Prognostic and health management for engineering systems: a review of the data-driven approach and algorithms

Thamo Sutharssan[1], Stoyan Stoyanov[2], Chris Bailey[2], Chunyan Yin[2]

[1]Sustainable Energy Technologies Group, University of Hertfordshire, Hatfield, Hertfordshire, UK
[2]Computational Mechanics and Reliability Group, Old Royal Naval College, University of Greenwich, Park Row, London, UK
E-mail: TSutharssan@iee.org

Abstract: Prognostics and health management (PHM) has become an important component of many engineering systems and products, where algorithms are used to detect anomalies, diagnose faults and predict remaining useful lifetime (RUL). PHM can provide many advantages to users and maintainers. Although primary goals are to ensure the safety, provide state of the health and estimate RUL of the components and systems, there are also financial benefits such as operational and maintenance cost reductions and extended lifetime. This study aims at reviewing the current status of algorithms and methods used to underpin different existing PHM approaches. The focus is on providing a structured and comprehensive classification of the existing state-of-the-art PHM approaches, data-driven approaches and algorithms.

1 Introduction

Prognostics and health management (PHM) is an engineering process of failure prevention, and predicting reliability and remaining useful lifetime (RUL). PHM of engineering systems has become very important as a malfunction or failure may cause severe damage to the system, environment and users, and may result in significant repair on un-scheduled maintenance costs. PHM is now widely recognised as an efficient and practical approach to these engineering challenges. Repair and maintenance cost can be reduced by converting the unscheduled maintenance tasks into evidence-based scheduled maintenance tasks. Evidence-based scheduled maintenance strategy reduces the inspection cost, required number of skilled labours, system down time, life-cycle cost of the system and emergency unscheduled maintenance [1]. PHM is identified as the best candidate to improve the maintenance cycle, reduce the maintenance cost and extend the overall lifetime through evidence-based scheduled maintenance strategies. PHM can also provide support to improve the qualification approach and improve the design of the future systems [1].

PHM applications can be classified into two main categories based on how the PHM is applied to the system or to the product:

(1) Real-time PHM (sometimes referred as online PHM).
(2) Off-line PHM.

Most of the safety critical and mission critical applications require real-time PHM (referred as on-board health monitoring). Modern aircrafts, automobiles and so on have substantial on-board monitoring capability that is based on the use of data from real-time sensors. For example, an electric car provides the range distance which can be achieved with the battery operation based on the real-time PHM of the battery. Another example is the autonomous unmanned vehicles which have embedded real-time on-board PHM used to re-plan the mission and reconfigure the controls based on the health diagnostic and prognostic information. Such capability requires the evaluation of the current state of the health and also a prediction of the future state of the product/ system's health [2]. Real-time PHM for electronic systems is sometimes known as built-in self-test (BIST) or self-scanning where the electronic system tests itself. BIST are generally used in all modern weapon systems, avionics, safety critical systems, automotive and other electronic hardware. Such embedded diagnostics and prognostics allow performing tests to verify if all parts of the electronic-based equipment operate as required.

Off-line PHM is deployed where the system's safety is not critical and likelihood of failures is very small. Data are collected from the system and they are used off-line to predict the RUL and to perform the maintenance. One main advantage of the off-line PHM is that complex systems models can be used to perform the PHM using computer simulations, whereas in the real-time PHM computer simulations may not be achievable as there might be limitation in the available on-board computational power and efficiency. Failure mode, effects and critically analysis (FMECA) which is an extension of failure modes and effects analysis, includes the probability of the failure modes against the severity of the consequences. It can be used in off-line PHM. FMECA is developed over the time by collecting the data for particular system or product; hence, the developed FMECA knowledge can be used to predict the failures and the RUL of a system.

PHM is also applied in development and deployment stages of a system or product. PHM methodology can be applied in the design stage to optimise the design to ensure the expected performance from the systems or products given certain reliability requirements. Physics of failure (PoF)-based models are used to optimise the product design based on failure modes, mechanisms and effective analysis. Life-cycle loads, for example, thermal, electrical, mechanical, chemical and so on, acting on the products at different stages and under different conditions of the product life, such as manufacturing, storage, shipment, harsh operating, non-operating and so on, are considered at the product design stage to optimise the product design and obtain the best performance from the product for a certain period of time without failure. These life-cycle loads are also monitored, and used with the PoF-based damage models to assess the reliability and degradation of the product in the field after it has been deployed [3]. Anomaly detection is the starting point of the PHM for systems and products when monitored in the field. Anomaly detection and failure prevention can be achieved effectively by monitoring the life-cycle loads and relevant performance parameters of the systems. The PHM will be more accurate if the life-cycle loads and parameters are monitored in real-time, especially in the case of critical applications. Many safety critical systems and mission critical systems consist of electronic hardware and software that control the electronic hardware and also interact with the user. Most of these electronic hardware devices use thousands of individual semiconductor components to perform their

operation. Malfunction or failure of any individual semiconductor component, electronic hardware or software module independently affects the system as a whole.

Health of a system can also be defined as the extent of deviation or degradation from its expected typical operating performance [4]. This extent of deviation or degradation of the expected typical operating performance has to be determined accurately to prevent the failures. It is also necessary to determine which operating parameters are contributing to this extent of deviation or degradation. There are two different approaches available to assess the degradation or extent of deviation from the expected performance, to assess reliability of systems and to predict the remaining useful life using PHM. They are:

(1) Data-driven approach.
(2) Model-driven approach.

Fusion approach, as illustrated in Fig. 1, incorporates the advance features from both data-driven and model-driven approaches to perform the PHM.

On the basis of the techniques used for data-driven and model-driven approaches, PHM can be further classified into different approaches. For example, data-driven approaches can be classified into statistical and machine learning approaches. Model-driven approach can be classified into PoF and system model approaches. Fig. 2 shows the classification of PHM approaches. All these approaches can be used as online or off-line prognostics techniques. The next sections provide detailed review of the different data-driven PHM classifications.

