Mechanical Systems and Signal Processing 58-59 (2015) 206–217

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

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

Evaluating covariance in prognostic and system health management applications

Sandeep Menon [a,*], Xiaohang Jin [b], Tommy W.S. Chow [b], Michael Pecht [a]

[a] *Center for Advanced Life Cycle Engineering, University of Maryland, College Park, MD 20742, United States*
[b] *Department of Electronic Engineering, City University of Hong Kong, Hong Kong*

ARTICLE INFO

Article history:
Received 14 July 2013
Received in revised form
28 June 2014
Accepted 29 October 2014
Available online 13 January 2015

Keywords:
Prognostics
System Health Management
Covariance estimation

ABSTRACT

Developing a diagnostic and prognostic health management system involves analyzing system parameters monitored during the lifetime of the system. This data analysis may involve multiple steps, including data reduction, feature extraction, clustering and classification, building control charts, identification of anomalies, and modeling and predicting parameter degradation in order to evaluate the state of health for the system under investigation. Evaluating the covariance between the monitored system parameters allows for better understanding of the trends in monitored system data, and therefore it is an integral part of the data analysis. Typically, a sample covariance matrix is used to evaluate the covariance between monitored system parameters. The monitored system data are often sensor data, which are inherently noisy. The noise in sensor data can lead to inaccurate evaluation of the covariance in data using a sample covariance matrix. This paper examines approaches to evaluate covariance, including the minimum volume ellipsoid, the minimum covariance determinant, and the nearest neighbor variance estimation. When the performance of these approaches was evaluated on datasets with increasing percentage of Gaussian noise, it was observed that the nearest neighbor variance estimation exhibited the most stable estimates of covariance. To improve the accuracy of covariance estimates using nearest neighbor-based methodology, a modified approach for the nearest neighbor variance estimation technique is developed in this paper. Case studies based on data analysis steps involved in prognostic solutions are developed in order to compare the performance of the covariance estimation methodologies discussed in the paper.

1. Introduction

Prognostics and health management (PHM) is an enabling discipline consisting of technologies and methods to assess the reliability of a product in its actual life cycle conditions to determine the advent of failure and mitigate system risk [1]. This methodology is employed by integrating sensor data and prediction models that enable in-situ assessment of the extent of deviation and degradation of a product from its expected normal operating conditions. PHM has been implemented using data-driven (DD), physics of failure (PoF), and fusion-based approaches. The DD approaches make a decision on anomalous behavior and predictions based on the data available [2] using numerical algorithms, such as

* Corresponding author. Tel.: +1 5714802128.
E-mail address: smenon85@calce.umd.edu (S. Menon).

http://dx.doi.org/10.1016/j.ymssp.2014.10.012
0888-3270/

regression, Kalman filters, or particle filters; or algorithms based on machine learning and data mining, such as neural networks, linear discriminant analysis (LDA), decision trees, and support vector machines [3]. PoF approaches to prognostics use underlying engineering and/or failure principles and models to predict the remaining useful life (RUL). RUL prediction is generally based on the identification of potential failure modes, failure mechanisms, and failure sites for the product as a function of the product's life cycle loading conditions. In such cases, the stress at each failure site is obtained as a function of the loading conditions, the product geometry, and the material properties. Damage models are then used to determine fault generation and propagation [4]. Fusion-based prognostic approaches combine the strengths of the PoF and data-driven approaches to estimate RUL under both operating and non-operating life cycle conditions, detect anomalous behavior or intermittent faults, identify precursors to failure for maintenance planning, identify the potential processes causing system failure, and determine the nature and extent of faults for maintenance strategies [5].

