Hidden Markov Model-Based Methods in Condition Monitoring of Machinery Systems

BY

Omid Geramifard

B. Sc., Isfahan University of Technology

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2013

To my dear parents,

Vaji & Hadi,

for their everlasting love and support.

To my lovely wife,

Maryam,

whose presence lights me up and lifts up my spirit.

Declaration

I hereby declare that the thesis is my original work and it is written by me in its entirety. I have duly acknowledged all the sources of information which has been used in this thesis.

This thesis has also not been submitted for any degree in any other university previously.

Omid Geramifard

16 January 2013

Acknowledgment

First and foremost, I would like to express my deepest appreciation to my supervisor, Professor Jian-Xin Xu for his invaluable guidance, patience and support in all aspects of this research. The enthusiasm he has for research, was greatly motivational for me during my Ph.D. pursuit. I am also grateful for the excellent example he has personified as a mentor and professor.

I would sincerely thank my oral Qualification Examination committee members, A/Professor Loh Ai Poh, A/Professor Geok Soon Hong and A/Professor Xiang Cheng, for their kindness to review my report and give encouraging feedback.

I would also like to express my gratitude to Dr. Junhong Zhou, Dr. Xiang Li and Dr. Oon Peen Gan from Singapore Institute of Manufacturing Technology (SIMTech), who contributed immensely to this research by providing the experimental data and their insightful advices.

I am truly thankful of all my friends and labmates for their companionship and support throughout my Ph.D. journey; especially Deng Xin, Sidath R. Liyanage, Ren Qinyuan, Zhaoqin Guo, Niu Xuelei, Deqing Huang, Yang Yue, Ramesh Bharath, Ehsan Keikha, and Yohanes Daud. Also, I am very thankful to lab officers at Control and Simulation lab, Zhang Hengwei and Aruchunan Sarasupathi as well as all the staff members at Department of Electrical and Computer Engineering and National University of Singapore for their kind support.

I would also like to specially thank my beloved wife Maryam Azh, my wonderful parents Vajiheh and Hadi, and my siblings Ordin, Golnar and Negar for their eternal love, support and encouragement; and my parents in-law, Parvin and Bahram for their understanding and support.

Lastly, I gratefully acknowledge the funding sources that made my Ph.D. work possible. My work has been supported by Singapore International Graduate Award (SINGA), funded by Singapore Agency of Science, Technology and Research (A*STAR).

Contents

Summary

Condition based maintenance (CBM) has become one of the main industrial challenges in the last decade. An early maintenance would reduce the efficiency of the production mainly by increasing the downtime of the machine, and a late maintenance would damage the quality of the production. Therefore, the goal of CBM is to do the maintenance whenever it is required. Early fault detection and diagnosis can help to increase the availability of the industrial machines and reduce the economical loss pertaining to the maintenance of the machinery systems. As the name of condition based maintenance implies the decision of maintenance in this system is based on the condition and the subsystem performing the condition monitoring is usually named tool condition monitoring (TCM) in the literature. This subsystem is responsible of assessing the health status of machinery system components and pieces based on direct or indirect acquired signals. However, direct methods are not usually favored as they involve stoppage of production for measurements contradicting with the goal of CBM. In the indirect TCM, using extracted features from non-intrusively sensed signals such as force, vibration, or acoustic emission, the health status of the tools are estimated.

The prediction process of health status can be dichotomized into diagnostics and prognostics. Diagnostics is to predict the current health status based on the data gathered from beginning of the task up to the current moment. Prognostics is to predict the future health status based on the data gathered from beginning till present. On the other hand, based on whether the predicted metric is continuous or discrete, the approaches can be divided into regression and classification. In this thesis, as the prediction approaches for the continuous tool condition monitoring were scarce yet important, the major focus is on this type of prediction. The developed continuous TCM approaches are evaluated based on the tool wear monitoring experimental data provided by Singapore Institute of Manufacturing Technology. Moreover, a semi-nonparametric temporal approach is also proposed for the fault detection and diagnostics (classification) in the rotary electric

motors and evaluated on the common faults in a synchronous motor.

In Chapter 1, the motivation of the research, relativeness of the research area to other prediction and forecasting areas and a literature review on the existing works of leading researchers in the field is introduced. Furthermore, the importance of temporal information in acquiring accurate predictions is highlighted and hidden Markov model (HMM) as a probabilistic model that can capture the temporal information in the sequential observations is briefed. In Chapter 2, a temporal probabilistic approach based on HMM is proposed to perform continuous tool wear monitoring. In Chapter 3, a more complex model called hidden semi-Markov model is then applied to improve the performance further and to study the tunability of the model based on a given loss function that may indicate the cost (loss) difference between an under- and over- estimation. Then in Chapter 4, a multi-modal HMM-based approach is proposed to improve the performance of the single HMM-based approach introduced in Chapter 2. Moreover, three weighting schemes and two switching strategies are proposed and compared along with the single HMM-based approach as benchmark. Chapter 5 studies the possible improvement of HMM-based fault detection and diagnosis (classification) using a semi-nonparametric approach. As the true model is usually not realizable for real world applications, it is attempted to increase the accuracy of the classification by using the training data more effectively. Finally, Chapter 6 summarizes the contributions of this thesis and gives possible directions for future work in this area.

Nomenclature

List of Notations and Abbreviations

Notation	Description
$P(.\|.)$	Conditional probability.
$O_{1:T}$	Observation sequence from time step 1 to T where observation at each time step is a vector.
$S_{1:T}$	State sequence from time step 1 to T.
A	Transition probability matrix.
$a_{i,j}$	probability of transition from ith state to jth.
π_0	Initial transition probability vector.
p_i	Self transition probability.
B	Emission probability matrix in discrete HMM.
λ	Parameter set.
D_i	ith data sequence (experiment).
m	number of hidden state values.
n	number of data sequences (experiments).
μ	Mean vector in multi-variate Gaussian distribution.
Σ	Covariance matrix in multi-variate Gaussian distribution.
χ	Dimensionality of the observation vector.
H_i	Continuous label of the ith health state.

Notation	Description
k_c^j	Number of samples belonging to the cth health state in the jth data sequence.
t_i^j	Starting time of the ith health state in the jth sequence.
$\alpha_t(.)$	Forward variable vector at time t.
$\beta_t(.)$	Backward variable vector at time t.
$\gamma_t(i)$	Joint probability of observing all input features through up to current time T while $S_t = H_i$ $(t \leq T)$.
$\xi_{t'}(i)$	Joint probability of observing all input features through up to current time T while $S_{t'} = H_i$ $(t' > T)$.
$\hat{y}_T$	expected tool wear at current time T.
y_t^j	Actual tool wear at time t in the jth experiment.
$x_i^j(t)$	value of the ith feature at time t in the jth experiment.
Sb	Scatter between.
Sw	Scatter within.
d_{max}	Maximum duration.
μ_{d_i}	Mean of ith health state duration distribution.
σ_{d_i}	Standard deviation of ith health state duration distribution.
(S_t, τ_t)	Pair of hidden state of the model at time step t and its remaining duration τ_t at that time step onward.
$\alpha_t(.,.)$	Forward variable in hidden semi-Markov model.
$\beta_t(.,.)$	Backward variable in hidden semi-Markov model.
$\zeta_t(i)$	Joint probability of observing $O_1 : T$ and transition from ith health state to the next health state at time t.
$\xi_{t'}(i,k)$	Joint probability of observing all the input features up to the current time T while $(S_{t'}, \tau_{t'}) = (i,k)$ where $t' > T$.
$loss(.)$	Loss function.
ρ	Asymmetry factor in the asymmetric Gaussian distribution.

Notation	Description
φ	General left side percentage factor in the asymmetric Gaussian distribution.
$\bar{d}_i$	Duration at the peak in the ith asymmetric duration distribution function.
N_{mode}	Number of modes considered in the multi-modal approach.
λ_i	Parameter set of the ith mode in the multi-modal approach.
λ'_T	Updated parameter set to be used at time step T.
$\pi'_{0,T}$	Updated initial state probability to be used at time step T.
L_w	Window length for the windowing algorithm.
$v_t(i)$	Highest probability obtained obtained by a single path up to time t that ends in state H_i.
$V_{1:T}$	Viterbi-path taken from time step 1 to T.
W_T^i	Weightage of the ith mode for the ultimate output calculation.
Δ	Bounded hindsight window length.
ϕ_t	Discount factor at time t in Discounted hindsight weighting scheme.
$I_q(.,.,.)$	Set of (starting,ending) time index pairs.
R^h	Observation segment in the reference sequence that its most likely health state index is h based on Viterbi-path.
$dist(.,.)$	Aligned distance between two matrices.
$Score(.,.,.,.)$	Score function.
v_i	ith discrete state value that hidden state variable can take in HMM.
F	Probabilistic transition frequency profile matrix with $f_{i,j}$ elements.
E	Average probabilistic emission matrix with $e_{i,j}$ elements.
$\delta(.,.)$	Similarity measure for two matrices.
$G(.\|.)$	similarity scoring function.
$Q_i(.)$	ith class similarity score computed for the given signature.
$C(.)$	Classification output of the HMMSNP approach.

Abbreviation	Description
CBM	Condition based maintenance.
TCM	Tool condition monitoring.
HMM	Hidden Markov model.
HSMM	Hidden semi-Markov model.
CNC	Computer numerically controlled.
REM	Rotary electric motor.
FDD	Fault detection and diagnosis.
PSHMCO	Physically segmented hidden Markov model with continuous output.
FDR	Fisher's discriminant ratio.
GMM	Gaussian mixture model.
BIC	Bayesian information criteria.
MLP	Multi-layer perceptron.
MSE	Mean squared error.
MRE	Mean relative error.
PSHsMCO	Physically segmented hidden semi-Markov model with continuous output.
CV	Cross-validation.
GLSP	General left side percentage.
m^2HMM	Multi-modal hidden Markov model.
SNPH	Semi-nonparametric hindsight.
DH	Discounted hindsight.
BH	Bounded hindsight.
TT	Training-testing ratio.
BRG	Bearing fault condition.
UBR	Unbalanced rotor bar condition.
HTY	Healthy condition.
SqS	Squeezing and stretching.
APPA	Average peak-to-peak amplitude.
PTFP	Probabilistic transition frequency profile.
APE	Average probabilistic emission.
HMMSNP	Hidden Markov model-based semi-nonparametric approach.
HMMSqS	Hidden Markov model with squeezing and stretching preprocessing.

List of Figures

List of Tables

Chapter 1

Introduction

As industrial machines started to grow more and more complex and sophisticated, their maintenance has become a major issue in the industry, therefore new methods have been developed to address this issue. The primarily developed maintenance approaches, were either *fault-driven* or *time-based*. In fault- driven approach, there wouldn't be any maintenance in the system till an apparent failure would occur which indicates this approach is reactive rather than being proactive. Furthermore, this strategy may cause a lot of physical and financial damage and it is not applicable to all machinery systems, specifically those in which the quality and precision of the product is greatly important. The other approach, which is time-based, is to do inspection and maintenance regularly and periodically. Although this strategy would increase the reliability of the machinery systems, it may lead to undesirable downtimes and unnecessary maintenance expenditures. Hence, the regular periodic maintenance should be advanced and shifted to the intelligent maintenance philosophy to satisfy the manufacturers' high reliability requirements. To address the disadvantages lying in both aforementioned approaches, the idea of a condition based approach was developed.

Condition Based Maintenance (CBM) has become one of the main industrial challenges in the last decade. An early maintenance would reduce the efficiency of the production mainly by increasing the downtime of the machine, and a late maintenance would damage the quality of the production. Therefore, the ultimate goal of CBM is to do the maintenance whenever it is required. As the industry grows, the importance

of fault detection and diagnostics in the machinery systems is also increasing. Early fault detection and diagnosis can help to increase the availability of the industrial machines and reduce the economical loss pertaining to the maintenance of the machinery systems [1]. As the name of condition based maintenance implies the decision of maintenance in this system is based on the condition and the subsystem performing the condition monitoring is usually named Tool Condition Monitoring (TCM) in the literature. This subsystem is responsible for assessing the health status of machinery system components and pieces based on either directly or indirectly acquired signals. However, direct methods are not usually favored as they involve stoppage of production for measurements, thus contradicting with the goal of CBM. TCM reduces the amount of unnecessary downtime for maintenance purposes, and consequently reduces the cost of maintenance [2, 3, 4, 6, 5]. Moreover, TCM improves the quality and precision of the product.

1.1 Background and Motivation of Research

In non-linear systems, acquiring perfect physical models may be a challenging task, as the interaction among various mechanisms such as electrical, mechanical, chemical, etc. and other properties of the system has to be completely comprehended. For example, in the tool wear progression, five wear mechanisms may be involved i.e. abrasion, adhesion, fatigue, dissolution, and tribo-chemical processes [7]. However, as stated in [8], it is difficult to predict their relative importance in various conditions. Thus, as a perfect physical model is not available in many real-world applications (such as tool wear monitoring), many researchers have focused on developing data-driven prediction approaches based on historical data. A survey on these approaches can be found in [9, 10]. Figure 1.1 schematizes components of a data-driven tool condition monitoring system.

A data-driven CBM system can be realized by integrating CBM's four essential components. These four components are as follows

1. Acquiring and collecting data in an indirect manner (non-intrusively) without

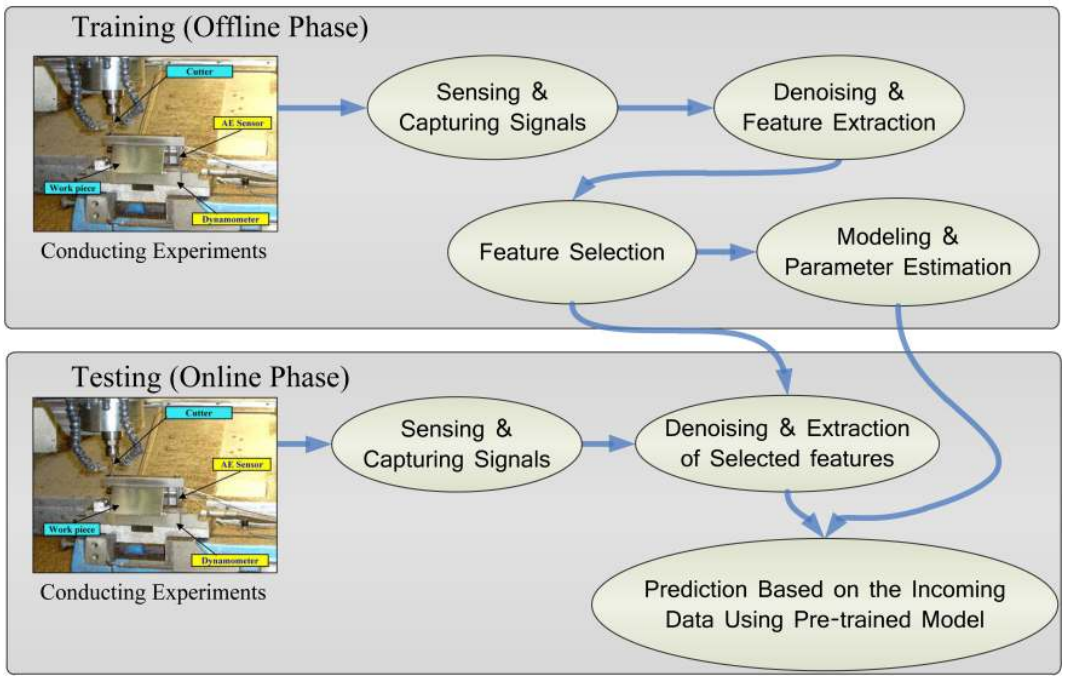


Figure 1.1: Components of a data-driven tool condition monitoring system.

causing machinery downtime (using sensors, etc.)

2. Preprocessing the acquired data as well as feature extraction and selection,

3. Modeling, condition monitoring (or fault detection and diagnosis),

4. Decision making.

The first three components materialize the TCM subsystem. After performing condition monitoring, assessing the health status of the components and providing the predicted health status for future time steps (remaining useful life), decision making can be performed by either experts (manually) or based on expert systems and automated decision making systems. In this research our focus is on the third component up to the decision making point where the outputs from the third component are provided either in continuous (e.g. tool wear monitoring) or discrete (e.g. fault detection) form.

Tool condition monitoring in a machinery system, means enabling a system to predict the health status (tool condition) in a machine based on the non-intrusively extracted features. The horizon of this prediction may be different depending on the application.

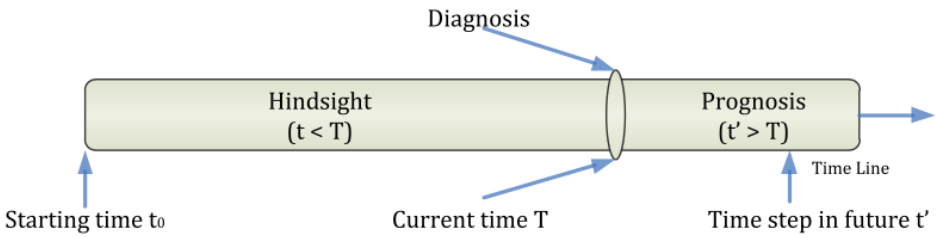


Figure 1.2: Schematizing the context of diagnosis, prognosis and hindsight. x-axis shows time line.

Basically, this prediction process is commonly dichotomized into two tasks, namely, *diagnostics* and *prognostics* [5, 6, 11, 12, 13, 14]. Figure 1.2 depicts the concept of these two tasks.

Diagnostics is to predict the current health status based on the data gathered from beginning of sampling up to the current moment. Prognostics is to predict the future health status based on the data gathered from beginning till present. Obviously, diagnostics is an easier task compared to prognostics and a good diagnostics algorithm is a necessary requirement and an initial step to achieve a sound prognostics algorithm. A survey on the diagnostics and prognostics approaches can be found in [9].

In [15], trend projection models are used, in which model parameters can be easily computed but may overfit the past degradation patterns. Fuzzy inference system (FIS)-based approaches are also extensively used in TCM [16, 17, 18, 19], which in general require a priori knowledge to be available when determining the rules and membership functions. The strategy exploiting fuzzy and neuro-fuzzy tools such as adaptive neuro fuzzy inference system (ANFIS) are also applied to TCM applications [20,21,22], which are data-driven and can be regarded as special classes of neural network methods. Artificial neural network (NN) is one of the most commonly used approaches in this domain. In [14, 23, 24, 25, 26, 27, 28, 29, 30], NNs are used in a time series prediction manner providing nonlinear projection without the need for prior knowledge. However their prediction horizon is short. Hidden Markov models (HMM) and hidden semi-Markov

models (HSMM) are used [31,32,33,34,35,36,37] to distinguish various wearing stages or machinery fault types.

Another way to categorize the prediction approaches is based on whether or not their predicted output is continuous. Consequently, the prediction approaches can be divided into regression (continuous output) and classification (discrete output) approaches. In this thesis, as the prediction approaches for the continuous tool condition monitoring were scarce yet important, the major focus is on this type of prediction approaches which are evaluated based on the experimental data. However, a semi-nonparametric temporal approach is also proposed for the fault detection and diagnosis (classification) in the rotary electric motors and evaluated on the common faults in a synchronous motor. As an illustrative example for the continuous TCM, tool wear monitoring in a computer numerically controlled (CNC)-milling machine is described, which has been used to evaluate the corresponding proposed approaches throughout this thesis. Here, the background on the Tool wear monitoring as well as fault detection and diagnosis in rotory electric motors are provided.

1.1.1 Tool Wear Monitoring

As the modern manufacturing industry develops, the question on how to improve the quality while reducing the production time-line and lowering its cost is more and more highlighted. Among various causes of poor production qualities, undetected amount of tool degradation and wearing that happens during the machining processes are one of the major issues. If the tool wear status would not be detected in time, it may lead to inefficient machining or destruction of the machine tool. Thus, it is necessary to perform accurate tool wear monitoring and integrate it as a part of CBM system.

As recognizing the accurate physical model of the tool wearing process turns to be infeasible in real-world applications, various researches are tended to data-driven approaches to perform tool wear monitoring. Many data-driven approaches are proposed so far for this purpose [40].

HMM is one of the commonly used approaches to perform tool wear monitoring for various machining processes such as grinding [41], milling [33, 42, 43, 44, 45],drilling

[37], turning [46, 47]. The cutting tool wear monitoring and prediction of useful life were modeled using hidden Markov model (HMM) and continuous HMM [33, 34, 35, 37, 36]. In all the existing HMM and HSMM-based approaches, the wearing data is discretized into several stages and then multiple HMMs are used to distinguish between those stages. Each HMM or HSMM is assigned to recognize one specific stage and an expectation-maximization method is utilized to estimate the parameters. However, training HMMs and HSMMs using expectation-maximization method is essentially a black-box approach, which does not provide explicit relationship between the wearing value and the hidden state values in the trained HMMs or HSMMs.

In most of the proposed approaches, tool wear monitoring is treated as a classification problem rather than regression. In contrast, in this thesis, tool wear monitoring is treated as a regression problem. The idea is to regularly assess the health status of the tool in the machinery system at each time step in terms of a continuous measure based on the past input data. In other words, instead of setting some thresholds and differentiating distinct health states as various (ordinal) classes, we would like to ultimately monitor the health state of the tool using a continuous measure. This allows us to have a smoother decision maker system for the condition based maintenance. It also enables us to incorporate different quality thresholds for different applications using the same condition based maintenance system e.g. to satisfy and guarantee different qualities in various products.

The continuous health state in each machinery system corresponds to different stages of deterioration. Tool wear monitoring in the cutting machinery systems is one of applications of continuous TCM. For example, in a milling machine, the features extracted from various signals such as force, vibration and acoustic emission are used as the inputs to predict the continuous wearing metric of the cutter [5, 6, 48].

In this thesis, tool wear monitoring in a CNC-milling machine is used as an illustrative example for continuous TCM. As indicated in [49], tool flank wear length or wear-land, is generally regarded as the tool wear criterion or an important index to evaluate the tool performance. Thus, it is adopted as the continuous tool wear indicator to be predicted in the CNC-milling machine.

1.1.2 Fault Detection and Diagnosis in Rotary Electric Motors

Rotary electric motors (REM) provide the basis for the electromechanical energy conversion in all industrial environments [39]. Thus, as the industry grows, the importance of fault detection and diagnosis (FDD) in the rotary electric motors is also increasing. Early fault detection and diagnosis can help to increase the availability of the industrial machines and reduce the economical loss pertaining to the maintenance of the machinery systems [1].

The task of FDD in REMs is to automatically detect the faulty condition from the healthy condition and furthermore recognize the specific type of fault such as bearing fault, unbalanced rotor bar, etc. that has occurred in order to reduce downtime and maintenance cost. The goal of automated FDDs is to detect the specific faults based on non-intrusively captured signals such as vibration and electrical signals over a wide range of operating speeds.

In the REMs, machine vibration arises due to action-reaction forces acting on the surface-to-surface contacts of moving machine parts. A healthy machine exhibits low level of vibrations. One the other hand, machine with bearing single-point defects and unbalanced rotor (possibly caused by breakage, wear and tear, accumulation of deposits, temperature changes, etc.) generates unique vibration signatures [51]. Among the common signatures analyzed during condition monitoring of REMs, vibration signature analysis seems to be the most responsive one [53, 54, 55]. Vibration is the most commonly measured signal used in monitoring machinery condition and an effective media for diagnosing mechanical faults [1, 51, 56, 57, 58, 59]. Another common signal that has been extensively used to diagnose faults in the motor is the motor current signal. Motor current signature analysis normally inspects the current spectrum for a specific fault spectrum [60, 61, 62, 63, 64, 65]. This requires a sufficiently high frequency resolution. Since the signature is non-stationary and non-linear, traditional Fourier transform using fast Fourier Transform (FFT) may not be able to capture the fault spectra, requiring other techniques, such as wavelet [60, 61, 62, 63, 64], high resolution techniques [66], or polynomial-phase transform [67]. Diagnosis methods using stator current by wavelet decomposition for bearing fault are reported in [64, 68, 69, 70]. Also in [71], wavelet

decomposition is applied on the inverter input current to identify the induction motor faults. However, the careful selection of wavelet is not trivial [72, 73]. In this thesis, instead of using motor current, vibration signatures are used, as it may be difficult to detect these faults using motor current signatures spectra especially under extreme low signal-to-noise ratio [74] and presence of varying load torque effect [60].

Hidden Markov models (HMM) are extensively used for fault detection and diagnosis in various rotary electrical motors [58, 75, 76, 77, 78, 79, 80, 81] as well as failure prognostics [82]. In all cases, the HMM-based approach is successful in distinguishing healthy condition from faulty conditions (fault detection). The challenging part is to diagnose the faults as the amplitude of the vibration signals from various faults may be similar between various operating speeds. That increases the chance of misclassification based on maximum likelihood strategy considering the fact that the true model is not practically realizable in real applications.

The most common fault in the REMs is bearing related faults which are responsible for about 50% of all rotary machine faults [50]. The second most common fault is the unbalanced rotor which causes excessive vibrations in the machines [51, 52]. Thus, as an illustrative example in this thesis, these two faults are tried to be classified along the healthy condition in a synchronous motor as one of the REM types that is widely used in all the industrial applications where constant speed is essential.

1.1.3 Necessity of Temporal Models for Diagnostics and Prognostics

Sequential data exist in every scientific and industrial domain. The sequential property of data in different domains is mainly imposed either by time (temporal sequential data) or space (spatial sequential data). Samples in the sequential training data, rather than being drawn independently and identically from a joint distribution of inputs (X) and outputs (y), consists of sequences of (X, y) pairs which have significant sequential correlations (patterns) [83]. That is, nearby X and y values are likely to be related to each other. These correlations and patterns are important because they can be exploited to improve the prediction accuracies in the utilized models. Therefore, the importance of capturing these temporal (spatial, or both) patterns, have become a major focus of

research in machine learning for various applications.

Various problems can be addressed and formulated as cases of sequential data analysis. Generally, these problems can be categorized as follows

- Time Series Prediction (e.g. stock market prediction [84], etc.)
- Sequential Supervised Learning (e.g. part of speech tagging [85], error control coding [86], DNA annotation [87], etc.)
- Sequence Classification or Labeling (e.g. hand-written identification [88], object-shape detection [89], sign language recognition [90], fault detection [91], etc.)
- Diagnostics & Prognostics (e.g. TCM in industrial machines [92])

These problems have similarities and differences in their specific formulation. Some of these similarities and differences between diagnostics & prognostics and the other three categories are listed in table 1.1.

Table 1.1: Some Similarities and differences between diagnostics & prognostics and the other sequential data analysis categories.

Category	Similarity	Difference
Time Series Prediction	The input data from beginning till current time is available.	y_{t+1} must be predicted while $y_{1:t}$ true values are available.
Sequential Supervised Learning	The true output values are not available.	The whole sequence is available.
Sequence Classification	The inputs are provided as sequential data similar to rest of categories.	Only one label must be predicted given the whole sequence.

Another aspect that sequential data analysis problems can be categorized based on, is the value of output that must be predicted whether it is continuous or discrete. On this basis, they can be categorized as regression or classification. By comparing the problem statements, it may be suggested that tool condition monitoring is more difficult than the other aforementioned problems specifically in case of continuous health assessment (Regression case such as continuous tool wear monitoring).

As mentioned earlier and it is stated in [83, 93], in order to achieve more accurate predictions, there is a necessity in capturing trends and modeling the sequential pat-

tern rather than treating the experimental data samples as if they are independently and identically distributed. Markov models and hidden Markov model can model the dependence (correlation) between the elements in a sequence [94]. Hidden Markov model (HMM) assumes that the system being modeled is a Markov process with unobserved (hidden) states. Since HMM satisfies the need to capture sequential patterns, in this thesis, HMM-based approaches are proposed and studied to fulfill the prediction tasks either for regression or classification.

1.1.4 Hidden Markov Model

Signal modeling methods can be broadly dichotomized into two classes, namely, deterministic models and statistical models [95]. Deterministic models exploit the known specific properties of the signal, for example when it is known that a signal is sinusoidal, then by identifying the amplitude, phase and frequency it can be modeled. In other type of the models classified as statistical models, which include Gaussian Processes, Markov Processes, and hidden Markov processes, only the statistical properties of the signals are characterized. The underlying assumption of the statistical model, is that the signal can be well characterized as a parametric random process, and that the parameters of the stochastic process can be determined (estimated) in a precise, well-defined manner [95].

Among the statistical models, hidden Markov model is one of the most popular models, since it is very rich in mathematical structure which helps researchers to form the theoretical basis required in different applications. In this Section, first the theory of discrete Markov chains is described and then it is shown how the concept of hidden states, where the observation is a probabilistic function of the state, can be used effectively.

Discrete Markov Process

Consider a system which may be described at any time as being in one state from the set of m distinct states $\{v_1, v_2, \ldots, v_m\}$. At regularly spaced discrete times, the system undergoes a change of state (self transition is also possible) according to a set of probabilities associated with the state. The time instants associated with the state changes are denoted as $t = 1, 2, \ldots$ and the actual state at time t is S_t.

A full probabilistic description of the aforementioned system in general requires specification of the current state at time t, as well as all the predecessor states [95]. However, for a special case of discrete first order Markov chain, this probabilistic description is truncated to just the current and the previous state as follows [96]

$$P(S_t = v_j | S_{t-1} = v_i, S_{t-2} = v_k, \dots) = P(S_t = v_j | S_{t-1} = v_i). \tag{1.1}$$

Furthermore, assuming stationarity, the right hand side of (1.1) is independent of time and thus leads to the state transition probabilities characterized as follows

$$A = [a_{i,j}]_{m\times m} = [P(S_t = v_j | S_{t-1} = v_i)]_{m\times m}, \quad 1 \le i, j \le m \tag{1.2}$$

with the state transition elements having the properties $a_{i,j} \ge 0$ and $\sum_{j=1}^{m} a_{i,j} = 1$ since they obey standard stochastic constraints.

The above stochastic process may be called an observable Markov model since the output of the process is the set of states at each time step, where each state corresponds to an observable event. As an example, consider a simple 3-state Markov model of the weather in Singapore. Assume that the weather once a day (e.g. at 12 pm) is observed as one and only one of the following states rainy, cloudy, or sunny that are denoted respectively as $\mathcal{R}$,$\mathcal{C}$, and $\mathcal{S}$. Figure 1.3 depicts the transition graph and gives the estimated transition probability matrix A in this example.

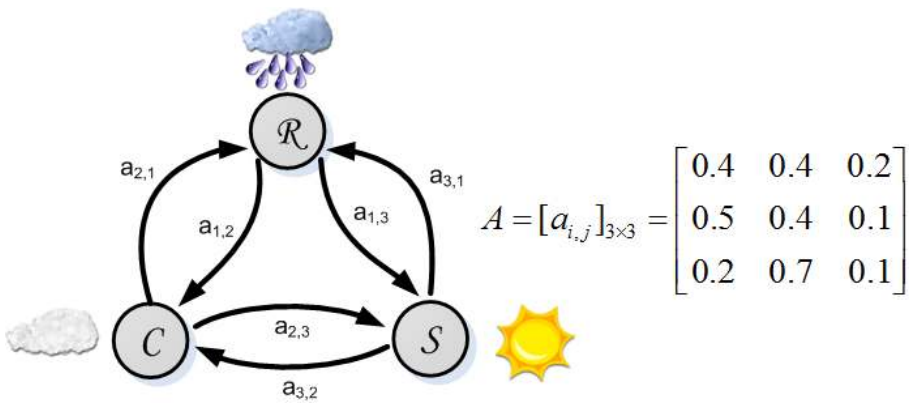


Figure 1.3: Illustrative 3-state discrete Markov Process for weather condition in Singapore.

Now as an example, we would like to calculate the probability (according to the

postulated model) that the weather sequence for the next 4 days will be "sunny, cloudy, cloudy, rainy", given that today's weather ($t = 0$) is rainy.

The corresponding observation sequence for $t = 1, 2, \ldots, 4$ can be defined as $O_{1:4} = \{\mathcal{S}, C, C, \mathcal{R}\}$, with initial state of $\mathcal{R}$.
Now given the model and the observation sequence, the probability $P(O_{1:4}|Model)$ can be computed as follows

$$\begin{aligned}
&P(O_{1:4}|Model) \\
&= P(\mathcal{R}) \times P(\mathcal{S}|\mathcal{R}) \times P(C|\mathcal{S}) \times P(C|C) \times P(\mathcal{R}|C) \\
&= \pi_0(1) \times a_{1,3} \times a_{3,2} \times a_{2,2} \times a_{2,1} \\
&= 1 \times 0.2 \times 0.7 \times 0.4 \times 0.5 = 0.028
\end{aligned}$$

where π_0 is the initial state probability distribution and $\pi_0(i)$ indicates the probability of initially being at ith state.

Hidden Markov Model

As mentioned, the described discrete Markov process may also be called observable Markov process as the states are observable at each time step. However, this may not be applicable to many real-world applications in which the actual physical states are not observable or hard to observe (hidden) and we may only have access to indirect observations that are probabilistic functions of those hidden physical states. Thus to address this issue in the applications, hidden Markov model may be utilized.

The hidden Markov model is a doubly stochastic process. This model has only one discrete hidden state variable, and a set of discrete or continuous observation nodes [96]. Fig. 1.4 depicts graphical model of HMM along with its transition graph. The basic theory of HMM was published in a series of papers by Baum and his colleagues in the late 1960s and early 1970s and was implemented for speech processing applications by Baker at CMU, and by Jelinek and his colleagues at IBM in the 1970s [95].

Here, the hidden Markov model is illustrated using a similar yet different weather condition example in Singapore. This time assume that there is a janitor in one of the buildings of Singapore (with no window and means of observing outside world) who never leaves the building and does not follow the weather reports. The only thing related

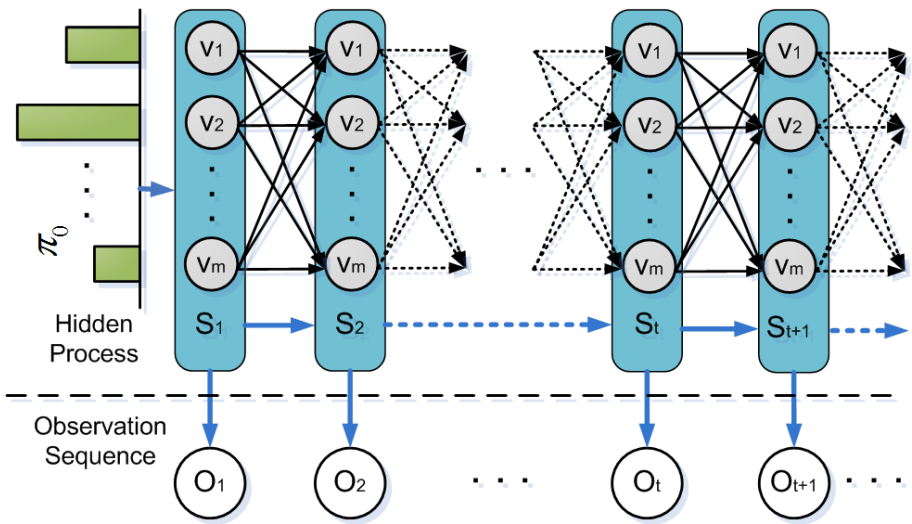


Figure 1.4: Graphical model of HMM including its transition graph.

to the weather condition that he observes is his boss who comes to the office at noon either carrying an umbrella or not. Thus, in this example, the observation is either boss carrying an umbrella or not $\{\mathcal{U}, \neg\mathcal{U}\}$ and the hidden state is the weather condition that can be rainy, cloudy or sunny, $\{\mathcal{R}, \mathcal{C}, \mathcal{S}\}$. Carrying an umbrella by the boss can be modeled as a probabilistic function of the weather condition (which is hidden to the janitor). The probabilistic function which connects the observations to the hidden state is named emission probability in the literature. Let's assume that the emission probabilities in this example are estimated to be as follows

$$B = [b_{i,j}] = \begin{bmatrix} P(\mathcal{U}|\mathcal{R}) & P(\neg\mathcal{U}|\mathcal{R}) \\ P(\mathcal{U}|\mathcal{C}) & P(\neg\mathcal{U}|\mathcal{C}) \\ P(\mathcal{U}|\mathcal{S}) & P(\neg\mathcal{U}|\mathcal{S}) \end{bmatrix} = \begin{bmatrix} 0.9 & 0.1 \\ 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix}$$

Now, the questions that we would like to answer is that firstly, what is the joint probability of having a specific sequence of hidden states while observing a set of observations (sequence of $\mathcal{U}$ and $\neg\mathcal{U}$), for that corresponding time line. Another question is that given a set of observations up to a point in time, what is the probability of being at ith hidden state for its next time step. These two questions are illustrated as follows.