2 Data-driven approach

Data-driven approaches are considered as a black box approaches to PHM as they do not require system models or systems specific knowledge to start the prognostics [3]. Monitored and historical data are used to learn the systems behaviours and to perform the prognostics. Hence, the data-driven approach is suitable for systems which are complex and with behaviours that cannot be assessed and derived from first principles. The implementation of data-driven techniques for the purpose of health monitoring and prognostics is generally based on the assumption that the statistical characteristics of system's performance will not change until a fault occurs [3]. Therefore the main advantage of data-driven approach is that the underlying algorithms are quicker to implement and computationally more efficient to run compared with other techniques. However, it is necessary to have historical data and knowledge of typical operational performance data, the associated critical threshold values and their margins. Data-driven techniques rely completely on the analysis of data obtained from sensors and exploit operational or performance related signals that can indicate the health of the monitored system. Data-driven strategies for

diagnostics and prognostics have been applied in a number of different PHM applications [5–12].

The principal disadvantage of the data-driven approach is that the confidence level in the predictions depends on the available historical and empirical data (i.e. healthy and failure data). Availability of run-to-failure data sets for a particular system or component is the main disadvantage of data-driven PHM, as running a system or a component to failure might be time consuming and expensive [13]. These data are required in the data-driven approach to define the respective threshold values. In some instances, it is difficult to obtain or have historical data available, for example, in the case of a new system or device that may require long time and/or expensive tests to failure to generate this data. However, there are techniques and procedures available that can be used to achieve this [14–17]. Three of the strategies used to address this challenge are based on the use of:

1. *Hardware-in-the-loop simulations (HiL):* HiL is a computer simulation which is used to test hardware under simulated loads as in the real application. Several failure parameters (i.e. operational and environmental) can be controlled independently. HiL can also be used to algorithm development, testing and validation, benchmarking and development of metrics for prognostics [14, 15].
2. *Accelerated life test (ALT):* Accelerated load test is designed to cause the product to fail more quickly than under normal conditions by applying accelerated (elevated) stress conditions resulting in the same failure mechanisms. ALT becomes an important methodology in the development of the PHM for electronics. Several environmental and loading conditions can be applied independently to accelerate the failure [16, 17].
3. *Online learning:* Online learning is based on the assumption that new systems do not fail for a certain period of time and hence the performance data during early stages of use (operation) can be used to define the 'healthy' status of the system. This approach is called semi-supervised learning approach as only healthy data are available.

Data-driven approaches for PHM can be also classified as falling within one of the following two classes [11]:

(1) Statistical approach.
(2) Machine learning approach.

The statistical approach uses statistical parameters, such as mean, variance, median and so on, to make predictions based on known or unknown underlying probabilistic distributions. Statistical approaches are generally considered to be simple if the underlying statistical property (i.e. probability distribution) is known. This type of approach is called parametric approach. Statistical parameter estimation techniques and hypothesis testing can be applied in this case to detect the presence of anomalies in the data [18]. Techniques based on the use of statistical distance measure are another simple way to estimate the distance of the new sample
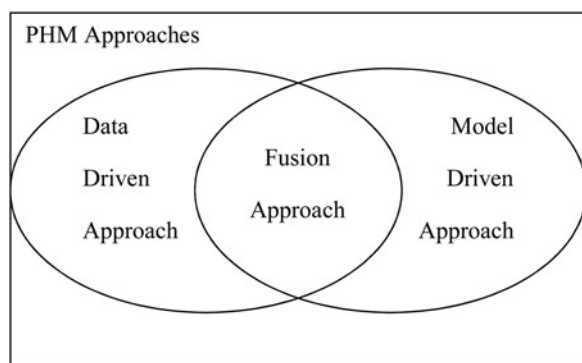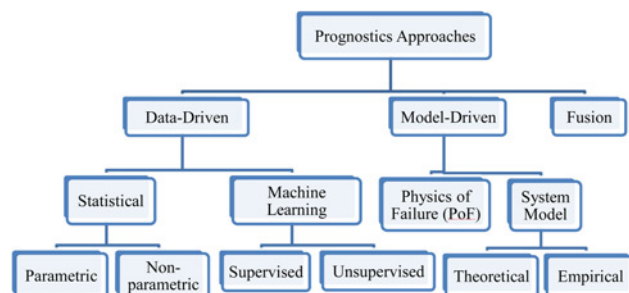

Fig. 1 *PHM approaches*

Fig. 2 *Classification of PHM approaches*

data from the expected mean data (i.e. how many standard deviations away from the mean) [18]. Outlier rejection techniques can also be exploited to detect the anomalies based on the box plot parameters such as lower extreme and upper extreme [18]. Unfortunately, most of the real-world reliability data's statistical properties are unknown and probability functions representing these data need to be constructed first. This type of approach is called non-parametric approach and it introduces more flexibility into computation. Therefore a non-parametric approach can be viewed as generalised approach. One of the widely used non-parametric approaches to PHM is through histogram analysis. A better way to estimate the density function is to use kernel methods [18].

Machine learning approaches make predictions based on acquired data (such as healthy and failure data) by converting the gathered data into useful information which can be used in conjunction with sensor data to provide future predictions. The machine learning approach is more data-driven approach and typically no statistical assumptions are made. One of the well-known PHM approaches in the field of machine learning is based on the use of neural networks [19]. Machine learning approach can also be realised using support vector machine (SVM). With this method the data are separated into different classes using hyper-planes, after they are transformed by a kernel function [20]. SVM uses linear combination of kernel functions centred on the subset of the training data which are known as support vectors [21].

PHM applications may require more than one algorithm for different tasks such as for the anomaly detection, parameter isolation, parameter trending, damage estimation, lifetime estimation and so on. Hence, different types of algorithms can be used to achieve these individual tasks based on the performance of those algorithms.

3 Statistical approach

The statistical approach to PHM is based on analysis of the underlying statistical property of the data such as type of distribution, stationary or non-stationary and so on. If new observation is not representing the statistical property of the data, then the observation is considered as an anomaly. Statistical techniques fit the typical expected operating condition data and then apply statistical inference test to determine if the new observation belongs to the fitted statistical model. For example, if the data representing the normal operating condition is modelled with probability density function (PDF) $p(x)$, then new observation data can be tested against the developed PDF (i.e. if $p(x_{\text{new}}) < \varepsilon \rightarrow$ flag anomaly and if $p(x_{\text{new}}) \geq \varepsilon \rightarrow$ flag normal). There are two different ways to fit the data into a statistical model, that is, to develop a PDF function ($p(x)$): (i) parametric approach and (ii) non-parametric approach [22]. Statistical models have different computational complexity and require different computational powers based on the complexity of the statistical models.

