Reliability Engineering and System Safety 132 (2014) 72–82

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

A copula-based sampling method for data-driven prognostics

Zhimin Xi [a,*], Rong Jing [a], Pingfeng Wang [b], Chao Hu [c,1]

[a] Deparment of Industrial and Manufacturing Systems Engineering, University of Michigan, Dearborn, MI 48128 USA
[b] Deparment of Industrial and Manufacturing Engineering, Wichita State University, Wichita, KS 67260 USA
[c] Department of Mechanical Engineering, the University of Maryland at College Park, College Park, MD 20742, USA

ARTICLE INFO

Article history:
Received 20 March 2013
Received in revised form
3 April 2014
Accepted 27 June 2014
Available online 10 July 2014

Keywords:
Prognostics and health management (PHM)
Data-driven prognostics
Remaining useful life
COPULA
Reliability

ABSTRACT

This paper develops a Copula-based sampling method for data-driven prognostics. The method essentially consists of an offline training process and an online prediction process: (i) the offline training process builds a statistical relationship between the failure time and the time realizations at specified degradation levels on the basis of off-line training data sets; and (ii) the online prediction process identifies probable failure times for online testing units based on the statistical model constructed in the offline process and the online testing data. Our contributions in this paper are three-fold, namely the definition of a generic health index system to quantify the health degradation of an engineering system, the construction of a Copula-based statistical model to learn the statistical relationship between the failure time and the time realizations at specified degradation levels, and the development of a simulation-based approach for the prediction of remaining useful life (RUL). Two engineering case studies, namely the electric cooling fan health prognostics and the 2008 IEEE PHM challenge problem, are employed to demonstrate the effectiveness of the proposed methodology.

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 engineering systems composed of multiple components, complex joints, and various materials. 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 concerns [1–3]. Research on real-time diagnosis and prognosis which interprets data acquired by distributed sensor networks, and utilizes these data streams in making critical decisions provides significant advancements across a wide range of applications for minimizing the cost [4–6], maximizing the availability [7] and extending the service life [8]. For instance, in nuclear power plants, unexpected breakdowns can be prohibitively expensive and disastrous since they immediately result in lost power production, increased maintenance cost, reduced public confidence, and, possibly, human injuries and deaths. Research has been conducted to study the role of prognostic and health management (PHM) in effective operation of nuclear power plants [9–11].

In general, prognostics approaches can be categorized into model-based approaches [12–15], data-driven approaches [16–44] and hybrid approaches [45–49]. Model-based approaches rely on the understanding of system physics-of-failure and underlying system degradation mechanisms. Li et al. [12] developed a stochastic defect-propagation model to predict the remaining useful life of a bearing. Qiu et al. [13] established a stiffness-based prognostic model for bearing systems on the basis of vibration responses and damage mechanics. Luo et al. [14] 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. Qi et al. [15] studied a model-based approach to predict fatigue life of solder joints under both thermal cycling and vibration conditions. As practical engineering 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. Some good review articles of data-driven prognostic approaches were given in [16–20]. Schwabacher [16] reported a brief review of data-driven prognosis using various artificial neural networks. Kumar and Pecht [17]

* Corresponding author. Tel.: +1 313 436 9130; fax: +1 313 593 3692.
E-mail address: zxi@umich.edu (Z. Xi).
[1] Currently at Medtronic Inc.