Question 1.

Five days back the Janitor has asked his boss about the weather that day and his boss has replied "It is raining", thus $S_0 = \mathcal{R}$. Consequently, the initial state probability is

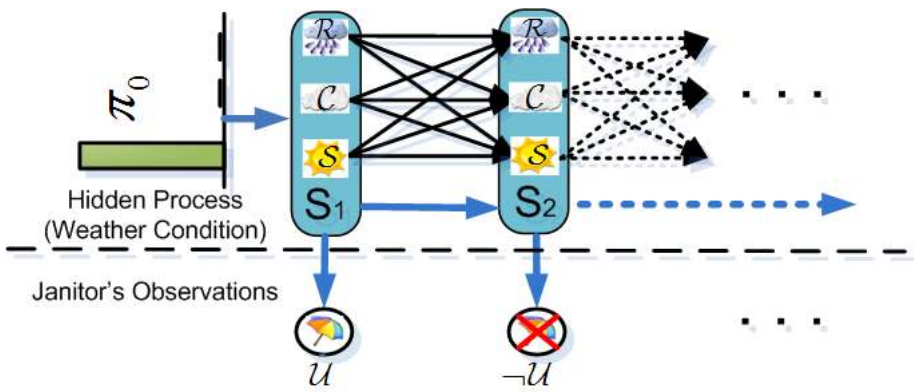


Figure 1.5: Illustrative graphical model of HMM for weather condition.

$\pi_0 = [1\ 0\ 0]^T$. Assuming that the Janitor's observations for the passed 4 days have been $O_{1:4} = \{\neg\mathcal{U}, \neg\mathcal{U}, \mathcal{U}, \mathcal{U}\}$ what is the joint probability that the weather condition for the passed four days has been $S_{1:4} = \{\mathcal{S}, \mathcal{C}, \mathcal{C}, \mathcal{R}\}$? This probability can be computed as follows

$$
\begin{aligned}
&P(O_{1:4}, S_{1:4}|Model) \\
&= P(S_0 = \mathcal{R}) \times P(S_1 = \mathcal{S}|S_0 = \mathcal{R}) \times P(O_1 = \neg\mathcal{U}|S_1 = \mathcal{S}) \\
&\times P(S_2 = \mathcal{C}|S_1 = \mathcal{S}) \times P(O_2 = \neg\mathcal{U}|S_2 = \mathcal{C}) \times P(S_3 = \mathcal{C}|S_2 = \mathcal{C}) \\
&\times P(O_3 = \mathcal{U}|S_3 = \mathcal{C}) \times P(S_4 = \mathcal{R}|S_3 = \mathcal{C}) \times P(O_4 = \mathcal{U}|S_4 = \mathcal{R}) \\
&= \pi_0(1) \times a_{1,3} \times b_{3,2} \times a_{3,2} \times b_{2,2} \times a_{2,2} \times b_{2,1} \times a_{2,1} \times b_{1,1} \\
&= 1 \times 0.2 \times 0.8 \times 0.7 \times 0.4 \times 0.4 \times 0.6 \times 0.5 \times 0.9 = 0.0048
\end{aligned}
$$

Question 2.

Two days back the boss has told the janitor that it is sunny. Assuming that the janitor has seen the boss carrying an umbrella yesterday, what is the probability of raining outside today given that the boss is not carrying an umbrella today?

Figure 1.5 depicts the graphical model of the HMM used in this question. Based on the assumptions, the problem can be written formally and calculated as follows

$$
P(S_2|O_{1:2}, Model) \propto P(O_2 = \neg\mathcal{U}|S_2) \times \sum_{S_1 \in \{\mathcal{R},\mathcal{C},\mathcal{S}\}} P(S_2|S_1) \times P(S_1|O_1) \tag{1.3}
$$

To compute the probability in (1.3), first $P(S_1|O_1)$ has to be computed as follows

$$P(S_1|O_1, Model) \propto P(O_1 = \mathcal{U}|S_1) \times P(S_1|S_0 = \mathcal{S}) \times P(S_0 = \mathcal{S})$$
$$\rightsquigarrow P(S_1 = \mathcal{R}|O_1, Model) = \mathcal{N}_1 \times P(O_1 = \mathcal{U}|S_1 = \mathcal{R}) \times P(S_1 = \mathcal{R}|S_0 = \mathcal{S}) \times P(S_0 = \mathcal{S})$$
$$\Rightarrow P(S_1 = \mathcal{R}|O_1, Model) = \mathcal{N}_1 \times b_{1,1} \times a_{3,1} \times 1 = 0.18\mathcal{N}_1$$

$$\rightsquigarrow P(S_1 = C|O_1, Model) = \mathcal{N}_1 \times P(O_1 = \mathcal{U}|S_1 = C) \times P(S_1 = C|S_0 = \mathcal{S}) \times P(S_0 = \mathcal{S})$$
$$\Rightarrow P(S_1 = C|O_1, Model) = \mathcal{N}_1 \times b_{2,1} \times a_{3,2} \times 1 = 0.42\mathcal{N}_1$$

$$\rightsquigarrow P(S_1 = \mathcal{S}|O_1, Model) = \mathcal{N}_1 \times P(O_1 = \mathcal{U}|S_1 = \mathcal{S}) \times P(S_1 = \mathcal{S}|S_0 = \mathcal{S}) \times P(S_0 = \mathcal{S})$$
$$\Rightarrow P(S_1 = \mathcal{S}|O_1, Model) = \mathcal{N}_1 \times b_{3,1} \times a_{3,3} \times 1 = 0.02\mathcal{N}_1$$

where $\mathcal{N}_1$ is a normalizing factor equal to $1/P(O_1)$. However, it is not required to be calculated, since based on the total probability principal

$$\sum_{S_1 \in \{\mathcal{R}, C, \mathcal{S}\}} P(S_1|O_1, Model) = 1 \Rightarrow 0.18\mathcal{N}_1 + 0.42\mathcal{N}_1 + 0.02\mathcal{N}_1 = 1 \Rightarrow \mathcal{N}_1 = 1.6129$$

Therefore, $P(S_1 = \mathcal{R}|O_1) = 0.2903$, $P(S_1 = C|O_1) = 0.6774$, and $P(S_1 = \mathcal{S}|O_1) = 0.0323$. Finally, $P_(S_2|O_{1:2}, Model)$ in (1.3) can be calculated as

$$\Rightarrow P(S_2 = \mathcal{R}|O_{1:2}, Model)$$
$$= \mathcal{N}_2 \times P(O_2 = \neg\mathcal{U}|S_2 = \mathcal{R}) \times \sum_{S_1 \in \{\mathcal{R}, C, \mathcal{S}\}} P(S_2 = \mathcal{R}|S_1) \times P(S_1|O_1)$$
$$= \mathcal{N}_2 \times b_{1,2} \times (a_{1,1} \times 0.2903 + a_{2,1} \times 0.6774 + a_{3,1} \times 0.0323) = 0.04613\mathcal{N}_2$$

$$\Rightarrow P(S_2 = C|O_{1:2}, Model)$$
$$= \mathcal{N}_2 \times P(O_2 = \neg\mathcal{U}|S_2 = C) \times \sum_{S_1 \in \{\mathcal{R}, C, \mathcal{S}\}} P(S_2 = C|S_1) \times P(S_1|O_1)$$
$$= \mathcal{N}_2 \times b_{2,2} \times (a_{1,2} \times 0.2903 + a_{2,2} \times 0.6774 + a_{3,2} \times 0.0323) = 0.1639\mathcal{N}_2$$

$$\Rightarrow P(S_2 = \mathcal{S}|O_{1:2}, Model)$$
$$= \mathcal{N}_2 \times P(O_2 = \neg\mathcal{U}|S_2 = \mathcal{S}) \times \sum_{S_1 \in \{\mathcal{R}, C, \mathcal{S}\}} P(S_2 = \mathcal{S}|S_1) \times P(S_1|O_1)$$
$$= \mathcal{N}_2 \times b_{3,2} \times (a_{1,3} \times 0.2903 + a_{2,3} \times 0.6774 + a_{3,3} \times 0.0323) = 0.1032\mathcal{N}_2$$

Similar to $\mathcal{N}_1$, $\mathcal{N}_2$ which is also a normalizing factor, can be computted and it is equal to 3.1925. Hence,

$$P(S_2 = \mathcal{R}|O_{1:2}, Model) = 0.04613 \times \mathcal{N}_2 = 0.1473.$$

Similar to the calculations used here, to find the probability of today being rainy based on today's and preceding observations, probabilities of being in different health status for the machinery system at each time step based on the indirect observations can be computed after estimating the parameters of HMM.

1.2 Objectives and Scope of Research

Among the aforementioned four components of the CBM, in this thesis, the focus is on the third component that is "Modeling and Condition Monitoring". Although the data acquisition process plays an important role in realization of an effective condition based maintenance system, it is out of scope of this thesis. Also, as the condition based maintenance systems can be operated by either experts (manually) or based on expert systems and automated decision making systems, the condition monitoring up to condition prediction is performed.

As mentioned in the previous Section, the intrinsic uncertainties which underlie the condition monitoring procedure and its temporal sequential nature, made hidden Markov model-based approaches a perfect option for this task. However, the HMM-based approaches that are implemented for this task in the literature are basically used as a black-box approach which are unable to depict a correspondence between the hidden state values in the HMM and the actual physical states. Thus, in this thesis, firstly an HMM-based approach for TCM is proposed which depicts the aforementioned correspondence. Furthermore, the proposed approach performs a continuous TCM in contrast to the previous HMM-based TCM approaches. Later on, more complex structures based on the similar idea developed in that approach are utilized to improve the prediction performance further. Using a hidden semi-Markov model-based approach, it is studied how to capture the trends in the training data more effectively by capturing the state-duration distributions with more realistic distributions. Also, it is investigated how to incorpo-

rate a given loss function into the proposed HSMM-based TCM approach. Moreover, to capture all possible underlying trends available in a given training data, a multi-modal TCM approach is proposed which uses parallel single models as various modes that are responsible for various captured trends. Various weighting schemes and switching strategies that can be incorporated in the multi-modal approach to unify the results from the multi modes into one, are proposed and studied.

Moreover, in this thesis, it is attempted to improve the performance of conventional HMM-based classification approach used for fault detection and diagnosis by incorporating the training data more effectively. To improve the performance of the existing HMM-based classification approach, an HMM-based semi-non parametric approach is proposed which takes the advantages of both parametric and nonparametric approaches.

1.3 Contribution and Outline of Thesis

This thesis is organized as follows. In Chapter 2, a temporal probabilistic approach based on the hidden Markov model (HMM), named physically segmented HMM with continuous output (PSHMCO), is introduced for continuous tool condition monitoring (TCM) in machinery systems. The proposed approach has the advantage of providing an explicit relationship between the actual health states and the hidden state values. The provided relationship is further exploited for formulation and parameter estimation in the proposed approach. The introduced approach is tested for continuous tool wear prediction in a computer numerical control (CNC)-milling machine and compared with two well-established neural network (NN) approaches, namely, multilayer perceptron and Elman network. In the experimental study, the prediction results are provided and compared after adopting appropriate hyper-parameter values for all the approaches by cross-validation. Based on the experimental results, physically segmented HMM approach outperforms the NN approaches. Moreover, the prognosis ability of the proposed approach is studied.

In Chapter 3, a more complex temporal probabilistic approach based on hidden semi-Markov model is proposed for continuous (real-valued) tool condition monitor-

ing (TCM) in machinery systems. Similar to Chapter 2, as an illustrative example, tool wear prediction in CNC-milling machine is conducted using the proposed approach. Results indicate that the additional flexibility provided in the new approach compared with the PSHMCO improves the performance. The prediction results are provided for three different cases i.e. cross-validation, diagnostics and prognostics. Possibility of incorporating an asymmetric loss function in the proposed approach in order to reflect and consider the cost differences between an under- and over-estimation in TCM is also explored and the simulation results are provided.

In Chapter 4, a novel multi-modal hidden Markov model-based approach is proposed for tool wear monitoring. The proposed approach improves the performance of the single hidden Markov model-based approach named PSHMCO (proposed in Chapter 2) by using multiple PSHMCOs in parallel. In this multi-modal approach, each PSHMCO captures and emphasizes on a different tool wear regiment. In this Chapter, three weighting schemes, namely, *bounded hindsight*, *discounted hindsight* and *semi-nonparametric hindsight* are proposed and two switching strategies named *soft-* and *hard-switching* are introduced to combine the outputs from multiple modes into one. Similar to preceding Chapters, the proposed approach is applied to tool wear monitoring in a CNC-milling machine. The performance of the multi-modal approach with various weighting schemes and switching strategies is reported and compared with PSHMCO.

In Chapter 5, a semi-nonparametric approach based on hidden Markov model is introduced for fault detection and diagnosis in Rotary Electric Motors. In this approach, after training the hidden Markov model classifiers (parametric stage), two matrices named *probabilistic transition frequency profile* and *average probabilistic emission* are computed based on the hidden Markov models for each signature (non-parametric stage) using probabilistic inference. These matrices are later used in forming a similarity scoring function, which is the basis of the classification in this approach. Moreover, a preprocessing method, named *squeezing and stretching* is proposed which rectifies the difficulty of dealing with various operating speeds in the classification process. The experimental results are provided and compared for a synchronous motor. Further investigations are carried out, providing sensitivity analysis on the length of signatures, the

number of hidden state values, as well as statistical performance evaluation and comparison with conventional hidden Markov model-based fault diagnosis approach.

Finally, the thesis is concluded in Chapter 6. This chapter summarizes the contributions of the research work reported in this thesis and outlines the future work directions.

Chapter 2

Physically Segmented Hidden Markov Model with Continuous Output

2.1 Introduction

Tool Condition Monitoring (TCM) has become one of the main industrial challenges in the last decade. TCM reduces the amount of unnecessary downtime for maintenance purposes, and consequently reduces the cost of maintenance [2, 3, 4, 6, 5]. Moreover, TCM improves the quality and precision of the product.

The idea of continuous tool condition monitoring is to monitor the health condition of the tool at each time step in terms of a continuous metric based on the available input data. In other words, instead of setting thresholds and differentiating distinct health states as various (ordinal) classes, we would like to ultimately monitor the health state of the tool in a continuous form. This task allows us to have smoother decision making systems in the condition based maintenance and it can incorporate different quality thresholds for different applications using the same condition based maintenance system e.g. to guarantee different qualities in various products. The input data in this task, is a set of selected features that are extracted from non-intrusively sensed and captured signals. Signals such as force, vibration and acoustic emission can be captured and recorded using various sensors mounted on the machinery systems.

Hidden Markov models (HMM) and hidden semi-Markov models (HSMM) are used

[31, 32, 33, 34, 35, 36, 37] to distinguish various wearing stages or machinery fault types. In all the existing HMM and HSMM-based approaches, the wearing data is discretized into several stages and then multiple HMMs are used to distinguish between those stages. Each HMM or HSMM is assigned to recognize one specific stage or fault. Training HMMs and HSMMs using expectation-maximization method, however is essentially a black-box approach, which does not provide explicit relationship between the wearing value and the hidden state values in the trained HMMs or HSMMs.

In this Chapter, a temporal probabilistic approach based on HMM, named physically segmented hidden Markov model with continuous output (PSHMCO), is proposed to tackle the problem of continuous health assessment of cutters in a CNC-milling machine. The proposed approach depicts the explicit relationship between the actual physical states and the hidden state values. Furthermore, the provided relationship is exploited for formulation and parameter estimation in PSHMCO. In addition, the suggested approach ultimately predicts the real-valued health state metric (tool wear) instead of discrete types or stages.

This Chapter is organized as follows. In Section 2, the proposed approach, PSHMCO, is introduced. Diagnostics and prognostics procedures are described in Section 3. Section 4 provides information about the experimental data and selected features from the acquired signals in the experiments. In Section 5, performance of PSHMCO is compared with two well-established neural networks, namely, Multi-Layer Perceptron (MLP) and Elman network. The Chapter is concluded in Section 6.

2.2 Physically Segmented Hidden Markov Model with Continuous Output

As the name implies, this approach is based on hidden Markov model. Contrary to the conventional use of HMMs which is in classification, here HMM is applied to a continuous problem (regression). In this approach, explicit relationship is provided between the tool conditions (physical health states) and the hidden state values of the HMM. Then, the relationship is further exploited to directly compute the parameters

using maximum likelihood method assuming to have a complete training set. Finally, the state estimation is described for different points in time.

Hidden Markov model is a simple dynamic Bayesian network [96]. This model has only one discrete hidden state variable, and a set of discrete or continuous observation nodes.

As mentioned in the preceding Chapter, a first-order temporal Markov model is characterized by the assumption that,

$$P(S_t = v_i | S_{t-1}, S_{t-2}, \ldots, S_1) = P(S_t = v_i | S_{t-1}), \tag{2.1}$$

where $P(.|.)$ is the conditional probability, S_t is the hidden state variable at time t and v_i is the ith hidden state value. Equation (2.1) indicates that the conditional probability of any current state, given knowledge of all previous states, is the same as the conditional probability of the current state, given its previous state only [97].

In order to use the HMM in the PSHMCO approach, firstly, the output space will be discretized into several hidden state values or health states as shown in Fig. 2.1. After that the outputs in the training set would be discretized and assigned to those health states (segmented). Then, the parameters of the hidden Markov model are directly estimated based on the complete training dataset with discretized outputs. Hence, when the new testing data is given to the HMM model, using the inference algorithms and the learned parameters, a probability distribution over real-valued health states (hidden state values) can be computed for each time step. Finally, using the calculated probability distribution and the discretized real-valued labels of health states, an expected real-valued output will be computed for each time step. In this and the following Sections, the PSHMCO approach for diagnostics and prognostics in TCM is introduced in details.

2.2.1 Discretization & Formulation

Using uniform discretization method on the continuous health metric (tool wear), continuous wearing values can be discretized into m classes or state values with real-valued labels, $\{H_1, ..., H_m\}$. These state values correspond to m ordinal wearing stages. As shown in Fig. 2.1, after discretization and segmenting the tool wear values in the

training set, segments are labeled ordinally from H_1 to H_m indicating explicit correspondence between the health states (hidden state values) in the HMM and the actual physical wearing metric.

An appropriate value for the number of health states can be adopted based on cross-validation results. Since wearing is a gradual process and sampling rate is relatively high comparing to the number of possible state values, at each time step only two possible transitions are available (excluding the last state which is modeled as an attraction point). The two possible transitions given that $S_{t-1} = H_i$ are either staying at the same condition H_i with probability p_i or going to the next degradation stage H_{i+1} with probability $1 - p_i$, as shown in Fig. 2.1. Therefore, the transition probability matrix, A, for the HMM used

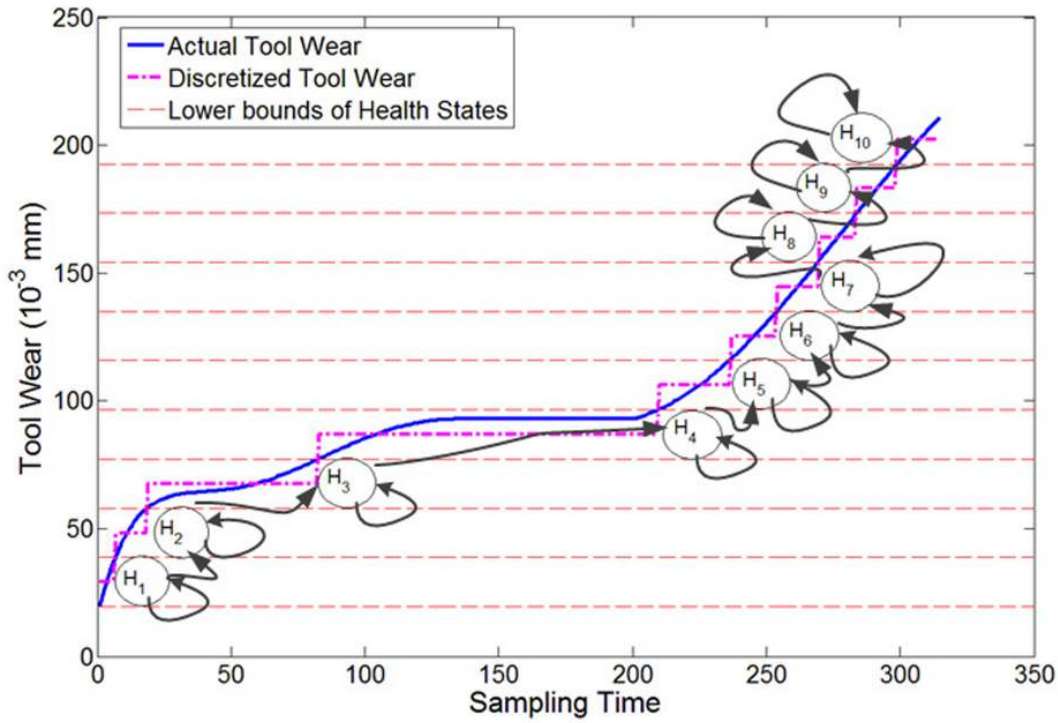


Figure 2.1: Illustrative example of tool wear discretization and the correspondence of the Actual tool wear with the hidden state values in the implemented HMM, $\{H_1, \ldots, H_{10}\}$. Arrows indicate possible transitions from each health state.

in the TCM can be formulated as,

$$A = [a_{i,j}]_{m\times m} = \begin{bmatrix} p_1 & 1-p_1 & 0 & 0 & ... & 0 \\ 0 & p_2 & 1-p_2 & 0 & ... & 0 \\ . & & . & & & . \\ . & & & . & & . \\ . & & & & . & . \\ 0 & 0 & 0 & ... & p_{m-1} & 1-p_{m-1} \\ 0 & 0 & 0 & ... & 0 & 1 \end{bmatrix}, \tag{2.2}$$

where the element at ith row and jth column, $a_{i,j}$, is the transition probability of going from ith health state to jth health state. m is the total number of state values that the hidden state can take and p_i is the probability of self-transition in the ith health state.

2.2.2 Parameter Estimation

Assuming stationarity, the only parameters to be identified are, the initial health state probabilities (prior probabilities), transition probability matrix in (2.2) and emission probabilities that connect the hidden states to the observations (input features). These parameters must be estimated for the HMM given the training data. The parameter estimation can be done using either maximum likelihood (ML) or maximum a posteriori method. Here, Maximum Likelihood learning method is adopted. Gradient ascent and expectation-maximization (Baum-Welch) are the two conventional algorithms to estimate the parameters based on ML [96]. However, in this work since the data is complete (without missing information) and taking the advantage of explicit relationship between the given actual physical states and the hidden state values, there is no need to use either of the two mentioned algorithms. Using the ML method, parameters can be directly estimated from the discretized health states based on the measured tool wear (as depicted in Fig. 2.1), and the input feature sequences that are extracted from the dataset.

The ML method calculates the parameters that maximize the likelihood probability of the training dataset. Therefore, in order to find the parameters, the joint probability distribution of the training dataset (likelihood probability) must be derived and then be maximized. Since the n experimental sequences included as the training set $\{D_1, \ldots, D_n\}$

are independent of each other, the joint probability distribution for HMM can be written as

$$P(D_1, D_2, \ldots, D_n|\lambda) = \prod_{j=1}^{n} P(D_j|\lambda), \tag{2.3}$$

where D_j is the experimental data sequence collected from the jth experiment and λ is the set of HMM parameters. In (2.3) the joint probability distribution for the jth experiment given the parameters, $P(D_j|\lambda)$, can be computed as

$$\begin{aligned} P(D_j|\lambda) &= P(S^j_{1:T_j}, O^j_{1:T_j}|\lambda) = \pi_o(S^j_1) \prod_{t=1}^{T_j} P(S^j_t|S^j_{t-1}) P(O^j_t|S^j_t), \\ O^j_{1:T_j} &= \{O^j_1, O^j_2, ..., O^j_{T_j}\}\ ,\ S^j_{1:T_j} = \{S^j_1, S^j_2, ..., S^j_{T_j}\}, \end{aligned} \tag{2.4}$$

where D_j includes extracted features $O^j_{1:T_j}$ and tool wear sequence $S^j_{1:T_j}$, and T_j is the length of the jth experimental sequence. Assuming that the emission probabilities $P(O^j_i|S^j_i)$ are Gaussian distributions [98] and considering the transition matrix in (2.2), joint probability distribution of the jth experimental data given in (2.4) can be rewritten as

$$\begin{aligned} P(D_j|\lambda) &= \pi_0(S^j_1) \times \prod_{i=1}^{m-1} p_i^{k^j_i - 1}(1 - p_i) \times \prod_{i=1}^{m} \prod_{t=t^j_i}^{t^j_{i+1}-1} \frac{1}{(2\pi)^{\chi/2}|\Sigma_i|^{1/2}} e^{(-\frac{1}{2}(O^j_t - \mu_i)^T \Sigma_i^{-1}(O^j_t - \mu_i))}, \\ t^j_i &= \begin{cases} 1 + \sum_{c=1}^{i-1} k^j_c & for\ i > 1 \\ 1 & for\ i = 1 \end{cases}, \end{aligned} \tag{2.5}$$

where $\lambda = \{\pi_0, p_1, ..., p_{m-1}, \mu_1, ..., \mu_m, \Sigma_1, ..., \Sigma_m\}$, π_0 is the prior probability distribution of the initial health state, m is the number of health states, μ_i and Σ_i are respectively mean and covariance matrix used in the Gaussian distribution to compute the emission probability at time t given the fact that $S^j_t = H_i$. χ is the dimension of observation vector or in other words the number of features being used for TCM. t^j_i is the starting time step of the ith health state and k^j_c is the number of samples belonging to the cth health state, both in the jth experimental data. From (2.5), the log likelihood for the implemented

HMM with real-valued observations can be computed as

$$\begin{aligned}
&\bar{L} = \sum_{j=1}^{n} L_j = \sum_{j=1}^{n} \log(P(D_j|\lambda)),\\
&L_j = \log(P(D_j|\lambda)) = \log(\pi_0(S_1^j)) - \frac{\chi T_j}{2}\log(2\pi)\\
&+ \sum_{i=1}^{m-1}(k_i^j - 1)\log(p_i) + \log(1-p_i) - \frac{k_i^j}{2}(|\Sigma_i|) - \sum_{i=1}^{m}\sum_{t=t_i^j}^{t_{i+1}^j-1}\frac{1}{2}(O_t^j-\mu_i)^T\Sigma_i^{-1}(O_t^j-\mu_i).
\end{aligned} \tag{2.6}$$

To find the parameter set, which maximizes the log likelihood in (2.6), the partial derivatives of the log likelihood are set to zero. Consequently, the parameters may be computed as follows,

$$\begin{aligned}
&\frac{\partial \bar{L}}{\partial p_i} = 0 \rightsquigarrow \sum_{j=1}^{n}[\frac{k_i^j - 1}{p_i} - \frac{1}{1-p_i}] = 0 \rightsquigarrow \frac{n\times(1-p_i) + n\times p_i}{p_i\times(1-p_i)} - \frac{\sum_{j=1}^{n} k_i^j}{p_i} = 0\\
&\rightsquigarrow p_i = 1 - \frac{n}{\sum_{j=1}^{n} k_i^j},
\end{aligned} \tag{2.7}$$

$$\begin{aligned}
&\frac{\partial \bar{L}}{\partial \mu_i} = 0 \rightsquigarrow -\sum_{j=1}^{n}\sum_{t=t_i^j+1}^{t_{i+1}^j}\frac{1}{2}[\Sigma_i^{-1} + (\Sigma_i^{-1})^T](O_t^j - \mu_i) = 0\\
&\rightsquigarrow \sum_{j=1}^{n}\sum_{t=t_i^j+1}^{t_{i+1}^j}(O_t^j - \mu_i) = 0 \rightsquigarrow \sum_{j=1}^{n}\sum_{t=t_i^j+1}^{t_{i+1}^j} O_t^j - \mu_i\sum_{j=1}^{n} k_i^j = 0\\
&\rightsquigarrow \mu_i = \frac{\sum_{j=1}^{n}\sum_{t=t_i^j+1}^{t_{i+1}^j} O_t^j}{\sum_{j=1}^{n} k_i^j},
\end{aligned} \tag{2.8}$$

$$\begin{aligned}
&\frac{\partial \bar{L}}{\partial \Sigma_i} = 0 \rightsquigarrow -\sum_{j=1}^{n}[\frac{k_i^j}{2}\frac{1}{|\Sigma_i|}|\Sigma_i|(\Sigma_i^{-1})^T - \sum_{t=t_i^j+1}^{t_{i+1}^j}\frac{1}{2}\Sigma_i^{-1}(O_t^j-\mu_i)(O_t^j-\mu_i)^T\Sigma_i^{-1}] = 0\\
&\rightsquigarrow \sum_{j=1}^{n}\frac{k_i^j}{2} = [\frac{1}{2}\sum_{j=1}^{n}\sum_{t=t_i^j+1}^{t_{i+1}^j}(O_t^j-\mu_i)(O_t^j-\mu_i)^T]\Sigma_i^{-1}\\
&\rightsquigarrow \Sigma_i = \frac{\sum_{j=1}^{n}\sum_{t=t_i^j+1}^{t_{i+1}^j}(O_t^j-\mu_i)(O_t^j-\mu_i)^T}{\sum_{j=1}^{n} k_i^j}.
\end{aligned} \tag{2.9}$$

Using the formulas in (2.7), (2.8) and (2.9), all of the required parameters can be estimated. The initial state probability π_0 is defined as the uniform probability distribution in case there are no prior knowledge about the general initial state value.

Next, a computationally efficient *forward-backward* algorithm which is required for diagnostics and prognostics using the implemented HMM, is described.

2.2.3 Forward-Backward Variables in PSHMCO

Forward-backward algorithm is a recursive algorithm used in Markov models to address inference problems. A simple implementation of this algorithm for TCM using PSHMCO is introduced in this Section. This algorithm uses two auxiliary variables, *forward* and *backward* variables, to compute the required probabilities, recursively.

Forward variable in HMM is defined as the joint probability of being at ith state H_i at time t while observing the inputs from time step 1 to t, $O_{1:t}$. As in [95], forward variable can be written as

$$\alpha_t(i) \triangleq P(O_{1:t}, S_t = H_i). \tag{2.10}$$

From law of total probability [95], a recursive formula to calculate α_t in (2.10) can be derived as follows,

$$\begin{aligned} \alpha_t(i) &= \sum_{j=1}^{m} \alpha_{t-1}(j) P(O_t|S_t = H_i) P(S_t = H_i|S_{t-1} = H_j) \\ &\textit{with initial condition } \alpha_1(i) = \pi_0(i) \times P(O_1|S_1 = H_i). \end{aligned} \tag{2.11}$$

Based on the graduality of the wearing process and the possible transitions in (2.2),the forward variable can be computationally simplified as

$$\alpha_t(i) = \begin{cases} p_i \alpha_{t-1}(i) P(O_t|s_t = H_i) + (1 - p_{i-1}) \alpha_{t-1}(i-1) P(O_t|s_t = H_i) & for\ i > 1 \\ p_i \alpha_{t-1}(i) P(O_t|s_t = H_i) & for\ i = 1 \end{cases}. \tag{2.12}$$

The second auxiliary variable, which is called backward variable, is the joint probability of observing all the inputs from time step $t + 1$ to T given the health state is at the ith stage (H_i) at time step t. It can be written as

$$\beta_t(i) \triangleq P(O_{t+1:T}|S_t = H_i). \tag{2.13}$$

According to [95], the backward variable can be computed recursively as

$$\beta_t(i) = \sum_{j=1}^{m} a_{ij} P(O_{t+1}|S_{t+1} = H_j)\beta_{t+1}(j) \quad \text{with initial condition } \beta_T(i) = 1, \tag{2.14}$$

where a_{ij} is the transition probability of going from ith health state to jth health state.

Based on the graduality of the wearing process and the transition matrix in PSHMCO, the backward variable can be computationally simplified further and rewritten as

$$\beta_t(i) = \begin{cases} (1 - p_i) \times P(O_{t+1}|S_{t+1} = H_{i+1}) \times \beta_{t+1}(i+1) & for\ i < m \\ +p_i \times P(O_{t+1}|S_{t+1} = H_i) \times \beta_{t+1}(i) & \\ \\ p_i \times P(O_{t+1}|S_{t+1} = H_i) \times \beta_{t+1}(i) & for\ i = m \end{cases}. \tag{2.15}$$

2.2.4 State Estimation

In order to estimate the state values of the state variables in the implemented HMM at each time step based on the observations, two variables are further defined. γ_t is the joint probability of observing all the input features up to the current time T while being at the ith health state at time t where $t < T$. γ_t is used to find the value of the state variable at t for diagnosis purpose based on all the observations from beginning of an experiment up to the current time, T. Based on [95], γ_t can be defined and computed as follows,

$$\gamma_t(i) \triangleq P(S_t = H_i, O_{1:T}) = P(O_{1:t}, S_t = H_i) \times P(O_{t+1:T}|S_t = H_i) = \alpha_t(i) \times \beta_t(i), \tag{2.16}$$

where $\alpha_t(i)$ and $\beta_t(i)$ are the forward and backward variables that can be computed based on (2.12) and (2.15).

In order to predict the state value of the state variable in future based on the available observations up to the current time T, another variable, $\xi_{t'}$, is defined. Similar to γ_t, $\xi_{t'}$ is the joint probability of observing all the input features from beginning of an experiment up to the current time T, but being at the ith health state at time t' in future ($t' > T$). $\xi_{t'}$

is defined and computed recursively based on transition probabilities and γ_T as follows,

$$\begin{aligned}
&\xi_{t'}(i) \triangleq P(s_{t'} = H_i, O_{1:T}),\ t' > T \\
&\xi_{t'}(i) = \sum_{S_{t'-1}} P(S_{t'-1}, O_{1:T}) \times P(S_{t'} = H_i | S_{t'-1}) = \xi_{t'-1}(i) \times p_i + (1 - p_{i-1}) \times \xi_{t'-1}(i-1) \\
&\textit{with initial condition } \xi_T(i) = \gamma_T(i).
\end{aligned} \tag{2.17}$$

In the succeeding Section, γ_t and $\xi_{t'}$ are used for diagnostics and prognostics.

2.3 Diagnostics & Prognostics

In this Section, the TCM approach (diagnostics & prognostics) is provided in a probabilistic manner based on the state estimation variables defined in the previous Section and the real valued labels determined in the discretization phase.

Diagnosis is the task of predicting the health state at the current time T given all the observations from time step 1 up to T [10]. In the realm of Bayesian Networks, this task is called *filtering* or *monitoring* and it can be written in a probabilistic form as follows,

$$P(S_T = H_i | O_{1:T}, \lambda) = \frac{P(O_{1:T}, S_T = H_i)}{P(O_{1:T})} = \frac{\gamma_T(i)}{P(O_{1:T})} \tag{2.18}$$

Probability of being at each health state at time T using the implemented HMM can be calculated using (2.16) and (2.18). It is worth mentioning that the denominator in (2.18) is not required to be calculated. It is a normalizing factor and can be calculated after finding $\gamma_T(i)$ for all the health states as follows,

$$\begin{aligned}
&\sum_{i=1}^{m} P(S_T = H_i | O_{1:T}, \lambda) = 1 \rightsquigarrow \frac{1}{P(O_{1:T})} \sum_{i=1}^{m} \gamma_T(i) = 1 \\
&\rightsquigarrow P(O_{1:T}) = \sum_{i=1}^{m} \gamma_T(i).
\end{aligned} \tag{2.19}$$

Hence, based on (2.16) and (2.19), (2.18) can be rewritten as

$$P(S_T = H_i | O_{1:T}, \lambda) = \frac{\gamma_T(i)}{\sum_{i=1}^{m} \gamma_T(i)} = \frac{\alpha_T(i)}{\sum_{i=1}^{m} \alpha_T(i)}. \tag{2.20}$$

where λ is included in the conditions as a reminder that the calculations are based on the parameter set of the trained HMM. Finally, the continuous output of the PSHMCO

model, $\hat{y}_T$, which corresponds to the expected amount of tool wear at the current time step T, can be calculated based on (2.20) as follows,

$$\hat{y}_T = \sum_{i=1}^{m} P(S_T = H_i | O_{1:T}, \lambda) \times H_i. \tag{2.21}$$

Fig. 2.2 depicts the procedure of diagnosis using PSHMCO approach (assuming that the covariance matrices are diagonal for simplicity in visualization).

