Methodology of Adaptive Prognostics and Health Management Using Streaming Data in Big Data Environment

Jianshe Feng[1]

[1] *NSF I/UCR Center for Intelligent Maintenance Systems, Department of Mechanical Engineering, University of Cincinnati, PO Box 210072, Cincinnati, Ohio 45221-0072, USA*

fengje@mail.uc.edu

ABSTRACT

Prognostics and health management (PHM) has gradually become an essential technique to improve the availability and efficiency of industrial systems. With the rapid advancement of sensor technology and communication technology, a huge amount of real-time data is generated from various applications industry, which brings new challenges to PHM in the context of big data streams. On one hand, high-volume stream data places a heavy demand on data storage, communication, and PHM modeling. On the other hand, continuous fluctuation and drift are essential properties of stream data in an online environment, which requires the PHM model to be capable to capture the new formation in stream data adaptively and continuously. This research proposes a systematic methodology to develop an effective online evolving PHM method with adaptive sampling mechanism against continuous stream data. An adaptive sample selection strategy is developed to effectively select the representative samples in both off-line and online environment. Meanwhile, a probabilistic theory-based modeling approach is developed to update the model with newly selected samples. Finally, the whole methodology is validated with real-world industrial cases. The result comparison between the proposed methodology and state-of-art methods verifies the superiority of the proposed method.

1. PROBLEM STATEMENT

With the development of sensing technology and networked monitoring system, condition-based maintenance (CBM) has been implemented as a maintenance strategy to significantly reduce the high maintenance expense (Ã, Lin, & Banjevic, 2006). CBM schedules maintenance based on the equipment's health status such that the unnecessary maintenance tasks and cost can be avoided(Lee, Yang, Lapira, Kao, & Yen, 2012). Therefore, the evaluation of devices' health status and prediction of future health status

are two key aspects in the development of effective operation and maintenance (O&M) activities. However, to date, most industrial applications of failure diagnosis and prognostics are problem-focused scenarios of a single unit, equipment, or manufacturing process in certain specific applications such as automotive, manufacturing, aerospace and so on

With the rapid advancement of sensor technology, a huge amount of real-time data is generated from various applications, which brings new challenges to data management, communication, and PHM modeling. What's more, continuous fluctuation and drift are essential properties of online streaming data, especially when industrial systems become much more complex than ever. This requires diagnosis and prognosis techniques to be capable to capture the new formation in stream data adaptively, continuously, and efficiently.

To address the gap between traditional PHM technologies and growing needs in modern industry, this research proposes a systematic framework of PHM methodology with adaptive learning capacity in big stream data environment. The feasibility of proposed methodology is demonstrated based on case studies from different industrial areas.

2. EXPECTED CONTRIBUTIONS

To deal with the problem stated above, the expected target of this research is to develop a systematic framework of adaptive PHM, which is capable to evolve with new information brought with streaming data during online process. Unlike the conventional static PHM models, which are usually trained based on historical data or design of experiments (DOE) data, the proposed methodology evolutionarily transfers the static PHM model to an adaptive model to accommodate the dynamics in manufacturing processes. The expected contributions of this research can be summarized as follows:

Contribution 1: Effective PHM modeling from abundant historical data. Traditionally, all available historical data are equally employed for PHM modeling. That turns to be

ineffective in modern manufacturing processes with data generated rapidly. Firstly, large volume of in-process data brings heavy computational burdens; secondly, information redundancy is an inevitable issue in big data environment. It is necessary to develop an efficient PHM modeling method to process a high volume of data with limited time and resources.

Contribution 2: Modeling of new information discovery in manufacturing processes. Manufacturing processes are subjected to disturbances and drifts caused by various issues such as the changes of ambient environment, device wear, adjustment of usage behaviors, etc. All these volatilities could undermine a static PHM model's performance. However, it is practically impossible to update PHM model every single time when new sample arrives. Therefore, to adaptively update the model efficiently, it is crucial to capture the volatility during the process monitoring.

