

TRAINING-FREE DATA-CENTRIC PREPROCESSING FOR ROBUST OPEN SET FAULT DIAGNOSIS

Youngjae Jeon⁽¹⁾, Dongjin Lee⁽²⁾

⁽¹⁾Department of Automotive Engineering (Automotive Computer Convergence) Hanyang University

⁽²⁾Department of Automotive Engineering Hanyang University

Keywords: Prognostics and health management, Fault diagnosis, Open-set recognition, Hyperparameter search, Short-Time Fourier Transform

Abstract

Reliable fault diagnosis in rotating machinery is critical for maintaining the stability and safety of industrial systems. Deep learning has demonstrated remarkable classification performance when every possible fault type is known in advance (the closed-set assumption), yet two problems limit its practical deployment: (1) misclassification of unseen anomalies as known fault types with high confidence and (2) heuristic selection of a deep learning model's hyperparameters for time-frequency preprocessing techniques.

We propose a comprehensive open-set recognition (OSR) framework that integrates class-specific autoencoders for unknown fault detection with a training-free data-centric preprocessing hyperparameter search strategy. As the autoencoders are trained on known fault types, class-specific boundaries are established in their latent feature space. At inference, a sample whose latent representation falls outside every class boundaries is classified as an unknown fault, providing a distribution-aware rejection mechanism. To overcome the computational bottleneck of preprocessing optimization, we introduce a training-free data-centric hyperparameter search: instead of retraining a model for every STFT configuration, we score each candidate using clustering metrics and show that these scores correlate strongly with downstream OSR classification accuracy. This reduces the search cost by orders of magnitude.

We evaluate the proposed method on the Case Western Reserve University (CWRU) bearing dataset. This example shows that latent feature boundaries provide a robust and intuitive criterion for detecting unknown faults. Furthermore, the strong correlation between clustering metrics and classification performance validates the effectiveness of our training-free preprocessing optimization. By reducing the computational time for hyperparameter search by a factor of 6 without compromising the classification accuracy (0.9658), the proposed approach offers a computationally efficient and reliable solution for fault diagnosis under open-set conditions.