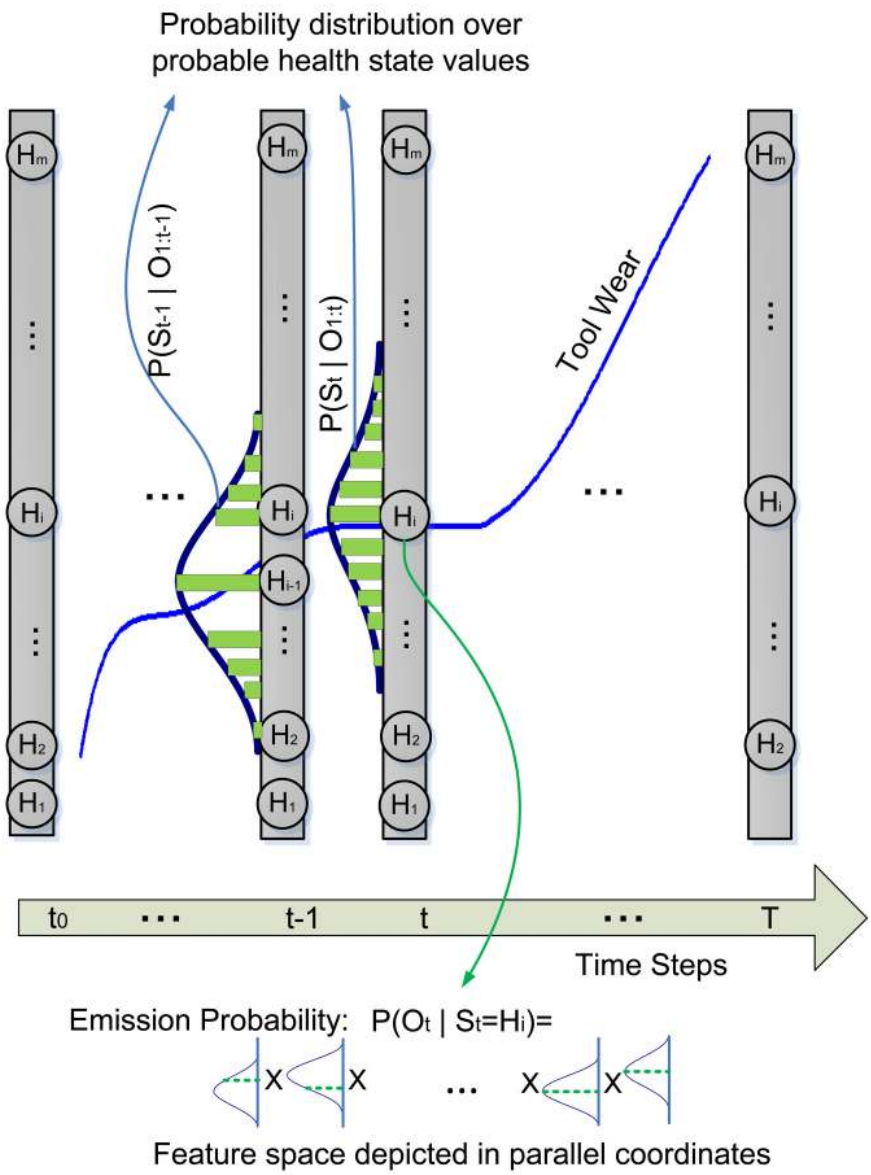


Figure 2.2: Schematic diagnosis procedure in PSHMCO approach (assuming that the covariance matrices are diagonal for simplicity in visualization).

Prognosis is the task of predicting the future health state at time t' ($t' > T$) while the observations are only available from the beginning up to the current time T [5,6,11]. In order to find the probability of being at each state in future, the model must be unrolled over the time horizon while there are no observations available from T onwards. Hence, prognosis for $t' > T$ can also be formulated similar to diagnosis case in a probabilistic manner as follows,

$$P(S_{t'} = H_i | O_{1:T}, \lambda) = \frac{P(S_{t'} = H_i, O_{1:T})}{P(O_{1:T})} = \frac{\xi_{t'}(i)}{\sum_{i=1}^{m} \xi_{t'}(i)}, \tag{2.22}$$

where λ is again included to indicate that calculations are based on the parameter set of the trained HMM and $P(O_{1:T})$ is a normalizing factor replaced by $\sum_{i=1}^{m} \xi_{t'}(i)$ similar to (2.19). In the end, the continuous output of the model, $\hat{y}_{t'}$, which corresponds to the expected amount of tool wear at time step t' in future, can be computed based on (2.22) as follows,

$$\hat{y}_{t'} = \sum_{i=1}^{m} P(S_{t'} = H_i | O_{1:T}, \lambda) \times H_i. \tag{2.23}$$

2.4 Experimental Data & Feature Selection

The experimental data is obtained through real-time sensing on a CNC-milling machine. The experimental setup and the extracted features that include both statistical features of the force signals and wavelet features extracted from force, vibration and acoustic emission (AE) signals are described in Appendix A. The total number of extracted features in the acquired experimental data is 482 features. The extracted features include both statistical and wavelet features. The statistical features comprise 16 features extracted from force signals in each direction, resulting in 48 statistical features in total. The wavelet features include features with 5 decomposition levels from every signal leading to 62 ($2^1 + 2^2 + 2^3 + 2^4 + 2^5 = 62$) extracted features from each signal, summing up to 434 wavelet features, 62×3 (force signals in three directions) + 62×3 (vibration signals in three directions) + 62 (AE signal).

Constructing the tool condition prediction model using only a selected subset of features prevents unnecessary complexity in the model and consequently improves the

prediction results. In addition, extracting only the selected features from the newly conducted experiments greatly reduces the feature extraction computation cost for online applications. In this study, the feature selection is performed offline. The features are determined based on Fisher's Discriminant Ratio (FDR) values on the training data and are not changed in the simulated online study. In the simulated online monitoring, only the selected features are extracted and utilized in order to save the computational effort and also for fast data processing.

The idea of feature selection in a classification domain is to find a subset of features that explicitly discriminate the classes based on the training set. Hence, the features to be selected must have similar values for the samples in one class and distinct values for the samples from different classes. FDR is such a metric that shows how discriminative a feature is. It is a ratio of scatter between (Sb) and the scatter within (Sw). A modified version of FDR introduced in [33] is as follows,

$$FDR(x_i) = \frac{Sb^{x_i}}{Sw^{x_i}} = \frac{\sum_{c=1}^{\kappa}\sum_{j=1}^{\kappa}(\mu_c^{x_i} - \mu_j^{x_i})^2}{\sum_{c=1}^{\kappa} Sw_c^{x_i}} \tag{2.24}$$

where κ is the number of classes, x_i is the ith feature (element) in the observation vector $O = [x_1\ x_2\ \ldots\ x_\chi]^T$, $\mu_c^{x_i}$ is the mean value of f_i in the cth class and $Sw_c^{x_i}$ is the scatter within (variance) the cth class measured for x_i. In order to use the FDR criterion for feature selection in the continuous TCM, samples must be clustered based on their output values. One of the conventional clustering algorithms is called Gaussian Mixture Model (GMM) [99, 100]. In this work, GMM is used for clustering of the samples based on their corresponding outputs. GMM tries to model the given data, which in this case is the output set (tool wear), by a mixture of Gaussian functions. The parameters of these Gaussian functions will be estimated by *expectation-maximization* technique. To find a proper number of Gaussian functions that fits the data, different number of Gaussian mixtures are explored from 1 to 10, and then the model with minimum Bayesian Information Criterion (BIC) is adopted as the appropriate GMM for clustering the output set. BIC is a measure of goodness of fit for an estimated statistical model [101]. The BIC results are depicted in Fig. 2.3. In this work, the data collected from three of the cutters which are *07BX1, 31PN4* and *34PT1*, is selected to be the training set and the feature se-

lection process is performed only on the data from these training cutters. From Fig. 2.3,

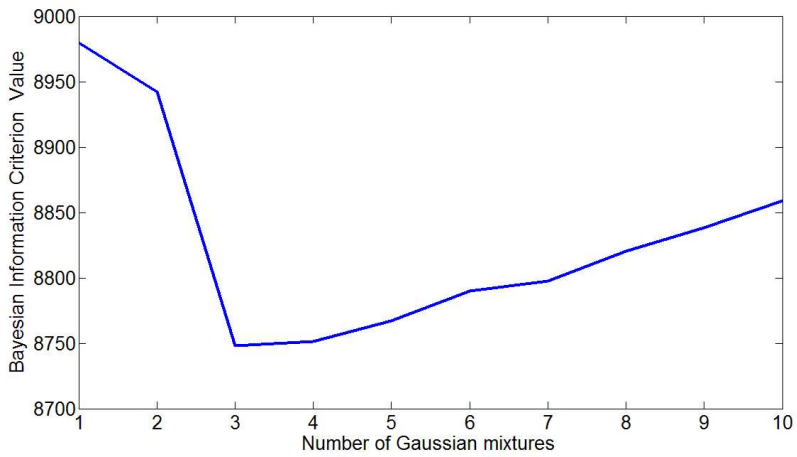


Figure 2.3: Bayesian Information Criterion for GMMs on the training outputs with various number of mixtures.

it can be seen that the minimum BIC is achieved by the GMM with three Gaussian functions. Hence, the number of Gaussian functions to be used is set to 3. After applying the GMM to the outputs, the posterior probability $p(M_c|y_t^j)$ can be computed for all samples in the training set, where $p(M_c|y_t^j)$ indicates the probability that the actual tool wear at time t in the jth experiment (y_t^j), is generated by the cth Gaussian function (M_c).

After finding the posterior probabilities, $\mu_c^{x_i}$ and $S\,w_c^{x_i}$ can be computed in a weighted form as follows,

$$\mu_c^{x_i} = \frac{\text{weighted sum of } i\text{th feature in } c\text{th cluster}}{\text{effective cardinality of } c\text{th cluster}} = \frac{\sum_{j=1}^{n}\sum_{t=1}^{T_j} p(M_c|y_t^j) \times x_i^j(t)}{\sum_{j=1}^{n}\sum_{t=1}^{T_j} p(M_c|y_t^j)}, \tag{2.25}$$

$$S\,w_c^{x_i} = \frac{\sum_{j=1}^{n}\sum_{t=1}^{T_j} [p(M_c|y_t^j) \times x_i^j(t) - \mu_c^{x_i}]^2}{\sum_{j=1}^{n}\sum_{t=1}^{T_j} p(M_c|y_t^j)}, \tag{2.26}$$

where $x_i^j(t)$ is the value of the ith feature at time t in the jth experimental sequence.

Then, the computed means and scatter within measures can be used in (2.24) to find FDR for ith feature. Fig. 2.4 shows the FDR values after being sorted in a descending manner.

The FDR value indicates how discriminant each feature is, hence it can be used to rank the features. As it can be seen in Fig. 2.4, there is a knee in the curve at 38 . This knee point is used as a rule of thumb to chose the number of features to be selected. Therefore, 38 features with the highest FDR value are chosen for the prediction model. Table 2.1 shows shares of extracted features from each signal in the set of selected features. Moreover, shares of different wavelet levels of each signal in the set of selected features are listed in Table 2.2.

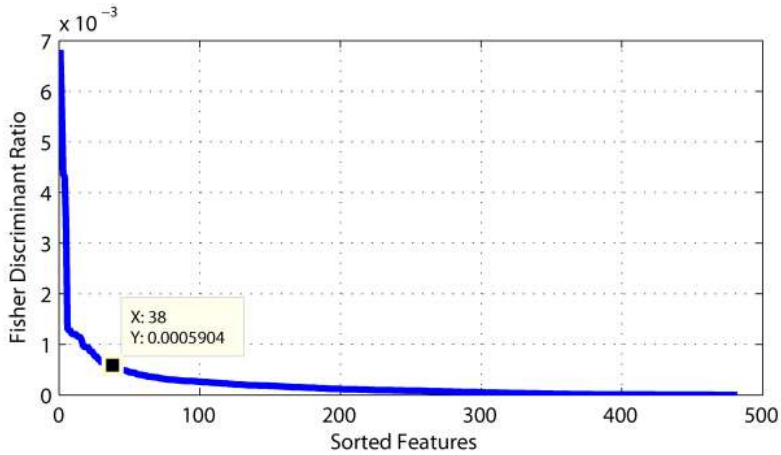


Figure 2.4: FDR values of features sorted in a descending manner.

Table 2.1: Shares of extracted features from each signal in the set of selected features.F_x,F_y and F_z are force signals in different directions.V_x,V_y and V_z are vibration signals in different directions and AE is the acoustic emission signal.

Signal	F_X	F_Y	F_Z	V_X	V_Y	V_Z	AE	**Total**
Statistical Features	6	2	6	-	-	-	-	14
Wavelet Features	11	5	5	0	3	0	0	24
Total Share	17	7	11	0	3	0	0	38

As it can be understood from Table 2.1, features extracted from the force signals are majorly selected as the most discriminant features. In contrast, none of the features extracted from acoustic emission are selected, which suggests that this signal may not

Table 2.2: Shares of different wavelet levels of each signal in the set of selected features.

Level	**Signal**						
	F_X	F_Y	F_Z	Vib_X	Vib_Y	Vib_Z	AE
$Level_1$	0	1	1	0	0	0	0
$Level_2$	1	1	1	0	0	0	0
$Level_3$	2	1	1	0	0	0	0
$Level_4$	3	1	1	0	1	0	0
$Level_5$	5	1	1	0	2	0	0

be useful for condition monitoring in CNC-milling machines because of its low signal to noise ratio. Feature selection results indicated that the most discriminant features are majorly selected from the force signals in X and Z directions. In X direction, total amplitude of cutting force, maximum force level and standard deviation are respectively the most discriminant signals. Average force in Z direction is also suggested to be important based on the FDR values.

2.5 Diagnostics & Prognostics Results

In this Section, performance of the proposed PSHMCO approach is compared with two artificial neural network (NN) approaches in diagnostics of the cutter. Moreover, the prognostics ability of the PSHMCO is tested.

As supervised learning algorithms, NNs are used in both regression and classification problems [102]. Among various NNs, multi-layer perceptron (MLP) and Elman network are two of the most commonly used neural networks that can be adopted for continuous tool condition monitoring. In this case, features are regarded as the inputs of the networks and the output is the measured wearing metric [14, 23, 24, 25].

First, optimal NNs and PSHMCO are determined through identifying their hyper-parameters, which are the number of hidden state values in PSHMCO as well as the structure and number of hidden neurons in both NNs. Here the cross-validation method

is applied. Then, performance of all three approaches are compared based on two accuracy indicators, namely, *mean square error* (MSE), *mean relative error* (MRE). Finally, the capability of PSHMCO in prognostics is studied. It is noteworthy that in all three cases, the collected data from the three cutters (*07BX1, 31PN4* and *34PT1*) that was used in feature selection process is regarded as the training set for all models.

2.5.1 Determination of Hyper-parameters

Two modes can be considered for cross-validation in case of condition monitoring for the cutter with multiple flutes i.e. *leave one flute out* and *leave one experiment out*. In this work, since the training set is limited to 3 experiments, *leave one flute out* is conducted. All models are trained on the data from two flutes out of three flutes, and then tested on the data from the excluded flute, all from the training experiments. Different hyper-parameter values for the three models are tested and the results are provided in Fig. 2.5 to Fig. 2.7.

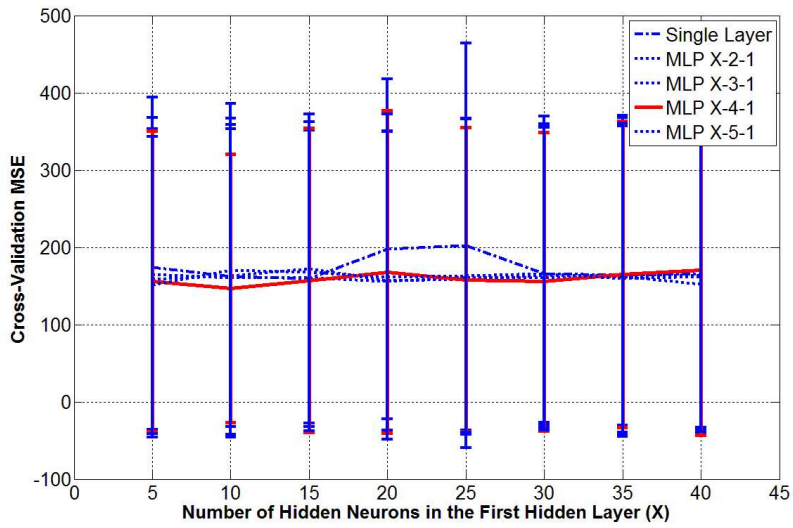


Figure 2.5: Cross-Validation results for MLP with different structures. The solid curve corresponds to MLP $X - 4 - 1$ structure which leads to minimum MSE at $X = 10$.

As shown in Fig. 2.5, the MLP structure $10 - 4 - 1$, where 10 and 4 are the number of hidden neurons in the first and second hidden layers respectively, leads to the best cross-validation MSE. Hence, the same structure for MLP is adopted in the succeeding

case. Furthermore, in order to prevent over-fitting problem in MLP, Bayesian regulation back-propagation [103, 104] is used as the learning algorithm for the implemented MLP.

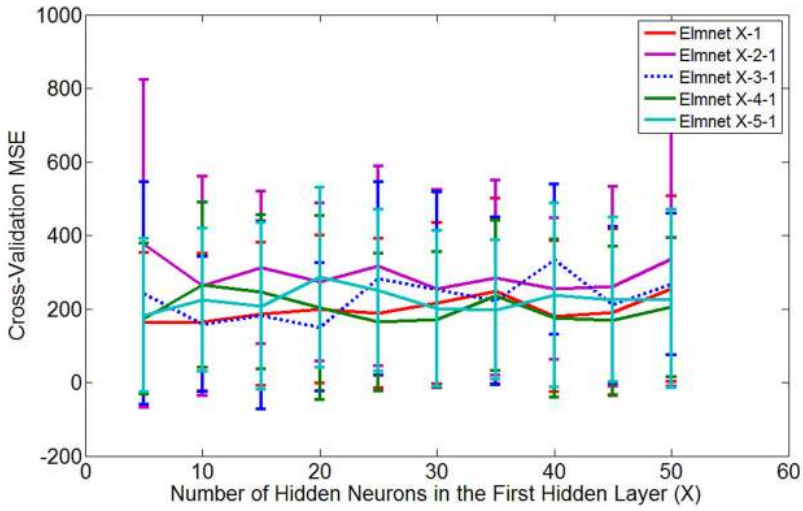


Figure 2.6: Cross-Validation results for Elman network with various structures. The dotted curve corresponds to Elman network with $X - 3 - 1$ structure which leads to minimum MSE at $X = 20$.

Fig. 2.6 indicates that the Elman network with structure $20 - 3 - 1$ with two hidden layers leads to the best performance in cross-validation comparing to other structures. Hence, the Elman network with structure $20 - 3 - 1$ is adopted to be used in the diagnostics case.

From Fig. 2.7, it can be seen that 14 may be used as an appropriate number of health states in the PSHMCO for the given dataset. Hence, the number of hidden state values (health states) is set to 14 in the succeeding cases.

2.5.2 Diagnostic Results

Now we compare diagnosis accuracy of the three approaches on the testing set using the implemented programs for each approach in MATLAB2010b. All models are trained using the training set (i.e. data collected from cutters *07BX1, 31PN4* and *34PT1*), and then tested for diagnostics on the testing set (i.e. data collected from cutters *09BX3,*

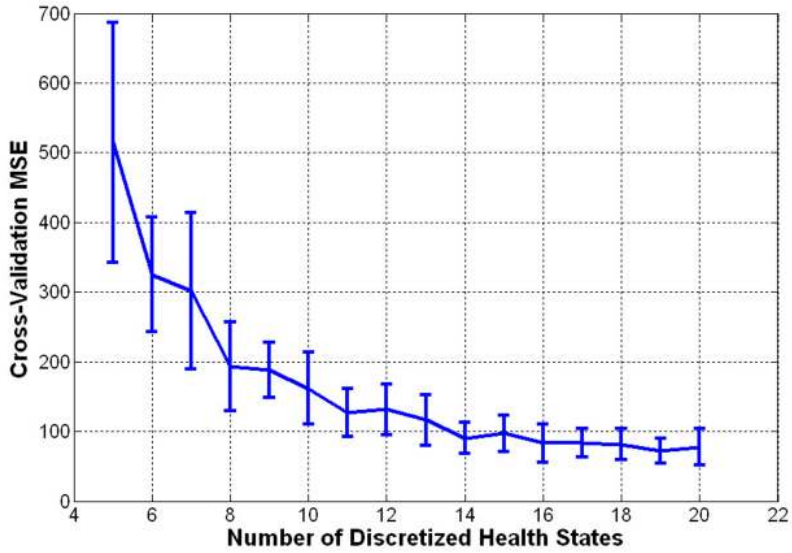


Figure 2.7: Cross-Validation results for PSHMCO with different number of discretized health states (hidden state values).

18SC3 and *33PN6*). In the diagnosis process, at each time step (cut), the input data from the beginning of the diagnosis process up to the current time step is available to all models in order to predict the tool wear of the three flutes at the current time step, T.

The required parameters for PSHMCO are estimated from the training dataset based on equations (2.7), (2.8) and (2.9). Given the observation sequences, $O_{1:T}$, of each flute, $\hat{y}_T$ is computed for each flute based on (2.21). The maximum value of the computed $\hat{y}_T$ among the three flutes at each time step T is regarded as the ultimate output in all approaches and is compared with the maximum actual wearing value of 3 flutes at T. Number of health states in PSHMCO are set to 14 as suggested in the cross-validation stage. Fig. 2.8 depicts the adopted parameter values in PSHMCO approach.

At each time step within each experiment, the maximum wearing value among 3 flutes is used as the desired outcome. The rationale is that, the quality of the ultimate work-piece is determined by the maximum tool wear of the three flutes. Figure 2.9 depicts the predicted ultimate output of all approaches along with the actual outcome (true output) for one of the testing experiments (*18SC3*). Table 2.3 shows the diagnosis error rate of all three approaches on the testing set in terms of MSE in details as well as the overall diagnosis accuracy of the approaches in terms of MSE and MRE. It is

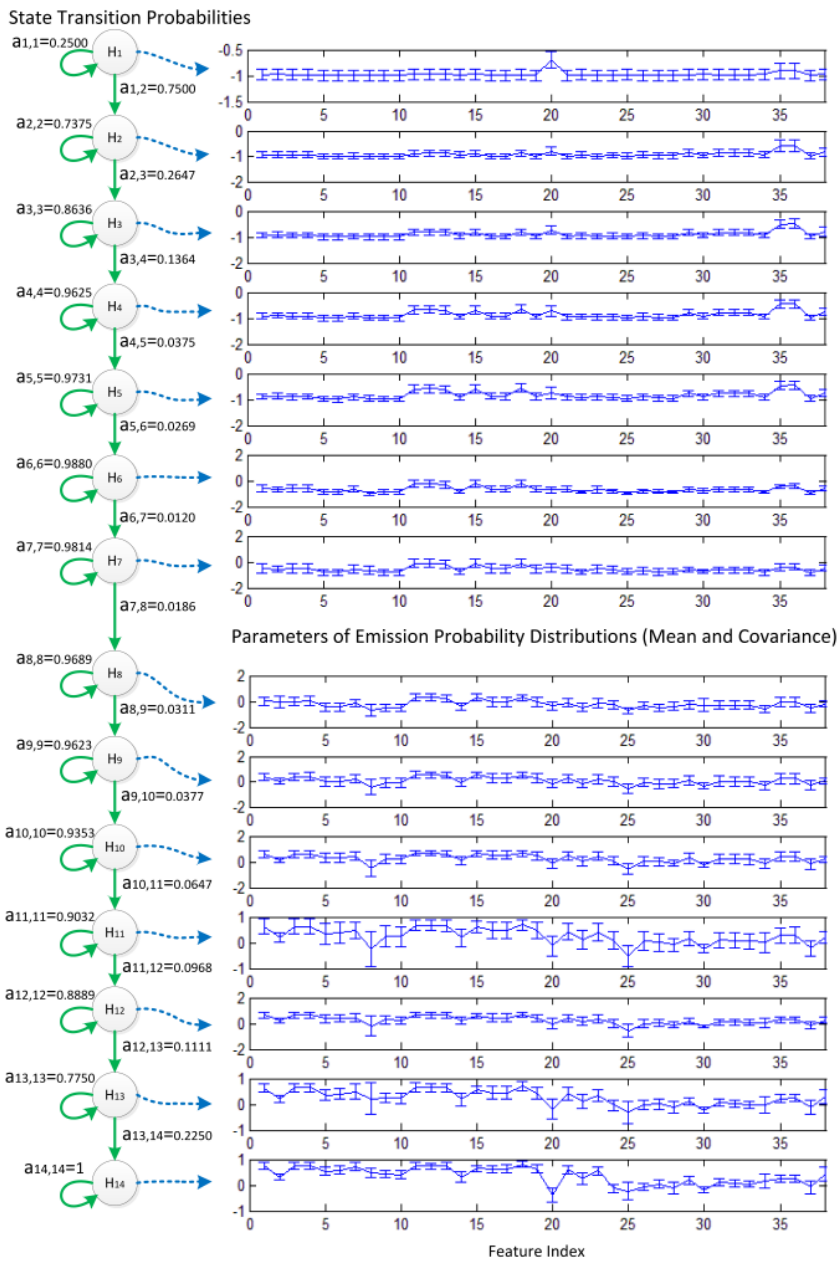


Figure 2.8: State transition probabilities and parameters of emission probability distributions depicted in parallel coordinates (assume that off-diagonal elements of the covariance matrices are zero so that the inter-feature correlations can be discarded and they can be simply visualized in parallel coordinates).

noteworthy that the provided accuracies of the NNs are the averaged accuracies over 10 trials.

As it can be seen in Table 2.3, Elman network outperforms the conventional MLP

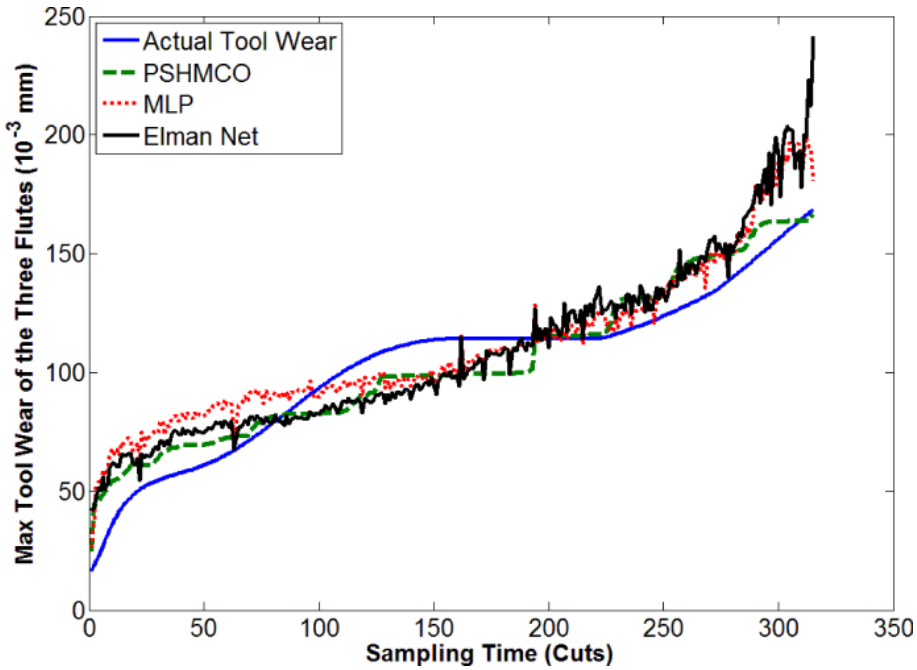


Figure 2.9: Predicted output of PSHMCO, MLP and Elman Network along with the true output for the data collected from cutter *18SC3*.

approach, which indicates that the underlying temporal information for TCM cannot be ignored. Moreover, the PSHMCO approach outperforms both MLP and Elman network, which suggests that PSHMCO approach is stronger in capturing the temporal information for TCM comparing to the Elman network.

It is noteworthy that the approaches are not applied online in this study. The data has collected and stored so that the same data can be provided to test different approaches

Table 2.3: Prediction error rate in diagnostics task using PSHMCO , multi-layer perceptron (MLP) and Elman network (Elm-Net) in terms of MSE and MRE.

Model	Mean Square Error			Total Testing Error Indicators	
	09BX3	*18SC3*	*33PN6*	MSE	MRE
PSHMCO	341.2336	135.3554	297.5691	258.0527 ± 108.4787	0.1164 ± 0.0082
MLP	778.47574	283.43268	436.13632	499.3482 ± 253.5029	0.1662 ± 0.0383
Elm-Net	580.4041	277.9204	324.8272	394.3839 ± 165.8576	0.1470 ± 0.0330

and conduct fair comparisons. However, the online application is simulated. In order to simulate time progression, data is sequentially provided to the prediction models each time up to a simulated current time step, and then the prediction is made. In this study, the average computation time for diagnostics through time by PSHMCO is measured to be 2.6ms, with 14 health states and 38 input features (observations) at each sampling time. Thus, it suggests that the PSHMCO approach is computationally feasible for the similar online applications.

2.5.3 Prognostic Results

In this test, PSHMCO is unrolled over the time axis for prognostics. Different prediction horizons are explored and the mean of corresponding prediction accuracies over the testing set are provided in Table 2.4 in terms of MSE.

Similar to diagnostics, the maximum value among the three flutes at each time step is used as the desired output. Having the prediction horizon, t_p, and given the observation sequences from beginning of the experiment up to T, the task is to predict the desired output at time $t' = T + t_p$ and this can be computed using (2.23). Fig. 2.10 shows the prediction results by PSHMCO along with the real tool wear on one of the test experiments (*18SC3*) with different prediction horizons, which are 1, 4, 7 and 10 time steps. The starting point is set to 11.

Table 2.4: Prognosis (mean) total accuracy for PSHMCO in terms of MSE with different prediction horizons on the testing set.

Model	Prediction Horizon (Time Steps Ahead)									
	1	2	3	4	5	6	7	8	9	10
PS-HMCO	422.31	491.46	602.73	680.12	732.52	771.17	802.26	829.00	852.78	874.34

From Table 2.4 and Fig. 2.10, it can be seen that, although the overall accuracy has reduced comparing to the diagnostics (which is expected), interestingly the resultant prognostics accuracy when the prediction horizon is equal to 1 or 2, still outperforms the

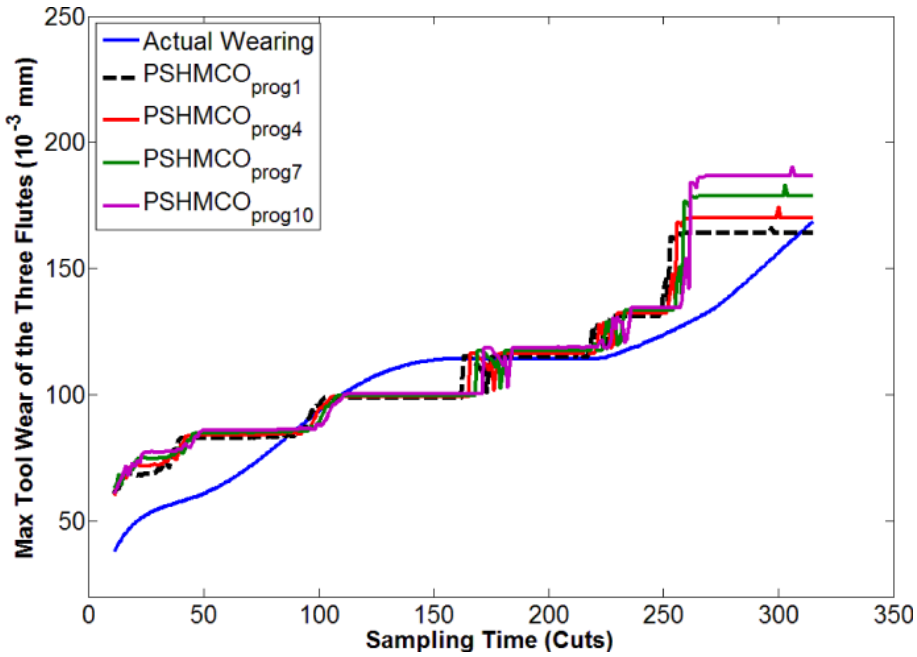


Figure 2.10: Prognostics results of PSHMCO model with different prediction horizons on the data collected from the cutter *18SC3*.

MLP in diagnostics, which shows the importance of the temporal information. Although the PSHMCO's prognosis accuracy is unable to beat the Elman network diagnostics result, since Elman network also captures temporal information, the results confirms the power of PSHMCO in capturing the temporal information, and the suitability of PSHMCO as a predictor.

2.6 Summary

In this Chapter, a temporal probabilistic approach is proposed for the continuous tool condition monitoring in machinery systems. The proposed PSHMCO approach is based on a physically segmented hidden Markov model that can handle continuous output. As an illustrative example, the proposed approach is applied to tool wear prediction in a CNC-milling machine.

The experimental study indicates that PSHMCO outperforms multi-layer perceptron and the Elman networks in tool condition monitoring. It is also shown that the proposed approach can be used for prognostics by unrolling the model over the time horizon. The

PSHMCO approach is found to be suitable for the applications in which the operating conditions are fixed. The fixed operating conditions property, similar to the conducted experiment, can be seen in applications with high volume of productions such as molding processes.

In succeeding Chapters, some of the restrictions that are enforced in the model must are lifted to improve the performance further. Moreover, as completely different regiments would arise when different operating conditions are being considered, multimodal approaches can be studied. Switching models that can each capture a specific regiment and increase the resolution within each trend can be applied to improve the prediction performance further.

Chapter 3

Hidden Semi-Markov Model-based Approach

3.1 Introduction

As it is stated in Chapter 2, to improve the prediction performance of the proposed PSHMCO approach even further, some of the restrictions that are imposed based on the Hidden Markov model (HMM) may be lifted. One of the deficiencies of hidden Markov model in modeling real-world applications, is its unrealistic fixed state-duration distribution, which is a *geometric* distribution. The state-duration distribution indicates the probability of staying in one state for different possible durations. Therefore in the HMM-based approach, having a fixed state-duration distribution may lead to unsatisfactory prediction results in cases that the assumption of having a geometric state-duration distribution does not hold. However, it is known that lots of processes in nature are not abiding geometric distribution [105].

In this chapter, a Hidden-Semi Markov Model (HSMM)-based approach is introduced to address the aforementioned deficiency in HMM-based approach and improve the prediction performance. Another issue that is addressed in this Chapter is how to incorporate an asymmetric loss function into the proposed approach, to consider the dramatic cost differences between under- and over-estimation of the tool condition. To this end, having flexible duration distributions in the HSMM, it is attempted to mod-

ify the overall skewness of the duration distributions based on a given asymmetric loss function and training dataset.

This Chapter is organized as follows. In Section 2, a more complex approach compared to PSHMCO, named a physically segmented hidden semi-Markov model with continuous output (PSHsMCO) is proposed for continuous TCM. Then a computationally efficient version of forward-backward algorithm for inference in the PSHsMCO as well as state estimation variables are described in the same Section. Diagnosis and Prognosis procedures based on efficient forward-backward algorithm in PSHsMCO are discussed in Section 3. Afterward, simulation results based on the experimental data are provided and compared in Section 4. Moreover, the possibility of incorporating an asymmetric loss function in PSHsMCO is explored and evaluated by simulation results in Section 5. Finally, the Chapter is concluded in Section 6.

3.2 Hidden Semi-Markov Model-Based Approach

The idea behind hidden semi-Markov model is to use temporal data in a more efficient way than in HMMs by keeping track of duration of staying in each state [106].