Contribution 3: Adaptive PHM modeling in on-line fashion. Thanks to the informative detection model mentioned by Contribution 2, the self-leaning PHM model does not evolve every time but updates only when a streaming-in sample is detected informative. We proposed an online self-learning PHM modeling method to enable the current model to learn from selected samples sequentially without model retraining.

Contribution 4: Development of a unified framework to combine the off-line PHM model and online PHM model, which could enhance the model's performance most. On one hand, an online model with self-learning ability processes samples and updates model incrementally, which itself has the efficiency property. On the other hand, the off-line model maintains the knowledge learned from historical data or DOE.

3. RESEARCH PLAN

As shown in Figure 1, the proposed methodology has 3 major modules. (1) Off-line/initial sampling and modeling: the goals of this module include: (a) selects important samples from historical database (DB) to formulate a representative subset called data library (DL); (b) trains an off-line PHM model based on the selected DL.

(2) On-line sample selection: this module determines the importance of each streaming sample and decides whether to import this new sample to DL and update PHM model or not.

(3) On-line model updating: this module enables the current model to learn from selected samples sequentially to adapt to dynamic changes. Considering new samples arrive in sequence, some sequence learning algorithms such as Bayesian-based methods, neural networks could be the potential candidates.

Besides the methodology development, some industrials case studies are also necessary to validate the effectiveness of the proposed method.

3.1. Work Performed

An interactive off-line sample selection and modeling is proposed as shown in Figure 2, the PHM model and DL are initialized with a small number of randomly selected samples from database. Then the algorithm updates DL and PHM model iteratively based on sample importance (SI) test, which is proposed to determine the importance of samples. At each iteration, the selected sample is added to the DL, and the PHM model is updated accordingly.
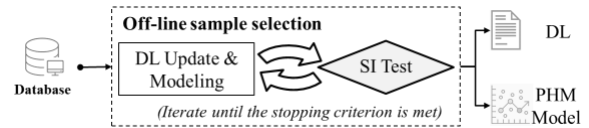

Figure 2. Proposed method for off-line sampling and modeling

SI test determines the sample importance based on the usefulness and the freshness. The usefulness of a sample is determined by a model-based importance test (MBIT), which identifies the sample with large prediction error or with large prediction uncertainty. The freshness of a sample is determined by a data-based importance test (DBIT) which judges the closeness of the sample to the existing samples in DL.
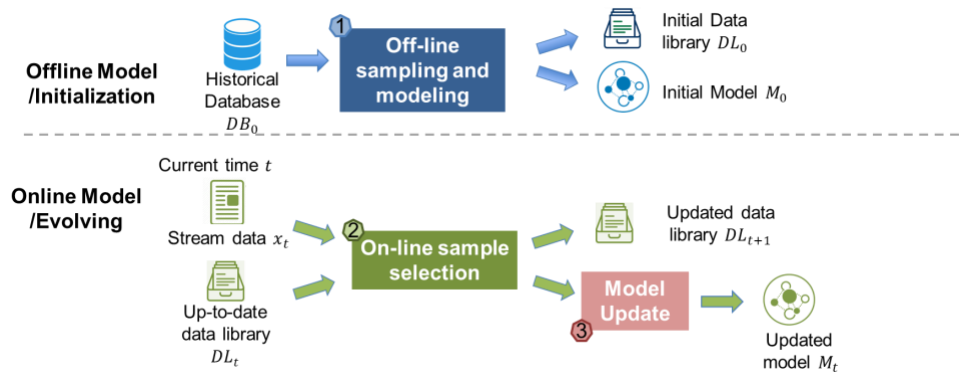

