Physics Informed Self Supervised Learning For Fault Diagnostics and Prognostics in the Context of Sparse and Noisy Data

Weikun Deng, Khanh T. P. Nguyen, and Kamal Medjaher

Laboratoire Génie de Production, LGP, Université de Toulouse, INP-ENIT, 47 Av. d'Azereix, 65016, Tarbes, France.
weikun.deng@enit.fr, tnguyen@enit.fr, kamal.medjaher@enit.fr

Abstract

Sparse & noisy monitoring data leads to numerous challenges in prognostic and health management (PHM). Big data volume but poor quality with scarce healthy states information limits the performance of training machine learning (ML) and physics based failure modeling. To address these challenges, this thesis aims to develop a new hybrid PHM framework with the ability to autonomously discover and exploit incomplete implicit physics knowledge in sparse & noisy monitoring data, providing a solution for deep physics knowledge-ML fusion by physics-informed machine learning algorithms. In addition, the developed hybrid framework also apply the self-supervised learning paradigm to significantly improve the learning performance under uncertain, sparse, and noisy data with lower requirements for specialist area knowledge. The performance of the developed algorithms will be investigated on the sparse and noise data generated by simulation data sets, public benchmark data sets, and the PHM platform to demonstrate its applicability.

Keywords—Prognostic and health management; Sparse & noisy data; Hybrid framework; Physics informed machine learning; Self-Supervised Learning.

1. Motivation and research problem statement

Prognostics and health management (PHM) plays a constructive role in ensuring the real-time health assessment of a system under its actual working conditions as well as the prediction of its future state based on up-to-date information (N. Kim, An, & Choi, 2017). Two mainstream methods, which are mainly used are Machine Learning (ML) and Physics-based methods (PBM). ML is proficient at automatically extracting features from data and building relationships between features based health indicators and system states. However, as a data-hungry and black-box method. ML meets dilemmas in processing sparse & noisy data. The pervasive monitoring instrument costs, the high run-to-failure operation costs, and the lack of data label are objective conditions that create sparse/noisy data that is insufficient for ML to learn a meaningful knowledge representation. Besides, PBM represent the degradation mechanisms by observing failure phenomena and then establishing mathematical equations or numerical laws, with the ability to infer hidden states from a limited sample (Chao, Kulkarni, Goebel, & Fink, 2019). However, modern engineering systems have simultaneous nonlinear interactions between their subsystems and their environment. Failure mechanisms and degradation processes are difficult to identify. With incomplete failure cognition, implementing detailed parametric or numerical degradation models for these systems in sparse & noisy data is challenging.

These challenges prompt PHM techniques into a hybrid framework. Hence, this thesis aims to explore the combination of PBMs and ML by physics informed machine learning (PIML). Providing a deep model & data-driven embedding fusion solution to assist trustworthy PHM deployment in " small data, small laws " contexts. The developed framework is hoped to be trained in self supervised learning training (SSL) paradigm to build the ability to autonomously discover and exploit implicitly incomplete physics knowledge in sparsenoise monitoring data.

2. Novelty and significance relative to the state of the art

To the best of our knowledge, the research about SSL-PIML hybrid framework is scarce, most of them are derived from reconstructive recognition of image data in the medical field, and physics-based loss functions are designed to test the effectiveness of the feature extractors in pretext (Yaman et al., 2020; Martín-González et al., 2021). A brief review of advanced research on PIML and SSL in PHM is performed. The bibliometrics results from Citespace analysis for 185 PIML hybrid methods -related and 35 SSL hybrid methods related papers are presented in Fig.1. In PHM field, the rotating machinery, grid, production lines, batteries, and materials are the main application scenarios of SSL and PIML techniques while the anomaly detection, fault diagnosis, and RUL prediction are their core objectives.

On one hand, the PIML methods have risen and are attracting more attention since 2017. According to the way of integrating physics knowledge in the ML pipeline, the PIML methods can be categorized into three groups: 1) Physics informed input space, 2) Physics informed structure, and 3) Physics informed loss function (Karniadakis et al., 2021). According to the Fig.1, the research has shifted from the integration of features and rules (expert systems) to the integration of algorithms structure and parametric models (physics informed neural network (Karniadakis et al., 2021)). It suggests that researchers seek to create an augmented input feature and a physics informed (PI) derivation process that can be interpretable (S. W. Kim, Kim, Lee, & Lee, 2021). In this process, a variety of refined laws and analytic relations such as linear damage accumulation laws, crack extension formula are incorporated by the different methods such as neural networks (Viana & Subramaniyan, 2021), Gaussian processes (Cury, Ribeiro, Ubertini, & Todd, n.d.). Besides, the research related to embedding the partial differential equations representing system behaviors into ML models is gradually becoming a popular method. The essence of PIML is to introduce physics constraints to ML data processing process. Its drawback is the high requirement of physics domain knowledge because the incorporation methods still relies heavily on manual designed explicit knowledge with parsed form.