3.2.1 HMM Fixed Duration Distribution

In the HMM used in PSHMCO approach proposed in the preceding chapter, as depicted in Fig. 3.1, there are only two possible transitions (excluding the last state which is modeled as an attraction point) from each health state H_i, either staying in the same wearing state H_i with probability of p_i or going to the next wearing state H_{i+1} with the probability of $1 - p_i$. Thus, the probability of staying in H_i for exact d time steps in the utilized HMM can be calculated as

$$P(\text{staying in H}_\text{i} \text{ for exact d time steps}) = p_i^d \times (1 - p_i) \tag{3.1}$$

in which p_i is the probability of self-transition at ith health state H_i. As it can be understood from (3.1), this probability has a *geometric* distribution. However, it is known that lots of processes in nature are not abiding geometric distribution [105]. Therefore, in

order to add more flexibility to the state-duration probability distribution function, duration factor is added to HMM and transition probabilities are redefined based on these durations.

3.2.2 Formulation and Parameter Estimation

In the proposed HSMM-based approach, duration of remaining in each health state will be treated as a random variable and its corresponding parameters would be added to the parameter set needed to be determined for the model in the training (parameter estimation) procedure. Furthermore, for simplicity, the duration variables are assumed to have normal distributions.

Similar to PSHMCO in Chapter 2, firstly the measured tool wearing of the cutters in the training set are uniformly discretized into m ordinal classes (clusters) $\{H_1, \ldots, H_m\}$. These ordinal classes corresponds to various health stages of the cutters from initial wearing conditions to severely worn out. Moreover, the labels of these classes are real-valued and correspond to the mean wearing value in each class. These real valued labels are later used as the values that hidden states of both HSMM and HMM can take at each time step.

Having all the HMM parameters except the transition probabilities as defined in Chapter 2 for realizing PSHMCO approach, the additional parameters that are required to be defined to formulate the HSMM utilized in PSHsMCO are as follows

- d_i is defined as a random variable having a normal distribution that corresponds to duration distribution in the ith health state (hidden state). Consequently, two more

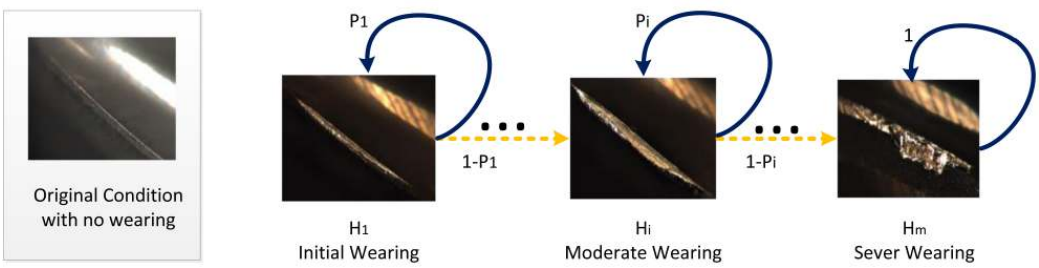


Figure 3.1: Schematic transition graph of the HMM utilized in PSHMCO approach.

parameters would be added to the parameter set for each health state, μ_{d_i} and σ_{d_i} as the mean and standard deviation of the duration variable d_i at each health state.

- For implementation purposes (d_{max}) is also needed to be defined as the overall maximum possible duration.

Similar to the HMM utilized in PSHMCO approach, in each time step, the implemented HSMM has only two possible options of either remaining at the same health state as the previous time step or proceeding to the next health state. The probability of remaining in the same state would be generated using the duration distribution, and the transition probability would only be used at the time that the duration of staying at one health state is over and the health state is going to change. That is why this model is called Hidden Semi Markov Model, since it only uses the Markov Transition when the duration of staying in one state is over [106]. Therefore, the transition matrix, A would be different from the PSHMCO transition matrix defined in Chapter 2. Out of two possible transitions at each time step, the only valid transition after the duration of staying in one state has passed is the option of going to the next state. This assumption corresponds to each row of the transition matrix, A, having only one non-zero element, which indicates the transition to the next wearing stage. Figure 3.2 schematizes the utilized HSMM transition graph.

Based on the described assumptions and the duration distributions to find the probability of staying at each health state, the parameter set and the transition matrix of the

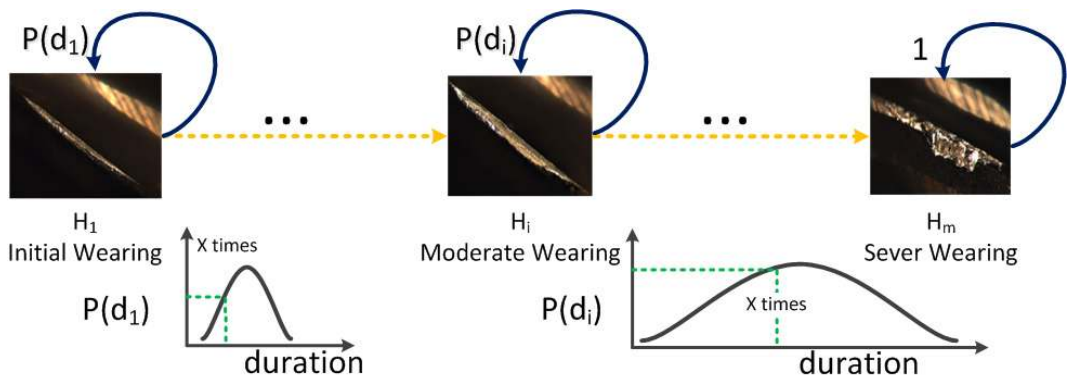


Figure 3.2: Schematic transition graph of the utilized HSMM in PSHsMCO where the ith state-duration probability distribution is denoted by $P(d_i)$.

HSMM utilized in PSHsMCO can be written as follows

$$A = \left[a_{ij}\right]_{m\times m} = \begin{bmatrix} 0 & 1 & 0 & 0 & ... & 0 \\ 0 & 0 & 1 & 0 & ... & 0 \\ . & & . & & & . \\ . & & & . & & . \\ . & & & & . & . \\ 0 & 0 & \dots & 0 & 0 & 1 \\ 0 & 0 & 0 & ... & 0 & 1 \end{bmatrix}, \tag{3.2}$$

$$\lambda = \{\underbrace{\pi_0, \mu_1, ..., \mu_m, \Sigma_1, ..., \Sigma_m}_{HMM\ Parameters}, \underbrace{\mu_{d_1}, \dots, \mu_{d_m}, \sigma_{d_1}, \dots, \sigma_{d_m}, d_{max}}_{Duration\ Distribution\ Parameters}\}, \tag{3.3}$$

where A is the transition matrix, π_0 is the prior probability distribution of the initial health state, μ_i and Σ_i are respectively mean and covariance matrices being used in the Gaussian distributions to compute the emission probability at time t given the fact that $S_t = H_i$, μ_{d_i} and σ_{d_i} are the parameters of the Gaussian distribution representing the duration variable at the ith health state denoted as d_i. All the required parameters can be identified using maximum log-likelihood approach on the complete dataset that is provided as the training set.

Assuming that all experiments are independently distributed (as the experiments are conducted independently), joint probability of observing training sequences given the parameters of the HSMM can be written as follows

$$\begin{aligned} &P(D_1, D_2, \dots, D_n|\lambda) = \prod_{q=1}^{n} P(D_q|\lambda) \\ &\text{where } D_q = \{O^q_{1:T_q}, S^q_{1:T_q}\},\ O^q_{1:T_q} = \{O^q_1, \dots, O^1_{T_q}\},\ S^q_{1:T_q} = \{S^q_1, \dots, S^1_{T_q}\} \\ &P(D_q|\lambda) = \pi_0(S^q_1) \prod_{i=1}^{m} P(S^q_{t^q_i}|S^q_{t^q_i-1}) P((S^q_{t^q_i}, \tau_{t^q_i}) = (H_i, k^q_i)) P(O^q_{t^q_i:t^q_{i+1}-1}|S^q_{t^q_i:t^q_{i+1}-1}) \\ &t^q_i = \begin{cases} 1 + \sum_{j=1}^{i-1} k^q_j & For\ i > 1 \\ 1 & For\ i = 1 \end{cases} \end{aligned} \tag{3.4}$$

where D_q is the qth experiment, n is the number of experiments, t^q_i is the starting time step of the ith health stage in each specific sequence, m is the number of health states, $\tau_{t^q_i}$ is the remaining duration of staying at the corresponding state in the pair $(S^q_{t^q_i}, \tau_{t^q_i})$, k^q_j

is the observed duration of the jth health state in each specific sequence, $O_{t_i^q:t_{i+1}^q-1}$ is the observations from the time step entering to the ith state till transition to the next state in the qth experimental data and $S_{t_i^q:t_{i+1}^q-1}$ indicates the state sequence during that time. Assuming conditional independence between the observations in one sequence given their corresponding states and the duration variables following Gaussian distributions, (3.4) can be written as

$$\begin{aligned} P(D_q|\lambda) &= \pi_0(S_1^q)\prod_{i=1}^{m} P(S_{t_i^q}^q|s_{t_i^q-1}^q)\frac{1}{\sqrt{2\pi\sigma_{d_i}^2}}\exp(-\frac{(k_i^q-\mu_{d_i})^2}{2\sigma_{d_i}^2})\prod_{r=t_i^q}^{t_{i+1}^q-1} P(O_r^q|S_r^q), \\ t_i^q &= \begin{cases} 1+\sum_{j=1}^{i-1}k_j^q & For\ i>1 \\ 1 & For\ i=1 \end{cases}. \end{aligned} \tag{3.5}$$

Furthermore, assuming the emission probability is a *multivariate Gaussian variable*, (3.5) can be written as,

$$\begin{aligned} P(D_q|\lambda) =& \pi_0(S_1^q)\prod_{i=1}^{m}\frac{1}{\sqrt{2\pi\sigma_{d_i}^2}}\exp(-\frac{(k_i^q-\mu_{d_i})^2}{2\sigma_{d_i}^2}) \\ &\times\prod_{r=t_i^q}^{t_{i+1}^q-1}\frac{1}{(2\pi)^{\chi/2}|\Sigma_i|^{1/2}}\exp(-\frac{1}{2}(O_r^q-\mu_i)^T\Sigma_i^{-1}(O_r^q-\mu_i)), \end{aligned} \tag{3.6}$$

where χ is the dimension of observation (input) space. Finally based on (3.4) and (3.6), the log-likelihood of the joint probability of $P(D_{1:n}|\lambda)$ can be written as follows

$$\begin{aligned} \bar{L} &= \sum_{q=1}^{n} L_q = \sum_{q=1}^{n}\ln(P(D_q|\lambda)), \\ L_q &= \ln(P(D_q|\lambda)) = \ln(\pi_0(S_1^q)) - \sum_{i=1}^{m}[\frac{1}{2}\ln(2\pi)+\ln(\sigma_{d_i})+\frac{(k_1^q-\mu_{d_1})^2}{2\sigma_{d_1}^2}] \\ &- \sum_{i=1}^{m}[\frac{k_i^q\chi}{2}\ln(2\pi)+\frac{k_i^q}{2}\ln(|\Sigma_i|)+\sum_{t=t_i^q}^{t_{i+1}^q-1}\frac{1}{2}(O_r^q-\mu_i)^T\Sigma_i^{-1}(O_r^q-\mu_i)]. \end{aligned} \tag{3.7}$$

Now, by setting the partial derivatives of the log-likelihood to zero, the parameters of the HSMM can be found as follows

$$\frac{\partial\bar{L}}{\partial\sigma_{d_i}} = 0 \rightsquigarrow \sum_{q=1}^{n}[\frac{-1}{\sigma_{d_i}}+\frac{(k_i^q-\mu_{d_i})^2}{\sigma_{d_i}^3}] = 0 \rightsquigarrow \sigma_{d_i} = \sqrt{\frac{\sum_{q=1}^{n}(k_i^q-\mu_{d_i})^2}{n}} \tag{3.8}$$

$$\frac{\partial \bar{L}}{\partial \mu_{d_i}} = 0 \rightsquigarrow \sum_{q=1}^{n} \frac{(k_i^q - \mu_{d_i})}{\sigma_{d_i}^2} = 0 \rightsquigarrow \mu_{d_i} = \frac{\sum_{q=1}^{n} k_i^q}{n} \tag{3.9}$$

$$\begin{aligned}
&\frac{\partial \bar{L}}{\partial \Sigma_i} = 0 \rightsquigarrow -\sum_{q=1}^{n}[\frac{k_i^q}{2}\frac{1}{|\Sigma_i|}|\Sigma_i|(\Sigma_i^{-1})^T - \sum_{t=t_i^q}^{t_{i+1}^q-1}\frac{1}{2}\Sigma_i^{-1}(O_t^q-\mu_i)(O_t^q-\mu_i)^T\Sigma_i^{-1}] \\
&\rightsquigarrow -\sum_{q=1}^{n}\frac{k_i^q}{2}\Sigma_i^{-1} + \Sigma_i^{-1}[\frac{1}{2}\sum_{q=1}^{n}\sum_{t=t_i^q}^{t_{i+1}^q-1}(O_t^q-\mu_i)(O_t^q-\mu_i)^T]\Sigma_i^{-1} = 0 \\
&\rightsquigarrow \sum_{q=1}^{n}\frac{k_i^q}{2} = [\frac{1}{2}\sum_{q=1}^{n}\sum_{t=t_i^q}^{t_{i+1}^q-1}(O_t^q-\mu_i)(O_t^q-\mu_i)^T]\Sigma_i^{-1} \\
&\rightsquigarrow \Sigma_i = \frac{\sum_{q=1}^{n}\sum_{t=t_i^q}^{t_{i+1}^q-1}(O_t^q-\mu_i)(O_t^q-\mu_i)^T}{\sum_{q=1}^{n}k_i^q}
\end{aligned} \tag{3.10}$$

$$\begin{aligned}
&\frac{\partial \bar{L}}{\partial \mu_i} = 0 \rightsquigarrow \sum_{q=1}^{n}\sum_{t=t_i^q}^{t_{i+1}^q-1}\frac{1}{2}[\Sigma_i^{-1} + \Sigma_i^{-T}](O_t^q-\mu_i)(-1) = 0 \\
&\rightsquigarrow \sum_{q=1}^{n}\sum_{t=t_i^q}^{t_{i+1}^q-1}\Sigma_i^{-1}(O_t^q-\mu_i) = 0 \rightsquigarrow \sum_{q=1}^{n}\sum_{t=t_i^q}^{t_{i+1}^q-1}O_t^q - \mu_i\sum_{q=1}^{n}k_i^q = 0 \\
&\rightsquigarrow \mu_i = \frac{\sum_{q=1}^{n}\sum_{t=t_i^q}^{t_{i+1}^q-1}O_t^q}{\sum_{q=1}^{n}k_i^q}
\end{aligned} \tag{3.11}$$

Using the formulas in (3.8), (3.9), (3.10) and (3.11) all of the required parameters can be estimated. In this implementation of HSMM, the initial probability π_0 and d_{max} are defined as follows

$$\pi_0(i) = \frac{1}{m},\ i = 1, \ldots, m,$$

$$d_{max} = \max_i\{\mu_{d_i} + 2\sigma_{d_i}\},$$

where d_{max} is used as the upper bound limit of duration variables and π_0 is defined as uniform distribution (similar to PSHMCO approach) for the case than no prior knowledge about the general initial state value is available. In this implementation of the

HSMM, an $m \times d_{max}$ probability distribution matrix is computed and stored based on the probability distributions parameterized by the mean and variance of the duration within each health state. The ith row, jth column element of the probability distribution matrix indicates the probability of staying in the ith health state for jth time steps computed by the relevant normal distribution. Each row of the probability distribution matrix is later normalized by the summation of its elements. Thus, the duration of staying in each health state is regarded as a bounded positive integer that may be modeled by a truncated normal distribution.

After estimating the parameters of the HSMM based on the training set using formulas in (3.8)-(3.11), the health states for a newly given data (test data) may be estimated by inference algorithms. The inference in HSMM can be done by means of Forward-Backward algorithm. An Efficient Forward-Backward algorithm for HSMM (Explicit-Duration Hidden Markov Model) is introduced in [107]. Here, a similar Forward-Backward algorithm is used to do inference in the implemented HSMM for TCM.

3.2.3 Forward-Backward variables in PSHsMCO

As mentioned before, Forward-Backward algorithm is a recursive algorithm that can be used in Markov models to answer inference problems in them. A computationally simplified implementation of this algorithm for tool condition monitoring using PSHsMCO is introduced in this part. This algorithm uses two auxiliary variables called Forward and Backward variables to compute the required probabilities, recursively.

Forward variable is defined as the joint probability of being at ith state at time t and remaining in that state for the next k steps while observing the inputs from time step 1 up to t. According to [107], forward variable can be formulated as

$$\alpha_t(i,k) \triangleq P(O_{1:t}, (S_t, \tau_t) = (H_i, k)) \tag{3.12}$$

where $O_{1:t} = \{O_1, O_2, \ldots, O_t\}$ are the newly given observations (inputs) from time step 1 to t, (S_t, τ_t) is the pair of hidden state of the model at time step t and its remaining duration τ_t at that state from time step t onward. H_i is the ith health state value that the state variable of the model can take, $H_i \in \{H_1, H_2, \ldots, H_m\}$. From law of total

probability, (3.12) can be computed as follows

$$\alpha_t(i,k) = \alpha_{t-1}(i,k+1)P(O_t|S_t = H_i) + \sum_{j=1,\ j\neq i}^{m} \alpha_{t-1}(j,1)a_{i,j}P(O_t|S_t = H_i)P((S_t,\tau_t) = (H_i,k)), \tag{3.13}$$

where m is the number of health state values, $a_{i,j}$ is the transition probability from the ith health state (H_i) to jth health state (H_j) and d_i is the duration of remaining in i health state (H_i). The initial condition for the recursive equation in (3.13) is

$$\alpha_1(i,k) = \pi_0(i)P(O_1|S_1 = H_i)P((S_1,\tau_1) = (H_i,k)). \tag{3.14}$$

Based on the assumptions that are made about graduality of the wearing process and consequently the possible transitions in the implemented model, (3.13) can be simplified as follows

$$\alpha_t(i,k) = \begin{cases} \alpha_{t-1}(i,k+1)P(O_t|S_t = H_i) & For\ i > 1 \\ +\alpha_{t-1}(i-1,1)P(O_t|S_t = H_i)P((S_t,\tau_t) = (H_i,k)) & \\ \\ \alpha_{t-1}(i,k+1)P(O_t|S_t = H_i) & For\ i = 1 \end{cases} \tag{3.15}$$

The second auxiliary variable to be defined is called Backward variable. According to [107], the backward variable is defined as

$$\beta_t(i,k) \triangleq P(O_{t+1:T}|(S_t,\tau_t) = (H_i,k)). \tag{3.16}$$

where $(S_t,\tau_t) = (H_i,k)$ indicates that the health status is supposed to remain in this state for the next k time steps and then will transit to another state $j,\ j \neq i$. Therefore, $\beta_t(i,k)$ can be written as

$$\beta_t(i,k) = \begin{cases} P(O_{t+1}|S_{t+1} = H_i)\beta_{t+1}(i,k-1) & For\ k > 1 \\ \\ \sum\limits_{j=1,\ j\neq i}^{m} a_{i,j}P(O_{t+1}|S_{t+1} = H_j) & For\ k = 1 \\ \times \sum\limits_{d\geq 1} P((S_{t+1},\tau_{t+1}) = (H_j,d))\beta_{t+1}(j,d)] & \end{cases}. \tag{3.17}$$

Since (3.17) is a backward recursive formula, its initial condition has been defined as,

$$\beta_T(i,k) = 1,\ \ k \geq 1. \tag{3.18}$$

Based on the assumptions that are made on the graduality of the wearing process and consequently the model's specific form of transition matrix in our implemented HSMM, the *backward* variable can be simplified further and be rewritten as

$$\beta_t(i,k) = \begin{cases} P(O_{t+1}|S_{t+1} = H_i)\beta_{t+1}(i,k-1) & For\ k > 1 \\ \\ P(O_{t+1}|S_{t+1} = H_{i+1}) & For\ k = 1 \\ \times \sum_{d=1}^{d_{max}} P((S_{t+1},\tau_{t+1}) = (H_{i+1},d))\beta_{t+1}(i+1,d) & \end{cases}. \tag{3.19}$$

3.2.4 State Estimation

In order to estimate the state values of the state variables at each time step based on the observations using the HSMM model, three auxiliary variables are further defined. Similar to auxiliary variables defined in [107], these variables would help to simplify the state estimation computation for either current time (diagnosis) or future (prognosis). The three variables are defined and further simplified based on the assumptions that are made in the TCM problem statement.

The first variable to be defined is $\zeta_t(i)$, which is the joint probability of observing $O_{1:T}$ and transition from ith health state (H_i) to its next health state at time t. $\zeta_t(i)$ can be written as

$$\zeta_t(i) \triangleq P(O_{1:T}, S_{t-1} = H_i, S_t = H_{i+1}) \tag{3.20}$$

The joint probability in (3.20) can be calculated as follows

$$\zeta_t(i) = \alpha_{t-1}(i,1)(1-a_{i,i})P(O_t|S_t = H_i)\sum_{k\geq 1} P((S_t,\tau_t) = (H_{i+1},k))\beta_t(i+1,k) \tag{3.21}$$

where $P((S_t,\tau_t) = (H_{i+1},k))$ is the probability of staying at H_{i+1} for the next k time steps.

The second auxiliary variable that can be used for state estimation of S_t based on the whole observation sequence $O_{1:T}$ is $\gamma_t(i)$. $\gamma_t(i)$ is the joint probability of $O_{1:T}$ and $S_t = H_i$ that can be written as

$$\gamma_t(i) \triangleq P(O_{1:T}, S_t = H_i). \tag{3.22}$$

Moreover, based on law of total probability and considering the possible transitions in

the implemented HSMM, $P(O_{1:T}, S_t = H_i, S_{t+1} = H_i)$ can be written as follows

$$\begin{aligned} P(O_{1:T}, S_t = H_i, S_{t+1} = H_i) &= P(O_{1:T}, S_t = H_i) - P(O_{1:T}, S_t = H_i, S_{t+1} = H_{i+1}) \\ &= P(O_{1:T}, S_{t+1} = H_i) - P(O_{1:T}, S_t = H_{i-1}, S_{t+1} = H_i) \end{aligned} \tag{3.23}$$

Consequently, based on (3.23) and using (3.21), a backward recursive formula can be derived for $\gamma_t(i)$ as

$$\gamma_t(i) = \gamma_{t+1}(i) + \zeta_{t+1}(i) - \zeta_{t+1}(i-1), \tag{3.24}$$

and its initial condition can be calculated as

$$\gamma_T(i) = \sum_{k=1}^{d_{max}} \alpha_t(i,k). \tag{3.25}$$

The third auxiliary variable is $\xi_{t'}(i,k)$. In order to find the probability of being at one state in future, the model must be unrolled on the time horizon while there are no more observations. Therefore, similar to the forward variable defined in (3.12), $\xi_{t'}(i,k)$ is defined as a form of forward variable for t' (where $t' > T$) that only considers the observations up to time step T, as the time steps further in time are not observed yet. It can be written in a recursive form using law of total probability as follows

$$\begin{aligned} &\xi_{t'}(i,k) \triangleq P(O_{1:T}, (S_{t'}, \tau_{t'}) = (H_i, k)), \quad \forall t' > T \\ &\xi_{t'}(i,k) = \xi_{t'-1}(i, k+1) + \xi_{t'-1}(i-1, 1) \times P((S_{t'}, \tau_{t'}) = (H_i, k)) \\ &\text{initial condition } \xi_T(i,k) = \alpha_T(i,k). \end{aligned} \tag{3.26}$$

3.3 Diagnostics & Prognostics

As mentioned in the preceding Chapters, diagnosis is the task of predicting the health state at time T given all the observations from time step 1 to T. In the realm of Bayesian Networks, this task is called *filtering* or *monitoring* and it can be written in a probabilistic manner as follows

$$P(S_T = H_i | O_{1:T}, \lambda) = \frac{P(O_{1:T}, S_T | \lambda)}{P(O_{1:T} | \lambda)} \rightsquigarrow P(S_T = H_i | O_{1:T}, \lambda) = \frac{\gamma_T(i)}{P(O_{1:T} | \lambda)}. \tag{3.27}$$

Therefore, probability of being at each health state at time T can be calculated using (3.27) and (3.25). It is also worth mentioning that the denominator in (3.27) is not

required to be calculated. It is a normalizing factor that can be calculated after finding $\gamma_T(i)$ for all the health states as

$$\sum_{i=1}^{m} P(S_T = H_i|O_{1:T}, \lambda) = 1 \rightsquigarrow \frac{1}{P(O_{1:T}|\lambda)} \sum_{i=1}^{m} \gamma_T(i) = 1 \rightsquigarrow P(O_{1:T}|\lambda) = \sum_{i=1}^{m} \gamma_T(i). \tag{3.28}$$

Consequently, based on (3.25) and (3.28), (3.27) can be rewritten as

$$P(S_T = H_i|O_{1:T}, \lambda) = \frac{\gamma_T(i)}{\sum_{i=1}^{m} \gamma_T(i)} = \frac{\sum_{k=1}^{d_{max}} \alpha_T(i,k)}{\sum_{i=1}^{m} \sum_{k=1}^{d_{max}} \alpha_T(i,k)}. \tag{3.29}$$

Finally, the continuous output of the model which corresponds to the expected amount of tool wear at the current time step T, can be calculated based on (3.29)as follows

$$\hat{y}_T = \sum_{i=1}^{m} P(S_T = H_i|O_{1:T}, \lambda) \times H_i. \tag{3.30}$$

Furthermore, prognosis is the task of predicting the future health state at time t' ($t' > T$) while the observation data is only available up to the current time T. Similar to the diagnosis case, prognosis can be written in a probabilistic manner as

$$P(S_{t'} = H_i|O_{1:T}, \lambda) = \frac{P(O_{1:T}, S_{t'} = H_i|\lambda)}{P(O_{1:T}|\lambda)} = \frac{\sum_{k=1}^{d_{max}} P(O_{1:T}, (S_{t'}, \tau_{t'}) = (H_i, k))}{P(O_{1:T}|\lambda)}. \tag{3.31}$$

Using $\xi_{t'}(i,k)$ defined in (3.26), (3.31) can be rewritten and computed as

$$P(S_{t'} = H_i|O_{1:T}, \lambda) = \frac{\sum_{k=1}^{d_{max}} \xi_{t'}(i,k)}{P(O_{1:T}|\lambda)} = \frac{\sum_{k=1}^{d_{max}} \xi_{t'}(i,k)}{\sum_{i=1}^{m} \sum_{k=1}^{d_{max}} \xi_t(i,k)}. \tag{3.32}$$

Similar to (3.28), $P(O_{1:T}|\lambda)$ is a normalizing factor that is replaced by $\sum_{i=1}^{m} \sum_{k=1}^{d_{max}} \xi_t(i,k)$.

After computing $P(S_{t'} = H_i|O_{1:T}, \lambda)$, the continuous output $\hat{y}_{t'}$ (where $t' > T$) of the model, which corresponds to the expected amount of tool wear at time step t' in future, can be calculated based on (3.32) as follows

$$\hat{y}_{t'} = \sum_{i=1}^{m} P(S_t = H_i|O_{1:T}, \lambda) \times H_i. \tag{3.33}$$

3.4 Diagnostics and Prognostics Results

In this section performance of single HSMM-based approach called PSHsMCO and single HMM-based approach named PSHMCO are compared in diagnostics and prognostics of the cutter's wearing metric in a CNC-milling machine. The dataset and the features that are used to conduct the comparative study in this section are identical to ones provided in Chapter 2.

The performance of the aforementioned approaches are compared in three cases. Case I, can be regarded as a cross-validation phase conducted to identify an appropriate number of hidden state values for both PSHMCO and PSHsMCO. After adopting an appropriate number of hidden state values, performance of the two approaches are compared for diagnostics and prognostics based on mean squared error (MSE) in Cases II and III, respectively. A short description of the three Cases are as follows

- Cross-validation: this can be done in two modes, i.e. *leave one flute out* and *leave one experiment out*.
- Testing diagnosis ability: testing the model for diagnostics on the experiment that are excluded from the training set.
- Testing prognostics: using the model for prognostics purposes by unrolling the model over the time horizon (this task can be done with different prediction horizons).

3.4.1 Cross-Validation Results

As mentioned before, two modes can be considered for cross-validation (CV) in case of condition monitoring for the cutters with multiple flutes i.e. *leave one flute out* and *leave one experiment out*. Similar to CV in Chapter 2, since the training set is limited to 3 experiments, *leave one flute out* is conducted. Which means in both approaches, both models are trained on two flutes out of three flutes in each experiment of the training set and then tested on the excluded flute. Different number of hidden state values are explored for the two approaches. Figure 3.3 depicts the cross validation error of the

excluded flute using the two approaches with various number of possible hidden state values.

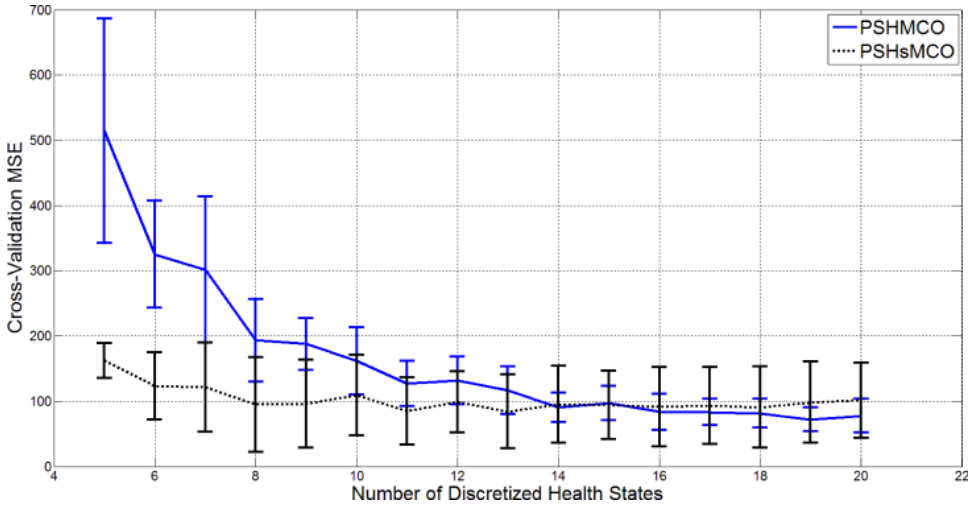


Figure 3.3: Cross-validation error rate in both PSHMCO and PSHsMCO with different number of hidden state values (health states).

From Fig. 3.3, it can be understood that 14 that is selected in Chapter 2 as an appropriate number of health states for PSHMCO sounds to be also an appropriate number of health states HSMM-based approach. Therefore, number of hidden state values (health states) of both approaches are set to 14 in the succeeding cases. In addition, Fig. 3.3 shows that PSHsMCO has generally outperformed PSHMCO approach in terms of average error rate in cross-validation case with various number of hidden state values.

3.4.2 Diagnostics Results

This case is to compare the diagnosis accuracy of the two approaches on the testing set (the data collected from the three cutters *09BX3*, *18SC3* and *33PN6*), which are excluded from the training set data collected from cutters (*07BX1*, *31PN4* and *34PT1*). As mentioned in Chapter 2, to simulate the online experiments using the stored experimental data, at each time step (cut), the data from the beginning of the experiment up to the current time step is given to both approaches in order to predict the tool wear of the cutter at that time step. As suggested in Case I, number of health states is adopted to be 14 for both approaches. Table 3.1 shows the diagnosis error of the two approaches on

the testing set.

Table 3.1: Prediction error rate in diagnosis using PSHMCO and PSHsMCO approaches in terms of MSE.

Approach	Mean Square Error			
	$09BX3$	$18SC3$	$33PN6$	**Total**
PSHsMCO	314.29	136.02	263.96	**238.09±91.90**
PSHMCO	341.23	135.35	297.56	258.05 ±108.47

All the parameters computed during the training phase of HSMM-based approach are depicted in Fig. 3.4. For visualization simplicity the covariance matrices are assumed to be diagonal in this figure. As can be seen in Fig. 3.4, the duration distributions for the initial wearing stages are more compressed towards the smaller values than the moderate wearing stages, which corresponds to the steepness of the wearing curve in the initial stages compared to the moderate wearing stages.

3.4.3 Prognostics Results

In this case, both models are unrolled over the future time steps for prognostics. Different prediction horizons are studied and the corresponding prediction accuracies are provided in Table 3.2.

From Table 3.2, it can be seen that the prediction error rate gets larger as the prediction horizon increases (which is expected). It is also noteworthy, that the prediction error rate of the HSMM-based approach with prediction horizon of 9 steps is less than the acquired average prognosis error rate from the HMM-based approach with prediction horizon of 1. This fact indicates that HSMM-based approach is more efficient in capturing the temporal information, which results into smaller error rates in both diagnostics and prognostics.

It is noteworthy that the error indicator (MSE) used in this section indicates predictions deviation from the exact actual wearing values. However, it does not recognize the

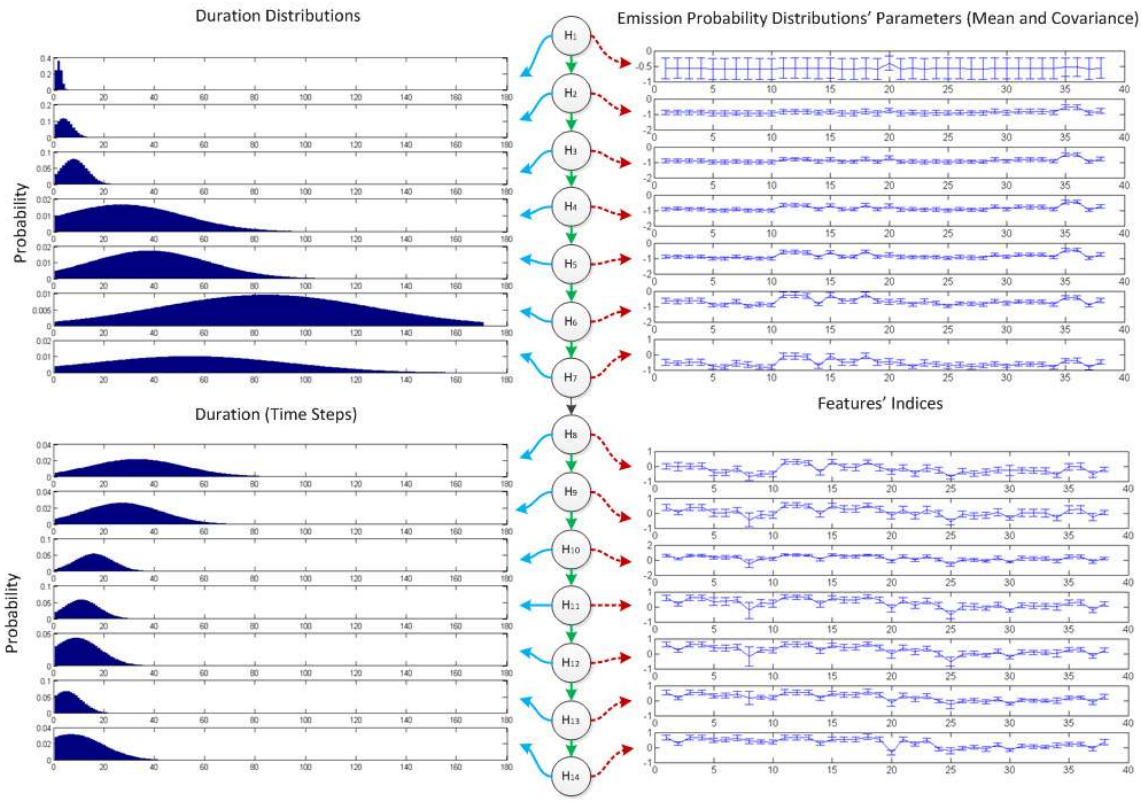


Figure 3.4: Computed parameters of the PSHsMCO approach in diagnostics case. The diagrams on the left depict all the duration distributions of various health states with truncated Gaussian distributions. On the right, the diagrams show the mean and variance of all observed selected features within each health state. The covariance matrices are assumed to be diagonal for simplicity in visualization.

difference between an under- and over-estimation and can be regarded as a symmetric loss function [108]. In the next section, an asymmetric loss function will be given and the performance of the approaches is compared based on their ability in minimizing the total loss.