http://dx.doi.org/10.1016/j.ress.2014.06.014
0951-8320/

categorized prognostic approaches into statistical reliability, life cycle loads, state estimation, and feature-extraction based models. Si et al. [18] broadly classified the approaches into two types of models (i) the model relies on directly observable state information and (ii) the model relies on estimated hidden state information. Kabir et al. [19] provided a review of data-driven prognosis in power electronics using the particle filter, Bayesian network, Mahalanobis Distance and Gaussian Process. An et al. [20] reviewed various prognostic algorithms in terms of their attributes, pros and cons using both simple and complex damage growth models. Recognizing that no unanimous categorization available for data-driven prognostic approaches, they are reviewed under several groups in this paper on the basis of above review articles and our best knowledge in the field. For the data-driven prognostics with directly measurable health state (e.g. vibration signal of the bearing system), regression-based approaches are the most popular methods, which include artificial neural networks [21,22], Gaussian Process [23,24], support vector machine/relevance vector machine (SVM/RVM) [25–28], Bayesian regression models [29–32], etc. These approaches build either a black-box or empirical model to predict the remaining useful life (RUL) of the system based on the current health state. The common property for using these approaches is that user must assume an underlying regression structure based on the training data sets, for example, number of layers and nodes in artificial neural networks, covariance function in the Gaussian Process, degradation functional form in Bayesian regression models, etc. The limitation is that different assumptions of the regression structure could affect prognostic solutions and it is tricky to identify the optimal regression structure especially when there are limited training date sets. Other than regression-based approaches, a similarity-based approach [33] was developed for estimating the RUL in prognostics where rich run-to-failure data are available and later was adopted in [34] with a novel fuzzy definition of trajectory pattern similarity for the RUL prediction of a nuclear system. The similarity-based approach predicts a deterministic RUL of the on-line test unit through weighted sum of RULs from all training data sets where the weights are determined by the similarity scores between the training data sets and the test unit. Disadvantage of the approach is that the RUL estimation is deterministic such that uncertainty of the RUL estimation is not considered. For the data-driven prognostics with non-directly measurable (or hidden) health state (e.g. state of charge of the battery cell), the system health state needs to be firstly estimated on the basis of directly measurable sensor signals, then the RUL can be predicted using the aforementioned data-driven methods. These state estimation approaches include Mahalanobis distance [35,36], Kalman filter [37,38], particle filter [39,40], hidden Markov models [41,42], etc. Other than standalone model and algorithm, an ensemble approach of various models and algorithms was also studied for data-driven prognostics [43,44]. The ensemble approach normally increases the RUL prediction robustness and accuracy with appropriate weight assignment for each member algorithm. However, it increases the algorithm complexity as well. In addition, there is lack of theoretical demonstration that the ensemble approach always performs better than each member algorithm for the online test units. Hybrid approaches attempt to take advantage of the strength from data-driven approaches as well as model-based approaches by fusing the information from both approaches [45–49]. 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.

This paper focuses on data-driven prognostics and proposes a Copula-based sampling method for the RUL prediction, in which the statistical relationship between the failure time and the time realizations at specified degradation levels is constructed using various Copulas. Generally, it is difficult to know which data-driven prognostic algorithm performs the best in a specific application because of the implicit relationship between the RUL and sensory signals. Available data-driven prognostic approaches intend to build such an empirical functional relationship with some assumptions. For example, aforementioned regression-based approaches assume certain functional structures and the similarity-based approach treats each training data set as an individual degradation model. The novelty of the proposed approach is to eliminate the assumptions for functional relationship between the RUL and sensory signals, but to build a general statistical relationship between them purely driven from the available training data sets. Two engineering case studies, namely the electric cooling fan health prognostics and the 2008 IEEE PHM challenge problem, are employed to demonstrate the effectiveness of the proposed methodology. The rest of the paper is organized as follows. Section 2 elaborates the Copula-based sampling method with three subsections: Subsection 2.1: a generic health index system, Section 2.2: Copula-based modeling, and Section 2.3: the RUL prediction. Section 3 demonstrates the proposed method with two engineering case studies. The paper is concluded in Section 4.

2. Copula-based sampling method

In a typical data-driven prognostics approach, a set of run-to-failure training units are required in order to build or fit a model for the system degradation, where the degradation feature or trend is extracted from raw sensory signals using signal processing techniques. This section explains the fundamentals of how the Copula-based sampling method can be employed to build the system degradation model and predict the RULs of on-line test units.

2.1. A generic health index system

Successful implementations of prognostic algorithms require the extraction of the health condition signatures and background health knowledge from massive training/testing sensory signals from engineered system units. To do so, this study uses a generic health index system that is composed of two distinguished health indices: physics health index (PHI) and virtual health index (VHI). The generic health index system is applicable for continuous state degradation process, which will be discretized in Copula-based modeling. For discrete state degradation process, Copula-based modeling can be applied directly described in the following subsection. In general, the PHI uses a dominant physical signal as a direct health metric and is thus applicable only if sensory signals are directly related to physics-of-failures. In the literature, most engineering practices of health prognostics are based on various PHIs, such as the battery impedance [50], the magnitude of the vibration signal [51] and the radio frequency (RF) impedance [52]. In contrast, the virtual health index (VHI) is applicable even if sensory signals are not directly related to system physics-of-failures. In this study, the VHI system is employed which transforms the multi-dimensional sensory signals to one-dimensional VHI with a linear data transformation method. The VHI system is elaborated according to the reference [33] for the sake of completeness of the proposed approach.

Suppose there are two multi-dimensional sensory data sets that represent the system healthy and failed states, $\mathbf{Q}_0$ of $M_0 \times D$ matrix and $\mathbf{Q}_1$ of $M_1 \times D$ matrix, respectively, where M_0 and M_1 are the data sizes for system healthy and failed states, respectively, and D is the dimension of each data set. Typically, signals for healthy and failed states could be collected from the training data sets at the beginning and at the end of the run-to-failure tests, respectively. One detail example for selecting the signals of healthy and failed state is illustrated in Section 3.2.2. With these two data matrices, a transformation matrix $\mathbf{T}$ can be obtained to transform the multi-dimensional sensory signal into the one-dimensional VHI as

$$\mathbf{T} = (\mathbf{Q}^T\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{S}_{off} \quad (1)$$