Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2010, Article ID 175936, 22 pages
doi:10.1155/2010/175936

Review Article

A Tutorial on Nonlinear Time-Series Data Mining in Engineering Asset Health and Reliability Prediction: Concepts, Models, and Algorithms

Ming Dong

College of Economics & Management, Shanghai Jiao Tong University, 200052 Shanghai, China

Correspondence should be addressed to Ming Dong, mdong@sjtu.edu.cn

Received 24 January 2010; Accepted 24 March 2010

Academic Editor: Ming Li

Copyright © 2010 Ming Dong. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The primary objective of engineering asset management is to optimize assets service delivery potential and to minimize the related risks and costs over their entire life through the development and application of asset health and usage management in which the health and reliability prediction plays an important role. In real-life situations where an engineering asset operates under dynamic operational and environmental conditions, the lifetime of an engineering asset is generally described as monitored nonlinear time-series data and subject to high levels of uncertainty and unpredictability. It has been proved that application of data mining techniques is very useful for extracting relevant features which can be used as parameters for assets diagnosis and prognosis. In this paper, a tutorial on nonlinear time-series data mining in engineering asset health and reliability prediction is given. Besides that an overview on health and reliability prediction techniques for engineering assets is covered, this tutorial will focus on concepts, models, algorithms, and applications of hidden Markov models (HMMs) and hidden semi-Markov models (HSMMs) in engineering asset health prognosis, which are representatives of recent engineering asset health prediction techniques.

1. Introduction

Dynamic behavior of real world systems can be represented by measurements along temporal dimension (time series). These time series are collected over long periods of time and such time series is usually a source of large number of interesting behaviors that the system may have undergone in past. Human beings will be overwhelmed by the high dimensionality of the measurements and the complex dynamics of the system. The task of forecasting the time series involves predicting the time series for next few steps, which can usually provide the trends into near future. Usually, such time series patterns are inherently nonstationary in nature. There will be nonlinear correlations between variables. The matching of such time

series patterns calls for the feature extraction/modeling methods which can explicitly capture the nonstationary behavior and nonlinear correlations among variables [1].

A fundamental problem encountered in many fields is to model data o_t given a discrete time-series data sequence $y = (o_1, \ldots, o_T)$. The data o_t can often be a multidimensional variable exhibiting stochastic activity. This problem is found in diverse fields, such as control systems, event detection, handwriting recognition, and engineering asset health and reliability prediction. To analyze a time-series data sequence, it is of practical importance to select an appropriate model for the data. Mathematical tools such as Fourier transform and spectral analysis are employed frequently in the analysis of numerical data sequences. For categorical data sequences, there are many situations that one would like to employ Markov models as a mathematical tool. A number of applications such as inventory control, bioinformatics, asset reliability prediction can be found in the literature [2]. In these applications and many others, one would like to (i) characterize categorical data sequences for the purpose of comparison and classification process or (ii) model categorical data sequences, and, hence to make predictions in the control and planning processes. It has been shown that Markov models can be a promising approach for these purposes. Frequently, observations from systems are made sequentially over time. Values in the future depend, usually in a stochastic and nonlinear manner, on the observations available at present. Such dependency makes it worthwhile to predict the future from its past. The underlying dynamics from which the observed data are generated will be depicted and therefore used to forecast and possibly control future events [3]. Nonlinear time series analysis is becoming a more and more reliable tool for the study of complicated dynamics from measurements. In this paper, a tutorial on nonlinear time-series data mining in engineering asset health and reliability prediction will be provided. In detail, the corresponding concepts, models, and algorithms of HSMM-based reliability prediction will also be discussed.

Engineering asset breakdowns in industrial manufacturing systems can have significant impact on the profitability of a business. Expensive production equipment is idled and labor is no longer optimized. Condition-based maintenance (CBM) was introduced to try to maintain the correct equipment at the right time. CBM is based on using real-time data to prioritize and optimize maintenance resources. A CBM program consists of three key steps: (1) time series data acquisition step (information collecting), to obtain data relevant to system health; (2) data processing step (information handling), to handle and analyze the data or signals collected in step (1) for better understanding and interpretation of the data; and (3) maintenance decision-making step (decision-making), to recommend efficient maintenance policies.

Observing the state of the system is known as condition monitoring (CM). Such a system will determine the equipment's health and act only when maintenance is actually necessary. With condition monitoring techniques being adopted in different industrial sectors, a large amount of observation data are typically collected from individual critical assets during their operation. Such CM data are used for troubleshooting (e.g., fault diagnosis) and short-term asset condition prediction (e.g., prognosis). Using condition data for the estimation of asset health reliability however has not been well explored. The idea is dependent on the belief that CM data are able to reflect the underlying degradation process of an asset, and that the variation of condition data manifests the reliability change of an asset. As a result, asset health reliability can be estimated from condition data [4]. Reliability estimation based on condition data produces a time series of reliability evaluations with respect to asset operation time. The time series of evaluations can then be projected into the future for prognosis or prediction. Developments in recent years have allowed extensive

instrumentation of equipment, and together with better tools for analyzing condition data, the maintenance personnel of today are more than ever able to decide what is the right time to perform maintenance on some piece of equipment. Engineering asset management (EAM) is the process of organizing, planning, and controlling the acquisition, use, care, refurbishment, and/or disposal of physical assets to optimize their service delivery potential and to minimize the related risks and costs over their entire life through the development and application of asset health and usage management in which the health and reliability prediction plays an important role. Modern EAM requires the accurate assessment of current and the prediction of future asset health condition. Diagnostics and prognostics are two important aspects in a CBM program. Diagnostics deals with fault detection, isolation, and identification when abnormity occurs. Prognostics deals with fault and degradation prediction before they occur. Appropriate mathematical models that are capable of estimating times to failures and the probability of failures in the future are essential in EAM. In real-life situations where an engineering asset operates under dynamic operational and environmental conditions, the lifetime of an engineering asset is generally governed by a large number of variables. These systems are nonlinear and are subject to high levels of uncertainty and unpredictability. Two major problems hamper the implementation of CBM in industrial applications: first, the lack of knowledge about the right features to be monitored and second, the required processing power for predicting the future evolution of features. Time series data mining techniques proved to be useful for relevant features extraction. It has been proved that application of data mining techniques is very useful for extracting relevant features which can be used as parameters for machine diagnosis and prognosis [5]. There are many studies and development on a variety of methods and technologies that can be regarded as the steps towards prognostics maintenance that are needed in order to support decision making and manage operational reliability. A CBM system usually comprises several functional modules such as feature extraction, diagnostics, prognostics, and decision support. Figure 1 illustrates the relationships between these modules.