3.5 Asymmetric Loss Function

All the error rates given in the preceding section are based on squared error that can be regarded as a symmetric loss function, not distinguishing between the over- and under-estimations. But in reality in most of tool condition monitoring systems, it is

Table 3.2: Prognosis error rate for PSHMCO and PSHsMCO approaches in terms of average MSE with different prediction horizons on the testing data.

Model	Prediction Horizon (Time Steps Ahead)								
	1	2	3	4	5	6	7	8	9
PSHsMCO	311.3	314.0	317.9	322.4	327.5	333.1	339.5	346.5	353.9
PSHMCO	422.3	491.4	602.7	680.1	732.5	771.1	802.2	829.0	852.7

preferred to be pessimistic rather than optimistic. For example, in the case of cutter condition monitoring in a CNC-milling machine, it is preferred to change the cutter sooner rather than using the cutter longer and having a defected work-piece as a result. Thus, it is better to over-estimate the tool wear of the cutter rather than under-estimate. This preference can be formulated using an asymmetric loss function. In this section, an arbitrary asymmetric function is given as follows

$$\begin{aligned} err\ (10^{-3}mm) &= \text{estimated tool wear} - \text{actual tool wear}, \\ loss(err) &= \begin{cases} e^{(-err/8)} - 1 & err < 0 \\ e^{(err/10)} - 1 & err \geq 0 \end{cases} . \end{aligned} \tag{3.34}$$

Now, the performance of the two approaches can be compared based on the given asymmetric loss function. In the example of CNC-milling machine given here, the values of estimated and actual tool wear are recorded and used in scale of $10^{-3}mm$. As desired in case of TCM for CNC-milling machine, the defined loss function in (3.34) penalizes more if the predicted tool wear is under-estimated ($err < 0$) compared to over-estimation to the same degree.

One of the main advantages of the proposed HSMM-based approach, which is its duration flexibility, can be exploited in this case. Since the given loss function is asymmetric and the HSMM duration distribution is flexible it can be made asymmetric to incorporate the asymmetricity of the loss function in the predictive model in order to improve the performance.

In the previous section, as we had a symmetric squared error loss function, we as-

sumed for simplicity that duration distributions may be modeled with Gaussian distributions. However, when we want to use an asymmetric loss function, consequently we would like different sides of the peak in the Gaussian functions to be also asymmetric corresponding to the difference indicated in the loss function.

Therefore, an asymmetric version of Gaussian function [109] which has a parameter called asymmetry to control the amount of asymmetricity in the function is adopted in this section. The newly defined duration distribution $P(d_i)$ is,

$$P(d_i = x) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma_i} \exp(-\frac{q^2}{2\sigma_L^2}) & q < 0 \\ \frac{1}{\sqrt{2\pi}\sigma_i} \exp(-\frac{q^2}{2\sigma_R^2}) & q \geq 0 \end{cases} \tag{3.35}$$

$$q = x - \bar{d}_i,\ \sigma_{L_i} = \sigma_i(1+\rho),\ \sigma_{R_i} = \sigma_i(1-\rho)$$

where ρ is the asymmetry factor, $\bar{d}_i$ is the duration at the peak in the ith duration distribution function, σ_i is the average root mean square width, σ_{L_i} and σ_{R_i} are left and right width (deviation), respectively.

In this chapter, since the right side of the Gaussian distribution corresponds to the longer probable durations which may cause under-estimation, the asymmetry parameter should favor the left side of the peak. Thus the asymmetry factor may vary between (0, 1). As indicated in (3.35), once the asymmetry parameter is set, then the right and left width can be computed based on asymmetry and the average width.

Without loss of generality, the duration samples from the training set can be sorted based on their values in an ascending order. Then, the log-likelihood of the ith duration distribution, L_{d_i} , can be written as

$$L_{d_i} = -\frac{n_{d_i}}{2} log(2\pi) - n_{d_i} log(\sigma_i) - \sum_{j=1}^{n_{L_i}} \frac{(x_j - \bar{d}_i)^2}{2\sigma_{L_i}^2} - \sum_{j=n_{L_i}+1}^{n_{d_i}} \frac{(x_j - \bar{d}_i)^2}{2\sigma_{R_i}^2} \tag{3.36}$$

where x_j is the jth duration sample, n_{d_i} is the total number of duration samples for the ith health state and

$$n_{L_i} = n_{d_i} \times \varphi, \tag{3.37}$$

is the number of samples assumed to be generated from the left side of the peak in the ith duration distribution. φ is an auxiliary hyper-parameter of the model that can be found through cross-validation and it indicates the general percentage of the samples

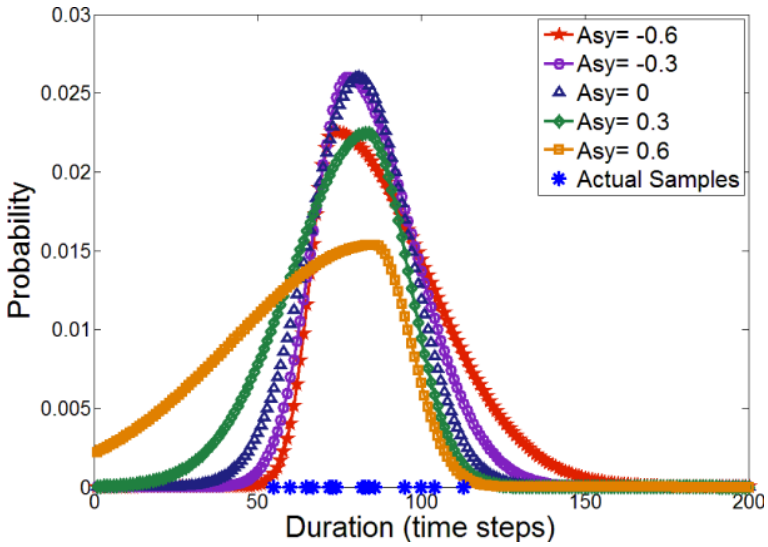


Figure 3.5: Effect of asymmetry factor on the estimated asymmetric Gaussian distribution while GLSP factor (φ) is arbitrarily set to 30%.

in the duration distributions generated from the left side of the peaks in the duration distributions. Figure 3.5 and 3.6 depicts the effects of asymmetry and general left side percentage (GLSP) factors, ρ and φ, on the estimated asymmetric Gaussian distributions.

Assuming to find proper values for ρ and φ as the hyper-parameters of the state-duration distributions based on cross-validation phase. $\bar{d}_i$ and σ_i may be estimated using maximum likelihood method. Substituting σ_{L_i} and σ_{R_i} in (3.36) with their parametric form in (3.35), (3.36) can be rewritten as

$$L_{d_i} = -\frac{n_{d_i}}{2} log(2\pi) - n_{d_i} log(\sigma_i) - \frac{\sum_{j=1}^{n_{L_i}} (x_j - \bar{d}_i)^2}{2\sigma_i^2 (1+\rho)^2} - \frac{\sum_{j=n_{L_i}+1}^{n_{d_i}} (x_j - \bar{d}_i)^2}{2\sigma_i^2 (1-\rho)^2} \tag{3.38}$$

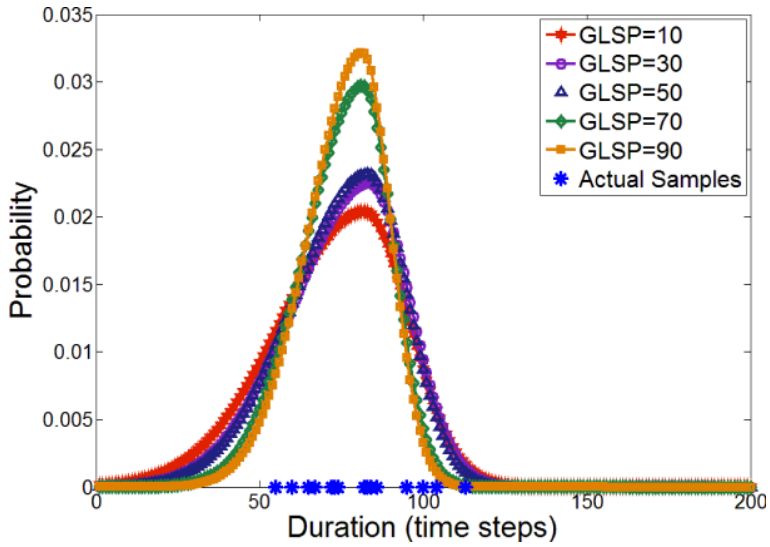


Figure 3.6: Effect of GLSP factor on the estimated asymmetric Gaussian distribution while asymmetry factor (ρ) is arbitrarily set to 0.3.

By setting the derivatives of (3.38) to zero, $\bar{d}_i$ and σ_i can be estimated as follows

$$\begin{aligned}
&\frac{\partial L_{d_i}}{\partial \bar{d}_i} = 0 \rightsquigarrow \frac{\sum_{j=1}^{n_{L_i}}(x_j - \bar{d}_i)}{\sigma_i^2(1+\rho)^2} + \frac{\sum_{j=n_{L_i}+1}^{n_{d_i}}(x_j - \bar{d}_i)}{\sigma_i^2(1-\rho)^2} = 0 \\
&\rightsquigarrow (1-\rho)\sum_{j=1}^{n_{L_i}}(x_j - \bar{d}_i) + (1+\rho)\sum_{j=n_{L_i}+1}^{n_{d_i}}(x_j - \bar{d}_i) = 0 \\
&\rightsquigarrow \sum_{j=1}^{n_{d_i}} x_j - n_{d_i}\bar{d}_i + \rho(n_{d_i} - 2n_{L_i})\bar{d}_i + \rho[\sum_{j=n_{L_i}+1}^{n_{d_i}} x_j - \sum_{j=1}^{n_{L_i}} x_j] = 0 \\
&\rightsquigarrow \bar{d}_i = \frac{C_{tot} + \rho(C_{tot} - 2C_L)}{n_{d_i} + \rho(n_{d_i} - 2n_{L_i})},\ C_{tot} = \sum_{j=1}^{n_{d_i}} x_j,\ C_L = \sum_{j=1}^{n_{L_i}} x_j,
\end{aligned} \tag{3.39}$$

$$\begin{aligned}
&\frac{\partial L_{d_i}}{\partial \sigma_i} = 0 \rightsquigarrow -\frac{n_{d_i}}{\sigma_i} + \frac{\sum_{j=1}^{n_{L_i}}(x_j - \bar{d}_i)^2}{{\sigma_i}^3(1+\rho)^2} + \frac{\sum_{j=n_{L_i}+1}^{n_{d_i}}(x_j - \bar{d}_i)^2}{{\sigma_i}^3(1-\rho)^2} = 0 \\
&\rightsquigarrow -\frac{n_{d_i}}{\sigma_i} + \frac{1}{{\sigma_i}^3} \times [\frac{\sum_{j=1}^{n_{L_i}}(x_j - \bar{d}_i)^2}{(1+\rho)^2} + \frac{\sum_{j=n_{L_i}+1}^{n_{d_i}}(x_j - \bar{d}_i)^2}{(1-\rho)^2}] = 0 \\
&\rightsquigarrow \sigma_i = \sqrt{\frac{1}{n_{d_i}} \times [\frac{\sum_{j=1}^{n_{L_i}}(x_j - \bar{d}_i)^2}{(1+\rho)^2} + \frac{\sum_{j=n_{L_i}+1}^{n_{d_i}}(x_j - \bar{d}_i)^2}{(1-\rho)^2}]}.
\end{aligned} \tag{3.40}$$

Now, diagnostics accuracy of the two approaches is re-examined based on the asymmetric loss function given in (3.34) to show the applicability of the HSMM-based approach

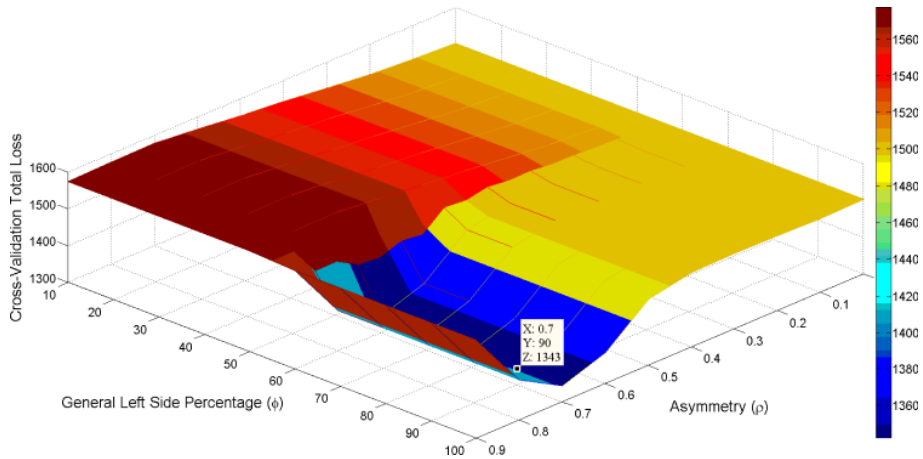


Figure 3.7: Mean of cross-validated total loss for every value taken by ρ and φ.

while having an asymmetric loss function. But, before proceeding to the diagnostics part, the appropriate values for the two newly added hyper-parameters must be adopted through cross-validation.

3.5.1 Asymmetric Cross-Validation

In this section, the number of hidden state values is set to 14 so that the results may be comparable with the models used in the previous sections. The hyper-parameters that are considered to be explored in this case are the asymmetry factor ρ and the general left side percentage φ. Figure 3.7 depicts the corresponding mean of the total loss on the cross-validated sets for every value taken by ρ and φ.

As can be seen in Fig. 3.7, $\rho = 0.7$ and $\varphi = 90\%$ are depicted to be appropriate values to be selected based on the cross-validation total loss. Therefore these values are adopted for the hyper-parameters of the HSMM-based approach in diagnostics.

3.5.2 Asymmetric Diagnostics

Here, diagnostics accuracy of the HMM and HSMM-based approaches (PSHMCO and PSHsMCO) are compared based on the given asymmetric loss function in (3.34). Similar to diagnostics case in the previous section, the testing set is the data collected from the three cutters $09BX3$, $18SC3$ and $33PN6$. All the models are trained based on

the data collected from cutters $07BX1$, $31PN4$ and $34PT1$.

In this simulated experiment, at each time step (cut), the data from the beginning of the experiment up to the current time step is given to both models in order to predict the tool wear of the three flutes at that time step. Number of health states is set to 14 in all models as suggested in the previous section. In this case, similar to the previous section, the maximum wearing value of all 3 flutes at each time step (cut) within each experiment is regarded as the desired outcome. As an indicator to show how much the asymmetric function adopted for the state-duration distributions in the HSMM-based approach has improved the performance, the results of HSMM with symmetric Gaussian function is also included along with the HMM-based approach. Table 3.3 shows the diagnosis accuracies of the approaches based on the total loss on the testing set.

Table 3.3: Prediction error rate for diagnostics using PSHMCO and PSHsMCO approaches in terms of total loss for the given loss function in (3.34).

Approach	Total Loss			
	$09BX3$	$18SC3$	$33PN6$	**Average Total Loss**
Asymmetric PSHsMCO	2147	705	1672	1508±734
Symmetric PSHsMCO	2719	843	1794	1785 ±938
PSHMCO	4826	883	2289	2666 ±1998

3.6 Summary

In this chapter, a more complex model compared to HMM used in Chapter 2 called hidden semi-Markov model, is utilized in the proposed approach named Physically Segmented Hidden Semi-Markov model with Continuous Output (PSHsMCO) to perform continuous diagnosis and prognosis in TCM applications. Also, a computationally efficient version of forward-backward algorithm for application of PSHsMCO in continuous health condition monitoring is described. Based on the simplified forward-backward algorithm, diagnostics and prognostics procedures are defined.

A comparative study is conducted between the suggested HSMM-based approach (PSHsMCO) and the HMM-based approach proposed in Chapter 2. Performances of the two approaches are compared in three cases i.e. cross-validation, diagnostics and prognostics. Based on the experimental results, HSMM-based approach outperforms the HMM-based approach in both diagnostics and prognostics. Interestingly, the error rate of the HSMM-based approach predicting 9 time steps ahead is less than the acquired one step ahead average prognosis error rate from the HMM-based approach which indicates how powerful HSMM is compared to HMM in capturing the underlying temporal information.

In the last section, applicability of the asymmetric Gaussian functions as the state-duration distributions in the proposed PSHsMCO approach is explored. Using the asymmetric Gaussian function, the asymmetric loss functions can be incorporated into the approach in order to recognize the preference between the under and over estimations. Experimental results also indicate that using asymmetric Gaussian functions as the duration distribution functions improve the performance substantially while using an asymmetric loss function.

Chapter 4

Multi-Modal Hidden Markov Model-Based Approach

4.1 Introduction

A hidden Markov model-based approach named physically segmented hidden Markov model with continuous output (PSHMCO) was proposed in Chapter 2 to estimate the continuous tool wear of cutters in a CNC-milling machine, where an explicit relationship between the physical states and the hidden state values is derived. The PSHMCO approach outperforms the conventional Artificial Neural Network approaches [111]. However, PSHMCO, that uses a single HMM, adopts a fixed regiment, hence may lose the desired generalization property.

In this Chapter, to improve the performance of PSHMCO, multiple PSHMCOs are used in parallel as multiple modes. In this multi-modal HMM-based (m^2HMM) approach, each PSHMCO captures and emphasizes on a different tool wear regiment. In this Chapter, three weighting schemes, namely, *bounded hindsight*, *discounted hindsight* and *semi-nonparametric hindsight* are proposed and two switching strategies named *soft-* and *hard-* switching are introduced to combine the outputs from multiple modes into one. The performance of the multi-modal approach with various weighting schemes and switching strategies is reported and compared with PSHMCO. Furthermore, the windowed variant of both PSHMCO and m^2HMM approaches are introduced that sub-

stantially reduces the computational cost without sacrificing the prediction accuracy.

This Chapter is organized as follows. Section 2 describes the proposed windowed variant of the single HMM-based prediction approach called PSHMCO. In Section 3, a multi-modal HMM-based approach is proposed with three weighting schemes and switching strategies along with the windowed variant of m^2HMM. The preliminary experimental results are provided and compared between the single and multi-modal approaches in Section 4. Furthermore, the switching strategies, weighting schemes and the windowed variants are compared in Section 5. Also, the robustness of the semi-nonparametric weighting scheme with respect to its hyper-parameter value is shown in this Section. Finally, the Chapter is summarized in Section 6.

4.2 Windowed Single HMM-based Approach

In this Chapter, the physically segmented hidden Markov model-based approach with continuous output (PSHMCO) proposed in Chapter 2 for tool wear monitoring is regarded as the single HMM-based approach for further discussions in this Chapter.

γ_t can be defined and computed as

Here, in order to reduce the computational cost in PSHMCO a windowing algorithm is introduced. In this algorithm, instead of providing the full observation sequence to the HMM up to current time, $O_{1:T}$, a windowed observation sequence of the recently past observations up to current time, $O_{T-L_w+1:T}$ is given to HMM where L_w is the window length. However, to preserve the general trend of the sequence and the path that HMM is taking through the sequence, the only parameter that has to be updated as the time proceeds is the initial probability.

As stated in Chapter 2, whenever the prediction starts on a newly given sequence, if no data is available on the initial degradation state, the initial probability is assumed to have a uniform distribution on all possible states. In the proposed windowed approach instead of preserving and using all the past observations and involving all of them into the current predictions, only the windowed observation sequence is used. However, the initial probability distribution at each time step is updated for the next sequence window

as follows

$$\begin{aligned} \pi'_{0,T+1} &= \gamma'_1 = P(S'_1, O'_{1:L_w}|\lambda'_T), \\ O'_{1:L_w} &= O_{T-L_w+1:T} = \{O_{T-L_w+1}, \ldots, O_T\}, \\ S'_1 &= S_{T-L_w+1}, \end{aligned} \tag{4.1}$$

where $\pi'_{0,T+1}$ is the updated initial probability at time T that must be used for the prediction at the next time step, $T+1$. λ'_T is the updated parameter set for time T. λ'_T is identical to the parameter set of the given HMM, λ, except for its initial probability distribution π_0 that is updated by $\pi'_{0,T}$. At each time step $T > L_w$, the windowed observations $O'_{1:L_w}$ is given to the HMM with the updated parameter set λ'_T based on (4.1).

The proposed windowing algorithm makes the length of observations used in the forward-backward algorithm bounded to a window length (L_w) smaller than the average full length sequence ($\bar{L}_f$). Consequently, the computation cost for the forward-backward algorithm reduces $\bar{L}_f/L_w$ times.

4.3 Multi Modal HMM-Based Approach

In this Chapter similar to the preceding Chapters, TCM of a cutter in a CNC-milling machine is used as an illustrative example. While usually all the cutters are made from the same material in one dataset, the geometrical difference of cutters in the training and testing set could deviate and cause inaccurate predictions.

The idea of multi-modal approach is to generalize the prediction model by capturing more possible trends using combinations of distinct experiments for training, and then integrating the prediction results from all the single models in a weighted form. Here, three weighting schemes are proposed along with two switching strategies based on the computed weightages. Moreover, the windowed algorithm introduced in Section 2 is extended for the multi-modal approach.

In the proposed multi-modal approach to improve performance of the single HMM-based approach called PSHMCO, given a set of n experiments as the training set, N_{mode} PSHMCOs are trained. Each PSHMCO is trained on one distinct combination of the experiments to capture the common trend between the combined experiments (mode).

It is noteworthy that the number of modes in this approach is less than or equal the number of possible combinations given n experiments, thus the number of modes is $N_{mode} \leq 2^n - 1$.

After training all the PSHMCOs as in Chapter 2, given a new set of observations as the testing set, we would like to assess the importance of each PSHMCO for the final prediction at each time step and reflect that as a weightage. Then, based on the switching strategy, the PSHMCOs' weightages are used to predict the ultimate output at each time step. Figure 4.1 schematizes the prediction process in the m^{2}HMM approach.

The weightages of the modes can be computed in various ways with each emphasizing on the importance of different factors. The common ground between the three weighting schemes that will be introduced here is that, the weight of each PSHMCO is computed based on the historical observations (hindsight) and the current time, in order to assess the usefulness and similarity of that PSHMCO for prediction at that point of time. However, the factors that are considered to assess the weightages in these three schemes are different. Two of the weighting schemes are solely based on the parametric models thus named parametric hindsight weighting schemes, namely, *bounded hindsight* and *discounted hindsight*. The third weighting scheme uses the parametric models to locate the positions in the corresponding training data used to train them (parametric phase), and then finds the most similar portion to the observations at hand given for prediction and ultimately assigns the similarity score to each model as the weightage (nonparametric phase). As the third weighting scheme has both parametric and nonparametric phases and furthermore its ultimate result is based on the nonparametric phase, it is named *Semi-Nonparametric* hindsight. As mentioned in Chapter 1, hindsight is the process of reviewing the historical data and calculating the probable health states at the past time steps [96, 112].

4.3.1 Most Probable Health States

All the proposed weighting schemes in this Chapter, use a dynamic programming method known as Viterbi algorithm. The Viterbi algorithm is used to find the most probable health states (Viterbi-path) taken at the past time steps within each model. As

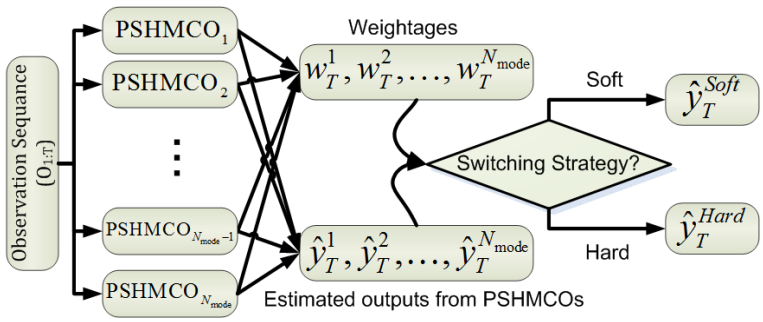


Figure 4.1: Illustration of multi-modal HMM-based approach.

described in [95], the Viterbi path can be formulated as follows

$$v_t(i) = \max_{S_{1:t-1}}\{P(S_{1:t-1}, S_t = H_i, O_{1:t}|\lambda)\}, \tag{4.2}$$

where $v_t(i)$ is the highest probability obtained by a single path up to time t that ends in state H_i. As indicated in [95], by induction a recursive formula can be obtained as follows

$$v_{t+1}(j) = \max_{i=1,\ldots,m}\{v_t(i) \times a_{i,j}\} \times P(O_{t+1}|S_{t+1} = H_j, \lambda), \tag{4.3}$$

where $a_{i,j}$ is the transition probability to go from H_i to H_j in the HMM with the λ parameter set. Furthermore because of the specific structure assumed in Chapter 2 for PSHMCO based on the gradual process of wearing and degradation, from each health state H_i only two possible transitions exist i.e. either staying at the same health state or going to the next health state H_{i+1} till the cutter completely worn out and enter the last wearing stage H_m. Thus, (4.3) can be computationally simplified as follows

$$\begin{aligned} v_{t+1}(j) = &\max\{v_t(j-1) \times (1 - p_{j-1}),\ v_t(j) \times p_j\} \\ &\times P(O_{t+1}|S_{t+1} = H_j, \lambda). \end{aligned} \tag{4.4}$$

Finally, $V_{1:T}$ that is the sequence of health state indices taken through the Viterbi-path and is required for the weighting schemes, can be computed by performing backtracking on the stored matrix that has kept track of all the arguments which has maximized (4.3) for every t and j. For more details on Viterbi algorithm refer to [95].

In the succeeding subsection, the three weighting schemes are described in details.

4.3.2 Weighting Schemes

As shown in Fig. 4.1, the observations (input features) are fed into all PSHMCOs in parallel. After finding the output of each PSHMCO based on (2.21), the next issue is to decide about the weightages of PSHMCOs outputs for polling (weighted averaging). Two methods, the parametric and semi-nonparametric hindsight, are adopted and compared for weighted polling. In all weighting schemes, the idea is to re-evaluate the likelihood of preceding time steps (hindsight) along with the current time step and use it as a metric to assess compatibility of each single model with the recent observations.

Parametric Hindsight

One of the motivations in using a multi-model approach is to use possible mix of trends captured by various models to improve prediction in new trends. These new trends may be relevant in parts to the modeled trends in multiple models. The aim is to give a higher weightage to the models which are more likely to have a similar trend with the new experiment at each time step. In this Section, two schemes are suggested for computation of weightages in parametric form. In both schemes, focus is more on the latest time steps rather than long past ones. This allows a more dynamic switching of influence on the outputs between the multiple models. The basic idea of the proposed parametric hindsight weighting schemes are similar to the sliding window approaches used in the image and signal processing applications such as spatio-temporal visual tracker [113], smoothing [114] as well as enhancing the temporal information [115]. Two parametric schemes are introduced as follows.

Bounded Hindsight

In this scheme, given the observations up to the current time step T, the maximum likelihood of the probable health states for $\Delta - 1$ (bounded) preceding time steps (hindsight) along with the current time step in each PSHMCO are calculated. Summation of these Δ likelihoods, w_T^i, is used as an indicator of how relevant each PSHMCO is to the currently on-going experiment at that time. Hence w_T^i is used as the corresponding weightage for ith PSHMCO. An appropriate value for Δ can be found using cross-

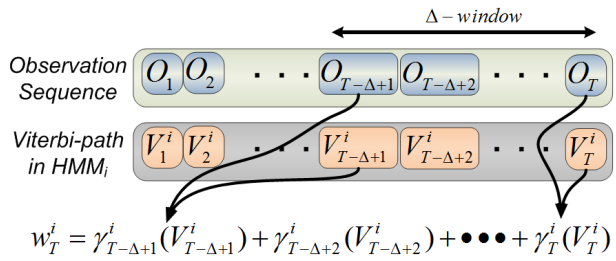


Figure 4.2: Illustration of bounded hindsight weighting scheme.

validation. The bounded weightage, w_T^i, for the ith HMM can be computed as follows

$$w_T^i = \sum_{t=T-\Delta+1}^{T} P(S_t = H_{V_t^i}, O_{1:T}|\lambda_i) = \sum_{t=T-\Delta+1}^{T} \gamma_t^i(V_t^i). \tag{4.5}$$

where Δ is the window size of the hindsight is bounded to, V_t^i is the health state index at time t of the viterbi-path taken by the ith PSHMCO. λ_i is the ith PSHMCO parameter set.

Discounted Hindsight

In this weighting scheme, similar to the bounded hindsight, given the observations up to the current time step T, the maximum likelihood of the Viterbi-path health states for preceding time steps (hindsight) in each PSHMCO are calculated. Instead of having a uniform summation on the likelihoods, first, every calculated likelihood is multiplied by a discount factor to increase the importance of the recent preceding time steps compared to the long past ones. Then, the summation of the discounted likelihoods is used as the corresponding weightage for each PSHMCO. In this work, the discount function is set to be a Gaussian distribution over the preceding time steps with its peak value at the current time step. An appropriate standard deviation for this function may be found based on cross-validation. The discounted weightage, w_T^i, for the ith PSHMCO can be computed as follows

$$\begin{aligned} w_T^i &= \sum_{t=1}^{T} \phi_t \times P(S_t = H_{V_t^i}, O_{1:T}|\lambda_i) = \sum_{t=1}^{T} \phi_t \times \gamma_t^i(V_t^i), \\ \phi_t &= N(t; T, \sigma^2) \end{aligned} \tag{4.6}$$

where ϕ_t is the value of the Gaussian distribution with mean value of T and standard deviation of σ at time step t. Figure 4.3 illustrates the discounted hindsight weighting

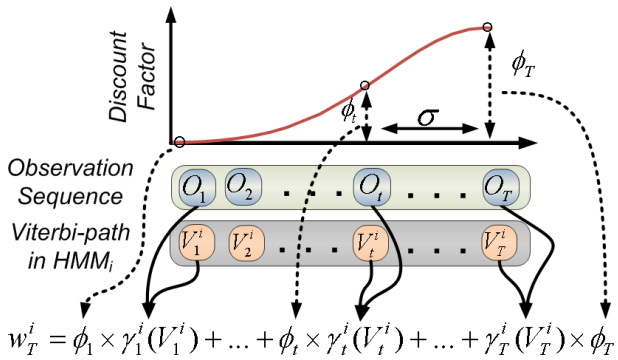


Figure 4.3: Illustration of discounted hindsight weighting scheme.

scheme.r

Semi-Nonparametric Hindsight

Here, another weighting scheme is introduced to compute appropriate weightages required in the m^{2}HMM approach to integrate the PSHMCOs outputs into its ultimate output. The basic idea of this weighting scheme is similar to the sequence alignment methods. Sequence alignment is a way of arranging sequences to identify regions of similarity originally developed in bioinformatics for aligning DNA, RNA or protein sequences [116].

Since the proposed scheme has both parametric and nonparametric phases and its final result comes from the nonparametric stage, this weighting scheme is named Semi-NonParametric Hindsight (SNPH). Given the new observations and the HMMs parameter sets, the procedure of semi-nonparametric hindsight is schematized as follows.

- For each HMM, using its parameter set and the given observations find the most probable path of health states based on Viterbi-path algorithm. Find $V_{1:T}^i$ based on λ_i and $O_{1:T}$.

- Find the intervals I_q in the qth training sequence of the ith training combination ($D^{i,q}$) that corresponds to the most likely path taken in the ith HMM based on the reference segment of the Viterbi-path $V_{1:T}^i$ that is denoted by $V_{T-r+1:T}^i$. I_q can be

identified as a set of starting, ending time index pairs as follows

$$I_q(V^i_{1:T}, D^{i,q}, r) = \bigcup_{h\in\Psi}\{(t_s^{i,q,h}, t_e^{i,q,h})\},$$
$$\Psi = \{V^i_{T-r+1}, V^i_{T-r+1}+1, \ldots, V^i_T\}, \tag{4.7}$$

where $D^{i,q} = \{O^{i,q}, Y^{i,q}\}$ is the qth training sequence of the ith training combination that includes observations $O^{i,q}$ and their actual corresponding tool wear values $Y^{i,q}$, Ψ is the set of health states' indices that have been taken in the reference sequence based on the computed Viterbi-path, $t_s^{i,q,h}$ and $t_e^{i,q,h}$ are respectively the starting and ending time steps of the hth health state in $D^{i,q}$.

- Align $O^{i,q,h} = O^{i,q}_{t_s^{1,q,h}:t_e^{1,q,h}}$ and the R^h which is the corresponding observation segment in the reference sequence that its most likely health state index is h based on Viterbi-path and is defined as follows

$$R^h = \{O_{t_s^h:t_e^h} | T-r+1 \le t_s^h \le t \le t_e^h \le T, V^i_{t_s^h+1} \ne h, V^i_{t_e^h-1} \ne h, V^i_t = h\}, \tag{4.8}$$

where the starting and ending time steps of the hth health state in the reference sequence are denoted by t_s^h and t_e^h, respectively.

- The Aligned distance between the two matrices can be computed as

$$dist(O^{i,q,h}, R^h) = \begin{cases} \min\limits_{i=1,\ldots,l_h-l'_h+1}\{\sqrt{\sum\limits_{k=1}^{K}\sum\limits_{j=1}^{l'_h}(O^{i,q,h}_{k,j+i-1} - R^h_{k,j})^2}\} & for\ l'_h \le l_h \\ \\ \min\limits_{i=1,\ldots,l'_h-l_h+1}\{\sqrt{\sum\limits_{k=1}^{K}\sum\limits_{j=1}^{l_h}(O^{i,q,h}_{k,j} - R^h_{k,j+i-1})^2}\} & for\ l'_h > l_h \end{cases}, \tag{4.9}$$

$$l_h = t_e^{i,q,h} - t_s^{i,q,h} + 1, \quad l'_h = t_e^h - t_s^h + 1,$$

where length of corresponding segments to the hth health state in the reference sequence (R) and training observation sequence $O^{i,q}$ are denoted by l'_h and l_h, respectively.

- Based on the aligned distance function $dist(.,.)$ defined in (4.9), the total score

function for qth sequence of the ith PSHMCO can be computed as follows

$$Score(D^{i,q}, O_{1:T}, V^i_{1:T}, r) = \left[\frac{\sum_{h=V^i_{T-r+1}}^{V^i_T} dist(O^{i,q,h}, R^h)}{\sum_{h=V^i_{T-r+1}}^{V^i_T} \min\{l_h, l'_h\}} \right]^{-1}, \tag{4.10}$$

where $O_{1:T}$ is the newly given observation sequence from time step 1 up to current time T and $V_{1:T}$ is the index sequence of the health states taken within the Viterbi-path.

- Finally, maximum of the scores obtained from all included sequences in the ith PSHMCO training set as the maximum aligned similarity, would be adopted for the weightage of ith PSHMCO in the final outcome. Thus, the weightage of the ith PSHMCO can be computed as

$$w^i_T = \max_q \{Score(D^{i,q}, O_{1:T}, V^i_{1:T}, r)\} \tag{4.11}$$

Figure 4.4 illustrates the semi-nonparametric hindsight weighting scheme. For illustration purpose, in Fig. 4.4, it is assumed that $V^i_{T-r+1} \neq V^i_T$ meaning that there are more than one distinct health index in the reference segment of the Viterbi-path.

After both parametric and semi-nonparametric hindsight schemes, the computed weightages can be used to either mix the outputs (soft-switching) or perform hard-switching between modes.

4.3.3 Switching Strategy

After assigning relevance weightages to all the modes based on any of the weighting schemes that are introduced, two strategies are considered for integrating various modes' outputs into the ultimate multi-modal output. These two strategies can be formulated as follows

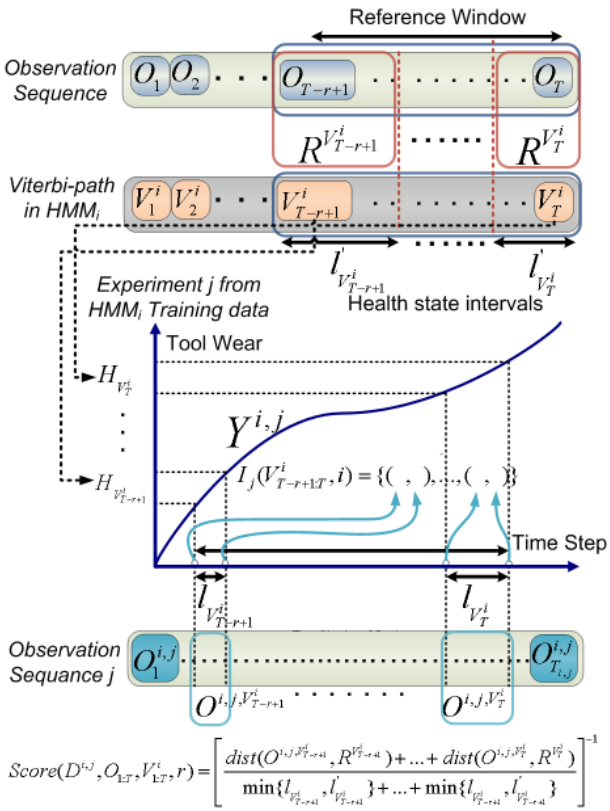


Figure 4.4: Illustration of semi-nonparametric hindsight weighting scheme.