The main advantage of the statistical approach to PHM is if the assumed statistical characteristics are true, then the result from the statistical inference test for the new observation will be statistically reasonable. In addition, the use of a statistical approach can provide a confident interval and this can be used in the decision making about the new observation data. Furthermore, this statistical approach can be underpinned by unsupervised techniques which do not require training data, as in the case of the robust statistical approach [22]. The main disadvantage of the statistical approach is that it depends on the assumed statistical characteristics of the data, and hence, if the assumption is not true, they will not detect the anomalies accurately. Typically, the assumption may not be true in particular high-dimensional real data sets. In addition, even if the assumed statistical characteristics are true, there are many statistical inference tests that are available, and selecting the suitable test itself might be difficult [22].

4 Parametric approach

Parametric approaches are based on assumed underlying statistical properties (typically in the form of known distribution, that is, normal, Weibull, exponential and so on.) of the expected data. On the basis of the assumed underlying probability distribution of the data, parameters associated with that probability distribution are calculated from the data. Typically, these data will represent a healthy systems performance under expected typical operating conditions. Healthy or normal operating data then will be defined by these parameters and assumed as being the representative probability distribution. This distribution can be used to detect anomalies and predict the RUL. Once the system's healthy or normal operating data are defined by a probability distribution, new monitored data can be classified using different methods based on the probability distributions used. Some of the methods are listed below:

1. *Hypothesis testing*: One of the simplest statistical procedures which can be used to test if the data comes from the same population as the training data [18]. It can also be used to test if the mean of a sample is equal to μ when the standard deviation σ is known [23]. Hypotheses are always statements about the sample population parameters instead of sample population data. There will be two types of errors which may occur in the hypothesis testing: (i) Type I Error and (ii) Type II Error. Type I Error (false positive – α) is defined as rejecting the null hypothesis when null hypothesis is actually true. Type II Error (false negative – β) is defined as accepting the null hypothesis when the hypothesis is actually false. It is not possible to eliminate these errors completely. Typically, the hypothesis test decision is taken by fixing an acceptable value for σ and by minimising the β. Standardised difference between the population and sample statistics is compared with the decision rules before making the decision. Most of the hypothesis tests use underlying PDF as normal.

2. *Analysis of variance (ANOVA):* It is a method to analyse the means of several groups of samples which can be affected by different types of factors. Simplest form is one-way analysis and it is an extension of *t*-test. Simple form of ANOVA can be used to compare different groups of sample data [23]. ANOVA can be applied to groups of data based on the assumptions: (i) values are normally distributed in every group and (ii) variance are equal. Decision will be made based on the variability among the groups. If the variability among the group is small compared with the variability within the group, then this will lead to the decision says groups can be treated as identical. If the variability among the groups is large compared with the variability within the groups, then the groups cannot be treated as identical.

3. *Extreme value theory (EVT):* EVT is a branch of statistics that deals with the analysis of data at the tails of a given distribution. EVT can be used to set the threshold values for anomaly detection, where EVT explicitly models the tails of the distribution of normal data [19, 24, 25].

4. *Maximum-likelihood (ML) estimation:* It is an approach to estimate the most likely value related to other values in the population data. Log-likelihood which is the logarithm of the likelihood function is typically used to estimate the MLE by maximising the log-likelihood [11]. If the ML is the mutually independent observations $x = \{x_x, x_2, x_3, \ldots, x_n\}$ which is an instance of the random sample $\{X_1, X_2, X_3, \ldots, X_n\}$, then the joint PDF is equal to the product of the marginal PDFs.

5. *Maximum-a-posteriori (MAP) estimation:* MAP estimation is considered as a Bayesian version of ML estimation [11]. This estimation technique can be used to estimate the parameters of a process or system based on prior knowledge of the system. This prior knowledge typically comes from the historical data that are available for the system. Such prior information can be included in the estimation in the form of PDF. Parameters θ which need to

be estimated are considered as random variable and the associate probabilities $P(\theta)$ are called the prior probabilities. Bayes' theorem can be applied to incorporate the prior information into the estimation [26].

6. *Expectation–maximisation (EM) algorithm:* EM algorithm is used to estimate the ML or MAP of parameters using an interactive process. These parameters are from a statistical model which depends on some hidden variable. The iterative process switches between two different steps in the process: (i) estimating the expectation ('E' step) and (ii) maximisation ('M' step). The 'E' step is used to compute the expectation of the log-likelihood based on the current estimate for the parameters. The 'M' step is used to compute parameters which maximise the computed log-likelihood in the 'E' step. Estimated parameters values are used to compute the expectation of the log-likelihood in the next 'E' step and these have to be repeated until the log-likelihood of the parameter becomes constant [11, 26].

7. *Gaussian mixture modelling (GMM):* It is widely used for density estimation and to form the hidden space of radical-basis function (RBF) networks [25]. GMM uses fewer kernels than the number of patterns in the training set and model parameters are estimated by maximising the log-likelihood of the training set with respect to the model. Optimisation algorithms such as conjugate gradients are used to maximise the log-likelihood of the training set with respect to the model. One of the disadvantages is the very large sample sizes that are required to train the model, particularly if the dimensionality of the data is high [18].

5 Non-parametric approach

The non-parametric approach is not based on any assumption of underlying statistical property of the population data. It gives more flexibility than the parametric approach and can be used to fit the data more accurately. The non-parametric approach is more suitable in the case when the underlying probability distribution is not known and also when the data cannot be modelled with standard distribution. For those reasons, most of the real-world data would typically require a non-parametric approach to estimate the density function. There are many approaches available to solve a non-parametric problem. Some of the non-parametric techniques and approaches are listed below:

1. *Parzen–Window density estimation:* This non-parametric density estimation technique was introduced by Emanuel Parzen in 1960s. Density is estimated such that all the observation data belong to a window function that can contribute to the density estimation based on selected window kernel function [11, 18, 27]. For a given data set $D = \{x_1, x_2, \ldots, x_n\}$ of n independent and identically distributed example drawn according to $p(x)$ which is the density function that needs to be estimated, the Parzen–window estimate of $p(x)$ based on the n example [27]. Typically Gaussian kernel is used in many situations as (i) they are very smooth and (ii) radially symmetrical Gaussian function is available. Hence the estimated density function will also be smooth and can be a mixture of radially symmetrical Gaussian kernel with a common variance σ^2 [27]. There are many different kernel functions and some of the commonly used functions are Gaussian, uniform, box, triangle, Epanechnikov and so on. The kernel function is generally selected based on the required property of the function and the available computational power [1].