In order to establish the nonlinear relationship between CM indices and actual asset health, CM data are commonly taken to indicate the health of a monitored unit. However, the measured condition indices do not always deterministically represent the actual health of the monitored unit. The challenges and opportunities here lie in developing prognostics models that recognize the nonlinear relationship between a unit's actual survival condition and the measured CM indices. CM indices are frequently used to represent the health of the monitored unit in the existing prognostics techniques and then regression or time series prediction is employed to estimate the unit's future health. In these techniques, a threshold for the CM data is predefined to represent a failure. In practice, it can be often seen that a system fails even when its condition measurement is still below a predefined failure threshold. Conversely a system may still be performing its required function when its condition measurements already fall outside the tolerance range. Missed alarms and false alarms are significant issues in practical applications of prognostics systems. Several methods have been proposed for determining thresholds for fault detection based on mathematical models instead of solely on maintenance personnel's past experiences [6]. In the field of prognostics, more attentions should be paid on developing prognostics models that can deduce the nonlinear relationship between a unit's actual survival condition and the measured CM indices. Artificial intelligence (AI) models can be trained to learn from past examples. Hence, there are research opportunities to use the past measured condition data as model training input and the actual unit health as target output. By repetitively presenting various pairs of training input and target to the intelligent models, the models may learn to recognize

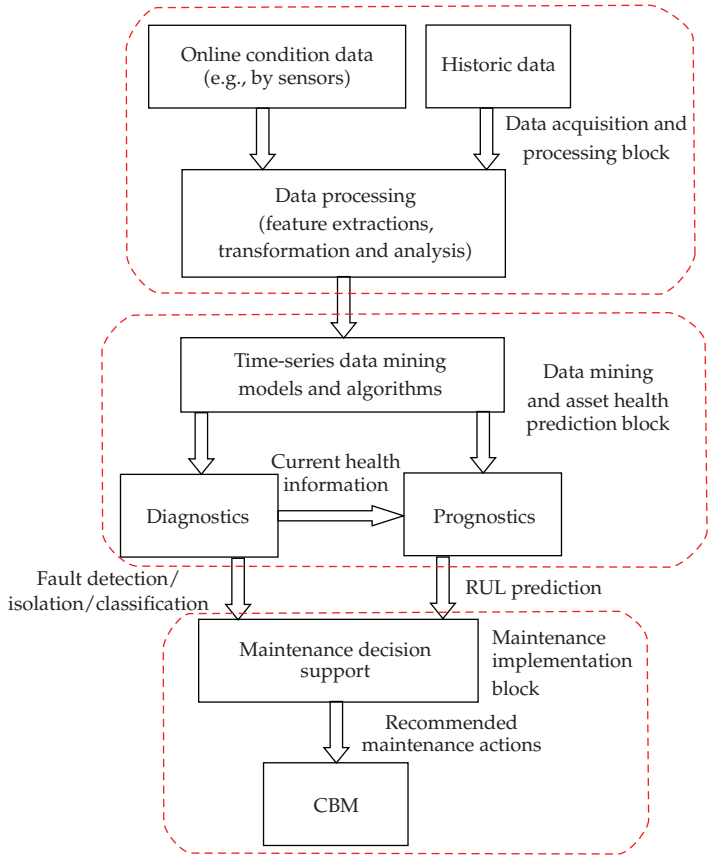

Figure 1: Functional modules and their relationships of a CBM program.

how unit degradation is veiled in the nondeterministic changes in CM measurements and disregard fluctuations caused by nondeterioration factors.

2. Health and Reliability Prediction Techniques for Engineering Assets

Health and reliability prediction is a complex process because of the numerous factors that affect the remaining useful life (RUL) levels such as the load, working condition, pressure, vibration, and temperature. The relationship between these factors has not been fully understood. Classical linear Gaussian time series (deterministic) models are inadequate in analysis and prediction of complex engineering asset reliability. Linear methods such as ARIMA (Autoregressive Integrated Moving Average) approach are unable to identify complex characteristics due to the goal of characterizing all time series observations, the necessity of time series stationarity, and the requirement of normality and independence of residuals [7]. Nonlinear time series approaches such as HMMs, artificial neural networks (ANNs), and nonlinear prediction (NLP) [8], applied to reliability forecasting, could produce accurate predictions for asset health.

Literature on prognostic methods is extremely limited but the concept has been gaining importance in recent years. Unlike numerous methods available for diagnostics,

prognostics is still in its infancy, and literature is yet to present a working model for effective prognostics [6]. Essentially, approaches for prognostics reasoning can be classified into four categories: (1) rule-based or case-based systems, (2) data-driven statistical learning models, and (3) model-driven statistical learning methods.

2.1. Rule-Based or Case-Based Systems

An example of rule-based or case-based systems is prognostic expert systems driven by data mining [9]. The application of data mining to prognostics involves identifying evolving patterns in historical data leading to failure, in order to predict and prevent imminent failures. The exploratory analysis of rules was performed using a Rule Induction algorithm to obtain rule sets along with cross-validation data to assess the strength and accuracy of the rules. The predication in the context of diagnostics is of the "If-Then" type. Prognostics involves predication of the "When" type. That is, prognosis necessitates consideration of time, or at the very least, the chronological sequence of events. Thus, prognostics needs time series and/or time sequence data prior to failure. Once sufficient amounts of such time-series data are available, one could apply a combination of techniques for time-series and time-sequence data mining to develop prognostic solutions. As indicated by Das et al. [10], extracting rules directly from time-series data involves two coupled problems. First, one must transform the low-level signal data into a more abstract symbolic alphabet. This can be achieved by data-driven clustering of signal windows in a similar way to that used in vector quantization data compression. The second problem is that of rule induction from symbolic sequences. Parameters such as cluster window width clustering methodology, number of clusters may affect the types of rules which are induced. Therefore, this technology is essentially intended as an exploratory method, and thus, iterative and interactive application of the method coupled with human interpretation of the rules is likely to lead to most useful results. A simple rule format for prognostics is "If A occurs, then B occurs within time T" (or briefly $A \overset{T}{\Rightarrow} B$). Here, A and B are letters from the alphabet produced by the discretization of a time series. The confidence of $A \overset{T}{\Rightarrow} B$, which is the fraction of occurrences of A that are followed by a B within T units, can also be derived. However, this method produces lots of rules, with varying confidences. An extension of the simple rule format is "If $A_1, A_2, \ldots, A_h$ occur within V time units, then B occurs within time T". Rules of this type have been studied under the name sequential patterns [11]. The problem with this extension is that the number of potential rules grows quickly.

2.2. Data-Driven Statistical Learning Models

Data-driven statistical learning models are developed from collected input/output data. Data-driven statistical learning models can process a wide variety of data types and exploit the nuances in the data that cannot be discovered by rule-based systems. Therefore, they potentially have superior to the rule-based systems. An example of data-driven statistical learning models is ANN. ANN is a data processing system that consists of three types of layer: input, hidden, and output (see Figure 2). In Figure 2, $r(k)$ and $e(k)$ are original input and error value, respectively, and $y(k)$ and $u(k)$ represent system's output and input, respectively. Each layer has a number of simple, neuron-like processing elements called "nodes" or "neurons" that interact with each other by using numerically weighted connections. ANN can be used to establish a complex regression function between a set of

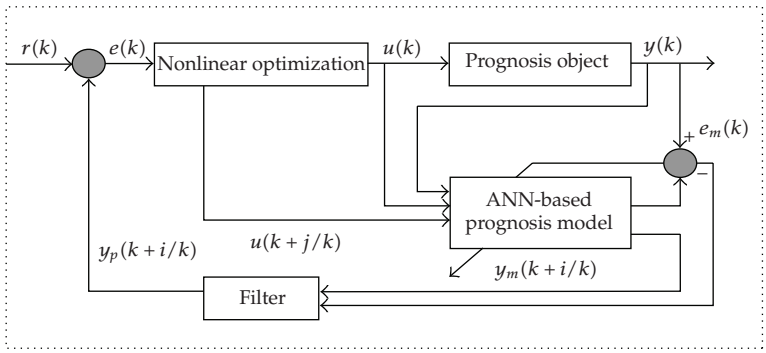

Figure 2: Structure of nonlinear prognosis model based on ANNs.

network inputs and outputs, which is achieved through a network training procedure. There are two main types of training methodologies: (1) supervised training where the network is trained using a specified sequence of inputs and outputs, and (2) unsupervised training where the primary function of the network is to classify network inputs. It is usually a tough problem for system designers to fit domain knowledge to ANN in practical applications. Besides, prognostic process itself is a "black box" for developers, which means that it is very difficult or even impossible to have physical explanations of the networks' outputs. And as ANN grows in size, training can become a complicated issue. For example, how many hidden layers should be included and what is the number of processing nodes that should be used for each of the layers are confused questions for model developers. Usually, there are five forms of ANNs: (1) multisteps prognosis model, (2) multiple back-propagation (BP) neural network model, (3) radial basis function neural network, (4) ANN Hopfield model, and (5) self-organizing maps neural network.