- Soft-Switching,

$$\hat{y}_T^{Soft} = \frac{\sum_{i=1}^{N_{mode}} w_T^i \times \hat{y}_T^i}{\sum_{i=1}^{N_{mode}} w_T^i}, \tag{4.12}$$

where $\hat{y}_T^{Soft}$ is the ultimate output for time step T using the soft-switching strategy, $\hat{y}_T^i$ is the output of the ith PSHMCO and w_T^i is its corresponding weightage.

- Hard-Switching,

$$\hat{y}_T^{Hard} = \hat{y}_T^{\arg\max_i \{w_T^i\}}, i = 1, \ldots, N_{mode}. \tag{4.13}$$

where $\hat{y}_T^{Hard}$ is the ultimate output for time step T using the hard-switching strategy.

As (4.12) indicates, the soft-switching strategy uses the weighted average of all PSHMCOs as the ultimate output. On the other hand, the hard-switching strategy as shown in (4.13) assigns the output of the PSHMCO with the highest weightage as the ultimate output.

4.3.4 Windowing Algorithm for m^{2}HMMs

Similar to the windowing algorithm introduced in Section 2, by using the windowed observations instead of full observation sequences, the computational cost can also be reduced drastically in the proposed m^{2}HMMs.

Here, instead of giving full observation sequence up to the current time T, $O_{1:T}$ to each PSHMCO, a windowed observation sequence $O_{T-L_w+1:T}$ is given to all PSHMCOs. Thus, similar to the windowing algorithm introduced in Section 2, the initial probability will be updated for all the HMMs based on their weightages as follows

$$\pi'_{0,T+1} = \frac{\sum_{i=1}^{N_{mode}} w_T^i \times \gamma'_{1,i}}{\sum_{i=1}^{N_{mode}} w_T^i} = \frac{\sum_{i=1}^{N_{mode}} w_T^i \times P(S'_1, O'_{1:L_w}|\lambda'_{i,T})}{\sum_{i=1}^{N_{mode}} w_T^i}, \tag{4.14}$$

$$O'_{1:L_w} = O_{T-L_w+1:T},\ S'_1 = S_{T-L_w+1},$$

where $\pi'_{0,T+1}$ is the updated initial probability that must be used for the prediction at next time step, $T+1$. $\lambda'_{i,T}$ is the same as parameter set of the ith HMM, λ_i, except that its initial probability distribution π_0 is updated by $\pi'_{0,T}$.

Similar to the windowing algorithm proposed in Section 2, the proposed windowing algorithm for m^{2}HMM approach will reduce the computation cost in the forward-backward algorithm by $\bar{L}_f/L_w$ times where $\bar{L}_f$ is the average length of the given full observation sequence.

4.4 Preliminary Experimental Results

In this Section, the results of cutter wear diagnostics in a CNC-milling machine obtained by the multi-modal HMM-based approach (m^{2}HMM) with various weighting schemes are compared with the single HMM-based approach (PSHMCO) proposed in Chapter 2. The m^{2}HMM is implemented with three weighting schemes and two switching strategies introduced in Section 3.

4.4.1 Experimental Data and Features

The experimental data comprises cutting process of 6 cutters, which are *07BX1, 31PN4, 09BX3, 18SC3, 34PT1* and *33PN6*. The cutters are different from one another by the cutter geometry and coating, but all are 6mm Alignment-Tool carbide ball-nose end with three flutes. In all the cutting processes, *Inconel 718*, which is used in Jet engines, is set as the work-piece material. During the cutting process, the upper face of the material is cut with horizontal lines from the top edge to the bottom edge. After 320 times of cutting, another cutter starts again at the top edge of the material for another round of experiment. A three-channel dynamometer is mounted on the CNC-milling machine that captures force signal in three directions. Moreover, for training and testing purposes, the tool wear data is directly measured and collected using microscope in the conducted experiments. Figure 4.5 schematizes the experimental setup used in this study. The detailed description of the experimental setup, data acquisition process can be found in Appendix A. The experimental setup is identical to the experimental setup used in previous Chapters. However, as indicated in Chapter 2 after feature selection, the vibration and acoustic emission signals, although are easier to be captured, the features extracted from them are not as discriminant as the features extracted from the force signals. Therefore to reduce the number of features further and to focus only on the force signals, after performing a comparative feature selection study in [117], 13 statistical features that shown to be the most salient features in that study are extracted to be utilized. Table 4.1 lists the extracted features and indicates the direction of the force signals that they are extracted from.

4.4.2 Preliminary Results

All prediction models are trained using data collected from three experiments, which are conducted using 07BX1, 33PN6, and 34PT1 type cutters. The trained prediction models are then tested on the experimental data acquired from the three remaining distinct cutters 09BX3, 18SC3, and 31PN4. In this study, similar to Chapter 2, the number of health states is adopted to be 14 for all the HMMs. Although the number of health states in all HMMs is adopted to be 14, the parameter sets of these HMMs which are

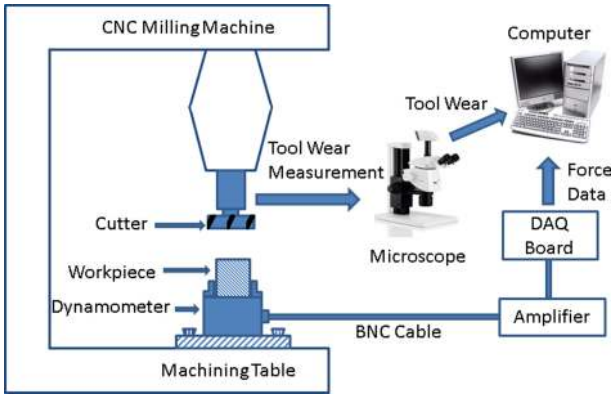


Figure 4.5: The tool wearing estimation experimental setup.

computed based on various combinations of training experiments would be different thus leading to various wearing regiments. Moreover, all possible combinations of the three training experiments are used here to train different modes in the m^{2}HMM approaches leading to 7 distinct modes. Based on the cross-validation, the hyper-parameters of the three weighting schemes which are the size of bounded hindsight window, the standard deviation of the discounted hindsight, and the size of the reference observation sequence in semi-nonparametric weighting scheme are adopted to be 10, 2.5, and 10, respectively. Figure 4.6 depicts the resultant relative weightage values by the three weighting schemes on the three cutters used for testing. As it can be seen in Fig. 4.6, the changes in the weightages through time using the Semi-nonparametric scheme is smoother than the other two.

Table 4.2 shows the prediction performance of PSHMCO compared with variants of multi HMM-based approach (considering the three suggested weighting schemes and the two switching strategies) in terms of mean squared error (MSE) and mean relative error (MRE).

In this experiment, at each time step (cut), the extracted features (observations) from the beginning of the experiment up to the current time step are given for diagnostics. It is noteworthy that in practice, the precision of a work-piece is ultimately determined by the flute with the maximum wearing value. Thus, during testing, the maximum estimated wearing value among the three flutes at each time step is used as the predicted outcome of the cutter wearing.

Table 4.1: List of statistical extracted features from force signals in X,Y and Z directions.

Feature	**Direction**
Amplitude Ratio	X, Y and Z
First Order Differencing	X and Y
Total Harmonic Power	X and Z
Maximum Force Level	X
Total Amplitude of Cutting Force	X
Standard Deviation	X
Kurtosis	Y
Average Force	Z
Skewness	Z

Table 4.2: Tool wear prediction error rate in CNC-milling machine using single HMM-based approach (PSHMCO), and variants of Multi-modal HMM-based (m^2-HMM) approach with three weighting schemes, namely, bounded, discounted and semi-nonparametric hindsight and two switching strategies (soft- and hard-switching).

Approach	Mean Squared Error				Mean Relative Error			
	09BX3	*18SC3*	*31PN4*	Total	*09BX3*	*18SC3*	*31PN4*	Total
PSHMCO	725.54	330.68	749.16	601.79 ± 235.09	0.1604	0.1168	0.1733	0.1502 ± 0.0296
Soft **m^2HMM Bounded**	522.33	**148.08**	354.01	341.47 ± 187.44	0.1386	**0.1012**	0.1439	0.1279 ± 0.0233
Soft **m^2HMM Discounted**	514.80	163.94	375.15	351.29 ± 176.64	0.1363	0.1045	0.1489	0.1299 ± 0.0229
Soft **m^2HMM Semi-Nonparametric**	**337.16**	156.73	**244.88**	**246.25 ± 90.21**	0.1314	0.1063	**0.1200**	**0.1193 ± 0.0125**
Hard **m^2HMM Bounded**	486.84	167.57	930.05	528.16 ± 382.91	**0.0959**	0.1111	0.2518	0.1529 ± 0.0860
Hard **m^2HMM Discounted**	487.18	188.43	916.77	530.79 ± 366.12	0.0961	0.1191	0.2497	0.1549 ± 0.0828
Hard **m^2HMM Semi-Nonparametric**	557.54	303.62	946.56	602.58 ± 323.82	0.1373	0.1555	0.2522	0.1816 ± 0.0617

It can be seen from Table 4.2 that the total performance of m^2HMM approach with parametric hindsight (either discounted or bounded) and soft-switching shows about 40% improvement in average precision and about 20% improvement in standard devi-

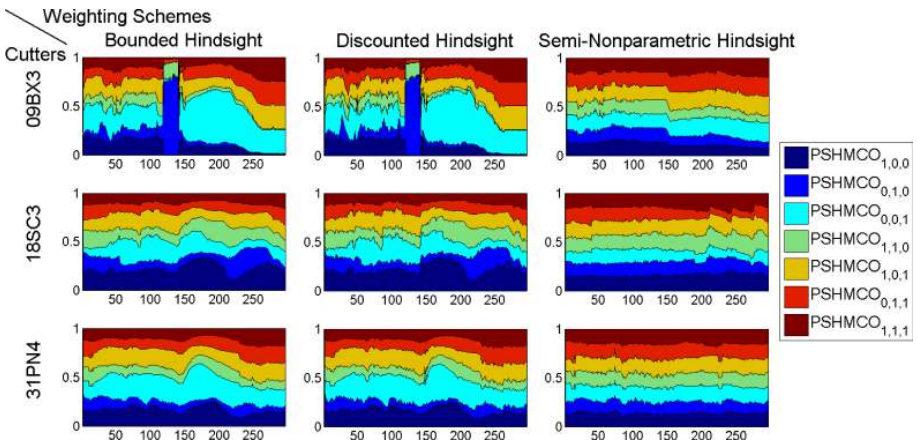


Figure 4.6: Resultant weightages for the three cutters using the three weighting schemes. Y-axes show the area value of each weightage and the X-axes show time steps. The weightages are normalized to sum up to one at each time step. The PSHMCO indices in the legend indicate which cutters' data are included for each mode in a binary manner. The three indices correspond to 07BX1,33PN6, and 34PT1 respectively from left to right.

ation. Furthermore, m^{2}HMM-based approach with semi-nonparametric hindsight and soft-switching strategy (Soft m^{2}HMM-SNPH) achieves the best average total performance. This suggests that semi-nonparametric weighting scheme can be an appropriate way of integrating outputs from various PSHMCOs. Moreover, m^{2}HMM approach variants with hard-switching have been unable to outperform PSHMCO and their counter-variants with soft-switching. Although, the results suggests that on average, Soft m^{2}HMM-SNPH outperforms PSHMCO and other variants of m^{2}HMM, these preliminary results may not be convincing. Thus, in order to assess the significance of improvements made by the m^{2}HMM variants and see whether the Soft m^{2}HMM-SNPH can outperform the other approaches in a statistically significant manner further investigations are conducted.

4.5 Further Investigations

In this Section, the efficiency of the proposed m^{2}HMM approach with various weighting schemes is further investigated to confirm whether it can statistically outperform

PSHMCO. Moreover, performance of the two switching strategies, namely *Soft-* and *Hard-* switching, are compared and the effect of reference length in the SNPH weighting scheme is studied. Furthermore, the windowing algorithms proposed in Sections 2 and 3 are examined to see whether the computational cost can be reduced while maintaining high prediction accuracy in comparison with the original PSHMCO and m^{2}HMM approaches.

For the purpose of having enough samples to assess the statistical significance in performance differences, various trials are generated by different partitionings of the whole acquired dataset from 6 cutters in Section 4 based on all possible combinations. Two scenarios are assumed in this Section,

- Easy Scenario: in which the data from 5 cutters is used for training and then the prediction models are tested on the one remaining cutter. The training-testing (TT) ratio in this case is 5:1.

- Difficult Scenario: in which the data from 3 cutters is used for training and then the prediction models are tested on the 3 remaining cutters. TT ratio in this case is 3:3.

The number of possible combinations and thus various trials for the easy scenario is $C_1^6 = 6$ and for the difficult scenario is $C_3^6 = 20$. In the succeeding subsections the prediction models are trained, tested and compared based on the various trials in the two scenarios. Moreover, every possible combination of experiments in the training partitions are used to train one mode in the m^{2}HMM approaches. Consequently, the N_{mode} of the m^{2}HMM approaches in the easy scenario is 31 ($2^5 - 1$) and for the difficult scenario is 7 ($2^3 - 1$).

It is noteworthy that in all testing experiments the predictions are made from time step 20 to 320 for the sake of fairness in comparisons and reporting the results while various windowing lengths or hyper-parameter values are considered. In the succeeding subsections, if not indicated differently the values of hyper-parameters are adopted to be identical with the ones in the Section 5.

Table 4.3: Overall performance comparison of hard- and soft- Switching strategies within all weighting schemes along with the average PSHMCO performance provided as the benchmark. Average performances are compared based on mean squared error and the P-value shows the significance of performance difference between the two strategies using the same weighting scheme.

Weighting Scheme	Easy (TT ratio 5:1)		Difficult (TT ratio 3:3)		P-Value (Hard vs. Soft)	
	Hard-Switching	*Soft-Switching*	*Hard-Switching*	*Soft-Switching*	Easy	Difficult
m^{2}HMM Bounded	543.96 ± 397.81	553.17 ± 436.31	806.00 ± 654.13	689.23 ± 556.64	0.8667	**0.0052**
m^{2}HMM Discounted	542.96 ± 308.54	551.85 ± 415.22	792.77 ± 625.78	692.31 ± 553.18	0.9050	**0.0143**
m^{2}HMM Semi-Nonparametric	556.64 ± 267.36	**486.59 ± 449.75**	836.94 ± 662.22	**631.89 ± 550.42**	0.6389	**0.0067**
PSHMCO	720.78 ± 362.45		842.24 ± 474.51			

4.5.1 Switching Strategy: Hard Vs. Soft

As mentioned in Section 3, the resultant weightages from three proposed schemes can be used to perform switching based on two strategies i.e. *Soft-* and *Hard-* switching. Table 4.3 and 4.4 show the results of the two strategies in the two scenarios based on various weighting schemes in terms of MSE and MRE, respectively. As it can be seen in both Tables, although m^{2}HMM with *hard*-switching average performances are better than the PSHMCO, the m^{2}HMM with *Soft*-switching outperforms all their *Hard*-switching counter-variants and the p-values in the difficult scenario indicate the significance of this matter. However, the p-values in the easy scenario indicate that there is not enough evidence to support or reject the idea in that scenario although the average performance of the soft-switching variants are better.

Thus, in general it is recommended to use the soft-switching disregarding the weighting scheme that has been used in the multi-modal approach as it leads to higher average performance without additional computation cost.

In the succeeding subsections soft-switching is adopted for all the m^{2}HMM approaches.

Table 4.4: Overall performance comparison of Hard- and Soft- Switching strategies in terms of mean relative error within all weighting schemes along with the average PSHMCO performance provided as the benchmark. Each P-value shows the significance of performance difference between the two strategies using the same weighting scheme.

Weighting Scheme	Easy (TT ratio 5:1)		Difficult (TT ratio 3:3)		P-Value (Hard vs. Soft)	
	Hard-Switching	*Soft-Switching*	*Hard-Switching*	*Soft-Switching*	Easy	Difficult
m^{2}HMM Bounded	0.1603 ± 0.0649	0.1593 ± 0.0617	0.1920 ± 0.0904	0.1772 ± 0.0804	0.9509	**0.0259**
m^{2}HMM Discounted	0.1631 ± 0.0592	0.1605 ± 0.0605	0.1910 ± 0.0893	0.1785 ± 0.0812	0.8996	**0.0701**
m^{2}HMM Semi-Nonparametric	0.1617 ± 0.0707	**0.1513 ± 0.0539**	0.1899 ± 0.0895	**0.1681 ± 0.0714**	0.7722	**0.0291**
PSHMCO	0.1890 ± 0.0560		0.1980 ± 0.0673			

4.5.2 Overall Performance Comparison

Here, the PSHMCO and the m^{2}HMMs with various weighting schemes are trained and tested using the two scenario trials that each has partitioned the whole dataset into training and testing sets. Table 4.5 shows the average performance of each approach in the two scenarios in terms of MSE and MRE. Moreover, the p-values from the pair-wise t-test performed between various approaches and PSHMCO as well as m^{2}HMM-SNPH are reported in Table 4.5.

From Table 4.5, it is suggested that all m^{2}HMM approaches disregarding their weighting schemes have outperformed PSHMCO in terms of both MSE and MRE in a statistically significant manner in both scenarios. Furthermore, the m^{2}HMM with SNPH weighting scheme significantly outperforms the rest in the difficult scenario. However, the p-values for the easy scenario indicate that there is not enough evidence to support or reject the hypothesis in that scenario.

4.5.3 Full Vs. Windowed Observations

Here, the windowing algorithms proposed in Sections 2 and 3 are applied on PSHMCO and the m^{2}HMM approaches. The results are reported and compared with the original PSHMCO which uses full observations. Table 4.6 shows the average performance of

Table 4.5: The pair-wise t-test results from comparing the performance in all the multi-modal approaches with various weighting schemes and PSHMCO as well as m^{2}HMM-SNPH with the rest. Each P-value shows the significance of performance difference between the two approaches either in terms of MSE or MRE based on pair-wise t-test.

Easy Scenario (TT ratio 5:1)	**Approaches**			
	PSHMCO	**m^{2}HMM-BH**	**m^{2}HMM-DH**	**m^{2}HMM-SNPH**
Mean Squared Error	720.78 ± 362.45	553.17 ± 436.31	551.85 ± 415.22	**486.59 ± 449.75**
P-Value (vs. PSHMCO)	*NA*	**0.0211**	**0.0100**	**0.0204**
P-Value (vs. m^{2}HMM-SNPH)	**0.0204**	0.1479	0.1970	*NA*
Mean Relative Error	0.1890 ± 0.0560	0.1593 ± 0.0617	0.1605 ± 0.0605	**0.1513 ± 0.0707**
P-Value (vs. PSHMCO)	*NA*	**0.0078**	**0.0037**	**0.0210**
P-Value (vs. m^{2}HMM-SNPH)	**0.0210**	0.2936	0.2891	*NA*
Difficult Scenario (TT ratio 3:3)	**PSHMCO**	**m^{2}HMM-BH**	**m^{2}HMM-DH**	**m^{2}HMM-SNPH**
Mean Squared Error	842.24 ± 474.51	689.23 ± 556.64	692.31 ± 553.18	**631.89 ± 550.42**
P-Value (vs. PSHMCO)	*NA*	**0.0014**	**0.0016**	**1.98e − 6**
P-Value (vs. m^{2}HMM-SNPH)	**1.98e − 6**	**0.0171**	**0.0154**	*NA*
Mean Relative Error	0.1980 ± 0.0673	0.1772 ± 0.0804	0.1785 ± 0.0812	**0.1681 ± 0.0714**
P-Value (vs. PSHMCO)	*NA*	**0.0028**	**0.0038**	**1.16e − 6**
P-Value (vs. m^{2}HMM-SNPH)	**1.16e − 6**	**0.0146**	**0.0144**	*NA*

windowed version of all approaches in the two scenarios in terms of MSE and MRE. Pair-wise t-test is conducted between the resultant performances from the original PSHMCO and the windowed version of all mentioned approaches on all trials in the two scenarios. Given the window length L_w, the hyper-parameters of bounded and semi-nonparametric hindsight are set equal to L_w and for the discounted hindsight the σ is set to $L_w/4$.

As it can be understood from Table 4.6, the windowed version of the multi-modal approaches disregarding the weighting scheme have significantly outperformed the original PSHMCO. Furthermore, the windowed PSHMCO has achieved a slightly better performance than its original form. Figure 4.7 and 4.8 show the average performance of

Table 4.6: The performance of windowed version of the three multi-modal approaches and PSHMCO in terms of MSE and MRE in two scenarios while $L_w = 13$. The average accuracies of the original PSHMCO with full observations are also given as the baseline. The P-values show significance of each windowed approach outperforming Full PSHMCO in terms of MSE in each case based on pair-wise t-test.

Windowed Approach	Easy (TT ratio 5:1)		Difficult (TT ratio 3:3)		P-Value (vs. Full PSHMCO)	
	MSE	*MRE*	*MSE*	*MRE*	Easy	Difficult
W m^{2}HMM-BH	558.99 ± 453.04	0.1593 ± 0.0626	668.95 ± 644.30	0.1709 ± 0.0839	**0.0370**	**0.0049**
W m^{2}HMM-DH	555.04 ± 425.54	0.1605 ± 0.0608	680.39 ± 636.94	0.1742 ± 0.0850	**0.0163**	**0.0077**
W m^{2}HMM-SNPH	479.86 ± 467.83	0.1497 ± 0.0530	643.71 ± 641.94	**0.1663 ± 0.0770**	**0.0299**	**0.0006**
W PSHMCO	636.16 ± 466.43	0.1784 ± 0.0624	759.74 ± 602.63	**0.1918 ± 0.0747**	0.3156	0.0778
Full PSHMCO	720.78 ± 362.45	0.1890 ± 0.0560	842.24 ± 474.51	**0.1980 ± 0.0673**	NA	NA

all approaches in the two scenarios in terms of MSE and MRE with respect to various window lengths. Based on Figure 4.7 and 4.8 , the m^{2}HMM-SNPH outperforms all the other weighting schemes as well as PSHMCO in the two scenarios.

Interestingly results indicate that the windowing algorithm not only reduces the computational time, but also improves the average performance of PSHMCO by reducing the unnecessary connection to the long past observations if adopted appropriately.

Table 4.6 shows the average computational time required to perform prediction using various approaches at each time step in the two given scenarios. Although the computational time required for m^{2}HMMs in the two scenarios are higher than PSHMCO, the table indicates feasibility of all approaches as all of them are preformed in a fraction of a second. It is also shown that using the windowed observations the required computational time has drastically reduced.

4.5.4 Reference Length Sensitivity Analysis

Here, the effect of reference length hyper-parameter on the performance of the m^{2}HMM with Semi-Nonparametric Hindsight (m^{2}HMM-SNPH) is studied. For this purpose, the average performance of the m^{2}HMM-SNPH approach is measured while varying its

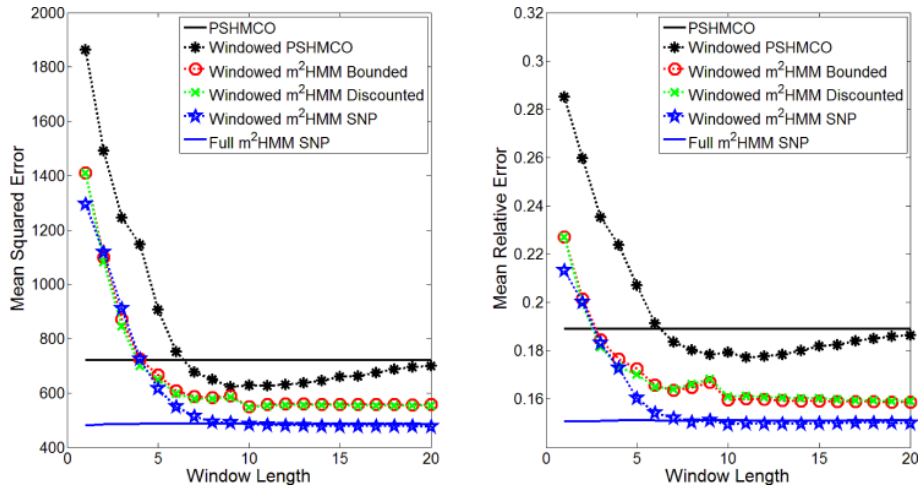


Figure 4.7: Average performance of windowed variants of m^{2}HMM approaches with various weighting schemes along with windowed PSHMCO with respect to window length in the easy Scenario (TT ratio 5:1). The x-axis in both graphs indicate the window length (L_w). Average performance of PSHMCO and the m^{2}HMM-SNPH are included for comparison.

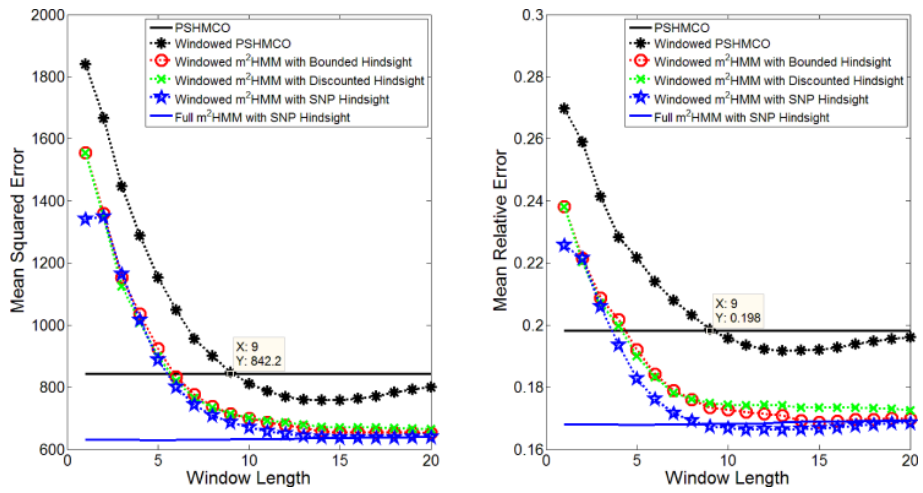


Figure 4.8: Average performance of windowed variants of m^{2}HMM approaches with various weighting schemes along with windowed PSHMCO with respect to window length in the difficult Scenario (TT ratio 3:3).

hyper-parameter value from 1 to 20 (shown in Fig. 4.9).

As it can be seen in Fig. 4.9, the average performance of m^{2}HMM-SNPH changes are small as its hyper-parameter value varies in the two cases, showing that m^{2}HMM

Table 4.7: The average computational time (in milliseconds) required to perform prediction in the two scenarios using each approach.

Approach	Easy Scenario (N_{mode} = 31)	Difficult Scenario (N_{mode} = 7)
Full m^{2}HMM-BH	375.5	85.5
Full m^{2}HMM-DH	380.8	87.6
Full m^{2}HMM-SNPH	410.8	95.0
Full PSHMCO	1.6	
Windowed m^{2}HMM-BH	23.1	5.4
Windowed m^{2}HMM-DH	26.3	6.0
Windowed m^{2}HMM-SNPH	133.1	23.8
Windowed PSHMCO	0.7	

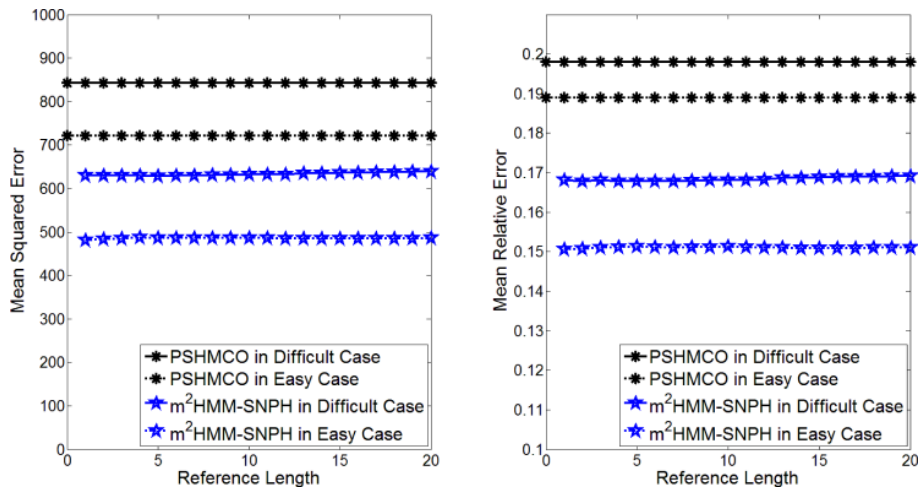


Figure 4.9: Reference length sensitivity analysis in m^{2}HMM-SNPH. The average performance of the m^{2}HMM-SNPH is depicted for the two scenarios Easy (TT ratio 5:1) and Difficult (TT ratio 3:3) while its hyper-parameter value varies. The PSHMCO average performance lines in the two cases are also depicted for comparison.

overall average performance is robust and disregarding its hyper-parameter value better than PSHMCO. Interestingly, the average performance achieved by m^{2}HMM-SNPH (disregarding it hyper-parameter value) in the difficult scenario is even outperforming the PSHMCO average performance obtained in the easy scenario .

4.6 Summary

A multi-modal HMM-based approach, which is an extension of existing physically segmented HMM with continuous output (PSHMCO) approach, is proposed for tool condition monitoring. Two switching strategies are introduced to combine the outputs from various modes into one. The relevance of each mode at each time is assessed based on the given observation at that time and its precedings. Three weighting schemes i.e. bounded-, discounted- and semi-nonparametric hindsight (SNPH) are proposed which quantify the relevance of modes as weightages to be used in switching strategies.

As an illustrative example, the proposed multi-modal approach is applied to cutter wear monitoring in the CNC-milling machine. Based on the experimental results, the m^2HMM approach with soft switching (disregarding its weighting scheme) outperforms the PSHMCO significantly. Furthermore, the m^2HMM with semi-nonparametric and soft switching proves to be the best among all combinations of strategies and weighting schemes.

Moreover, to reduce the computational burden, a windowing algorithm is proposed for both PSHMCO and m^2HMM that dramatically increases the speed. Results indicate that if an appropriate window length is adopted, the windowed version of both approaches can reach the same level of performance as their original and in some cases even outperform them.

Chapter 5

Hidden Markov Model-Based Fault Detection and Diagnosis

5.1 Introduction

Rotary electric motors (REM) provide the basis for the electromechanical energy conversion in all industrial environments [39]. Thus, as the industry grows, the importance of fault detection and condition based maintenance in the rotary electric motors also increases. Early fault detection and diagnosis can help to increase the availability of the industrial machines and reduce the economical loss pertaining to the maintenance of the machinery systems [1].

As mentioned in Chapter 1, the most common fault in the REMs is bearing related faults which are responsible for about 50% of all rotary machine faults [50]. The second most common fault is the unbalanced rotor which causes excessive vibrations in the machines [51, 52] . Among the REMs, synchronous motors are one of the motor types that are widely used in all the industrial applications where constant speed is essential. In this Chapter, a new hidden Markov model (HMM)-based fault diagnosis approach is introduced to distinguish the two major faults namely bearing fault and unbalanced rotor bar from the healthy condition in synchronous motors over a wide range of operating speeds. In this work, vibration signatures are used for fault detection and diagnosis.

As mentioned in Chapter 1, the vibration signature is the most responsive and most

commonly measured signal used for mechanical fault diagnosis in REMs. Machine vibration arises due to action-reaction forces acting on the surface-to-surface contacts of moving machine parts. A healthy machine exhibits low level of vibrations and a machine with unbalanced rotor or bearing defects generates unique vibration signatures [51]. However, the detection of bearing faults using vibration signals is affected by the machine speed [118]. In this Chapter, vibration signals are used as the observation sequences through time which are fed to the HMMs used in the proposed approach.

Hidden Markov models (HMM) are extensively used for fault detection and diagnosis [58, 75, 76, 77, 78, 79, 80, 81] as well as failure prognostics [82] in various rotary electrical motors. In all cases, the HMM-based approaches are successful in distinguishing healthy condition from faulty conditions (fault detection). The challenging part is to diagnose the faults as the amplitude of the vibration signals from various faults may be similar with various operating speeds. That increases the chance of misclassification based on maximum likelihood strategy considering the fact that the true model is not completely realizable in real-world applications. In this Chapter, to improve the performance of the existing HMM-based approach, a semi-nonparametric approach is proposed. In this approach, after training the HMM classifiers (parameter estimation stage), two matrices named *probabilistic transition frequency profile* and *average probabilistic emission* are computed and stored based on the HMMs for each signature (nonparametric stage) in the training phase using probabilistic inference. When a new signature is given for fault diagnosis, after applying all HMM classifiers on the given signature and computing the required matrices based on each HMM, the similarity between the given signature and the training signatures in each class is assessed with a scoring function based on the stored matrices. The introduced approach is named *HMM-based Semi-NonParametric* (HMMSNP) since firstly it uses both HMM (parametric) and similarity of the resultant matrices with the corresponding matrices from the training data (nonparametric), and ultimately the classification result is based on the nonparametric stage on top of the conventional (parametric) HMM classifiers. Performance of the proposed approach is compared with the conventional HMM-based approach. Furthermore, a preprocessing method named *squeezing and stretching* is introduced to rectify the difficulty

of dealing with various operating speeds in the classification process.

The rest of this Chapter is organized as follows. In Section 2, the mechanics of vibration for bearing and unbalanced rotor faults are introduced. Section 3 describes the rationale and the procedure of the proposed preprocessing method. In Section 4, the conventional HMM-based fault diagnosis approach is briefed, and the proposed HMMSNP approach is introduced. Preliminary experimental results are provided and compared for the mentioned approaches in Section 5. In Section 6, further investigations are carried out to show the statistical significance of the performance improvements. Sensitivity analysis is also conducted on the number of hidden state values and the length of input signatures. Finally, this Chapter is concluded in Section 7.

5.2 Rotary Machine Fault Mechanics

Machines with rolling elements have moving bearings, e.g. spherical balls, tapered rollers, or cylindrical rollers, to support the rotating shaft. These rolling elements are always in metal-to-metal contact with the inner and outer raceway, and as a result are subject to constant wear and tear. Bearing and raceway wear and tear present initially as general roughness and progresses to metal fatigue, and ultimately spall and chip on the surface of the rolling elements [50].

Defective surfaces on these components are a source of machine vibration. A chipped rolling element spins as it revolves around the raceway. When it is in contact with the defective surface of the raceway, an impact pulse is produced, creating a free vibration. In the absence of significant damping medium in the bearing assembly, the impact pulses decay exponentially.