Figure 1. Proposed framework of adaptive PHM

In online stage, the SI test keeps checking the importance of the incoming sample. If it passes the SI test, DL and PHM model is incrementally updated. In this study, Bayesian-related algorithms and online SVM are adopted for modeling as they can easily be updated when new observations are available.
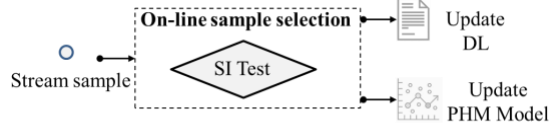

Figure 3. Proposed method for on-line sampling and modeling

To validate the effectiveness of the proposed method, an industrial data set of hard drive disk (HDD) fault detection is employed. One challenging issue in this dataset is the strong uncertainty of the records with a long sampling interval (one sample per hour), which makes the failure related data behaviors cannot be timely and effectively captured. The data partition and modeling process is shown in Figure 4, where off-line data means the available data prior to model training, online stream data means the data arrived during on-line monitoring process, and testing data are used to verify the performance of both off-line model and on-line adaptive model.
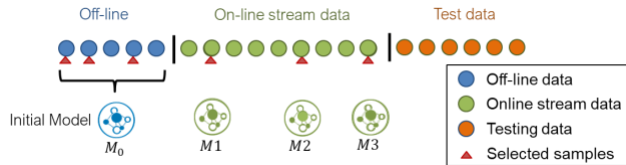

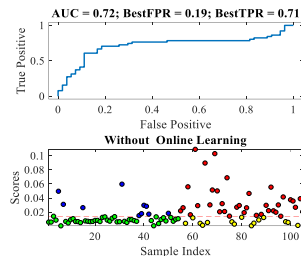Figure 4. Data partition and modeling
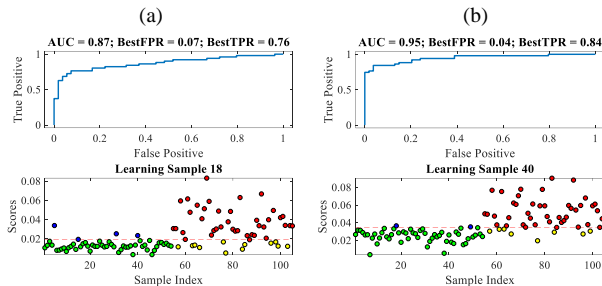

Figure 5. Static model prediction result

Figure 6. Proposed model prediction result. (a) after model learned 18 informative samples; (b) after model learned 40 informative samples.

Figure 5 and Figure 6 present the prediction results based on static model and proposed model. In the static model, only

historical samples are utilized for model training. It is validated that the proposed method improves the fault detection accuracy.

3.2. Remaining Work

Currently, first 3 tasks in research plan are almost done, and a case study about fault detection has been finished. The future work will focus on the development and implementation of degradation assessment and remaining useful life (RUL) prediction under this proposed adaptive PHM framework. One or two real world industrial cases will be also investigated to further enhance the methodology in practice.

4. CONCLUSION

This study proposes an adaptive PHM methodology with offline model initialization, adaptive sample selection, and online model learning mechanism. The off-line part of the proposed method aims to identify the important samples from a large database for modeling. With the selected data, the PHM model is trained with better efficiency yet the model performance is maintained. The importance of new observations collected in monitoring process is dynamically identified through the proposed sample importance test and the model is updated. The effectiveness of the proposed method is verified in industrial case study. Compared with static PHM modeling approach, the proposed method could provide more accurate and adaptive solution.

REFERENCES

Ã, A. K. S. J., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance, *20*, 1483–1510.

Lee, J., Yang, S., Lapira, E., Kao, A., & Yen, N. (2012). Recent advances and trends on cloud-based machinery prognostics and health management. *PECCS 2012 - Proceedings of the 2nd International Conference on Pervasive Embedded Computing and Communication Systems*,

Liu, B., Wu, S., Xie, M., & Kuo, W. (2017). A condition-based maintenance policy for degrading systems with age- and state-dependent operating cost. *European Journal of Operational Research*, *263*(3), 879–887.