ANNs can be used to recognize the nonlinear relationship between actual asset health and measured condition data. A variant of the conventional neural network model, called the stochastic neural network, is used to approximate complex nonlinear stochastic systems. Lai and Wong [12] show that the expectation-maximization algorithm can be used to develop efficient estimation schemes that have much lower computational complexity than those for conventional neural networks. This enables users to carry out model selection procedures, such as the Bayesian information criterion, to choose the number of hidden units and the input variables for each hidden unit. And model-based multistep-ahead forecasts are provided. Results show that the fitted models improve postsample forecasts over conventional neural networks and other nonlinear and nonparametric models.

While ANN is being widely used to predict and forecast highly nonlinear systems, wavelet networks (WNs) have been shown to be a promising alternative to traditional neural networks. A family of wavelets can be constructed by translating and dilating the mother wavelet. Hence, in WNs, along with weights and bias, the translation and dilation factors need to be optimized. Most of the WN models make use of back-propagation algorithm to optimize their parameters. In Parasuraman and Elshorbagy's work [13], performance of ANNs and WNs in modeling two distinct time-series is investigated. The first time-series represents a chaotic system (Henon map) and the second time-series represents a geophysical time-series (streamflows). While the first time-series can be considered to be a high-frequency signal, the later time-series can be considered as a low-frequency signal. Results from the study indicate that, in modeling Henon map, WNs perform better than

ANNs. WNs are also shown to have better generalization property than ANNs. However, in modeling streamflows, ANNs are found to perform slightly better than WNs. In general, WNs are more appropriate for modeling high-frequency signals like Henon map. Moreover, WNs are computationally faster than ANNs. The performance of the models can further be improved by combining a local search technique with genetic algorithm (GA).

Li [14] gives a tutorial review about fractal time series that are substantially differs from conventional one in its statistic properties such as heavy-tailed probability distribution function and slowly decayed autocorrelation function. The concepts such as the statistical dependence, power law, and global or local self-similarity are explained. The long-range dependence (LRD) series considerably differ from the conventional series. M. Li and J. Li [15] address the particularity of the predictability of LRD series. Currently, suitable mean-square error (MSE) used for predicting LRD series may be overlooked, leaving a pitfall in this respect. Therefore, they present a generalized MSE in the domain of generalized functions for the purpose of proving the existence of LRD series prediction.

Vachtsevanos and Wang [16] attempt to address the prognosis with dynamic wavelet neural networks (DWNNs). DWNNs incorporate temporal information and storage capacity into their functionality so that they can predict into the future, carrying out fault prognostic tasks. The prognostic architecture in [16] is based on two constructs: a static "virtual sensor" that relates known measurements to fault data and a predictor which attempts to project the current state of the faulted component into the future thus revealing the time evolution of the failure mode and allowing the estimation of the component's remaining useful lifetime. A virtual sensor takes as inputs measurable quantities or features and outputs the time evolution of the fault pattern. Both constructs rely upon a wavelet neural network (WNN) model acting as the mapping tool. The WNN belongs to a new class of neural networks with unique capabilities in addressing identification and classification problems. Wavelets are a class of basic elements with oscillations of effectively finite-duration that makes them look like "little waves". The self-similar, multiple resolution nature of wavelets offers a natural framework for the analysis of physical signals and images. DWNNs have recently been proposed to address the prediction/classification issues. The DWNNs can be trained in a time-dependent way, using either a gradient-descent technique like the Levenberg-Marquardt algorithm or an evolutionary one such as the genetic algorithm. Vachtsevanos and Wang [16] point out that the notion of Time-To-Failure (TTF) is the most important measure in prognosis. The data used to train the predictor must be recorded with time information, which is the basis for the prognosis-oriented prediction task. The features are extracted in temporal series and are dynamic in the sense that the DWNN processes them in a dynamic fashion. Then, the obtained features are fused into the time-dependent feature vector that characterizes the process at the designated time instants. In the case of a bearing fault, the predictor could take the fault dimensions, failure rates, trending information, temperature, component ID, and so forth as its inputs and generate the fault growth as the output. The DWNN must be trained and validated before any online implementation and use. Such algorithms as the BP or GA can be used to train the network. Once trained, the DWNN, along with the TTF calculation mechanism, can act as an online prognostic operator. A drawback of this fault prognosis architecture consisting of a virtual sensor and a dynamic wavelet neural network is that a substantially large database is required for feature extraction, training, validation, and optimization. Since neural networks work like a black box, users do not know what features in the input data have led to the net's performance [17]. Particle filters, also known as sequential Monte Carlo (SMC) methods, are sophisticated model estimation techniques based on simulation. Particle filtering has also been employed

to provide nonlinear projection in forecasting the growth of a crack on a turbine engine blade [18]. The current fault dimension was estimated based on the knowledge of the previous state of the process model. The a priori state estimate was then updated using new CM data. To extend this state estimation to multistep-ahead prediction, a recursive integration process based on both importance sampling and kernel probability density function approximation was applied to generate state predictions to the desired prediction horizon.

2.3. Model-Driven Statistical Learning Methods

The model-driven statistical learning methods assume that both operational data and a mathematical model are available. Bayesian technique is a model-driven statistical method. A recursive Bayesian technique is proposed to calculate failure probability based on the joint density function of different CM data features [19]. This method enabled reliability analysis and prediction based on the degradation process of historical units, rather than on failure event data. The prediction accuracy of this model relied strongly on the correct determination of thresholds for the various trending features. Another widely used technique is regression, which is a generic term for all methods attempting to fit a model to observed data in order to quantify the relationship between two groups of variables. In statistics, regression analysis refers to techniques for the modeling and analysis of numerical data consisting of values of a dependent variable (also called a response variable) and of one or more independent variables (also known as explanatory variables or predictors). The fitted model may then be used either to merely describe the relationship between the two groups of variables or to predict new values. Machine learning methods have been shown to be successful for several pattern classification, regression, and data-based latent variable modeling tasks. It should be noted that the i.i.d. assumption is implicit in developing these methods. Hence, temporal aspect in the data is ignored. The state-of-the-art kernel methods proposed in [20] are no different. These methods include the kernel formulation of the latent variable models such as Kernel Principal Component Analysis (KPCA) and Kernel Partial Least Squares (KPLS). However, an important advantage of the kernel methods is that they are capable of solving nonlinear problems mainly due to implicit nonlinear mapping of data from the input space to a higher-dimensional feature space efficiently. Wavelets are mathematical tools for analyzing time series. They have two advantages when applied to analyze time series: the wavelets are shown to approximately decorrelate the time series temporally for quite general classes of time series [21]. Usually, the interesting events in time series will happen at different scales. There may be abrupt changes and steady portions. These kinds of patterns can be easily localized using multiresolution analysis capability of wavelets [1].