Another major fault in the rotary machines is unbalanced rotor. Unbalanced rotor is the most common source of excessive vibration. Possible causes are, asymmetrical mass distribution of the rotating element as a result of wear, erosion, material build-up, thermal expansion or contraction, causing shaft bending or misalignment. As a result, the center of gravity of the rotating element does not coincide with the center of rotation, and at the point of unbalanced mass creates a synchronous radial perturbation

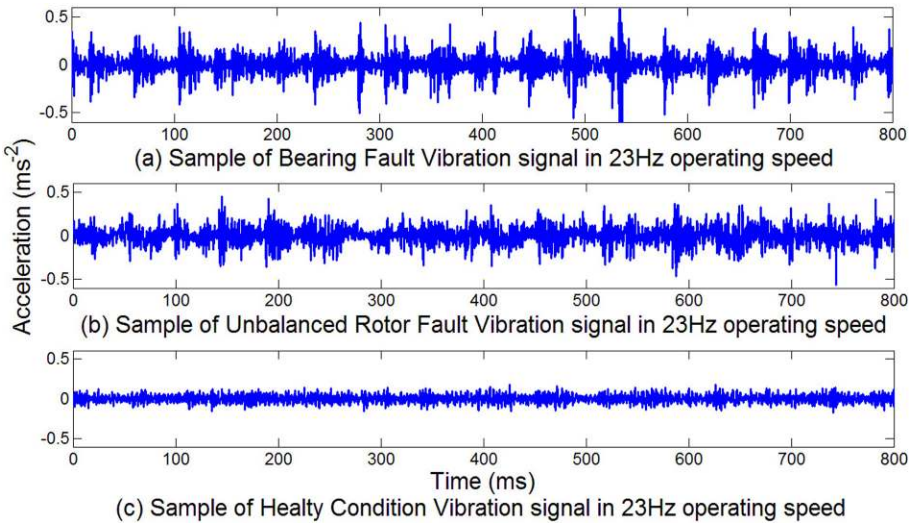


Figure 5.1: Samples from three conditions at 23Hz operating speed. (a) Bearing fault, (b) unbalanced rotor and (c) healthy condition.

force F_c, causing a forced vibration. Assuming a rigid isotropic rotor system [52], this phenomenon is described as

$$F_c = m_{ubr} \times r_{ubr} \times \Omega^2 \times e^{\jmath(\Omega t+\theta)}, \tag{5.1}$$

where m_{ubr} is the unbalanced rotor mass, r_{ubr} is the distance between the unbalanced mass and the center of gravity of the rotor, t is time, Ω is the shaft rotational speed, θ is the angular position of r_{ubr} and $\jmath$ is the complex operator. As a reference, healthy machine signatures are also recorded. Fig. 5.1 shows three sample vibration sequences under different conditions, i.e. bearing fault (BRG), unbalanced rotor (UBR) and healthy (HTY) conditions at 23Hz as a medium rotating speed. As can be seen in Fig. 5.1, a healthy machine has the lowest level of vibration compared to other conditions.

Note that the signatures corresponding to the same fault are different in period and amplitude when the operating speeds are different. Figure 5.2 depicts the unbalanced rotor fault signatures with the operating speed starting from 15Hz and increased up to 32Hz. As the operating speed increases from 15 Hz to 32 Hz, signatures are getting squeezed along the time horizon and increased in amplitude. Thus, a preprocessing procedure is required to map all the signatures acquired from various operating speeds to a common speed. For this purpose, a preprocessing method is proposed in the succeeding

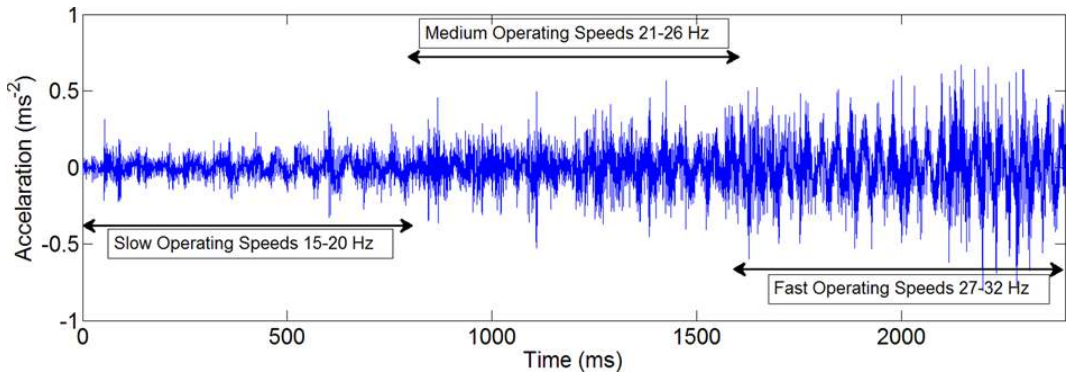


Figure 5.2: Unbalanced rotor fault signature generated at various speeds ranging from 15 to 32 Hz. The Operating frequency is increased by 1Hz starting from 15Hz. The time span for data recorded at each operating speed (frequency) is 66.8 ms. As the operating frequency increases, the signal squeezes and the vibration in the y-axis increases. This observation is the basis of the Squeezing and Stretching preprocessing.

section.

5.3 Signature Squeezing & Stretching

In this Section, a signature preprocessing method named *Squeezing and Stretching* (SqS) is introduced. The idea of SqS method is similar to Dynamic Time Warping (DTW), that is used extensively in data-mining and sequence labeling applications [81]. In DTW, the similar subsequences between the reference sequence and the testing sequence will be dynamically matched [119, 120]. This method warps the time axis at various points to match the two sequences. However, the length of these warpings are not fixed and they have to be computed dynamically with the objective of achieving the highest similarity matching. In signature squeezing, as the operating speeds are known, thus the conversion of the time series to a common time line is simply possible through a fixed warping and appropriate resampling. The vibration amplitude in the signals acquired from lower operating speeds can also be scaled to match a common higher operating speed through stretching in amplitude.

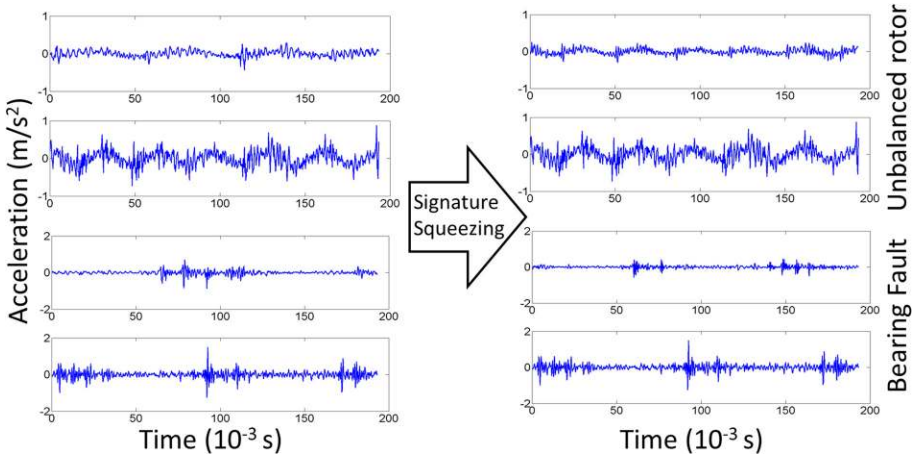


Figure 5.3: Signature squeezing application scheme. In both examples, the upper signature is squeezed to match the speed of the other signature at the bottom and the results after squeezing process are shown and compared on the right.

5.3.1 Squeezing in Time

In this phase of preprocessing, all the signatures will be squeezed to a common high operating speed. After adopting a high operating speed (the highest operating speed available in the training set) as the common operating speed, since the operating speed is known for each signature, signatures will be squeezed based on their ratio of speed to the adopted common operating speed. The squeezing in time mappings for time (x-axis) can be formulated as

$$t_s = \frac{\Omega_o}{\Omega_c} \times t_o, \tag{5.2}$$

where t_s is the squeezed time, t_o is the time line in the original sequence, Ω_o is the operating speed of the original signature and Ω_c is the common operating speed. Figure 5.3 shows examples of applying squeezing method on UBR and BRG signatures.

As can be seen in Fig. 5.3, after applying the squeezing preprocessing, the period of the sequences in both cases matches although the signatures are asynchronous and have relative phase difference.

5.3.2 Stretching in Amplitude

After squeezing procedure, the period differences between the signatures are remedied. However, as can be seen in Fig. 5.3, the amplitude differences between sequences acquired from various operating speeds are obvious. Thus, further preprocessing is done to map the amplitude of the given signatures to the common operating speed. Here, the signatures are stretched in the y-axis, based on the average peak-to-peak amplitude (APPA) in the signatures gathered from that operating speed and the common speed. After windowing the signatures, the minimum and maximum amplitudes within each window is found and used to calculate APPA for each signature. The ratio of average APPA in common speed to the average APPA is used for scaling the signature signal in amplitude. It is noteworthy that in this procedure all the signatures, no matter which fault and condition they are indicating, are stretched based on the average APPA in the collected speed w.r.t. the common speed. Thus the stretching procedure can be formulated as

$$y_s = \frac{\overline{APPA}_c}{\overline{APPA}_o} \times y_o, \tag{5.3}$$

where y_s is the stretched vibration signal, y_o is the original vibration signal before transformation, $\overline{APPA}_c$ is the average APPA in the common operating speed, and $\overline{APPA}_o$ is the average APPA in the collected operating speed. Figure 5.4 depicts the effect of signature stretching in matching the amplitudes from various operating speeds by relative magnification of the vibration amplitudes.

5.4 HMM-based Fault Diagnosis

Hidden Markov models can be used as generative models in fault diagnostics. In its conventional form, each HMM is trained on a specific fault. Whenever a new data is given to the model, each HMM returns the log-likelihood that the new data is generated by a similar fault as the training data. This approach is used in many fault detection and diagnostics applications [75, 76, 77, 80]. In this Chapter, the conventional HMM-based fault diagnostics is used as a benchmark to assess the proposed HMM-based approach.

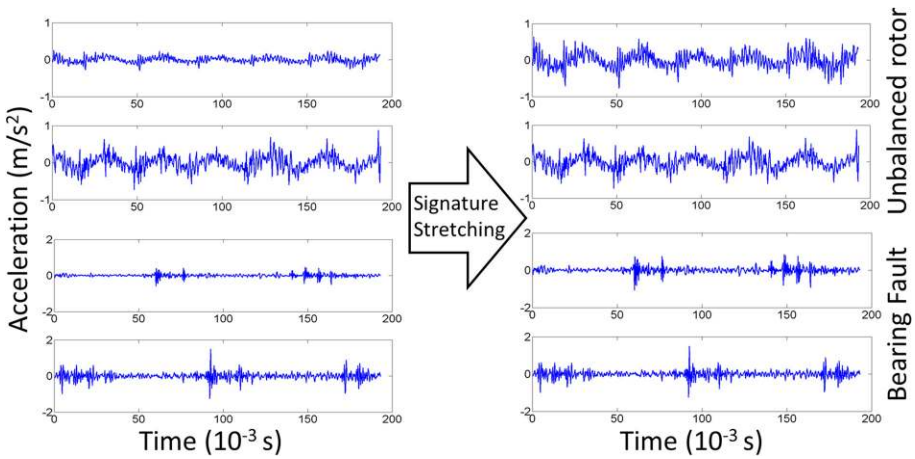


Figure 5.4: Signature stretching application on the pre-squeezed signatures. In both examples, the upper signature is stretched in the amplitude based on (5.3) to match the other signature at the bottom and the results after stretching process are shown and compared on the right.

5.4.1 Conventional HMM-Based Classification

As mentioned in Chapter 1, hidden Markov model is a simple dynamic Bayesian network. This model has only one discrete hidden state variable, and a set of discrete or continuous observation nodes. The HMM parameters can be estimated using expectation-maximization techniques. The hidden state variable in the hidden Markov model can take different state values. As the number of these values increases, the computational cost for estimating the parameters and performing inference on the HMM becomes more and more expensive. Thus, an appropriate number of hidden state values for the HMM must be adopted and it can be done by considering either the average log-likelihood of the trained models with various number of hidden state values on the training set or cross-validation on the training set.

Assume the number of hidden state values adopted is m and the state values are $\{v_1, v_2, \ldots, v_m\}$. Then, a first order temporal Markov model can be characterized by the assumption that

$$P(S_t = v_i | S_{t-1}, S_{t-2}, \ldots, S_1) = P(S_t = v_i | S_{t-1}), \qquad i = 1, \ldots, m \tag{5.4}$$

where $P(\cdot|\cdot)$ is a conditional probability, S_t is the hidden state variable at time t and v_i is

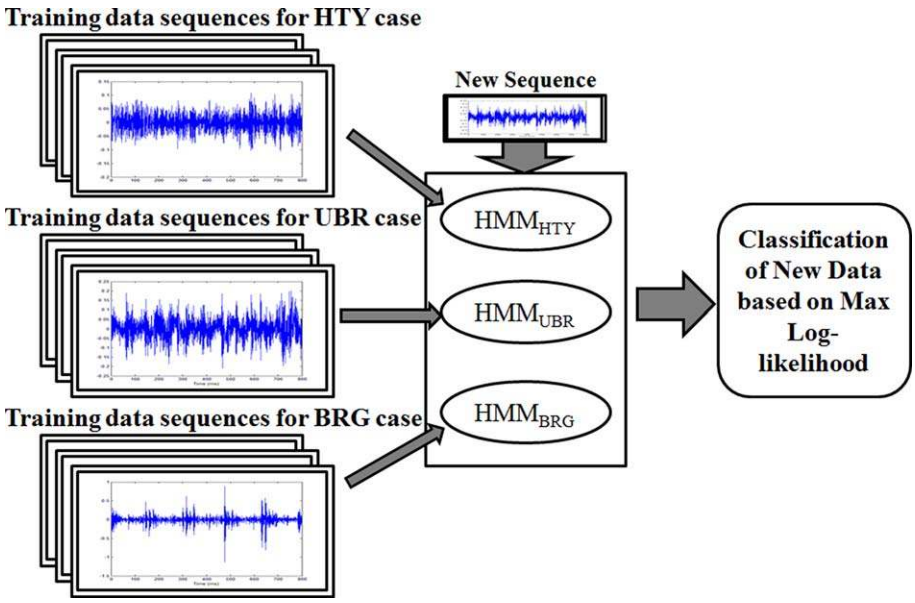


Figure 5.5: Conventional HMM-based fault diagnostics scheme with three classes i.e. healthy condition (HTY), unbalanced rotor (UBR), and bearing fault (BRG).

the ith hidden state value.

Assuming stationarity, the initial state probabilities (prior probabilities), transition probabilities and emission probabilities, which connect the hidden states to the observations (vibration signals) are the only parameters to be identified. Given the training data, these parameters can be estimated using expectation-maximization technique [80].

Task of fault diagnostics in REMs is to correctly classify a given signature (sequence of observations through time) into one of the predefined classes such as healthy condition or various types of faults. This task is conventionally done by comparing the provided log-likelihoods from different HMMs each trained on the samples having specific faulty or healthy condition [80]. The newly given signature is classified into the condition whose corresponding HMM achieves the maximum log-likelihood among all HMMs. Based on a trained HMM with estimated parameter set of λ, likelihood of a given sequence being generated by that HMM can be computed using forward-backward algorithm [95]. As the signatures are periodic and acquired asynchronously, no prior information can help us to identify the initial probability for the HMM and all the states are equally probable with a uniform distribution. Thus, the initial probability can be ignored in the calculations.

As illustrated in Fig. 5.5, in the conventional HMM-based fault diagnostics, the newly given signature is classified as the corresponding class of the HMM that achieves the maximum log-likelihood.

5.4.2 HMM-based Semi-Nonparametric Approach

As the true model in real-world applications is not always realizable, we will try to improve the classification performance by integrating and taking the advantages of both parametric (model based) and nonparametric classification approaches.

Here, an *HMM-based semi-nonparametric* (HMMSNP) approach is developed to improve the performance of the conventional HMM-based approach that relies on the log-likelihood. In this approach, after performing the *squeezing and stretching* preprocessing, the parameters are learned from the training sequences. The parameter estimation procedure is identical with the conventional HMM-based approach. When the expectation-maximization training procedure is over, the trained HMMs are applied on their corresponding training sequences. Using the probabilistic inference, two matrices named *probabilistic transition frequency profile* and *average probabilistic emission* which are denoted as F and E, respectively, are computed and stored for each training sequence. These matrices will be used to form a scoring function that is used as the classification basis. The procedure of computing F and E matrices and their computation rationale are explained in the succeeding subsections.

After training the HMMs based on the training signatures, given a signature, probability of state values that the state variable of an HMM takes at each time step can be estimated using forward-backward algorithm [95]. As described in the preceding Chapters, γ_t is defined as the probability distribution of the state values at time t and it can be written as

$$\begin{aligned} &\gamma_t = \Big[\gamma_t(i)\Big]_{m\times 1}, \quad o_{1:T} = \{o_1, o_2, \ldots, o_t, o_{t+1}, \ldots, o_T\} = o_{1:t} \cup o_{t+1:T} \\ &\gamma_t(i) \triangleq P(S_t = v_i, o_{1:T}) = P(o_{1:t}, S_t = v_i) \times P(o_{t+1:T}|S_t = v_i) = \alpha_t(i) \times \beta_t(i), \end{aligned} \tag{5.5}$$

where $o_{1:T}$ is the given sequence of observations, which in this case is the vibration signal after SqS preprocessing in terms of acceleration, T is the length of the given sequence,

and S_t is the hidden state at time step t. $\alpha_t(i)$ and $\beta_t(i)$ are the forward and backward variables that can be computed using the forward-backward algorithm [95].

Probabilistic Transition Frequency Profile

Using the computed probabilities of each state at every time step, the probabilistic transitions between the states can be computed by the outer product of consequent time steps probabilities. After computing all the probabilistic transitions given the sequence and the parameters of the HMM, the average of the computed matrices called *probabilistic transition frequency profile* (PTFP) is stored and used as the overall transition pattern for the given signature based on a given HMM. Here, PTFP is denoted by F and it can be formulated as

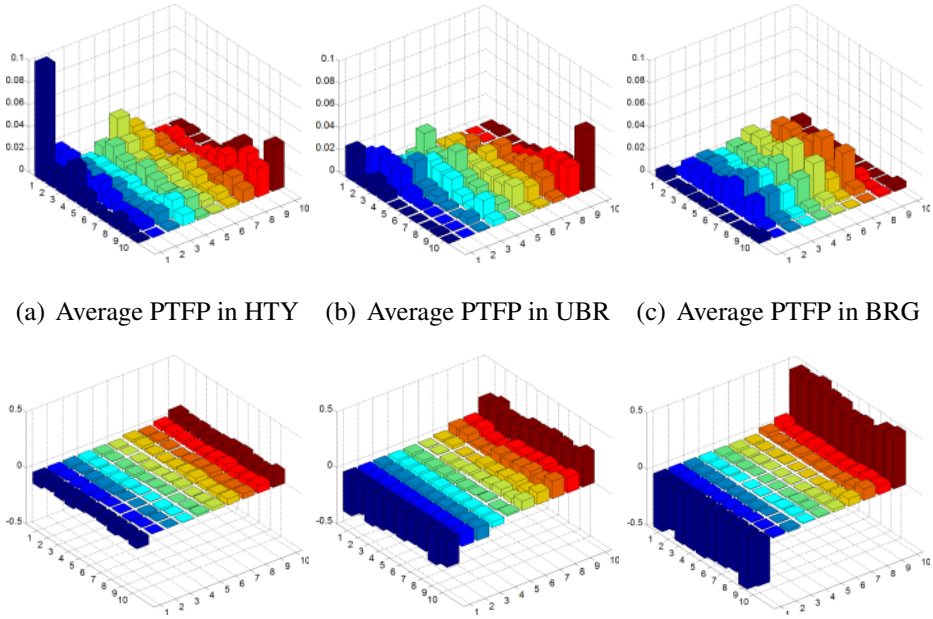
$$F = \left[f_{i,j}\right]_{m\times m} = \frac{1}{T-1} \times \sum_{t=2}^{T} \gamma_{t-1} \otimes \gamma_t, \tag{5.6}$$

where T is the length of the signature, $\gamma_{t-1} \otimes \gamma_t$ is the outer product of the two vectors γ_{t-1} and γ_t. $f_{i,j}$ is the ith row, jth column element in matrix F, which indicates the probabilistic frequency that transition from ith state to jth state has occurred in that signature (normalized by the total number of transitions throughout the sequence). In other words, matrix F captures a normalized 3-dimensional *transition frequency map* for each signature.

The transition frequency map can be used to recognize differences and distinguish between signatures which have distinct trend of transition frequency. Figure 5.6 visualizes the average F matrices in the HTY, UBR, and BRG conditions (classes). It is noteworthy that although the number of state values in all the classes are set to be identical, the state values vary between the classes. However, in order to be able to visually compare the overall trends between matrices from different classes, the states are rearranged and sorted based on their average emission value in the vibration domain.

Average Probabilistic Emission

It is noteworthy that the map provided by F only recognizes the differences between sequences of transitions in the state values after applying HMM on the signatures. However, as the probabilistic transition frequencies are normalized to become comparable



(a) Average PTFP in HTY (b) Average PTFP in UBR (c) Average PTFP in BRG

(d) Average APE in HTY (e) Average APE in UBR (f) Average APE in BRG

Figure 5.6: schematizing PTFP (F) and APE (E) matrices as a 3-dimensional map. Average PTFPs from the training data for the HTY, UBR, BRG classes are shown in the first row from left to right, respectively. Similarly, the APEs are shown in the second row. For the sake of better visual comparison, the states are sorted based on their average emitted vibration value within each class.

between various classes, there may be cases that identical transitions correspond to very different real-valued observations in the vibration domain (e.g. when an HMM from HTY class is applied on a signature from UBR class while performing testing). Thus, another metric is required to assess how similar two identical transitions within two signatures are in their real-valued observed sequences (emitted vibrations). For this purpose, the average probabilistic emission (APE) matrix denoted by E is introduced as a metric to be used in the similarity scoring function. Given an observation sequence (signature), E can be computed as follows

$$\begin{aligned} E &= \left[e_{i,j}\right]_{m\times m}, \\ e_{i,j} &= \frac{\sum_{t=2}^{T} o_t.\gamma_{t-1}(i).\gamma_t(j)}{\sum_{t=2}^{T} \gamma_{t-1}(i).\gamma_t(j)}, \end{aligned} \tag{5.7}$$

where o_t is the observed vibration at time t in the signature (after performing SqS preprocessing), and $e_{i,j}$ is the ith row, jth column element in matrix E. $e_{i,j}$ indicates the

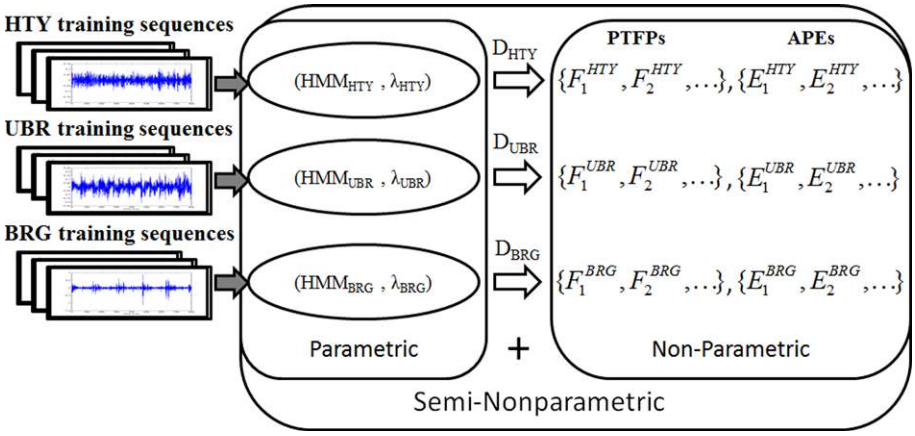


Figure 5.7: Training phase illustration in the HMM-based semi-nonparametric approach. After estimating the parameters of the HMMs (parametric phase) for healthy (HTY), unbalanced rotor (UBR) and Bearing fault conditions, the HMMs are applied on their corresponding training sequences (D_{HTY}, D_{UBR}, and D_{BRG}) and their corresponding PTFPs and APEs are computed and stored (nonparametric phase).

average observed vibration through the sequence with the probabilistic transition from ith state to jth state. Similar to F matrix, E can also be recognized as a 3-dimensional graph. This graph indicates the average observed vibration through the sequence while the transition has taken place from i to jth state. Here, our goal is to firstly, find the most similar training sequence to the given sequence in terms of PTFP and APE within each class. Then use the highest similarity scores achieved from each class and compare them to finally classify the given sequence. Figure 5.7 schematizes the training stage in the HMMSNP approach implemented to classify the three conditions (HTY, UBR, and BRG) in this application.

Classification

When a new sequence is given to the system for fault diagnostics (classification), its PTFP and APE are first computed based on each HMM corresponding to a specific class (condition). In the testing phase, similar to the nonparametric nearest neighbor approach [98, 121], we will try to find the nearest training data in terms of F and E matrices and classify the newly given sequence as the same class that the most similar

training data belongs to. To this end, first we define a similarity measure $\delta(.,.)$ for two given E (or F) matrices as follows

$$\delta(E_1, E_2) = \sum_{i=1}^{m} \sum_{j=1}^{m} \left| \frac{E_1(i,j) - E_2(i,j)}{E_2(i,j)} \right|^{-1}, \tag{5.8}$$

where E_1 and E_2 are two arbitrary E matrices. $\delta(.,.)$ computes the summation of element-wise inverted relative difference between the two given matrices as an indicator of similarity. Then, the similarity of the signature with the training sequences used to train those HMMs is assessed based on a similarity scoring function that uses $\delta(.,.)$ on the computed PTFP and APE matrices for the given sequence and those matrices previously computed and stored during the training phase. The goal of the similarity scoring function is to assess how similar two sequences are based on their PTFP and APE matrices taking into account that two signatures with high similarity scores must have relatively high similarity measure values in terms of both APE and PTFP. Thus, the similarity scoring function, G, is defined as multiplication of the two similarity measures as follows

$$G(o'_{1:T}|\lambda_i, O^{i,q} \in D_i) = G(F^i_{new}, E^i_{new}|F^i_q, E^i_q) = \delta(F^i_{new}, F^i_q) \times \delta(E^i_{new}, E^i_q), \tag{5.9}$$

where $o'_{1:T}$ is the new observation sequence (signature) that has to be classified, λ_i is the parameter set of the ith HMM, $D_i = \{O^{i,1}, O^{i,2}, \ldots, O^{i,n_i}\}$ is the set of training signatures that corresponds to the ith condition. F^i_q, E^i_q, F^i_{new} and E^i_{new} are the F and E matrices calculated based on ith HMM from the qth training signature, $O^{i,q}$, used to train the ith HMM and the given sequence, respectively. After applying the HMMs (parametric phase) on the newly given sequence and computing the corresponding F_{new} and E_{new} matrices, the most similar sequence (signature) to the new sequence based on (5.9) is found from each class (nonparametric phase). The highest similarity score within each class is regarded as the representative similarity score of that class as follows

$$Q_i(o'_{1:T}) = \max_{q=1,\ldots,n_i} \{G(o'_{1:T}|\lambda_i, O^{i,q} \in D_i)\} = \max_{q=1,\ldots,n_i} \{G(F^i_{new}, E^i_{new}|F^i_q, E^i_q)\}, \tag{5.10}$$

where $Q_i(o'_{1:T})$ is the ith class similarity score computed for the given signature $o'_{1:T}$, and n_i is the number of training sequences for the ith class. Finally, the new sequence is classified as the class with the highest score. Thus, the classification output can be

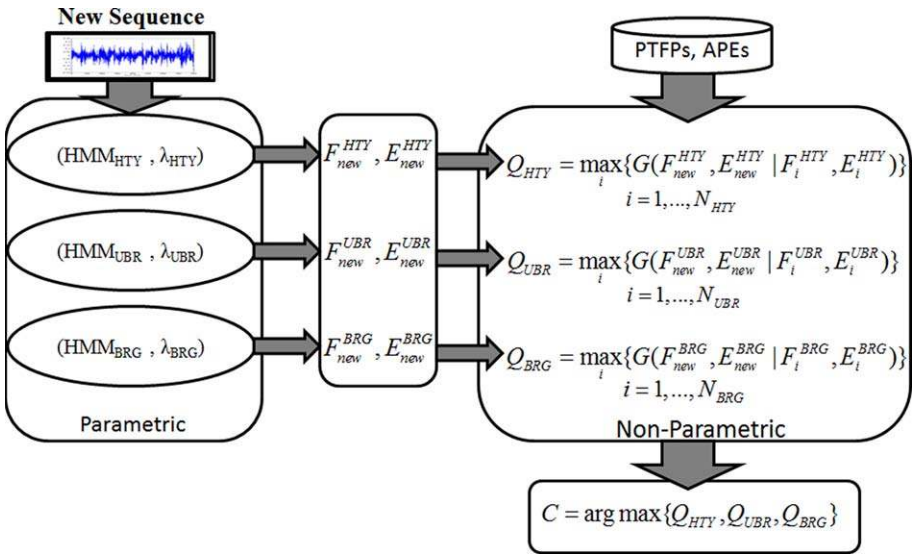


Figure 5.8: Testing phase illustration in the HMM-based semi-nonparametric approach. After applying the HMMs (parametric phase) on the newly given sequence and computing the corresponding F and E matrices, the most similar signature from each class is found based on their PTFP and APE matrices (nonparametric phase). The highest similarity score within each class is regarded as the representative similarity score. Finally, the new sequence is classified as the class with the highest score.

written as

$$C(o'_{1:T}) = \arg\max\{Q_1(o'_{1:T}), Q_2(o'_{1:T}), \ldots, Q_K(o'_{1:T})\}, \tag{5.11}$$

where K is the number of classes (conditions) and $C(o'_{1:T})$ is the class index that $o'_{1:T}$ is ultimately classified as based on HMMSNP approach. Figure 5.8 schematizes the HMMSNP classification (testing phase) implemented for fault diagnostics.

5.5 Preliminary Experimental results

In this Section, three fault diagnosis approaches which are HMMSNP, the conventional HMM-based approach without SqS preprocessing and HMM-based approach with SqS preprocessing (HMMSqS) are evaluated and compared on a fixed dataset. The HMMs in all three approaches are trained on one training set and then tested on another testing set. These approaches are evaluated based on their accuracies and the confu-

sion matrices shown in Table 5.1 as well as the exemplified average incurred cost using each approach. To have a fair comparison between the three approaches, the number of hidden state values in all the HMMs is fixed to 10.

In this Section the acquired dataset mentioned in Section 3 is partitioned into training and testing set based on the machine operating speed, the data from all the machine conditions that has been acquired from 15Hz,17Hz,...,31Hz is considered as the training set and the rest of the dataset which is acquired from 16Hz,18Hz,...,32Hz are regarded as the testing set.

5.5.1 Classification Accuracy

As expected, the conventional HMM-based approach without the use of *Squeezing & Stretching* preprocessing, achieves the least accuracy since it does not consider the differences between the signatures with various rotary speeds within the classes. The second best performance is achieved by the conventional HMM-based approach with the use of SqS preprocessing (HMMSqS), which unifies the speed within the signatures by mapping all the different rotary speeds to the highest rotary speed available in the set and scaling their vibration amplitude. Using the SqS preprocessing, the accuracy is significantly improved comparing to the conventional HMM-based approach. Finally, the HMM-based Semi-nonparametric approach with the use of SqS preprocessing (HMM-SNP) has achieved the highest accuracy on the testing set.

Table 5.1: Classification accuracy and the confusion matrices using HMM, HMMSqS, and HMMSNP evaluated on the testing set.

Approach	**HMM**			**HMMSqS**			**HMMSNP**		
	HTY	**UBR**	**BRG**	**HTY**	**UBR**	**BRG**	**HTY**	**UBR**	**BRG**
HTY	**73**	17	0	**90**	0	0	**90**	0	0
UBR	19	**66**	5	0	**80**	10	0	**90**	0
BRG	9	21	**60**	9	14	**67**	7	12	**71**
Total Diagnostics Accuracy	**73.70%**			**87.78%**			**92.96%**		

As can be seen from Table 5.1, the semi-nonparametric approach has achieved the highest accuracy and shows considerable improvement to the conventional HMM-based approaches with or without the SqS preprocessing. Although, it is shown that HMM-SNP has improved the classification performance on the given testing set, further investigations are carried out in the succeeding Section to see whether the improvement is repeatable and statistically significant.

5.5.2 Cost Analysis

To compare the performance of the three mentioned approaches in terms of financial benefit, here we define an average incurred cost matrix which exemplifies the actual costs that various classifications/misclassifications can lead to. Table 5.2 shows the postulated costs and required hours of maintenance. The average incurred cost C, which takes into account material cost, downtime cost and man power cost is defined as follows

$$\begin{aligned}
C &= \left[C_{i,j}\right]_{3\times 3}, \\
C_{1,1} &= C_{HTY|HTY} = 0 + 0 + 0, \\
C_{1,2} &= C_{UBR|HTY} = P^{UBR} + {t_{dd}}^{UBR} \times V + {t_m}^{UBR} \times M, \\
C_{1,3} &= C_{BRG|HTY} = P^{BRG} + {t_{dd}}^{BRG} \times V + {t_m}^{BRG} \times M, \\
C_{2,1} &= C_{HTY|UBR} = P^{UBR} + t_{ud} \times V + (t_I + {t_m}^{UBR}) \times M, \\
C_{2,2} &= C_{UBR|UBR} = P^{UBR} + {t_{dd}}^{UBR} \times V + {t_m}^{UBR} \times M, \\
C_{2,3} &= C_{BRG|UBR} = C_{2,2} + C_{3,3} + t_{aI} \times M, \\
C_{3,1} &= C_{HTY|BRG} = P^{BRG} + t_{ud} \times V + (t_I + {t_m}^{BRG}) \times M, \\
C_{3,2} &= C_{UBR|BRG} = C_{2,2} + C_{3,3} + t_{aI} \times M, \\
C_{3,3} &= C_{BRG|BRG} = P^{BRG} + {t_{dd}}^{BRG} \times V + {t_m}^{BRG} \times M,
\end{aligned} \tag{5.12}$$

where all the parameters and their postulated values are listed in Table 5.2.

Based on Table 5.2, the cost matrix C given in (5.12) can be evaluated as

$$C = \begin{bmatrix} 0 & 850 & 800 \\ 15050 & 850 & 1750 \\ 15000 & 1750 & 800 \end{bmatrix}. \tag{5.13}$$

Table 5.2: List of assumed material and human resource costs.

Notion	Description	Value
V	Value generated by the machine per hour	200$
M	Maintenance fee given to technician per hour	50$
P^{BRG}	Average Material Cost for BRG Fault	50 $
P^{UBR}	Average Material Cost for UBR Fault	100$
${t_{dd}}^{UBR}$	Downtime for detected UBR fault	3 hours
${t_m}^{UBR}$	Maintenance time for UBR fault	3 hours
${t_{dd}}^{BRG}$	Downtime for detected BRG fault	3 hours
${t_m}^{BRG}$	Maintenance time for BRG fault	3 hours
t_{aI}	Additional Inspection time	2 hours
t_{ud}	Downtime for undetected fault	3 days
t_I	Inspection time	8 hours

Table 5.3 shows the average incurred cost resulted from the three approaches. The cost matrices shown in Table 5.3 are calculated by element-wise multiplication of the confusion matrices and the average incurred cost matrix C in (5.13). It is also shown that the incurred cost would be $148,500$ if all the samples were detected correctly. And in case of taking a failure-driven approach (assuming that machine is healthy for all signatures) would incur a cost equal to $2,704,500$. As it can be understood from Table 5.3, in this example, the HMMSNP saves 325,400 and 39,000 compared to HMM and HMMSqS, respectively.

5.6 Further Investigations and Sensitivity Analysis

In this Section, the efficiency of the proposed approach is further investigated to confirm whether it can statistically improve the performance. Moreover, the effect of the signature length and the hyper-parameter value on the classification accuracy is studied.

Table 5.3: Cost Analysis for HMM, HMMSqS, and HMMSNP approaches evaluated on the testing set.

Approach	HMM			HMMSqS			HMMSNP		
	HTY	**UBR**	**BRG**	**HTY**	**UBR**	**BRG**	**HTY**	**UBR**	**BRG**
HTY	0	14,450	0	0	0	0	0	0	0
UBR	285,950	56,100	8,750	0	68,000	17,500	0	76,500	0
BRG	135,000	36,750	48,000	135,000	24,500	53,600	105,000	21,000	56,800
Total Cost	**585,000**			**298,600**			**259,600**		
Misclassification Cost⋆	**436,500**			**150,100**			**111,100**		

⋆ Cost Due to Misclassification = Total Incurred Cost - Optimum Cost

For this purpose, different trials are generated by randomly partitioning the whole dataset and the aforementioned three approaches are compared based on the randomly partitioned trials. Assume that the operating speeds are categorized into three, that is, Slow: 15-20 Hz, Medium: 21-26 Hz, Fast: 27-32 Hz. In each trial, 2 operating speeds out of 6 within each category are randomly adopted as the testing data and the rest as the training set. The categorization of speeds is done to make sure that all three ranges of speeds would be included within both testing and training set, although they are selected randomly. This investigation is carried out by randomly generating 30 distinct trials out of $(C_2^6)^3 = 15^3$ possible combinations (nearly 1 percent of all possible combinations).

