Mechanical Systems and Signal Processing 62-63 (2015) 75–90

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

A co-training-based approach for prediction of remaining useful life utilizing both failure and suspension data

Chao Hu [a,1], Byeng D. Youn [b,*], Taejin Kim [b], Pingfeng Wang [c]

[a] Department of Mechanical Engineering, the University of Maryland at College Park, College Park, MD 20742, USA
[b] School of Mechanical and Aerospace Engineering, the Seoul National University, Seoul 151–742, Korea
[c] Department of Industrial and Manufacturing Engineering, the Wichita State University, Wichita, KS 67260, USA

ARTICLE INFO

Article history:
Received 2 September 2014
Received in revised form
29 December 2014
Accepted 5 March 2015
Available online 31 March 2015

Keywords:
Co-training
Semi-supervised learning
Suspension data
Data-driven prognostics
RUL prediction

ABSTRACT

Traditional data-driven prognostics often requires some amount of failure data for the offline training in order to achieve good accuracy for the online prediction. Failure data refer to condition monitoring data collected from the very beginning of an engineered system's lifetime till the occurrence of its failure. However, in many engineered systems, failure data are fairly expensive and time-consuming to obtain while suspension data are readily available. Suspension data refer to condition monitoring data acquired from the very beginning of an engineered system's lifetime till planned inspection or maintenance when the system is taken out of service. In such cases, it becomes essentially critical to utilize suspension data which may carry rich information regarding the degradation trend and help achieve more accurate remaining useful life (RUL) prediction. To this end, this paper proposes a co-training-based data-driven prognostic approach, denoted by COPROG, which uses two data-driven algorithms with each predicting RULs of suspension units for the other. After a suspension unit is chosen and its RUL is predicted by an individual algorithm, it becomes a virtual failure unit that is added to the training data set of the other individual algorithm. Results obtained from two case studies suggest that COPROG gives more accurate RUL prediction, as compared to any individual algorithm with no use of suspension data, and that COPROG can effectively exploit suspension data to improve the prognostic accuracy.

1. Introduction

To support critical decision-making processes such as maintenance replacement and system design, activities of health monitoring and life prediction are of great importance to engineered systems composed of multiple components, complex joints, and various materials, such as aerospace systems, nuclear power plants, chemical plants, advanced military systems and so on. Stressful conditions (e.g., high pressure, high temperature and high irradiation field) imposed on these systems are the direct causes of damage in their integrity and functionality, which necessitates the continuous monitoring of these systems due to the health and safety implications [1–5]. Currently, there are mainly three paradigms for health prognostics, that is, model-based approaches [6–11], data-driven approaches [12–20] and hybrid approaches [21–23]. The application of

* Corresponding author. Tel.: +82 2 880 1919; fax: +82 2 883 1315.
E-mail addresses: huchaostu@gmail.com (C. Hu), bdyoun@snu.ac.kr (B.D. Youn), godori16@snu.ac.kr (T. Kim), pingfeng.wang@wichita.edu (P. Wang).
[1] Current address: Medtronic Energy and Component Center, Medtronic Inc., Minneapolis, MN 55430, USA

http://dx.doi.org/10.1016/j.ymssp.2015.03.004
0888-3270/

general model-based prognostic approaches relies on the understanding of system physics-of-failure and underlying system degradation models. Myötyri et al. [6] proposed the use of a stochastic filtering technique for real-time remaining useful life (RUL) prediction in case of fatigue crack growth while considering the uncertainties in both degradation processes and condition monitoring measures. A similar particle filtering approach was later applied to condition-based component replacement in the context of fatigue crack growth [7]. Luo et al. [8] developed a model-based prognostic technique that relies on an accurate simulation model for system degradation prediction and applied this technique to a vehicle suspension system. Gebraeel et al. presented a degradation modeling framework for RUL predictions of rolling element bearings under time-varying operational conditions [9] or in the absence of prior degradation information [10]. Si et al. presented a drift coefficient model for a nonlinear Wiener degradation process and employed a recursive filter algorithm to derive an approximate RUL distribution [11]. As complex engineered systems generally consist of multiple components with multiple failure modes, understanding all potential physics-of-failures and their interactions for a complex system is almost impossible. With the advance of modern sensor systems as well as data storage and processing technologies, the data-driven approaches for system health prognostics, which are mainly based on the massive sensory data with less requirement of knowing inherent system failure mechanisms, have been widely used and become popular. Two good reviews of data-driven prognostic approaches were given in [12] and [13]. Data-driven prognostic approaches generally require the sensory data fusion and feature extraction, statistical pattern recognition, and, for the life prediction, the interpolation [13–16], extrapolation [17], and machine learning [18–20]. Hybrid approaches leverage the strengths of model-based and data-driven approaches by fusing the information from both approaches. Kozlowski et al. [21] described a data fusion approach where domain knowledge and predictor performance are used to determine weights for different state-of-charge predictors. Goebel et al. [22] employed a Dempster–Shafer regression to fuse a physics-based model and an experience-based model for prognostics. Saha et al. [23] combined an offline relevance vector machine with an online particle filter for battery prognostics. Similar to model-based approaches, the application of hybrid approaches is limited to the cases where sufficient knowledge on system physics-of-failures is available.