HMM and its varieties also belong to this category. Since the changes in feature vector are closely related to model parameters, a mathematical functional mapping between the drifting parameters and the selected prognostic features can be established. Moreover, if understanding of the system degradation improves, the model can be adapted to increase its accuracy and to address subtle performance problems. Consequently, model-driven methods can significantly outperform data-driven approaches. Being able to perform reliable prognostics is the key to CBM since prognostics are critical for improving safety, planning missions, scheduling maintenance, and reducing maintenance costs and down time. Prognostics and health management (PHM) system architectures must allow for the integration of anomaly, diagnostic, and prognostic technologies from the component level all the way up through the system level [22]. Therefore, a framework that is able to integrate

diagnostics and prognostics is desired. As indicated above, a number of approaches to the problem have been reported in the technical literature. However, these methods have yet to produce a systematic, efficient, and integrated approach to the prognostic problem. Damle and Yalcin [23] propose a novel approach to river flood prediction using time series data mining which combines chaos theory and data mining to characterize and predict events in complex, nonperiodic, and chaotic time series. Geophysical phenomena, including earthquakes, floods, and rainfall, represent a class of nonlinear systems termed chaotic, in which the relationships between variables in a system are dynamic and disproportionate, however completely deterministic. Chaos theory provides a structured explanation for irregular behavior and anomalies in systems that are not inherently stochastic. On the other hand, nonlinear approaches such as ANN, HMM, and NLP are useful in forecasting of daily discharge values in a river. The drawbacks of HMM approach are that the initial structure of the Markov model may not be certain at the time of model construction and it is very difficult to change the transition probabilities as the model itself changes with time. It was also observed that the HMMs have a higher error for longer prediction periods as well as for prediction of events with sudden occurrences.

Bunks et al. [24] and Baruah and Chinnam [6] first point out that HMM-based models could be applied in the area of prognostics in machining processes. However, only standard HMM-based approaches are proposed in their studies. The principle of HMM-based prognostics in [6] is as follows: first, build and train N HMMs for all component health states. Between N-trained HMMs, the authors assume that the estimated vectors of state transition times follow some multivariate distribution. Once the distribution is assessed, the conditional probability distribution of a distinct state transition given the previous state transition points can be estimated. In diagnostics of machining processes, tool wear is a time-related process. In prognostics of components, the objective is to predict the progression of a fault condition to component failure and estimate the remaining-useful-life of the component. Component aging process is the critical point in this issue. Therefore, it is natural to use explicit state duration models. In the new HSMM-based framework, for each health state of components, a HSMM is built and trained. Here, each health state of a component corresponds to a segment of the HSMM. These trained HSMMs can be used in the classification of a component failure mechanism given an observation sequence in diagnostics. For prognostics, another HSMM is used to model a component's life cycle. After training, the duration time in each health state can be estimated. From the estimated duration time, the proposed macrostate-based prognostic approach can be used to predict the remaining useful time for a component. Compared to the approach given in [6], the new approach provides a unified HSMM-based framework for both diagnostics and prognostics. In [6], the coordinates of the points of intersection of the log-likelihood trajectories for different HMMs along the life/usage axis represent the estimated "state transition time instants". That is, the probability distribution for state transition times in [6] is estimated from the estimation of "state transition time instants" while in HSMM, the macrostate durations are estimated directly from the training data. Also as indicated in [6], the overall shapes of actual log-likelihood plots do not resemble the ideal plots on which the "state transition time instants" are estimated. This makes the estimations of "state transition time instants" more difficult. And, the duration-based approach is more flexible than the method suggested by Baruah and Chinnam [6] and could be used in the multiple failure mode situations more efficiently. The major drawback of HSMMs is that the computational complexity may increase for the inference procedures and parameter estimations. In this regard, some approaches could be adopted to alleviate the computational burden. For example, parametric probability distributions have been used in

the variable duration HMMs. In [25], to overcome this problem, Gamma distributions are used to model state durations. In summary, the advantage of segment models is that there are many alternatives for representing a family of distributions, allowing for explicit trajectory and correlation modeling. As recent representative techniques for engineering asset reliability prediction, this tutorial will focus on models, algorithms, and applications of HMMs, and HSMMs-based approaches.

3. Concepts and Theoretical Background

3.1. Remaining Useful Life

RUL, also called remaining service life, residual life, or remnant life, refers to the time left before observing a failure given the current machine age and condition and the past operation profile [4]. It is defined as the conditional random variable:

$$T - t \mid T > t,\ Z(t), \tag{3.1}$$

where T denotes the random variable of time to failure, t is the current age, and $Z(t)$ is the past condition profile up to the current time. Since RUL is a random variable, the distribution of RUL would be of interest for full understanding of the RUL. In the literature, a term "remaining useful life estimate (RULE)" is used with double meanings. In some cases, it means finding the distribution of RUL. In some other cases, however, it just means the expectation of RUL, that is,

$$E[T - t \mid T > t, Z(t)]. \tag{3.2}$$

3.2. Description of Fault Diagnostic Process Using HMMs

The failure mechanisms of mechanical systems usually involve several degraded health states. For example, a small change in a bearing's alignment could cause a small nick in the bearing, which over time could cause scratches in the bearing race, which could then cause additional nicks, which could lead to complete bearing failure. This process can be ideally described by a mathematical model known as hidden Markov model since it can be used to estimate the unobservable health states using observable sensor signals. The word "hidden" means that the HMM states are hidden from direct observations. In other words, the HMM states manifest themselves via some probabilistic behavior. HMM can exactly capture the characteristics of each stage of the failure process, which is the basis of using HMM for failure diagnosis and prognosis [26, 27].

3.3. Elements of a Hidden Markov Model

A Markov chain is a sequence of events, usually called states, the probability of each of which is dependent only on the event immediately preceding it. An HMM represents stochastic sequences as Markov chains where the states are not directly observed but are associated with a probability function.

In the HMM framework, the time-series data sequence (observation data sequence) $y = (o_1, \ldots, o_T)$ and the hidden variable sequence $z = (s_1, \ldots, s_T)$ must be considered. The terms o_t and s_t represent the time-series data and the hidden variable at time t, and T is the sequence length. The hidden variable s_t is a variable that takes finite values among the available N states (i.e., $s_t \in \{1, \ldots, N\}$), whereas the data o_t are a discrete variable. An HMM has the following elements [28]:

(1) The first is N, the number of states in the model. Although the states are hidden, there is often some physical signal attached to the states of the model. We denote the individual states by $\{1, 2, \ldots, N\}$ and the state at time t by s_t.

(2) The second is M, the number of distinct observations for each state. The observation symbols correspond to the physical output of the system being modeled. The individual observation symbols are denoted by $V = \{v_1, v_2, \ldots, v_M\}$.

(3) The third is the state transition probability distribution $A = \{a_{ij}\}$, where

$$a_{ij} = P\left[s_{t+1} = j \mid s_t = i\right], \quad 1 \le i, j \le N. \tag{3.3}$$

(4) The fourth is the observation probability distribution in state i, $B = \{b_i(k)\}$, where

$$b_i(k) = P[v_k \mid s_t = i], \quad 1 \le i \le N, 1 \le k \le M. \tag{3.4}$$

(5) The fifth is the initial state distribution $\pi = \{\pi_i\}$ where

$$\pi_i = P[s_1 = i], \quad 1 \le i \le N. \tag{3.5}$$

It can be seen that a complete HMM requires the specifications of N, M, A, B, and π. For convenience, a compact notation is often used in the literature to indicate the complete parameter set of the model: $\lambda = (\pi, A, B)$.

The durational behavior of an HMM is usually characterized by a durational pdf $P(d)$. For a single state i, the value $P(d)$ is the probability of the event of staying in i for exactly d time units. This event is in fact the joint event of taking the self-loop for $(d-1)$ times and taking the out-going transition (with probability $1 - a_{ii}$) just once. Given the Markovian assumption, and from probability theory, $P(d)$ is simply the product of all the d probabilities:

$$P_i(d) = a_{ii}^{d-1}(1 - a_{ii}). \tag{3.6}$$

Here, $P_i(d)$ denotes the probability of staying in state i for exactly d time steps, and a_{ii} is the self-loop probability of state i. It can be seen that this is a geometrically decaying function of d. It has been argued that this is a source of inaccurate duration modeling with the HMMs since most real-life applications will not obey this function [29].

3.4. The Three Basic Problems for HMMs

In real applications, there are three basic problems associated with HMMs.