2. *Histogram-based approach:* The simplest non-parametric approach is the histogram-based approach. It involves two steps: (i) building the histogram based on available data typically under normal operating conditions and (ii) test the new observation data against the developed histogram. If the data does not belong to any of the bins of the histogram, then the observed data are judged to be an anomaly. The size of the bins plays a critical role in this approach. If the size of the bins is small, then many normal test instances will fall in empty or rare bins, which would lead to a high rate of false alarms. If the size of the bins is large, then many fault instances will fall in frequent bins. This will lead to high false negative rate. An optimum size for the bin is necessary to construct a suitable histogram which will maintain a low false alarm rate and a low false negative rate [22]. Accuracy of the histogram-based approach can be estimated using integrated square mean error [1].

3. *Nearest neighbour (NN) approach:* It is another technique which can be used to estimate the density function. It does not require a smoothing parameter. Instead, it requires a width parameter which sets the position of the data point in relation to other data points. Main disadvantage of this method is the large number of computations that are required [11, 18]. NN approach assumes the normal operating instances occurring in the dense neighbourhoods, whereas the anomalies occur far from their closest neighbourhoods [22]. This approach requires a similarity evaluation or a form of distance measure between two data points. This distance or similarity can be calculated in many different ways, for example, using Euclidean distance, Mahalanobis distance, Manhattan distance, cosine angle distance and so on. Distance measures are also used in many other tasks such as clustering (K-mean), distance-based outlier detection, classification (SVM) and several other machine learning techniques [28]. NN anomaly detection approach can be divided into two groups: (i) NN which uses the distance of a data instance to its kth NN as the anomaly and (ii) NN which computes the relative density of each data instance [28]. The basic NN approach is based on the assumption that the anomaly score of a data instance is defined as its distance to its kth NN in a given data set. NN approach based on relative density estimates the density of the neighbourhood of all data instances. A new data instance (observation) with low density is marked as anomaly and a data instance with high density is marked as normal [28]. The main disadvantage of the NN approach is the computational complexity which is $O(n^2)$. Although sampling techniques try to address the computational complexity associated with considering the NNs within a limited sample of the data set, they might end up in incorrect anomaly scores if the sample size is very limited [22]. The main advantage of this approach is that it does not require any assumptions about the distribution of the data.

4. *Wilcoxon–Mann–Whitney test:* The test is used to compare two groups of sample data. Wilcoxon–Mann–Whitney test is also called as Wilcoxon rank sum test [11, 29]. This is a hypothesis test on the two different samples. Main advantage is that the rank can be estimated in advance; hence, computational run time is small. In addition, noise effects are reduced using the rank instead of the raw data [11].

6 Machine learning approach

Although there is no explicit definition for machine learning, Samuel [30] has defined the machine learning as a field of study that gives computers the ability to learn without being explicitly programmed. Tom Mitchell defines the machine learning problem as a computer programme that learns from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improve with experience E [31]. Machine learning is an established approach in many different fields, such as speech recognition, computer vision (i.e. face, hand writing and object recognitions), information retrieval, robotics, medical diagnosis, financial prediction, target tracking, biological predictions and so on. There are mainly three types of learning approaches: (i) supervised learning (ii) unsupervised learning and (3) reinforcement learning. Machine learning approach can be used in PHM applications to learn the behaviours of the systems and make predictions based on the generated predictive models.

Since PHM problems can be formulated as classification or clustering techniques of machine learning approach, machine learning approach can be used to classify or cluster the data into different groups (i.e. healthy, anomaly and so on). Hence with the help of machine learning, new data can be classified into healthy or anomaly and then can be used to isolate the anomalies and faults. Further this information can be fed into prediction algorithm to predict the reaming useful lifetime of a system [11].

7 Supervised learning approach

If the algorithm is given the labelled outputs for a set of input, then the learning is called supervised learning. Its goal is to predict a correct output for a new input data. Most of the PHM problems can be treated as supervised learning problems where set of healthy and failure data are available. Some of the frequently used supervised learning techniques are discussed below:

1. *Neural networks*: Many data-driven PHM approaches are based on artificial neural networks [8, 32, 33]. Neural network is a graph based on some interconnected numerical values attached to each node. It has also a set of input nodes, output nodes and hidden layers. Neural networks are trained for a set of training data to optimise the network parameters to obtain the desired output. This can be achieved by minimising the output error. For the PHM application, neural network can be used as a statistical modelling and prediction algorithm which can be achieved in two different ways: (i) density estimation and prediction and (ii) classification and regression. For a statistical modelling and prediction problem, a neural network can be trained to produce a statistical model which can be used to predict the output for a new input data. Density estimation is achieved by modelling the unconditional distribution of the training data. In the case of an input vector X, the neural network is trained to model the density function $p(X)$. On the basis of the labelled target variable threshold, value of the probability of anomalies will be determined. Classification is achieved by classifying the input data into different groups based on the output classes. In the case of an input vector X, the neural network classifies the input vector into one of the C classes $C_1, C_2, …, C_n$ represented by the labels of the output variable [34].