On the other hand, SSL methods mainly focuses on mining its own supervised information from large-scale unlabelled monitoring data using an auxiliary task (pretext), and training ML with this constructed supervised information to build valuable representations for downstream detection, diagnostic, and prediction tasks. It is clear in Fig.1 that SSL methods in PHM are in the stage of self-supervised feature engineering. They focus on signal reconstruction and feature extraction through principal component analysis (PCA) (Wang, Qiao, Zhang, Yang, & Snoussi, 2020), Deep Clustering and Auto-encoder (Zhang, Chen, He, & Zhou, 2022), Generative Adversarial Network (Ding, Zhuang, Ding, & Jia, 2022). In these studies, self-supervised (SS) features construct bounds for different health states by fine-tuning valuable representations for downstream tasks, such as bounds for reconstruction error as a normal-abnormal watershed and bounds for similarity as a distinguishing representation for different fault states. Particularly, SSL based RUL predictions are rarely studied. Moreover, only generative schemes are widely used compared to the other SSL architectures, e.g., contrastive or generative-contrastive strategies.

In summary, the focus of the hybrid framework proposed in this study is autonomously incorporating the implicit incomplete physical knowledge into ML, under sparse/noisy monitoring data. It is an issue that is hardly mentioned in existing studies but indeed needs to be addressed by original and innovative research in the development of PHM without delay.

3. Work progress and future direction

Motivated by the philosophical concept of " constructivism learning", it is hoped to build PIML-SSL hybrid framework based on conformity and assimilation. In conformity, ML transforms the original data-driven reasoning process by incorporating physics constraints. In assimilation, ML trains feature extractors in self-supervised way for downstream PHM tasks without changing the PIML framework. Currently, the literature review has completed and based on it, this thesis is at the beginning of the methodological development. Particularly, the developed hybrid framework using PI-SSL paradigm consists of the following techniques in Fig.2:

a) Knowledge - ML module inter-conversion mechanisms

The inter-conversion mechanisms are dedicated to embedding the mathematical relations, i.e., Input-output (IO) model (analytical function) or physics operator (differential relationships) of the failure to a part of the ML calculation diagram in layer functions, regression formulas, coefficient distribution, etc. Based on generic mathematical relations, ML will infer uncertain parameters and automatic search for hidden representations of the possible formation of degradation relations for these units of embedded physical knowledge.

b) Physics informed metric learning

It aims to establish boundaries metric distances for failure states and the corresponding HI. This enables the ML's results to respect the basic physics consistency such as physics-informed similarity, principle of cumulative energy dissipation for wear behavior, etc. In particular, distance measures between different health states based on comparative learning will be investigated in depth.

c) Boundary condition exploration pretext task design

In SSL, the inter-conversion mechanisms helps to establish a downstream data-driven health indicator (HI) according to PHM tasks. In detail, an appropriate PI computational structures will be constructed to complete the assimilation process, e.g., Siamese, Codec, Graph, etc. These structures seek to maximize the difference between the boundaries of different health states while satisfying physics consistency. Through training on a "pretext task", these structures automatically generates pseudo labels. Their parameters are frozen as a supervised feature extractor, connecting with the different functional ML module in the fine tuning process when it is transferred into the downstream PHM tasks.

d) Hybrid framework design and validation

Based on the previous research, we will construct the PI-SSL hybrid framework. The quantitative and sensitivity analyses of data quality on the its performance will be performed to probe the lower limit of tolerance and upper limit of its applicability in sparse & noise and incomplete physics knowledge. Relevant metrics on sparsity, noise, and knowledge completeness will be defined and quantified through masking and selective cropping of public data sets, mechanistic models, and experimental data. The influence of the above indicators to