(1) *Evaluation* (also called *Classification*). Given the observation sequence $O = o_1 o_2 \cdots o_T$, and an HMM λ, what is the probability of the observation sequence given the model, that is, $P(O \mid \lambda)$.

(2) *Decoding* (also called *Recognition*). Given the observation sequence $O = o_1 o_2 \cdots o_T$, and an HMM λ, what sequence of hidden states $S = s_1 s_2 \cdots s_T$ most probably generates the given sequence of observations.

(3) *Learning* (also called *Training*). How do we adjust the model parameters $\lambda = (A, B, \pi)$ to maximize $P(O \mid \lambda)$?

Different algorithms have been developed for the above three problems. The most straightforward way of solving the evaluation problem is through enumerating every possible state sequence of length T (the number of observations). However, the computation burden for this exhaustive enumeration is prohibitively high. Fortunately, a more efficient algorithm that is based on dynamic programming exists. This algorithm is called forward-backward procedure [30]. The goal for decoding problem is to find the optimal state sequence associated with the given observation sequence. The most widely used optimality criterion is to find the single best state sequence (path), that is, to maximize $P(S \mid O, \lambda)$ that is equivalent to maximizing $P(S \mid O, \lambda)$. A formal technique for finding this single best state sequence exists, based on dynamic programming methods, and is called Viterbi algorithm [31]. For learning problem, there is no known way to obtain analytical solution. However, the model parameters $\lambda = (A, B, \pi)$ can be adjusted such that $P(O \mid \lambda)$ is locally maximized using an iterative procedure such as the Baum-Welch method (or equivalently the Expectation-Maximization algorithm) [32].

4. HSMM-Based Modeling Framework for Reliability Diagnostics and Prognostics

4.1. Macrostates and Microstates

For a component, it usually evolves through several distinct health-statuses prior to reaching failure. For example, mechanics of drilling processes suggest that a typical drill-bit may go through four health-states: good, medium, bad, and worst. In general, for a component, we can identify N distinct sequential states for a failure mechanism, that is, determination of health status of a component: no-defect (i.e., health state 1, denoted by h_1), level-1 defect (denoted by h_2), level-2 defect (denoted by h_3),..., level-$(N-1)$ defect (denoted by h_N). Here, the level-$(N-1)$ defect means failure. Let d_i be the duration staying at a health state h_i and T be the life time of a component. Then, $T = \sum_{i=1}^{N} d_i$.

Unlike a state in a standard HMM, a state in a segmental semi-Markov model generates a segment of observations, as opposed to a single observation in the HMM. In this study, the states in a segmental semi-Markov model are called *macrostates* (i.e., segments). Each macrostate consists of several single states, which are called *microstates*. Suppose that a macro-state sequence S has N segments, and let q_n be the time index of the end-point of the nth segment $(1 \le n \le N)$. The segments are as follows (see Figure 3):

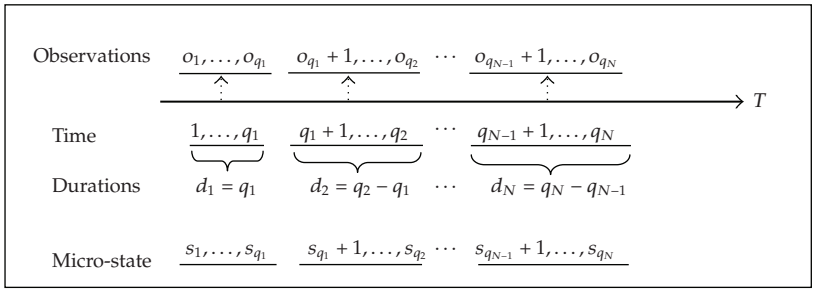

Figure 3: Segments of an HSMM.

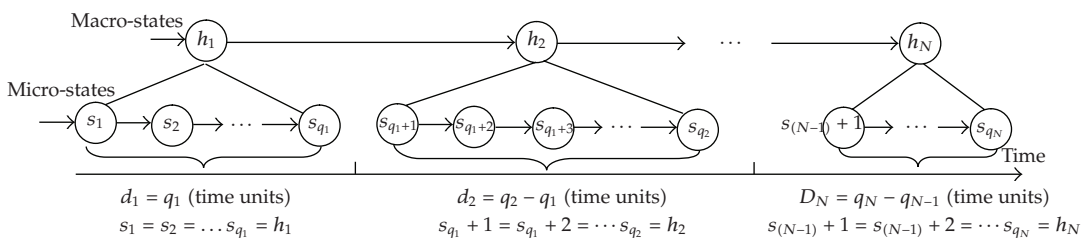

Figure 4: Segmental HSMM-based modeling framework for component diagnostics and prognostics.

For the nth macro-state, the observations are $o_{q_{n-1}+1}, \ldots, o_{q_n}$, and they have the same microstate label:

$$s_{q_{n-1}+1} = s_{q_{n-1}+2} = \cdots = s_{q_n} \equiv h_n. \tag{4.1}$$

The segmental HSMM-based modeling framework for component diagnostics and prognostics is described in Figure 4.

4.2. Model Structure

Let s_t be the hidden state at time t and let O be the observation sequence. Characterization of an HSMM is through its parameters. The parameters for an HSMM are the initial state distribution (denoted by π), the transition model (denoted by A), state duration distribution (denoted by D), and the observation model (denoted by B). Thus, an HSMM can be written as $\lambda = (\pi, A, D, B)$.

In the segmental HSMM, there are N states, and the transitions between the states are according to the transition matrix A, that is, $P(i \rightarrow j) = a_{ij}$. Similar to standard HMMs, we assume that the state s_0 at time $t = 0$ is a special state "START". This initial state distribution is denoted by π.

Although the macro-state transition $s_{q_{n-1}} \rightarrow s_{q_n}$ is Markov

$$P\left(s_{q_n} = j s_{q_{n-1}} = i\right) = a_{ij}, \tag{4.2}$$

the microstate transition $s_{t-1} \rightarrow s_t$ is usually not Markov. This is the reason why the model is called "semi-Markov" [27]. That is, in the HSMM case, the conditional independence between the past and the future is only ensured when the process moves from one state to another distinct state.

Another extension in segmental HSMM from the HMM is the segmental observation distribution. The observations $o_{(t_1,t_2]}$ in a segment with state i and duration d are produced by

$$P\left(o_{(t_1,t_2]} \mid i, d\right), \tag{4.3}$$

where $d = t_2 - t_1$.

5. Inference and Leaning Mechanisms for HSMM-Based Reliability Prediction

5.1. Inference Procedures

Similar to HMMs, HSMMs also have three basic problems to deal with, that is, evaluation, recognition, and training problems. To facilitate the computation in the HSMM-based diagnostics and prognostics framework, in the following, forward-backward variables are defined and modified forward-backward algorithm is developed [33].

A dynamic programming scheme is employed for the efficient computation of the inference procedures. To implement the inference procedures, a *forward variable* $\alpha_t(i)$ is defined as the probability of generating $o_1 o_2 \cdots o_t$ and ending in state i:

$$\alpha_t(i) = P(o_1 o_2 \cdots o_t, i \text{ ends at } t \mid \lambda) = \sum_{i=1}^{N} \sum_{d=1}^{\min(D,t)} \alpha_{t-d}(i) a_{ij} P(d \mid j) b_j\left(O_{t-d+1}^{t}\right), \tag{5.1}$$

where D is the maximum duration within any state. $b_j(O_{t-d+1}^{t})$ is the joint density of d consecutive observations $(o_{t-d+1} o_{t-d+2} \cdots o_t)$.