For example, in the case of PHM application the labels of the output variable can be healthy, anomaly and so on. Then the regression can be used to extrapolate the damage or failure precursor to estimate the RUL of a system. Main advantage of the neural network is that a very small number of parameters need to be
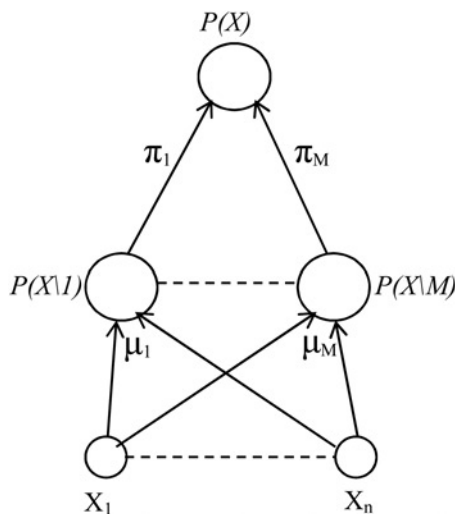
optimised for training networks and no prior assumptions on the property of the data are made. There are many different types of architectures available for neural networks such as multi-layer perceptron networks, self-organised maps (SOMs), RBF networks, SVMs, Hopfield networks, oscillatory networks and so on [18]. Fig. 3 illustrates a neural network for density estimation. Fig. 4 shows a neural network for classification problem.

2. *SVMs*: SVM was introduced by Vapnik in 1998. SVM is described as a function estimation problem for a given set of measurement data with noise. The idea behind this approach is to map the low-dimensional data (input space vector X) into high-dimensional vectors of the features space (feature space Z) such that the input vectors can be grouped based on the label of the target variable by an optimal unique hyper-plane [20]. Initially, SVM was applied for pattern recognition problems but became a popular approach in many different fields because of its performance. SVM has been applied to anomaly detection problems. A set of normal data is used to learn a region using kernel functions. This region can be defined as a normal operating region. If the new observation data belongs to the normal region, it will be flagged as normal and as anomaly otherwise [22].

For a set of training data $\{x_i, y_i\}$, $i = 0, 1, …, n$ $y_i \in \{-1, 1\}$, $x_i \in R^d$, there are some hyper-planes which separate the positive (1) from the negative (−1) training data. Fig. 5 illustrates some of the hyper-planes which can be used to separate two classes of the sample data. Shortest distances to the closest negative and closest positive points from the hyper-plane are d_- and d_+, and these distances are defined as the margin of separating the hyper-plane. In the case of linearly separable, the hyper-plane with the largest margin will be selected by the SVM. The closest points from the hyper-plane are called the support vectors [35]. Fig. 6 shows the hyper-plane with the largest margin and the support vectors.

Relevance vector machine (RVM) is a Bayesian treatment model of identical functional form of the SVM. RVM overcomes a number of practical disadvantages associated with the SVM. In addition, RVM uses dramatically fewer kernel functions while still showing a performance comparable with the equivalent SVM [8, 36].

3. *Gaussian process (GP) regression:* A GP is a collection of random variables, any finite number of which has (consistent) joint Gaussian distributions [37]. GP can be used with more flexibility for the non-linear regression problem [38].

4. *Bayesian networks (BN):* BNs is a directed acyclic graph which represents the joint probability distribution of the variables [39]. It is a directed graph which does not have any closed paths within the graph, that is, paths giving connection back to the starting node. In the graph of Fig. 7 there are three nodes which represent three variables. Node c has two parent nodes (i.e. a, b), node b has only one parent (i.e. a) and node 'a' does not have any parent nodes.
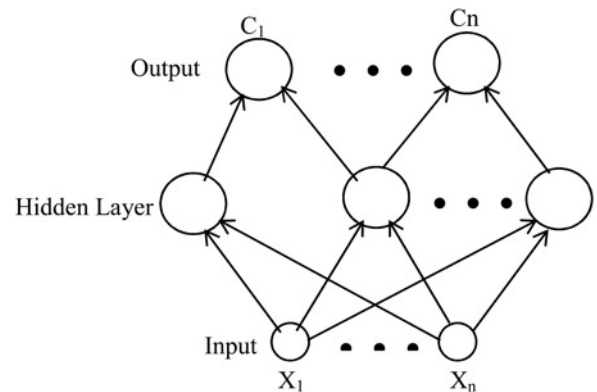

Fig. 3 *Neural network for density estimation problem based on Gaussian mixture distribution*

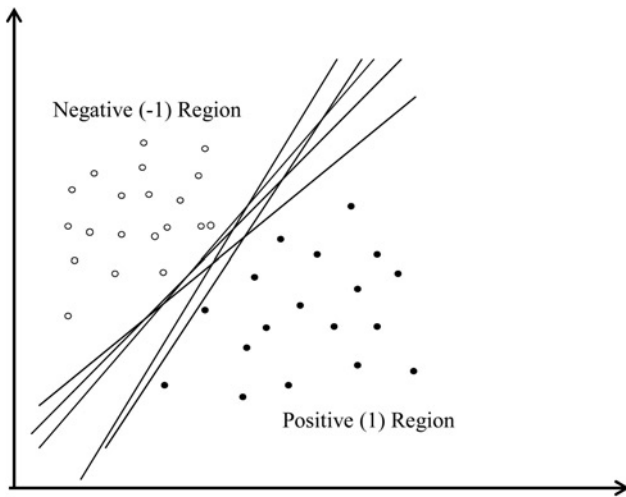**Fig. 4** *Neural network for classification problem*

Fig. 5 *Linearly separable sample data with some hyper-planes*

Joint distribution for the above BN can be formulated using the product rule of probability [39]. The equation for the BN is the factorisation property of the joint distribution. Generally, BN is used to estimate the conditional probability of one node, given values for other nodes. Since BN is used to estimate the posterior probability of one node given the values for other nodes, BN can be used as a classifier. Nodes represent the data set attributes when BN learns from the data sets [40].

Naïve–Bayes (NB) classifier is a simple BN where the classification node represented by the parent node to all the other nodes and no other connections are allowed in the NB classifier. The main advantages of the NB classifier are that they are easy to construct and the respective classification process is very efficient [40]. A general NB network is shown in Fig. 8.