The PHM approaches, especially the DD approaches, incorporate within their frameworks a variety of data analysis steps, including data reduction and feature extraction, pattern recognition, clustering and classification, building control charts, anomaly detection, machine learning, and forecasting, to evaluate the state of health (SOH) of the system under investigation. These data analysis steps in DD approaches often evaluate the covariance between the monitored system parameters. For example, Gebraeel and Pan [6] developed a degradation modeling framework that estimated the variance and covariance of the monitored system parameters. Covariance estimates have been used for state classification in a hidden Markov-based model for diagnostics and RUL prediction for metal cutting tools [7]. Covariance estimates have also been utilized in a particle filter-based prognostics approach to represent the uncertainty in prediction based on vibration features extracted from a gear box [8] and to assess degradation in bearing performance using linear discriminant analysis (LDA) and coupled hidden Markov models (HMM) [9]. DD approaches, such as Gaussian Process Regression (GPR), require accurate estimates of covariance, as shown by Goebel et al. [10]. Evaluating the covariance between system parameters also provides the ability to reduce the dimensionality of the problem by using techniques such as principal component analysis (PCA) and factor analysis. The reduction in dimensionality provides an efficient way to model system behavior when a large number of system parameters are being monitored [2,11,12]. Typically, such estimates of covariance are evaluated based on a sample covariance. However, the monitored system data are often sensor data, which inherently includes noise. Sample covariance estimates do not account for such irregularities, thus reducing the accuracy of the estimated covariance. For example, the presence of noise (outliers) in the data set often results in an overestimated covariance matrix, which reduces the accuracy of the data analysis [13]. This occurs as a result of the masking effect, where the outliers mask their existence by overestimating the covariance matrix. In turn, this decreases the accuracy of the diagnostic and prognostic solutions.

Approaches have been introduced to account for outliers in data, including trimming and “Winsorizing” [14–16]. The Winsorizing procedure begins by ordering the sample data by magnitude. Then the outlier is replaced by the value next to it. The effect of Winsorizing is to give less weight to the values in the tails while at the same time allowing more attention to be paid to the data in the middle. The trimming procedure begins similar to Winsoriszing by ordering the sample data in ascending order. The desired percentage of data is then trimmed or removed from both ends of the sample distribution of data. For example, 10% trimming means that 10% of the largest data points are removed from the data points and 10% of the smallest data points are also removed from the data points. Research has also presented alternative approaches to estimate covariance in order to develop improved diagnostic and prognostic solutions. For example, Lee et al. [17] used estimates of covariance based on the minimum covariance determinant technique for removing outliers to improve the accuracy of diagnostic features in a PHM system. Covariance estimation based on a minimum volume ellipsoid technique has also been used to develop control charts for early detection of broken rotor bars in induction motors [18]. Sample estimates of covariance have also been used for the classification of data. However, with limited training data, the sample estimates of covariance increased the misclassification error [19]. In order to reduce misclassification errors when there are limited training data as compared to the number of variables being monitored, a mixture model to estimate covariance was developed by Hoffbeck et al. [19] as an alternative to the sample covariance.

This paper discusses the sample covariance and then describes alternative approaches to estimating covariance, including the minimum volume ellipsoid (MVE), the minimum covariance determinant (MCD), and the nearest neighbor variance estimation (NNVE). A case study is then used to compare the performance between the sample covariance and these alternative approaches to covariance estimation. The case study demonstrates that in the presence of Gaussian noise, the NNVE estimates consistently show the lowest error percentages in estimated covariance among the discussed methodologies. Further, as a part of this study, a modification to the NNVE methodology is developed in order to improve the estimated covariance.

Two case studies are presented to compare the performance of the discussed methodologies in prognostic applications. The first case study focuses on LDA-based classification. The misclassification errors obtained in this case study are seen to be dependent on the type of covariance estimate used, with the modified NNVE estimates having the best performance. The second case study investigates the degradation in fan bearings by analyzing the monitored accelerations, voltage and current data. A Mahaloanobis distance (MD) measure is used determine anomalous behavior in the monitored bearing data. As in the first case study, the modified NNVE methodology had the best performance, providing the earliest detection of anomalous behavior.

The results shown in this paper demonstrate that the methodologies used for estimation of covariance between monitored parameters are important for developing robust prognostic solutions, with the modified NNVE methodology displaying the best performance.