It can be seen that the probability of O given the model λ can be written as

$$P(O\lambda) = \sum_{i=1}^{N} \alpha_T(i). \tag{5.2}$$

5.2. Forward-Backward Algorithm for HSMMs

Similar to forward variable, the backward variable can be written as

$$\beta_t(i) = \sum_{j=1}^{N} \sum_{d=1}^{\min(D,t)} a_{ij} P(dj) b_j\left(O_{t+1}^{t+d}\right) \beta_{t+d}(j). \tag{5.3}$$

In order to give reestimation formulas for all variables of the HSMM, three more segment-featured forward-backward variables are defined:

$$\begin{gathered}\alpha_{t,t'}(i,j) = P(o_1 o_2 \cdots o_{t'}, t = q_n, s_t = i, t' = q_{n+1}, s_{t'} = j \mid \lambda),\\ \phi_{t,t'}(i,j) = \sum_{d=1}^{D}\left[P(d = t' - t \mid j) \cdot P\left(O_{t+1}^{t'} \mid t = q_n, s_t = i, t' = q_{n+1}, s_{t'} = j, \lambda\right)\right],\\ \xi_{t,t'}(i,j) = P\left(t = q_n, s_t = i, t' = q_{n+1}, s_{t'} = j \mid O_1^T, \lambda\right),\end{gathered} \tag{5.4}$$

Here, $O_{t+1}^{t'} = o_{t+1} o_{t+2} \cdots o_{t'}$ and $O_1^T = o_1 o_2 \cdots o_T$.

$\phi_{t,t'}(i,j)$ is the probability of the system being in state i for d time units and then moving to the next state j. $\alpha_{t,t'}(i,j)$ can be described, in terms of $\phi_{t,t'}(i,j)$, as follows:

$$\alpha_{t,t'}(i,j) = P(o_1 o_2 \cdots o_t o_{t+1} \cdots o_{t'}, t = q_n, s_t = i, t' = q_{n+1}, s_{t'} = j\lambda) = \alpha_t(i) a_{ij} \phi_{t,t'}(i,j). \tag{5.5}$$

The relationship between $\alpha_t(i)$ and $\alpha_{t,t'}(i,j)$ is given in the following:

$$\alpha_{t'}(j) = P(o_1 o_2 \cdots o_{t'}, t' = q_{n+1}, s_{t'} = j\lambda) = \sum_{i=1}^{N}\sum_{d=1}^{D} P(d = t' - tj)\alpha_{t,t'}(i,j). \tag{5.6}$$

From the definitions of the forward-backward variables, $\xi_{t,t'}(i,j)$ can be obtained as follows:

$$\xi_{t,t'}(i,j) = P\left(t = q_n, s_t = i, t' = q_{n+1}, s_{t'} = jO_1^T, \lambda\right) = \frac{\sum_{d=1}^{D} \alpha_t(i) a_{ij} \phi_{t,t'}(i,j) \beta_{t'}(j)}{\beta_0(i = \text{“START”})}. \tag{5.7}$$

The Forward-Backward algorithm computes the following probabilities.

Forward Pass

The forward pass of the algorithm computes $\alpha_t(i)$, $\alpha_{t,t'}(i,j)$, and $\phi_{t,t'}(i,j)$.

Step 1. Initialization ($t = 0$):

$$\alpha_{t=0}(i) = \begin{cases} 1, & \text{if } i = \text{“START”},\\ 0, & \text{otherwise.}\end{cases} \tag{5.8}$$

Step 2. Forward recursion ($t > 0$). For $t = 1, 2, \ldots, T, 1 \le i, j \le N$, and $1 \le d \le D$:

$$\phi_{t,t'}(i,j) = \sum_{d=1}^{D}\left[P(d = t' - t \mid j) \cdot P\left(O_{t+1}^{t'} t = q_n, s_t = i, t' = q_{n+1}, s_{t'} = j, \lambda\right),\right],$$
$$\alpha_{t,t'}(i,j) = \alpha_t(i) a_{ij} \phi_{t,t'}(i,j), \tag{5.9}$$
$$\alpha_{t'}(j) = \sum_{i=1}^{N}\sum_{d=1}^{D} P(d = t' - tj)\alpha_{t,t'}(i,j).$$

Backward Pass

The backward pass computes $\beta_t(i)$ and $\xi_{t,t'}(i,j)$.

Step 1. Initialization ($t = T$ and $1 \le i, j \le N$):

$$\beta_T(i) \;=\; 1. \tag{5.10}$$

Step 2. Backward recursion ($t < T$). For $t = 1, 2, \ldots, T$, $1 \le i, j \le N$, and $1 \le d \le D$:

$$\beta_t(i) = \sum_{j=1}^{N}\sum_{d=1}^{\min(D,t)} a_{ij} P(dj) b_j\left(O_{t+1}^{t+d}\right)\beta_{t+d}(j) = \sum_{j=1}^{N} a_{ij}\phi_{t,t'}(i,j)\beta_{t'}(j),$$
$$\xi_{t,t'}(i,j) = \frac{\sum_{d=1}^{D}\alpha_t(i)a_{ij}\phi_{t,t'}(i,j)\beta_{t'}(j)}{\beta_0(i = \text{“START”})}. \tag{5.11}$$

Let D_i be the maximum duration for state i. The total computational complexity for the forward-backward algorithm is $O(N^2DT)$, where $D = \sum_{i=1}^{N} D_i$.

5.3. Parameter Reestimation for HSMM-Based Reliability Prediction

The reestimation formula for initial state distribution is the probability that state i was the first state, given O:

$$\overline{\pi}_i = \frac{\pi_i\left[\sum_{d=1}^{D}\beta_d(i)P(d \mid i)b_j\left(O_1^d\right)\right]}{P(O \mid \lambda)}. \tag{5.12}$$

The reestimation formula of state transition probabilities is the ratio of the expected number of transitions from state i to state j, to the expected number of transitions from state i:

$$\overline{a}_{ij} = \frac{\sum_{t=1}^{T}\xi_{t,t'}(i,j)}{\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\xi_{t,t'}(i,j)}. \tag{5.13}$$

The formula of state duration distributions is the ratio of the expected number of times state i occurred with duration d to the expected number of times state i occurred with any duration:

$$\overline{p}_i(d) = \frac{\sum_{t=1}^{T} \alpha_{t,t'}(i,j) P(d = t' - tj) b_j\left(O_{t+1}^{t'}\right)}{\sum_{d=1}^{D} \sum_{t=1}^{T} \alpha_{t,t'}(i,j) P(d = t' - tj) b_j\left(O_{t+1}^{t'}\right)}. \tag{5.14}$$

The reestimation formula for segmental observation distributions is the expected number of times that observation $o_t = v_k$ occurred in state i, normalized by the expected number of times that any observation occurred in state i. Since $\alpha_t(i)$ accounts for the partial observation sequence $o_1 o_2 \cdots o_t$ and state i at t, $\beta_t(i)$ accounts for the partial observation sequence $o_{t'} o_{t'+1} \cdots o_T$, given state i at t. The remainder of the observation sequence $o_t o_{t+1} \cdots o_{t'}$ given state i at t and state j at t' is accounted by $P(O_{t+1}^{t'} \mid t = q_n, s_t = i, t' = q_{n+1}, s_{t'} = j)$. Therefore, the reestimation of segmental observation distributions can be calculated as follows:

$$\overline{b}_i(k) = \frac{\sum_{t=1 \text{ s.t. } O_t = v_k}^{T} \alpha_t(i) \left[\phi_{t,t'}(i,j) / \sum_{d=1}^{D} P(d = t' - ti)\right] \beta_t(i)}{\sum_{t=1}^{T} \alpha_t(i) \left[\phi_{t,t'}(i,j) / \sum_{d=1}^{D} P(d = t' - ti)\right] \beta_t(i)}. \tag{5.15}$$

5.4. Training of Macrostate Duration Models Using Parametric Probability Distributions

State duration densities could be modeled by single Gaussian distribution estimated from training data. The existing state duration estimation method is through the simultaneous training HSMMs and their state duration densities. However, these techniques are inefficient because it requires huge storage and computational load. Therefore, a new approach for training state duration models is adopted. In this approach, state duration probabilities are estimated on the lattice (or trellis) of observations and states which is obtained in the HSMM training stage.