In the context of machine learning, traditional data-driven prognostic approaches mentioned in the literature survey above belong to the category of supervised learning which relies on some amount of failure data for the offline training in order to achieve good accuracy in the online prediction. Here, failure data refer to condition monitoring data collected from the very beginning of an engineered system's lifetime till the occurrence of its failure. Unfortunately, in many engineered systems, only very limited failure data are available since running systems to failure can be a fairly expensive and lengthy process. In contrast, we often can easily obtain a large amount of suspension data. Suspension data refer to condition monitoring data acquired from the very beginning of an engineered system's lifetime till planned inspection or maintenance when the system is taken out of service. The lack of failure data and the large amount of suspension data carrying rich information on the degradation trend make it essentially critical to exploit suspension data in order to improve the accuracy in RUL prediction. However, the utilization of both failure and suspension data for data-driven prognostics, which can be treated as semi-supervised learning in the context of machine learning, is still in infancy. The very few relevant works we are aware of are the survival probability-based approaches [24–26] and life-percentage-based approach [7]. The former approaches use condition monitoring data as inputs to an artificial neural network (ANN) [24] or relevance vector machine [25,26] which then produces the survival probability as the output. As pointed out in [27], the drawback of these approaches lies in the fact that the outputs cannot easily be converted to equivalent RULs for practical use. The latter approach employs condition monitoring data and an age value as inputs to an ANN which then produces the life percentage as the output. Although this approach is capable of enhancing the accuracy in RUL prediction, it still suffers from the follows drawbacks: (i) it simply uses all suspension data regardless of the quality and usefulness; and (ii) the only criteria to determine the RUL of a suspension unit is the minimization of a validation error in the offline training, which could lead to a largely incorrect RUL estimate or even a physically unreasonable estimate (i.e., less than or equal to zero) of that unit.

Recently, the co-training regression has been recognized as one of the main paradigms of semi-supervised learning [28–30], but its usefulness in data-driven prognostics has not been investigated. In this paper, a co-training-based data-driven prognostic approach, named COPROG (i.e., Co-training PROGnostics), is proposed as the first attempt to derive a semi-supervised learning framework for data-driven prognostics. This approach employs two individual data-driven algorithms, each of which predicts the RULs of suspension units iteratively for the other during the training process. After the RUL of a suspension unit is predicted by an individual algorithm, it becomes a virtual failure unit that is added to the training data set. In order to choose the appropriate suspension unit to use, COPROG quantifies the confidence of an algorithm in predicting the RUL of a suspension unit by how much the inclusion of that unit in the training data set reduces the sum of squared errors (SSE) in RUL prediction on the training data set. The process of iterative training is repeated until there is no suspension unit that is capable of reducing the SSE of any individual algorithm on the training data set or the maximum number of co-training iterations is reached. The final RUL prediction is performed by combining the RUL estimates produced by both individual algorithms. The underlying idea of COPROG is to facilitate an effective exploitation of the degradation trend information carried by suspension data in order to improve the generalization ability of each individual algorithm and achieve more accurate RUL prediction.

The remainder of this paper is organized as follows. Section 2 gives a brief introduction to the two data-driven prognostic algorithms, the feed-forward neural network (FFNN) and the radial basis network (RBN), that are used in this study. Section 3 presents the proposed co-training approach. Applications of the proposed approach are presented in Section 4. The paper is concluded in Section 5.