Title	音色関連特徴量に着目した教師なし機械学習による異常音検知
Author(s)	小倉, 稜也
Citation	
Issue Date	2025-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/19788
Rights	
Description	Supervisor: 鵜木 祐史, 先端科学技術研究科, 修士 (情報科学)

Unsupervised learning for anomalous sound detection focused on acoustical features related to timbral attributes

2310029 Ryoya Ogura

Experienced factory inspectors can identify abnormalities in equipment by listening to the operating sounds of the machines. However, due to the aging workforce, there is a lack of successors. Therefore, investigating computer-based anomalous sound detection (ASD) is necessary. ASD system consists of a feature extractor and a discriminator. The feature is mainly log Mel-spectrogram. The discriminators primarily use deep learning models such as autoencoder (AE) and convolutional neural network (CNN). Recent research has proposed methods using timbral features for ASD. Ota & Unoki reported high performance using acoustical features related to timbral attributes. However, this method employs supervised learning, where anomalous sounds are used together with normal sounds. Covering all anomalous sounds for learning is challenging, and collecting sufficient anomaly samples proves difficult. Therefore, unsupervised learning that learns only normal sounds is desirable. This paper proposes an ASD system based on unsupervised learning using acoustical features related to timbral attributes, addressing three limitations in Ota & Unoki's previous method.

First, discriminators other than support vector machines (SVM) are explored. Better models than SVM for acoustical features related to timbral attributes could improve performance. However, these features are 6-dimensional or 7-dimensional data. Deep learning models such as CNNs typically handle high-dimensional data. These models are difficult to apply to low-dimensional features, such as acoustical features related to timbral attributes. Therefore, the investigation focused on supervised and unsupervised learning models that can handle relatively low-dimensional data. For supervised learning, models such as LightGBM and TabPFN were investigated. These models handle tabular data. For unsupervised learning, outlier detection methods such as the gaussian mixture model (GMM) and k-Nearest Neighbors (k-NN) were investigated. The models were evaluated using MIMII dataset employed in previous method. MIMII dataset contains four Machine Types: Fan, Pump, Slider, and Valve. Each Machine Type contains four Machine ID sounds. The evaluation results showed that TabPFN achieved the best performance, surpassing the previous method. GMM demonstrated the best performance in unsupervised learning.

The second limitation addresses performance degradation under low signal-to-noise ratio (SNR) conditions. In industrial equipment ASD, sounds from surrounding machines become environmental noise. Therefore, robust system performance is required even under low SNR conditions. The performance of the previous method significantly deteriorates as machine noise SNR decreases. The previous method employ the average of acoustical features related to timbral attributes calculated from multiple frames of machine sound. However, the frames are averaged in non-stationary machine sounds, including those containing only environmental noise without machine sounds. This averaging significantly impacts performance for non-stationary sounds. Therefore, a feature extraction method called Timbral frames selection (TFS) was developed as a robust feature extraction for low SNR conditions. This method selects frames only containing non-stationary machine sound events. Evaluation of the MIMII dataset showed that TFS improved the performance of non-stationary sounds such as Slider and Valve. The effect of TFS was particularly significant under low SNR conditions.

The third limitation is implementing unsupervised learning. The previous method used supervised learning, requiring normal and anomalous sounds. However, collecting anomalous sounds for learning is challenging due to their rare occurrence in machines. This necessitates an unsupervised machine-learning system capable of learning from normal sounds alone. The solution estimates the distribution of anomalous sounds using various machine sounds prepared in advance. The anomalous sound distribution is defined as the distribution obtained by subtracting the normal sound distribution of the target machine from the distribution of various machine sounds. The estimated anomalous sound distributions were then sampled to generate pseudo-anomalous sound data. This approach enables model training with authentic normal sounds and pseudo-anomalous sounds. It satisfies unsupervised learning conditions while enabling supervised learning.

The proposed system incorporates all three solutions described above. The evaluation used the DCASE 2020 Task 2 development dataset, which includes six Machine Types: Fan, Pump, Slider, Valve, ToyCar, and ToyConveyor. The area under the curve (AUC) was used as the evaluation metric. These results showed that the proposed system using TabPFN achieved higher overall AUC than the AE baseline. The AUC scores for Fan and Valve improved by over 20%. However, the ToyConveyor system exhibited decreased AUC. Therefore, further investigation of timbral features is necessary.

In conclusion, the three solutions presented effectively address limitations in ASD systems based on unsupervised learning using acoustical features related to timbral attributes. The proposed system demonstrates superior discriminative performance compared to baseline methods. Future work should address domain generalization and first-shot anomalous sound detection.