Although the vector quantization (VQ) can be used to quantize signals via codebook, there might be serious degradation associated with such quantization [28]. Hence it would be advantageous to use the HSMMs with continuous observation densities. In this research, mixture Gaussian distribution is used. The most general representation of the pdf is a finite mixture of the following form:

$$b_j(O) = \sum_{m=1}^{M_j} c_{jm} \cdot \eta[O, \mu_{jm}, U_{jm}] = \sum_{m=1}^{M_j} P(M_m = m \mid s_t = j) \cdot \eta[O, \mu_{jm}, U_{jm}], \quad 1 \le j \le N, \tag{5.16}$$

where η represents a Gaussian distribution with mean vector μ_{jm} and covariance matrix U_{jm} for the mth mixture component in the state j, O is the vector being modeled, M_j is the number of Gaussian component η in state j, and $c_{jm} = P(M_m = m \mid s_t = j)$ is the conditional weight

for the mth mixture component in the state j. The mixture gains c_{jm} satisfy the stochastic constraint:

$$\sum_{m=1}^{M_j} c_{jm} = 1, \quad 1 \le j \le N, \tag{5.17}$$

where $c_{jm} \ge 0, \quad 1 \le j \le N, \quad 1 \le m \le M_j$, so that the pdf is properly normalized, that is,

$$\int_{-\infty}^{\infty} b_j(x)dx = 1, \quad 1 \le j \le N. \tag{5.18}$$

As pointed out in [28], the pdf of (5.18) can be used to approximate, arbitrarily closely, any finite continuous density function of practical importance. Hence, it can be applied to a wide range of problems.

6. HSMM-Based Engineering Asset Prognosis

6.1. Macrostate Duration Model-Based Prognostics

The objective of prognostics is to predict the progression of a fault condition to component failure and estimate the RUL of the component. In the following, the procedure using a macro-state duration model-based approach is provided [34].

Since each macro-state duration density $P(d_n \mid h_i)$ is modeled by a single Gaussian distribution, state durations, which maximize $\log P(S \mid \lambda, T) = \sum_{i=1}^{N} \log P(d_n \mid h_i)$ under the constraint $T = \sum_{i=1}^{N} D(h_i)$, are given by

$$\begin{gathered} D(h_i) = \mu(h_i) + \rho \cdot \sigma^2(h_i), \\ \rho = \frac{\left(T - \sum_{i=1}^{N} \mu(h_i)\right)}{\sum_{i=1}^{N} \sigma^2(h_i)}. \end{gathered} \tag{6.1}$$

6.2. Prognostics Procedure

The macro-state duration model-based component prognostics procedure is given as follows.

Step 1. From the HSMM training procedure (i.e., parameter estimation), we can obtain the state transition probability for HSMM.

Step 2. Through the HSMM parameter estimation, the duration pdf for each macro-state can be obtained. Therefore, the duration mean and variance can be calculated.

Step 3. By classification, identify the current health status of the component.

Step 5. The RUL of the system can be computed by the following backward recursive equations (suppose that the system currently stays at health state i; RUL_i indicates the remaining useful life starting from state i).

At state $N-1$,

$$\mathrm{RUL}_{N-1} = a_{N-1,N-1}[D(h_{N-1}) + D(h_N)] + a_{N-1,N}[D(h_N)]. \tag{6.2}$$

At state $N-2$,

$$\mathrm{RUL}_{N-2} = a_{N-2,N-2}[D(h_{N-2}) + \mathrm{RUL}_{N-1}] + a_{N-2,N-1}[\mathrm{RUL}_{N-1}]. \tag{6.3}$$

At state i,

$$\mathrm{RUL}_i = \widehat{t}_{ic} + a_{i,I+1}[\mathrm{RUL}_i + 1]. \tag{6.4}$$

7. Demonstration of Engineering Asset Diagnostic and Prognostic Processes

7.1. Diagnostics for Pumps

In this demonstration, a real hydraulic pump health monitoring application is provided by using HSMM-based reliability prediction. In the test experiments, three pumps (pump 6, pump 24, and pump 82) were worn to various percent decreases in flow by running them using oil containing dust. Each pump experienced four states: baseline (normal state), contamination 1 (5 mg of 20-micron dust injected into the oil reservoir), contamination 2 (10 mg of 20-micron dust injected into the oil reservoir), and contamination 3 (15 mg of 20-micron dust injected into the oil reservoir). The contamination stages in this hydraulic pump wear test case study correspond to different stages of flow loss in the pumps. As flow rate of a pump clearly indicates the heath state of a pump, therefore, the contamination stages corresponding to different degrees of flow loss in a pump were defined as the health states of the pump in the pump wear test. The data collected were processed using wavelet packet with Daubechies wavelet 10 (db10) and five decomposition levels as the db10 wavelet with five decomposition levels provides the most effective way to capture the fault information in the pump vibration data [35–37]. The wavelet coefficients obtained by the wavelet packet decomposition were used as the inputs to the HMMs and HSMMs. In this test, we wanted to see how the HSMMs could classify the health conditions of the pumps in comparison with the HMMs. From the diagnosis results, it can be seen that the classification rates for all three pumps reach 100%. For individual pump's diagnostics, it can be seen that the correct recognition rate is increased by 29.3%, which shows that the HSMM is superior to currently used HMM-based approach. In addition, experiments show that both HMM-based diagnosis and HSMM-based diagnosis have almost the same computational time. This means that HSMM-based method is efficient and could be used in the real applications with large data sets.

7.2. Prognostics for Pumps

For prognostics, the life time training data from pump 6, pump 24, and pump 82 are used. By training, an HSMM with four health states can be obtained. And, the mean and variance

of the duration time in each state are also available through the training process. Then, the mean value of the remaining useful life of a pump can be calculated as follows (in terms of (6.4) and suppose that the component currently stays at state "Contamination1"):

$$\mathrm{RUL}_{\mathrm{mean}} = 28.0829 \quad \text{(time units)}. \tag{7.1}$$

Similarly, the variance of the remaining useful life of a pump can be obtained as follows:

$$\mathrm{RUL}_{\mathrm{variance}} = 1.7846 \quad \text{(time units)}. \tag{7.2}$$

That is, if the component is currently at state "Contamination1", then its expected remaining useful life is 28.0829 time units with confidence interval 1.7846 time units.

8. Conclusions

The concepts, models, algorithms, and applications of nonlinear time-series data mining in engineering asset health and reliability prediction are discussed. Various techniques and algorithms for engineering asset reliability prediction have been reviewed and categorized depending on what models are usually adopted. In order to obtain the insights of the engineering asset health and reliability prediction, the detailed models, algorithms-and applications of HSMM-based asset health prognosis are given. The health states of assets are modeled by state transition probability matrix and observation probability. The duration of each health segment is described by the state duration probability. As a whole, they are modeled as a hidden semi-Markov chain.

Although prognostics is still in its infancy and literature is yet to present a working model for effective prognostics, a new trend is that more combination models are designed to deal with data extraction, data processing, and modeling for prognostics. From simple heuristic-based models to complex HSMM models that impose artificial intelligence knowledge, these methodologies have their own advantages and disadvantages. Since single-approach models have some difficulties in achieving satisfied results, it is a very challenging work to develop prognostics applications that can provide precise prediction. A well designed combination model usually combines two or more theories and algorithms to model the system in order to eliminate the disadvantages of each individual theory and utilize the advantages of all combined methods. On the other hand, it is also a challenging work to choose appropriate methods and combine them together for engineering asset health and reliability prediction modeling.