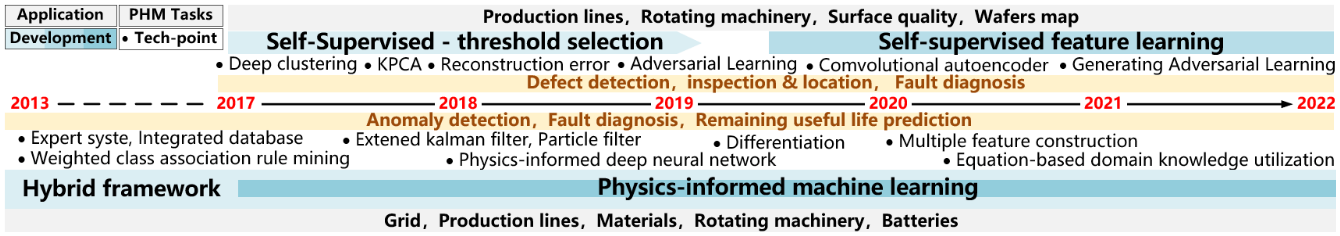

Figure 1. SSl and PIML tech-development analysis

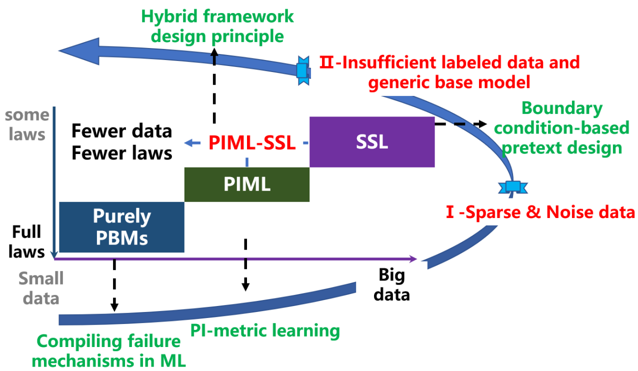

Figure 2. Technology road-map for the thesis.

varying degrees under different working conditions will be studied deeply in the rotating machine PHM test platform. The results from different data sources will be compared to find the difference in their feature representation.

4. Discussion on the applications

This thesis develop a PIML hybrid framework equipped with SSL training paradigm for fault diagnostics and prognostics purposes in the context of sparse & noisy data with incomplete and implicit failure knowledge. It design the physics informed operator or ML module to completes the seamless methods integration, establishing fault boundary metric distance in the objective function to improve the physics consistency as well as to reduce the data dependency of ML. In fact, the need of large labeled and high-quality data is too difficult or costly to satisfy. Meanwhile, the ability to correctly interpret the output of a PHM model is essential in high-value devices. The developed hybrid PHM framework with physics consistency and excellent exploitation of sparse & noise data allows better understanding of the system state and maintenance supports. Its algorithms potentially extended to the large-scale and low-cost deployment.

References

Chao, M. A., Kulkarni, C., Goebel, K., & Fink, O. (2019). Hybrid deep fault detection and isolation: Combining deep neural networks and system performance models. *arXiv preprint arXiv:1908.01529*.

Cury, A., Ribeiro, D., Ubertini, F., & Todd, M. D. (n.d.). *Structural health monitoring based on data science techniques.* Springer.

Ding, Y., Zhuang, J., Ding, P., & Jia, M. (2022). Self-supervised pretraining via contrast learning for intelligent incipient fault detection of bearings. *Reliability Engineering & System Safety*, *218*, 108126.

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics*, *3*(6), 422–440.

Kim, N., An, D., & Choi, J.-H. (2017, 01). Prognostics and health management of engineering systems..

Kim, S. W., Kim, I., Lee, J., & Lee, S. (2021). Knowledge integration into deep learning in dynamical systems: an overview and taxonomy. *Journal of Mechanical Science and Technology*, 1–12.

Martín-González, E., Alskaf, E., Chiribiri, A., Casaseca-de-la Higuera, P., Alberola-López, C., Nunes, R. G., & Correia, T. (2021). Physics-informed self-supervised deep learning reconstruction for accelerated first-pass perfusion cardiac mri. In *International workshop on machine learning for medical image reconstruction* (pp. 86–95).

Viana, F. A., & Subramaniyan, A. K. (2021). A survey of bayesian calibration and physics-informed neural networks in scientific modeling. *Archives of Computational Methods in Engineering*, *28*(5), 3801–3830.

Wang, T., Qiao, M., Zhang, M., Yang, Y., & Snoussi, H. (2020). Data-driven prognostic method based on self-supervised learning approaches for fault detection. *Journal of Intelligent Manufacturing*, *31*(7), 1611–1619.

Yaman, B., Hosseini, S. A. H., Moeller, S., Ellermann, J., Uğurbil, K., & Akçakaya, M. (2020). Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data. *Magnetic resonance in medicine*, *84*(6), 3172–3191.

Zhang, T., Chen, J., He, S., & Zhou, Z. (2022). Prior knowledge-augmented self-supervised feature learning for few-shot intelligent fault diagnosis of machines. *IEEE Transactions on Industrial Electronics*.