	

Ensembling

- total of 5×4=20 subsystems
  - network for extracting embeddings is trained with 2^0 to 2^4 sub-clusters
  - networks are trained for 400 epochs, training is stopped after every 100 epochs
- Ensemble trained a second time, only using one sub-cluster AdaCos loss for sections + machine types
  - take mean of both ensembles or best performing ensemble per machine type

Conclusions

- system significantly outperforms baseline systems
- system ranked 3rd among all teams' submissions
- future work:
  - reduce size of ensemble
  - also utilize autoencoder structure by using additional reconstruction loss[6]

1 Kawaguchi et al, **Description and discussion on DCASE 2021 challenge task 2: Unsupervised anomalous sound detection for machine condition monitoring under domain shifted conditions**, DCASE 2021

2 Wilkinghoff, **Sub-Cluster AdaCos: Learning Representations for Anomalous Sound Detection**, IJCNN 2021

3 Zhang et al, **AdaCos: Adaptively scaling cosine logits for effectively learning deep face representations**, CVPR 2019

4 Koizumi et al, **Description and discussion on DCASE2020 challenge task 2: Unsupervised anomalous sound detection for machine condition monitoring**, DCASE 2020

5 Zhang et al, **Mixup: Beyond empirical risk minimization**, ICLR

6 Narita et al, **Unsupervised anomalous sound detection using intermediate representation of trained models and metric learning based variational autoencoder**, DCASE 2021 Challenge Report