Acknowledgment

This work was partly supported by Grants from the National High Technology Research and Development Program ("863" Program) of China (2008AA04Z104).

References

[1] B. V. Kini and C. C. Sekhar, "Multi-scale kernel latent variable models for nonlinear time series pattern matching," in *Proceedings of the 14th International Conference on Neural Information Processing (ICONIP '08)*, vol. 4985 of *Lecture Notes in Computer Science*, pp. 11–20, 2008.

[2] W.-K. Ching and M. K. Ng, *Markov Chains: Models, Algorithms and Applications*, International Series in Operations Research & Management Science, 83, Springer, New York, NY, USA, 2006.

[3] J. Fan and Q. Yao, *Nonlinear Time Series: Nonparametric and Parametric Method*, Springer, New York, NY, USA, 2005.

[4] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical Systems and Signal Processing*, vol. 20, no. 7, pp. 1483–1510, 2006.

[5] G. Van Dijck, *Information theoretic approach to feature selection and redundancy assessment*, Ph.D. dissertation, Katholieke Universiteit Leuven, 2008.

[6] P. Baruah and R. B. Chinnam, "HMMs for diagnostics and prognostics in machining processes," in *Proceedings of the 57th Society for Machine Failure Prevention Technology Conference*, Virginia Beach, Va, USA, April 2003.

[7] R. J. Povinelli, *Time series data mining: identifying temporal patterns for characterization and prediction of time series events*, Ph.D. dissertation, Marquette University, Milwaukee, Wis, USA, 1999.

[8] F. Laio, A. Porporato, R. Revelli, and L. Ridolfi, "A comparison of nonlinear forecasting methods," *Water Resources Research*, vol. 39, no. 5, pp. TNN21–TNN24, 2003.

[9] A. Mathur, "Data mining of aviation data for advancing health management," in *Component and Systems Diagnostics, Prognostics, and Health Management II*, vol. 4733 of *Proceedings of SPIE*, pp. 61–71, 2002.

[10] G. Das, K. P. Lin, H. Mannila, G. Renganathan, and P. Smyth, "Rule discovery from time series," in *Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining (KDD '98)*, pp. 16–22, AAAI Press, 1998.

[11] R. Agrawal and R. Srikant, "Mining sequential patterns," in *Proceedings of the International Conference on Data Engineering (ICDE '95)*, pp. 3–14, Taipei, Taiwan, 1995.

[12] T. L. Lai and S. P.-S. Wong, "Stochastic neural networks with applications to nonlinear time series," *Journal of the American Statistical Association*, vol. 96, no. 455, pp. 968–981, 2001.

[13] K. Parasuraman and A. Elshorbagy, "Wavelet networks: an alternative to classical neural networks," in *Proceedings of the International Joint Conference on Neural Networks*, vol. 5, pp. 2674–2679, Montreal, Canada, 2005.

[14] M. Li, "Fractal time series—a tutorial review," *Mathematical Problems in Engineering*, vol. 2010, Article ID 157264, 26 pages, 2010.

[15] M. Li and J. Li, "On the predictability of long-range dependent series," *Mathematical Problems in Engineering*, vol. 2010, Article ID 397454, 9 pages, 2010.

[16] G. Vachtsevanos and P. Wang, "Fault prognosis using dynamic wavelet neural networks," in *Proceedings of IEEE Systems Readiness Technology Conference (AUTOTESTCON '01)*, pp. 857–870, 2001.

[17] T. Brotherton, G. Jahns, J. Jacobs, and D. Wroblewski, "Prognosis of faults in gas turbine engines," in *Proceedings of the IEEE Aerospace Conference*, vol. 6, pp. 163–172, 2000.

[18] M. Orchard, B. Wu, and G. Vachtsevanos, "A particle filter framework for failure prognosis," in *Proceedings of the World Tribology Congress*, Washington, DC, USA, 2005.

[19] S. Zhang, L. Ma, Y. Sun, and J. Mathew, "Asset health reliability estimation based on condition data," in *Proceedings of the 2nd World Congress on Engineering Asset Management and the 4th International Congress of Chinese Mathematicians (ICCM '07)*, pp. 2195–2204, Harrogate, UK, 2007.

[20] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 1995.

[21] D. B. Percival and A. T. Walden, *Wavelet Methods for Time Series Analysis*, vol. 4 of *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge University Press, Cambridge, UK, 2000.

[22] M. J. Roemer, E. O. Nwadiogbu, and G. Bloor, "Development of diagnostic and prognostic technologies for aerospace health management applications," in *Proceedings of the IEEE Aerospace Conference*, vol. 6, pp. 63139–63147, Big Sky, Mont, USA, March 2001.

[23] C. Damle and A. Yalcin, "Flood prediction using time series data mining," *Journal of Hydrology*, vol. 333, no. 2–4, pp. 305–316, 2007.

[24] C. Bunks, D. McCarthy, and T. Al-Ani, "Condition-based maintenance of machines using hidden Markov models," *Mechanical Systems and Signal Processing*, vol. 14, no. 4, pp. 597–612, 2000.

[25] A. Ljolje and S. E. Levinson, "Development of an acoustic-phonetic hidden Markov model for continuous speech recognition," *IEEE Transactions on Signal Processing*, vol. 39, no. 1, pp. 29–39, 1991.

[26] C. Kwan, X. Zhang, R. Xu, and L. Haynes, "A novel approach to fault diagnostics and prognostics," in *Proceedings of the IEEE International Conference on Robotics and Automation*, vol. 1, pp. 604–609, Taipei, Taiwan, 2003.

[27] P. D. McFadden and J. D. Smith, "Vibration monitoring of rolling element bearings by the high-frequency resonance technique—a review," *Tribology International*, vol. 17, no. 1, pp. 3–10, 1984.

[28] L. R. Rabiner, "Tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[29] M. J. Russell and R. K. Moore, "Explicit modeling of state occupancy in hidden Markov models for automatic speech recognition," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '85)*, pp. 5–8, Tampa, Fla, USA, 1985.

[30] L. E. Baum and J. A. Eagon, "An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology," *Bulletin of the American Mathematical Society*, vol. 73, pp. 360–363, 1967.

[31] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimal decoding algorithm," *IEEE Transaction on Information Theory*, vol. 13, pp. 260–269, 1967.

[32] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society*, vol. 39, no. 1, pp. 1–38, 1977.

[33] M. Dong and D. He, "Hidden semi-Markov model-based methodology for multi-sensor equipment health diagnosis and prognosis," *European Journal of Operational Research*, vol. 178, no. 3, pp. 858–878, 2007.

[34] M. Dong, D. He, P. Banerjee, and J. Keller, "Equipment health diagnosis and prognosis using hidden semi-Markov models," *International Journal of Advanced Manufacturing Technology*, vol. 30, no. 7-8, pp. 738–749, 2006.

[35] D. He, D. Wang, A. Babayan, and Q. Zhang, "Intelligent equipment health diagnosis and prognosis using wavelet," in *Proceedings of the Conference on Automation Technology for Off-Road Equipment*, pp. 77–88, Chicago, Ill, USA, 2002.

[36] P. Pawelski and D. He, "Vibration based pump health monitoring," *Journal of Commercial Vehicles*, vol. 2, pp. 636–639, 2005.

[37] M. Dong and D. He, "A segmental hidden semi-Markov model (HSMM)-based diagnostics and prognostics framework and methodology," *Mechanical Systems and Signal Processing*, vol. 21, no. 5, pp. 2248–2266, 2007.