5. *Hidden Markov model (HMM):* Markov models (MM) assume the future predictions are independent of all but the most recent observation [39]. HMMs are one of MMs in which the latent variables are discrete. HMM is widely used to model the sequential data. Fig. 9 shows an HMM as a specific instance of the state space model. It can be viewed as a mixture model with component densities given by $p(x|z)$. The state of the latent variable depends on the state of the previous latent variable, and therefore $p(z_n|z_{n-1})$. Initial latent node is unique as it does not have a parent node, and therefore it has a marginal distribution $p(z_1)$. Another important distribution
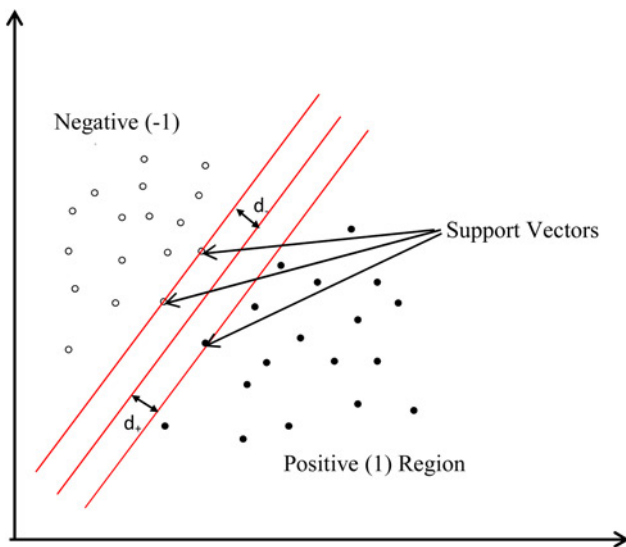

Fig. 6 *Linearly separable sample data with support vectors and hyperplane with largest margin*
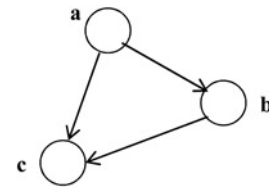

Fig. 7 *BN with three variables*

is the conditional distribution of the observed variables $p(x_n|z_n)$. Sometimes these are known as emission probabilities. This is a special case of BN called as dynamic BN.

8 Unsupervised learning approach

Unsupervised learning is used where there are no labelled data available (i.e. target variable). It is used to discover similar groups within the data based on clustering techniques or estimates the distribution of the data within the input space. It can be also used to map the high-dimensional input space into a low-dimensional space for the purpose of visualisation [39]. In the case of PHM applications, the unsupervised learning approach can be used to classify the data into different groups and identify the healthy and normal data. For most new systems, only normal operating data will be available and these data can be used to learn recognising a healthy system under different settings. Then this learned information can be used to detect the anomalies in the new observations and to predict the reliability and remaining lifetime. Some of the supervised learning approaches can also be used under unsupervised setting. Some of the frequently used techniques for unsupervised learning approach are listed below:

1. *Principal component analysis (PCA):* PCA is a widely used method for dimensionality reduction, data compression, feature extraction and data visualisation via mapping the data into a lower-dimensional linear space also called principal subspace. The goal of this approach is to map the higher-dimensional data into a lower dimension while maximising the variance of the mapped data. Alternatively, PCA can be performed by minimising the sum-of-squares of the projection errors [39]. Fig. 10 illustrates an example of mapping two-dimensional (2D) data into a 1D data.

2. K-means clustering: *K*-means clustering refers to grouping the data into *K* number of clusters such that the inter-point distance are small compared with the distance to the points outside of the cluster. Every cluster centre is given with a centre point μ_k where
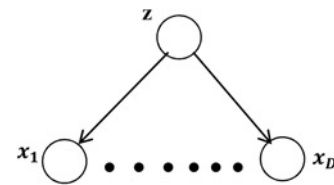
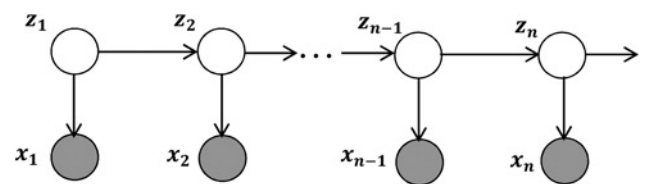
Fig. 8 *Graphical representation of NB classifier*

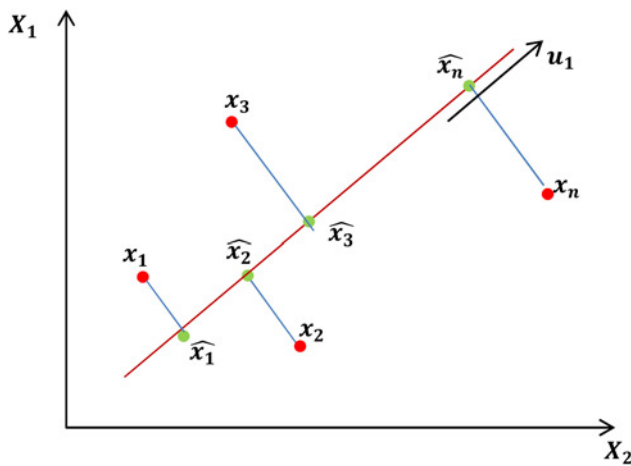**Fig. 9** *Graphical structure of HMM*

Fig. 10 *PCA for 2D data*

$k = 1, 2, …, K$. Data point needs to be assigned to these clusters such that the sum of the squares of the distances of each data point to its closest centre is the minimum one [39].

3. *Neural networks:* Neutral network can also be used in the un-supervised setting where labelled data are not available. Self-organising maps (SOMs) are the type of neural networks used for unsupervised learning. It is an alternative approach to the statistical clustering. In most of the SOMs, every cluster is iden-tified by a threshold value and based on these thresholds data points are assigned to the particular cluster [18].

4. *Kalman filters (KFs):* Kalman [41] proposed a technique to solve the problems such as: (i) prediction of random signal, (ii) sep-aration of random signal from the random noise and (iii) detection of signals of known form (i.e. pulses, sinusoids and so on) in the presence of random noise. The KF is based on the assumption that the posterior density at every time step is Gaussian, and hence parameterised by the mean and covariance [42]. The KF is frequently used as an optimised estimation technique for systems state. It is a recursive approach to estimate the systems state based on the prior knowledge of the state of the system and the mea-sured information. The KF is also used to fuse the measurements for same variable from different sensors. The KF is used in PHM appli-cation of electrical components based on changes in resistance [43]. There are different versions of KF available.

5. *Particle filter (PF):* PFs also referred as sequential Monte Carlo are used to handle model non-linearity or non-GP or observation noise [44]. PF was developed based on the concept of sequential important sampling and the Bayesian theory. PF has been applied in many fields, such as economics, biostatistics, target tracking, time series analysis, signal processing and so on [42]. There are many different PFs based on different sampling techniques, such as sampling important resampling PF, auxiliary PF, regularised PF and so on PFs have been applied successfully in number of PHM applications [44–49].

9 Conclusion

A typical PHM application consists of many different tasks from sensing to prediction. Each task benefited from different techniques; hence, the real-world PHM application does not necessarily depend on a single approach. Filtering techniques such as KF and PF can also be used to sequentially estimate the system state based on a model and sensor data. In particular, they are capable of correcting the predictions based on their outer feedback correction loops. PF demonstrated its robustness in online (real-time) estimation of the RUL of a system [48].

The fusion approach is based on the advance features of both the data-driven and model-based approach. This approach will require an accurate mathematical model of the system for physics-based failure approach and enough historical data and knowledge of typical operational performance data for data-driven approach. The aim of the fusion approach is to overcome the limitations of both the model and data-driven approach to estimate the remaining useful life (RUL) [11]. Therefore the accuracy of the fusion ap-proach should be high [11], although for real-time analysis it may not be suitable because of the computational resource required.

This paper discussed different algorithms and mathematical models under different data-driven PHM approaches. These approaches and algorithms have their own advantages and disad-vantages depending on the application, availability of the historical data, system specific knowledge, programmability and so on PHM applications also have many different individual processes such as noise reduction, anomaly detection, fault isolation and monitoring, state estimation, lifetime prediction and so on All these processes may require different approaches and different algorithms. Hence, the selection of the approach and algorithm for each process of a PHM application plays a key role and deciding factor of the accur-acy of overall PHM methodology. One has to investigate the selec-tion process properly to come up with a successful PHM application.

10 References

[1] Vichare N.M., Rodgers P., Pecht M.G.: 'Methods for binning density estimation of load parameters for prognostics and health manage-ment', *Int. J. Performability Eng.*, 2006, **2**, (2), pp. 149–161

[2] Tang L., Kacprzynski G.J., Goebel K., Saxena A., Saha B., Vachtsevanos G.: 'Prognostics-enhanced automated contingency management for advanced autonomous systems'. Proc. of the Annual Conf. of the Prognostics and Health Management Society (PHM 2008), Denver, 6–9 October 2008, pp. 1–9

[3] Pecht M.G., Jaai R.: 'A prognostics and health management roadmap for information and electronics-rich systems', *Microelectron. Reliab.*, 2010, **50**, (3), pp. 317–323

[4] Sotiris V.A., Tse P.W., Pecht M.G.: 'Anomaly detection through a Bayesian support vector machine', *IEEE Trans. Reliab.*, 2010, **59**, (2), pp. 277–286

[5] Alam M.A., Azarian M.H., Osterman M., Pecht M.: 'Prognostics of failures in embedded planar capacitors using model-based and data-driven approaches', *J. Intell. Mater. Syst. Struct.*, 2011, **22**, (12), pp. 1293–1304

[6] Byington C.S., Watson M., Edwards D.: 'Data-driven neural network methodology to remaining life predictions for aircraft actuator com-ponents'. Proc. of the IEEE Aerospace Conf., Big Sky, 6–13 March 2004, vol. 6, pp. 3581–3589

[7] Celaya J.R., Saxena A., Wysocki P., Saha S., Goebel K.: 'Towards prognostics of power MOSFETs: accelerated aging and precursors of failure'. Proc. of the Annual Conf. of the Prognostics and Health Management Society (PHM 2011), Montreal, 25–29 September 2011, pp. 1–10

[8] Goebel K., Saha B., Saxena A.: 'A comparison of three data-driven techniques for prognostics'. Proc. of the 62nd Meeting of the Society for Machinery Failure Prevention Technology (MFPT), Virginia Beach, 6–8 May 2008, pp. 119–131

[9] Ling B., Khonsari M., Hathaway R.: 'Data-driven roller bearing diag-nosis using degree of randomness and Laplace test'. Proc. of the Annual Conf. of the Prognostics and Health Management Society (PHM 2009), San Diego, 27 September–01 October 2009, pp. 1–8

[10] Pecht M., Kumar S.: 'Data analysis approach for system reliability, diagnostics, and prognostics'. Proc. of the Pan Pacific Microelectronics Symp., Hawaii, 22–24 January 2008, pp. 1–9

[11] Pecht M.G.: 'Prognostics and health management of electronics' (John Wiley & Sons, New York, 2008)

[12] Sankavaram C., Pattipati B., Kodali A., ET AL.: 'Model-based and data-driven prognosis of automotive and electronic systems'. Proc. of the IEEE Int. Conf. on Automation Science and Engineering (CASE 2009), Bangalore, 22–25 August 2009, pp. 96–101

[13] Saxena A., Goebel K., Simon D., Eklund N.: 'Damage propagation modelling for aircraft engine run-to-failure simulation'. Proc. of the

Annual Conf. of the Prognostics and Health Management Society (PHM 2008), Denver, 6–9 October 2008, pp. 1–9

[14] Saha B., Goebel K.: 'Prognostics HIL testbed'. Proc. of the Aviation Safety IVHM Posters, Virginia, 17–19 November 2009, pp. 1–1

[15] Saha B., Goebel K.: 'Model adaptation for prognostics in a particle filtering framework', *Int. J. Prognostics Health Manage.*, 2011, **6**, (6), pp. 1–10

[16] Celaya J.R., Saxena A., Saha S., Goebel K.F.: 'Prognostics of power MOSFETs under thermal stress accelerated aging using data-driven and model-based methodologies'. Proc. of the Annual Conf. of the Prognostics and Health Management Society (PHM 2011), Montreal, 25–29 September 2011, pp. 1–10

[17] Celaya J.R., Kulkarni C., Biswas G., Saha S., Goebel K.: 'A model-based prognostics methodology for electrolytic capacitors based on electrical overstress accelerated aging'. Proc. of the Annual Conf. of the Prognostics and Health Management Society (PHM 2011), Montreal, 25–29 September 2011, pp. 1–9

[18] Markou M., Singh S.: 'Novelty detection: a review – part 1: statistical approaches', *Signal Process.*, 2003, **83**, (12), pp. 2481–2497

[19] Clifton D.A., Clifton L.A., Bannister P.R., Tarassenko L.: 'Automated novelty detection in industrial systems', *Stud. Comput. Intell.*, 2008, **116**, (1), pp. 269–296

[20] Vapnik V.N.: 'Statistical learning theory' (Wiley-Interscience, 1998)

[21] Bishop C.M., Tipping M.E.: 'Vibrational relevance vector machines'. Proc. of the 16th Conf. on Uncertainty in Artificial Intelligence, Stanford, 30 June–3 July 2000, pp. 46–53

[22] Chandola V., Banerjee A., Kumar V.: 'Anomaly detection: a survey', *ACM Comput. Surv.*, 2009, **41**, (3), pp. 15:1–58

[23] Box G.E.P., Andersen S.L.: 'Permutation theory in the derivation of robust criteria and the study of departures from assumption', *J. R. Stat. Soc.*, 1955, **17**, pp. 1–34

[24] Lee H.-J., Roberts S.J.: 'On-line novelty detection using the Kalman filter and extreme value theory'. Proc. of the 19th Int. Conf. on Pattern Recognition (ICPR 2008), FL, 8–11 December 2008, pp. 1–4

[25] Roberts S.J.: 'Novelty detection using extreme value statistics', *IEE J. Vis. Image Signal Process.*, 1999, **146**, (3), pp. 124–129

[26] Awate S.P.: 'Adaptive, nonparametric Markov models and information-theoretic methods for image restoration and segmentation'. PhD dissertation, University of Utah, Salt Lake City, 2006

[27] Yeung D.-Y., Chow C.: 'Parzen–window network intrusion detectors'. Proc. of the 16th Int. Conf. on Pattern Recognition, 2002, vol. 4, pp. 385–388

[28] Boriah S., Chandola V., Kumar V.: 'Similarity measures for categorical data: a comparative evaluation'. Proc. of the Eighth SIAM Int. Conf. on Data Mining, GA, 2–26 April 2008, pp. 243–254

[29] Tang Y.: 'Size and power estimation for the Wilcoxon–Mann–Whitney test for ordered categorical data', *Stat. Med.*, 2011, **30**, (29), pp. 3461–3470

[30] Samuel A.L.: 'Some studies in machine learning using the game of checkers. II – recent progress', *IBM J. Res. Dev.*, 1967, **11**, (6), pp. 601–617

[31] Blum A., Mitchell T.: 'Combining labeled and unlabeled data with co-training'. Proc. of the 11th Annual Conf. on Computational Learning Theory (COLT 98), Madisson, 24–26 July 1998, pp. 92–100

[32] Schwabacher M.: 'A survey of data-driven prognostics'. Proc. of the AIAA Infotech@Aerospace Conf., Virginia, 26–29 September 2005, pp. 1–5

[33] Schwabacher M., Goebel K.: 'A survey of artificial intelligence for prognostics'. Proc. of the AAAI Fall Symp., 9–11 November 2007, pp. 1–8

[34] Jordan M.I., Bishop C.M.: 'Neural networks', in Tucker, A. (Ed.): 'CRC handbook of computer science' (CRC Press, Boca Raton, FL, 1996), pp. 536–556

[35] Burges C.J.C.: 'A tutorial on support vector machine for pattern recognition', *Data Min. Knowl. Discov.*, 1998, **2**, (2), pp. 121–167

[36] Tipping M.E.: 'Sparse Bayesian learning and the relevance vector machine', *J. Mach. Learn. Res.*, 2001, **1**, pp. 211–244

[37] Rasmussen C.E., Williams C.K.I.: 'Gaussian processes for machine learning' (The MIT Press, MA, 2006)

[38] Williams C.K.I.: 'Prediction with Gaussian processes: from linear regression to linear prediction and beyond'. Learning and Inference in Graphical Models, Kluwer, Netherlands, 1998, pp. 599–621

[39] Bishop C.M.: 'Pattern recognition and machine learning' (Springer, New York, 2007)

[40] Cheng J., Greiner R.: 'Learning Bayesian belief network classifiers: algorithms and system'. Proc. of the 14th Biennial Conf. of the Canadian Society on Computational Studies of Intelligence: Advances in Artificial Intelligence, Ottawa, 7–9 June 2001, pp. 141–151

[41] Kalman R.E.: 'A new approach to linear filtering and prediction problem', *Trans. ASME, J. Basic Eng.*, 1960, **82**, pp. 35–45

[42] Arulampalam M.S., Maskell S., Gordon N., Clapp T.: 'A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking', *IEEE Trans. Signal Process.*, 2002, **50**, (2), pp. 174–188

[43] Lall P., Lowe R., Goebel K.: 'Prognostics using Kalman-filter models and metrics for risk assessment in BGAs under shock and vibration loads'. Proc. of the 60th Electronic Components and Technology Conf. (ECTC), Las Vegas, 1–4 June 2010, pp. 889–901

[44] Orchard M., Wu B., Vachtsevanos G.: 'A particle filter framework for failure prognosis'. Proc. of the World Tribology Congress III (WTC2005), Washington DC, 12–16 September 2005, pp. 1–2

[45] Orchard M.E.: 'A particle filtering-based framework for on-line fault diagnosis and failure prognosis'. PhD dissertation, Georgia Institute of Technology, Atlanta, 2007

[46] Orchard M.E., Vachtsevanos G.J.: 'A particle filtering-based framework for real-time fault diagnosis and failure prognosis in a turbine engine'. Proc. of the 15th Mediterranean Conf. on Control & Automation, Athens, 27–29 July 2007, pp. 1–6

[47] Orchard M., Kacprzynski G., Goebel K., Saha B., Vachtsevanos G.: 'Advances in uncertainty representation and management for particle filtering applied to prognostics'. Proc. of the Int. Conf. on Prognostics and Health Management (PHM 2008), Denver, 6–9 October 2008, pp. 1–6

[48] Orchard M.E., Vachtsevanos G.J.: 'A particle filtering approach for online fault diagnosis and failure prognosis', *Trans. Inst. Meas. Control*, 2009, **31**, (3–4), pp. 221–246

[49] Butler S., Ringwood J.: 'Particle filters for remaining useful life estimation of abatement equipment used in semiconductor manufacturing'. Proc. of the 2010 Conf. on Control and Fault Tolerant Systems, Nice, 6–8 October 2010, pp. 436–441