5.6.1 Overall Performance

Here, the three fault diagnostics approaches, which are the conventional HMM classifier, HMM-based Classifier with SqS preprocessing, and the HMMSNP approach, are examined and compared on the 30 randomly partitioned trials. In all three approaches, the number of the hidden state values in the HMMs is adopted to be 10 and the length of the signatures is fixed to 6 motor rotations (6 periods).

Figure 5.9 shows the resultant classification accuracy of the aforementioned ap-

proaches on the 30 trials. It can be seen, that the accuracy has been significantly improved by the use of the SqS preprocessing. It can also be identified that the semi-nonparametric approach has achieved an overall better performance in comparison to the other two conventional HMM classifiers. A pair-wise t-test on the HMMSqS and the semi-nonparametric approaches indicates that the mean of the resultant accuracies from these two approaches are significantly different (P-value=1.74e-9, where average accuracy using HMMSNP is 92.8% comparing to 87.6% in case of HMMSqS). Thus, the HMMSNP approach improves the classification accuracy nearly 5% in a statistically significant manner.

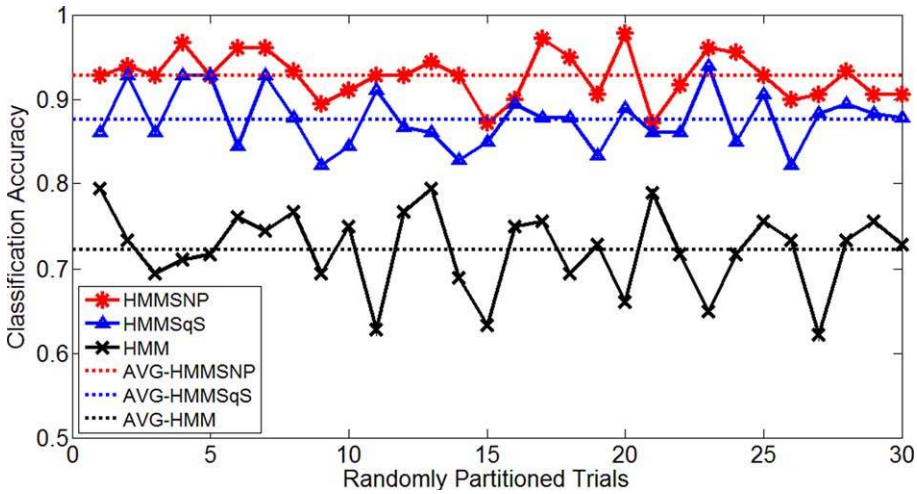


Figure 5.9: Resultant classification accuracies on 30 random trials using conventional HMM, the HMM (HMMSqS) and HMM-based Semi-nonparametric (HMMSNP) approaches with SqS preprocessing.

Table 5.4 shows the average computation time required to classify a given signature using each approach in terms of milliseconds. All the approaches are implemented and run on a same regular PC (Dell-OPTIPLEX 980). As expected, by adding the nonparametric phase to the HMM-based classification approach, the computation time increases. However, the computation time required to classify a new given signature in all three approaches are less than a second that makes them computationally feasible for this task.

Table 5.4: Computation time in various fault diagnostics approaches given a new signature for classification in milliseconds.

	HMM	**HMMSqS**	**HMMSNP**
Computation Time (milliseconds)	247.6 ± 4.1	247.7 ± 0.6	321.1 ± 1.3

5.6.2 Hyper-parameter Sensitivity

Here, the sensitivity of the aforementioned approaches w.r.t. the value of the hyper-parameter i.e. number of the hidden state values is studied in terms of accuracy and computational cost. Figure 5.10 shows the performance of the three approaches w.r.t. various hyper-parameter values ranged from 2 to 20. Each point of the error bars in Fig. 5.10 indicates the mean and variance of the classification accuracies achieved by the approach in the 30 randomly partitioned trials given the specified number of hidden state values.

As can be seen in Fig. 5.10, the performance of all three approaches would generally improves as the number of hidden state values increases. However, this increment in the accuracy is steeper for the smaller values and it starts to reduce for larger values. It can also be seen from Fig. 5.10, that except for the case of two hidden state values, HMMSNP outperforms the other two approaches. This indicates that HMMSNP outperforms HMMSqS and conventional HMM classifier in fault diagnostics given that an appropriate number of hidden state values is adopted. It is noteworthy that in all three approaches, as the number of hidden state values (m) increases the computational complexity increases exponentially with $O(m^2)$. Thus, it is important to adopt the smallest number of states that can satisfy the required accuracy. In this study, based on Fig. 5.10, the number of hidden state values used for HMMs in the primary results Section, which is 10, shows to be an appropriate value. Thus, the same value is adopted in the succeeding signature length sensitivity analysis.

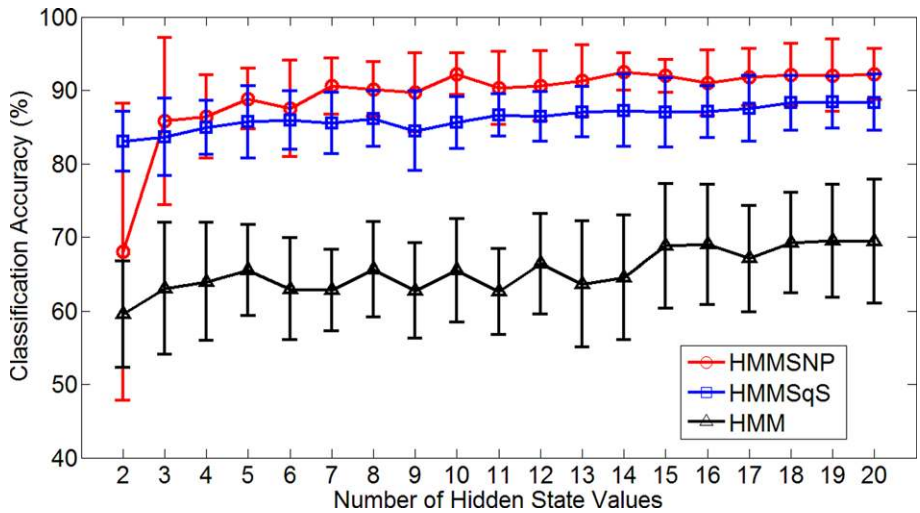


Figure 5.10: Resultant classification accuracies from various number of hidden state values using conventional HMM compared with the HMM (HMMSqS) and HMM-based Semi-nonparametric (HMMSNP) approaches with SqS preprocessing.

5.6.3 Signature Length Sensitivity

In this part, the sensitivity of the aforementioned approaches w.r.t. the length of the given sequences (signatures) is studied. In Fig. 5.11, the performance of the three approaches are shown in terms of classification accuracy for various adopted signature lengths. Each point of the error bars in Fig. 5.11 indicates the mean and variance of the classification accuracies achieved by the approach over the 30 randomly partitioned trials given the specified length of signatures. As can be seen in Fig. 5.11, the classification accuracy increases in all the three approaches up to a certain length and then starts to fluctuate. This indicates that all three approaches are sensitive to the length of signatures while it is ranged between 1 to 6 rotations, but after passing that range, they become insensitive to the length. In this experiment, the classification accuracy of all three approaches monotonically increases up to the point that signature length is equal to 6 rotations. To have a fair comparison between the three approaches, the number of the hidden state values is fixed and set to 10.

It can be seen from Fig. 5.11, that except for the 1-rotation length case that is considered as inadequate information since the signature is too short, the HMMSNP outperforms the other two approaches in all the other cases. This indicates that HMMSNP can

use the underlying information within the data more effectively. It is also noteworthy that the computational cost of all the three approaches increases linearly by increasing the length of signature.

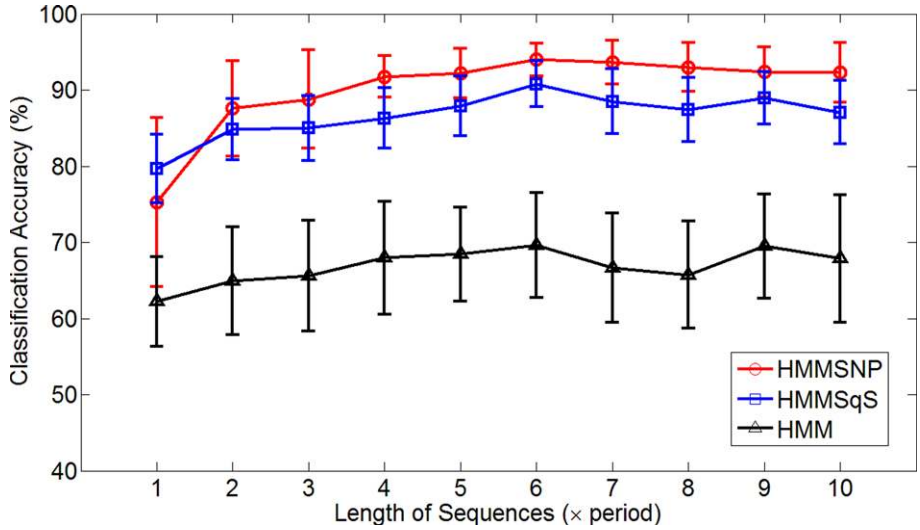


Figure 5.11: Resultant classification accuracies with respect to various signature lengths in terms of number of periods using conventional HMM compared with the HMM (HMMSqS) and HMM-based Semi-nonparametric (HMMSNP) approaches with SqS preprocessing.

5.7 Summary

In this Chapter, an HMM-based semi-nonparametric (HMMSNP) fault diagnostics approach is proposed to improve further the classification accuracy of the conventional HMM-based approach that relies on the log-likelihood. The HMMSNP approach improves the fault diagnostics performance by integrating and taking the advantages of both parametric and nonparametric classification approaches. In HMMSNP, based on the parametric HMMs, two matrices named *probabilistic transition frequency profile* and *average probabilistic emission* are computed for each signature which can represent the trend of that sequence in states and vibration domains. Furthermore to remedy the difficulty of dealing with various operating speeds in the fault diagnostics a preprocessing method named *squeezing and stretching* (SqS) is introduced. The experimental

results indicate that using the SqS processing increases the classification accuracy significantly.

The proposed approach is applied to fault diagnostics in a synchronous motor with two types of faults i.e. bearing fault and unbalanced rotor. Experimental results indicate that the HMMSNP achieves a higher classification accuracy compared to the conventional HMM-based classification approach and the HMM-based classification with SqS preprocessing (HMMSqS). The results from sensitivity analysis indicate that although the number of hidden state values and the length of signature affects the performance of all the approaches, but generally HMMSNP outperforms the HMM and HMMSqS.

Chapter 6

Conclusion and Future Work

This chapter summarizes the contributions of the research work reported in this thesis and outlines the future work directions.

6.1 Contributions

6.1.1 PSHMCO

Tool condition monitoring (TCM) has become one of the main challenges in the industrial environment. As the trend in TCM is changing from determining different classes (discrete condition states) to monitoring the continuous condition metrics (such as continuous tool wear index), Firstly, in Chapter 2, a temporal probabilistic approach called PSHMCO is proposed for the continuous tool condition monitoring in machinery systems. The proposed PSHMCO approach is based on a physically segmented hidden Markov model that can handle continuous output. The PSHMCO has the advantage of providing explicit relationship between the actual health states and the hidden state values of the HMM. The provided relationship is further exploited for formulation and parameter estimation of PSHMCO.

As an illustrative example, the proposed approach is applied to tool wear prediction in a CNC-milling machine. The experimental study indicates that PSHMCO outperforms multi-layer perceptron and the Elman networks in tool condition monitoring. It is also shown that the proposed approach can be utilized for prognostics by unrolling

the model over the time horizon. The PSHMCO approach is found to be suitable for the applications in which the operating conditions are fixed. The fixed operating conditions property, similar to the conducted experiment in Appendix A, can be seen in applications with high volume of productions such as molding processes. Later on, the prediction performance of the PSHMCO approach is further improved by changing its core model from HMM to Hidden Semi-Markov Model (HSMM), which has a flexible state-duration distribution in contrast with HMM.

6.1.2 HSMM-based Approach

In Chapter 3, A hidden semi-Markov model based approach called PSHsMCO was introduced for continuous diagnosis and prognosis. Also, a computationally efficient version of forward-backward algorithm for application of HSMM in continuous tool condition monitoring is described. Based on the simplified forward-backward algorithm, diagnostics and prognostics procedures are defined.

A comparative study is conducted based on the experimental data in Appendix A between PSHMCO and PSHsMCO approaches. Based on the experimental results, PSHsMCO approach outperforms the PSHMCO approach in both diagnostics and prognostics. The results indicate that (as expected) by changing the unrealistic fixed state-duration distribution (geometric) in PSHMCO to a more realistic Gaussian state-duration distribution in PSHsMCO, the efficiency of the prediction model in capturing the underlying temporal information in the experimental sequential data increases.

Furthermore, in order to recognize the preference between the under- and over- estimation. In the last section, applicability of the asymmetric Gaussian functions as the state-duration distributions in the proposed HSMM-based approach is studied. Using the asymmetric Gaussian function, the asymmetric loss functions can be incorporated into the PSHsMCO approach. Finally, the performance improvement in case of utilizing asymmetric Gaussian functions as the state-duration distribution while given asymmetric loss function, is verified on the experimental data.

6.1.3 Multi-modal HMM-Based Approach

Although the first, two single model approaches proposed in Chapters 2 and 3, outperformed other conventional approaches such as Multi-layer perceptron and Elman network in the continuous tool condition monitoring applications. A single model may not be adequate to capture all possible trends when different wearing regiments are available because of various operating conditions, different cutter shapes and geometries, etc. Therefore, a multi-modal HMM-based (m^2HMM) approach, which is an extension of PSHMCO approach proposed in chapter 2, is proposed for continuous tool condition monitoring. The m^2HMM uses multiple PSHMCOs in parallel. In this approach, each PSHMCO captures and emphasizes on a different tool wear regiment. Moreover, two switching strategies namely, soft- and hard- switching, are introduced to combine the outputs from various modes into one. The relevance of each mode at each time is assessed based on the given observation at that time and its precedings. Three weighting schemes i.e. bounded-, discounted- and semi-nonparametric hindsight (SNPH) are proposed which quantify the relevance of modes as weightages to be used in the switching strategies.

The proposed multi-modal approach is applied to cutter wear monitoring in the CNC-milling machine and based on the experimental results, the m^2HMM approach with soft switching (disregarding its weighting scheme) outperforms the PSHMCO in a statistically significant manner. Furthermore, the m^2HMM with semi-nonparametric and soft switching proves to be the best among all combinations of the switching strategies and weighting schemes.

As utilizing multiple PSHMCOs in parallel increases the computational cost, a windowing algorithm is proposed for both PSHMCO and m^2HMM that dramatically increases the speed. Results indicate that if an appropriate window length is adopted, the windowed variant of both approaches can reach the same level of performance as their original and in some cases even outperform them.

6.1.4 Semi-Nonparametric HMM-based Classification

In Chapter 5, in contrast with the previous Chapters which focus on the HMM-based continuous tool condition monitoring approaches and trying to formulate and directly estimate the parameters based on their correspondence to the physical states, an HMM-based semi-nonparametric (HMMSNP) fault detection and diagnosis approach is proposed to improve further the accuracy of the conventional HMM-based classification approach that relies on the log-likelihood.

The HMMSNP approach improves the fault diagnosis performance by integrating and taking the advantages of both parametric and nonparametric classification approaches. In HMMSNP, based on the parametric HMMs, two matrices named *probabilistic transition frequency profile* and *average probabilistic emission* are computed for each signature which can represent the trend of that sequence in states and observation (such as vibration) domains. Furthermore to remedy the difficulty of dealing with various operating speeds in the fault diagnosis, a pre-processing method named *squeezing and stretching* (SqS) is introduced. The experimental results indicate that using the SqS pre-processing increases the classification accuracy significantly.

The proposed approach is applied to fault detection and diagnosis in a synchronous motor with two types of faults i.e. bearing fault and unbalanced rotor (Appendix B). Experimental results indicate that the HMMSNP achieves a higher classification accuracy compared to the conventional HMM-based classification approach and the HMM-based classification with SqS pre-processing (HMMSqS). Also, the results from sensitivity analysis indicate that although the number of hidden state values and the length of signature affects the performance of all the approaches, but generally HMMSNP outperforms the HMM and HMMSqS.

Finally the advantages, disadvantages and some comments on the general approaches developed through Chapters 2 to 5 in this thesis are summarized in Table 6.1.

6.2 Future Work

In this Section, some of the future work directions that can be taken are suggested.

In the case of continuous TCM, to improve the prediction performance further, more complex dynamic Bayesian networks (DBN) that can reflect more inter (observation, state) dependencies may be applied. However, while increasing the complexity of the DBNs, the computational feasibility of the proposed approaches have to be considered for possible online applications.

Furthermore, to improve the asymmetric loss function incorporation and responsiveness of the HSMM to a given loss function, multiple asymmetricity factors can be identified for various partition of the states instead of one for all states (as in Chapter 3). However, this will increase the complexity and challenge associated with the identification of the newly added parameters.

Also the windowing approximation can be employed in the HSMM-based approach to reduce the computational burden and then a multi-modal form of the HSMM-based approach can be utilized to increase the prediction capability of multi-modal approach and incorporate the loss-function applicability in that approach as well.

As the number of possible combinations in the training data exponentially increases with the number of provided experiments, it would be interesting to assess the importance of the modes created using the combination of training data, and see whether the results from combined training sets or from individual training experiments are more important while being used in a multi-modal approach. Thus, being able to give rules of thumb on mode selection while constructing the multi-modal approach as the computation burden significantly increases if all of possible combinations will be considered in general.

In future, the HMM-based semi-nonparametric fault diagnosis approach proposed in Chapter 5, can be extended and examined on various types of electromechanical devices. Moreover, the number of classes considered for diagnosis may be increased to make the experiments more realistic. Also, as the perfect ground truth model is not realizable in most of real-world applications, it would be interesting to use the proposed semi-nonparametric approach in other time series or sequence classification applications such as brain-computer Interaction, etc.

Finally, considering the limited labeled information, and training data available in

the real industrial applications, in contrast to the required amount for the data-driven approaches in general, it would be interesting to pursue intelligent ways to synthesize pseudo-data based on the limited actual experimental data to increase the generalization power of the prediction models. Although this area has been explored in machine learning domain as developing upsampling methods, an upsampling method that preserves the locality and dependencies in the sequential data is required to generate effective pseudo-data for the industrial applications.

Table 6.1: Advanteges, disadvantages and some comments on the approaches developed in this thesis.

Approach	Advantages	Disadvantages	Comments
PSHMCO	- Correspondence between actual physical state and hidden state values. - Direct parameter estimation.	- Unrealistic (fixed) geometric state-duration distribution.	- It outperforms Multi-Layer Perceptron and Elman Network in tool wear monitoring. - Only has one hyper-parameter i.e. number of hidden state values (health states) that can be set via cross-validation.
PSHsMCO	- Correspondence between actual physical state and hidden state values. - Direct parameter estimation. - flexible state-duration distribution. - asymmetric loss function can be incorporated by utilizing asymmetric state-duration distributions.	- Computational cost increases compared to PSHMCO by d_{max} times.	- It outperforms PSHMCO in tool wear monitoring. - Its symmetric state-duration distribution variant has one hyper-parameter i.e. number of hidden state values (health states) but its asymmetric variant has more hyper-parameters to be set.
m^{2}HMM	-Capturing multiple trends by utilizing multiple modes. - Direct parameter estimation. - Correspondence between actual physical states and the hidden state values.	- Computationally is more expensive than single model approaches.	- It outperforms single HMM-based approach (PSHMCO). - Various switching strategies and weighting schemes can be considered. - A windowed variant is proposed which reduces the computational cost.
HMMSNP	- It outperforms the conventional HMM-based approach. - Its non-parametric stage enables it to utilize the training data more effectively. - The similarity matrices generated based on PTFP and APE specifies various trends corresponding to various conditions or classes without a need for a priori knowledge.	- Iterative parameter estimation using Expectation-Maximization. - Computationally more expensive than the conventional HMM-based approach.	- Squeezing & Stretching preprocessing approach utilized in HMMSNP makes the acquired signals from various operating speeds comparable. - The ultimate classification output is based on its nonparametric stage rather than maximum likelihood.

Appendices

Appendix A

Tool Wear in CNC-milling machine Dataset and Experimental Setup

A.1 Introduction

Ball-nose milling cutters have been used extensively in CNC machining of critical parts in the aerospace and motor industries. In the milling process, these changes are closely linked with the cutting forces acting on the edge of the ball-end milling cutter. The cutting forces that are developed during the milling process, can directly or indirectly estimate process parameters such as tool wear, tool life, surface finish, etc.

A.2 Dataset & Features

The experimental data is obtained through real-time sensing on a CNC-milling machine using a force sensor and a vibration sensor both with 3 channels for different directions and an AE sensor. The data comprises cutting process of 6 cutters which are *07BX1, 09BX3, 18SC3, 31PN4, 33PN6* and *34PT1*. The cutters are different with one another in terms of cutter geometry and coating but they are all 6*mm* alignment-tool carbide ball-nose end with three flutes. Figure A.1 depicts the tool wear regiments in the 6 experimented cutters.

In the conducted experiment, a röders TEC vertical milling machine is used as the

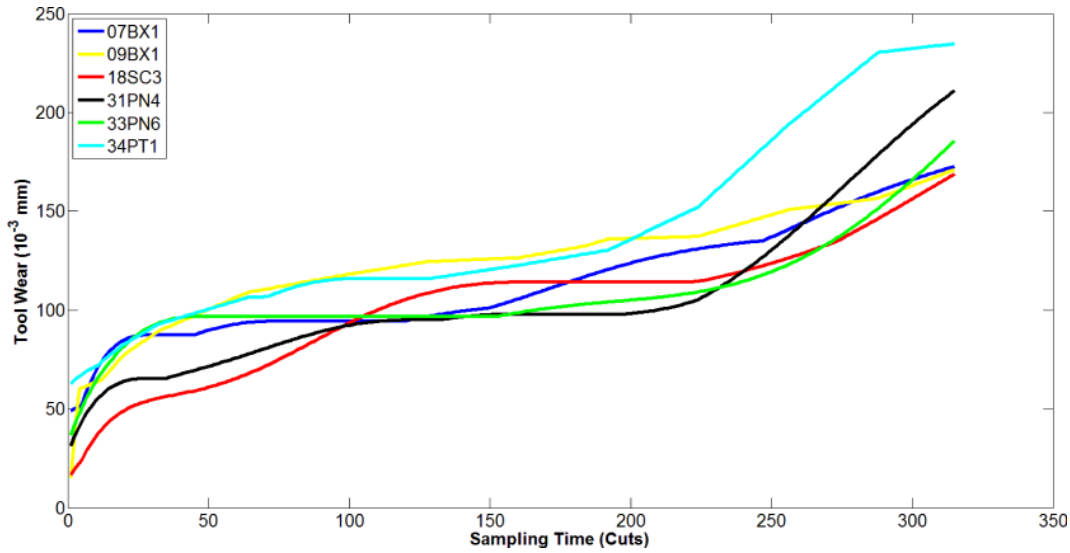


Figure A.1: Tool wear regiment in the 6 experimented cutters.

testbed. For all the cutting processes, *Inconel 718*, which is used in Jet engines, is adopted as the work-piece material. During the cutting process, the upper face of the material is cut with horizontal lines from the top edge to the bottom edge. After each cut, tool snapshots were taken to measure the amount of tool wear. An LECIA MZ12.5 high performance stereo microscope is used to measure the tool wear of the cutting tool. After 320 cutting times, another cutter process will start at the top edge of the material. Table B.1 shows the operating condition parameters of the CNC-milling machine and the experimental components. According to Table B.1, each cutter will travel for $112.5 \times 320 = 36000\ mm = 36\ m$. Fig. A.2 shows the experimental setup.

As it is depicted in Fig. A.2, the tool wear is measured after each cut and stored on the computer along with the sensed signals which are captured using LabView® software running on the computer. After data acquisition, these signals are used for feature extraction and selection using FDR.

A.2.1 Statistical Features

16 statistical features are extracted from force signal available in each direction (X, Y and Z), resulting in 48 features available in total as a 3-channel dynamometer is mounted on the CNC-milling machine. A list of these features is shown in Table A.2.

Table A.1: List of operating condition parameters for the experimental setup and the required components.

Parameter	**Value**
spindle speed	10360 rpm
Feed rate	1.555 m/min
Width of cuts	0.125 mm
Hight of cuts	78mm
Depth of cuts	0.25 mm
Duration of cut	≈ 4s
Number of cuts per experiment	320
Components	
röders TEC vertical milling machine	
6mm ball nose tungsten carbide cutters	
Inconel 718 workpiece	
Kistler 8152B211 Piezotron® (AE sensor)	
Kistler 8636C PiezoBEAM® accelerometers (vibration sensor)	
Kistler 9265B Quartz 3-component dynamometer (force sensor)	
Kistler 5019A multichannel charge amplifier (force amplifier)	
NI-DAQ PCI 6250 M series	
LECIA MZ12.5 stereomicroscope	
Computer	

A.2.2 Wavelet Features

Signals that have been captured using the sensors mounted on the milling machine have non-stationary characteristics, therefore wavelet or multi-resolution approaches are used for feature extraction along with the extracted statistical features. In this work, Daubechies wavelet 8 (DB8) is applied to three force signals, and discrete Meyer wavelet is used for three vibrations and AE signals, all wavelets are with 5 decomposition levels. Hence, 62 ($2^1+2^2+2^3+2^4+2^5 = 62$) coefficients are acquired for each signal, summing up to 434 for all the 7 signals. The average energy of these coefficients are used as the

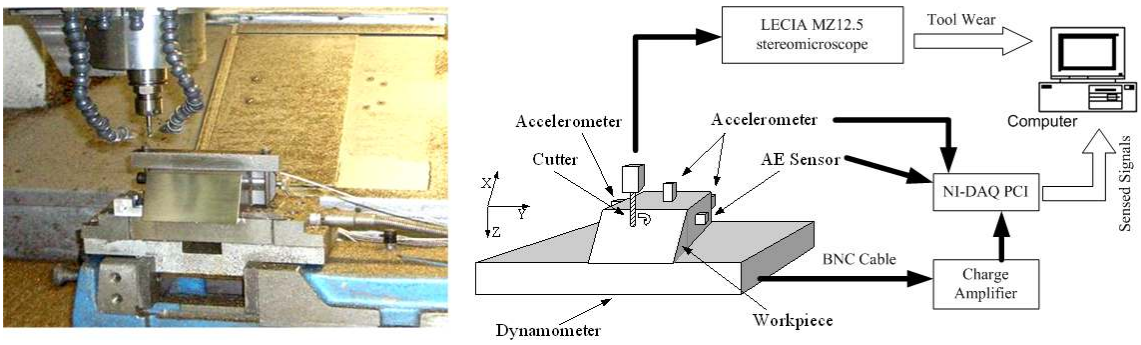


Figure A.2: Experimental setup.

Table A.2: List of extracted statistical features from each force signal channel [4].

No.	Feature	No.	Feature
1	Residual Error	9	Sum of the Squares of Residual Errors
2	First Order Differencing	10	Peak Rate of Cutting Forces
3	Second Order Differencing	11	Total Harmonic Power
4	Maximum Force Level	12	Average Force
5	Total Amplitude of Cutting Force	13	Variable Force
6	Combined Incremental Force changes	14	Standard Deviation
7	Amplitude Ratio	15	Skewness
8	Standard Deviation of the Force Components in Tool Breakage Zone	16	Kurtosis

extracted features. According to [33], the average energy can be written as,

$$E_j = \frac{1}{N_j} \sum_{k=1}^{N_j} [d^n_{j,k}(t)]^2 \tag{A.1}$$

where j is the scale, $d^n_{j,k}(t)$ is the wavelet packet coefficient of the signal, N_j is the number of coefficients at each scale and t is the discrete time.

Acknowledgment

I would like to acknowledge Dr. Junhong Zhou and Dr. Xiang Li from Singapore Institute of Manufacturing Technology (SIMTech) who have helped with the data acquisition process and provided the experimental data on the CNC-milling machine.

Appendix B

Synchronous Motor Fault Generating Setup and Dataset

In this work, the two common fault modes in the synchronous motors, which are unbalanced rotor and bearing faults, are generated using a SpectraQuest® fault simulator machine shown in Fig. B.1. Table B.1 lists the operating parameters and the required components for the experiments.

Table B.1: List of operating condition parameters for the experimental setup and the required components.

Parameter	**Value**
Room Ambient Temperature	$\approx 28^oC$
Duration of Each Signature	800ms
Operating Rotational Speed	15, 16, . . . , 32Hz
Sampling Frequency	5KHz
Number of Signatures per Rotational Speed	10
Components	
DEWETRON® digital data acquisition system	
1/2 HP 230V/50Hz/3-phase AC variable speed synchronous machine	
Piezoelectric Accelerometer with 10kHz bandwidth	

The accumulated data comprises data sequences for the bearing fault (BRG), unbalanced rotor fault (UBR), and healthy machine (HTY), with 180 sequences (signatures)

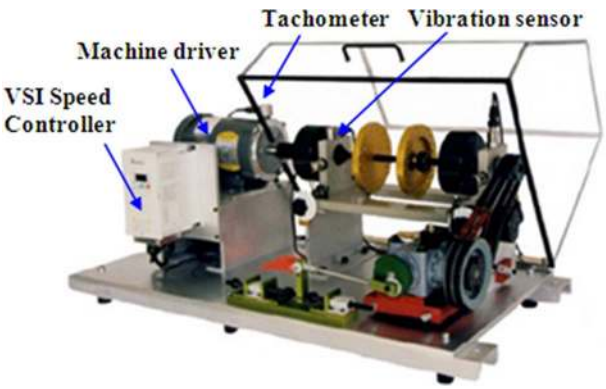
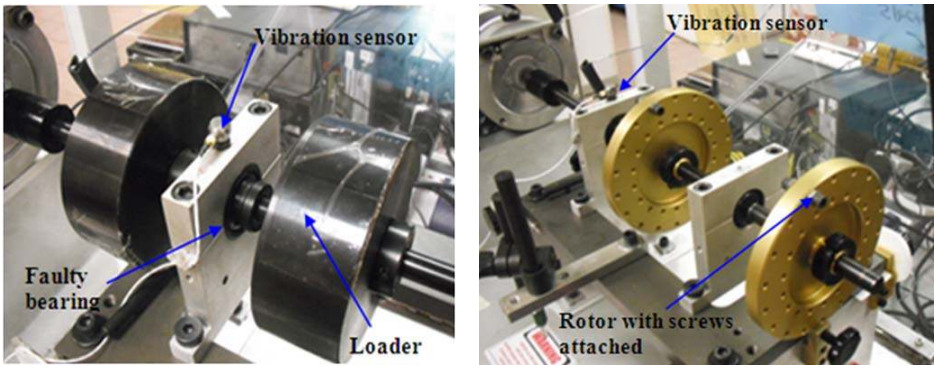


Figure B.1: Machinery Fault Simulator by SpectraQuest®, Inc.



(a) (b)

Figure B.2: Experimental setups used to generate bearing faults (a) and unbalanced rotor (b).

for each class, where 10 signatures are recorded at each machine operating speed of 15Hz up to 32Hz. The BRG class consists of a mixture of three possible faults associated with bearings i.e. rolling element fault, inner raceway fault and outer raceway fault.

The acceleration of the machine vibration is measured by a stud-mounted piezoelectric accelerometer of 10kHz bandwidth above the bearing journal. Vibration signatures are recorded by DEWETRON® digital data acquisition system with a sampling frequency of 5kHz using Hanning window. Each data sequence (vibration signature) is 4000 samples in length, representing a snap-shot time window of 800ms. Fig. B.2 shows the experimental setups used to generate the various bearing faults and the unbalanced rotor.

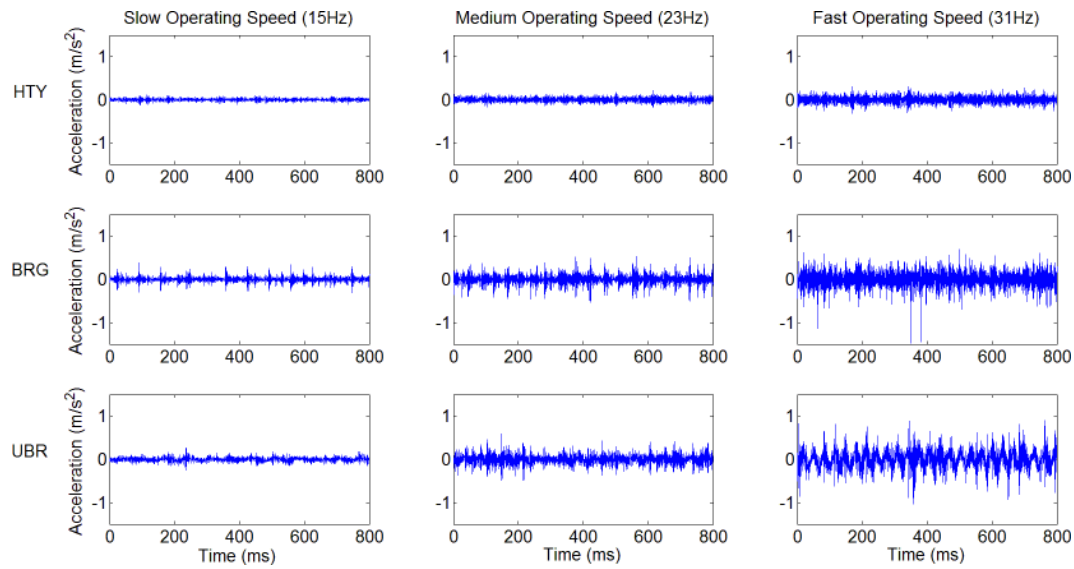


Figure B.3: Samples from three conditions namely, Healthy, Bearing fault and Unbalanced rotor fault at three operating speeds i.e. 15hz, 23Hz and 31Hz.

As can be seen in Fig. B.2(b), two rotors with screws attached are mounted on the shaft to simulate unbalanced rotor fault. The weight of these screws generates unequal centrifugal forces when the rotor spins, thus creating vibrations due to unbalanced rotor.

Fig. B.3 shows three sample vibration sequences under different conditions, i.e. bearing fault, unbalanced rotor and healthy conditions at various rotating speeds. As can be seen in Fig. B.3, a healthy machine has the lowest level of vibration compared to other conditions.

Acknowledgment

I would like to acknowledge Mr. Wee Yee Chen and A/Professor Sanjib Kumar Panda from Electrical Machines and Drives Laboratory at National University of Singapore who have helped with the data acquisition process and provided the experimental data.

Bibliography

[1] Y. Lei, Z. He, Y. Zi, "Application of an intelligent classification method to mechanical fault diagnosis," *Expert Syst Appl*, vol. 36, no. 6, pp. 9941-9948, 2009.

[2] B. Akin, S. Choi, U. Orguner, H.A. Toliyat, "A Simple Real-Time Fault Signature Monitoring Tool for Motor-Drive-Embedded Fault Diagnosis Systems ," *IEEE Trans. on Industrial Electronics*, vol. 58, no. 5, pp. 1990 - 2001 , May 2011.

[3] B. Zhang, C. Sconyers, C. Byington, R. Patrick, M.E. Orchard, G. Vachtsevanos, "A Probabilistic Fault Detection Approach: Application to Bearing Fault Detection ," *IEEE Trans. on Industrial Electronics*, vol. 58, no. 5, pp. 2011 - 2018 , May 2011.

[4] S. Jie, G.S. Hong, M. Rahman, Y.S. Wong, "Feature Extraction and Selection in Tool Condition Monitoring System," Proceedings of the 15th Australian Joint Conference on Artificial Intelligence: Advances in Artificial Intelligence, pp. 487-497, 2002.

[5] J.H. Zhou, C.K. Pang, F.L. Lewis, Z.W. Zhong, "Intelligent Diagnosis and Prognosis of Tool Wear Using Dominant Feature Identification," *IEEE Trans. Ind. Informat.*, vol. 5, no. 4, pp. 454-464, Nov. 2009.

[6] J.H. Zhou, C.K. Pang, Z.W. Zhong, F.L. Lewis, "Tool Wear Monitoring Using Acoustic Emissions by Dominant-Feature Identification," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no.2, pp. 547-559, Feb. 2011.

[7] X. Wang, W. Wang, Y. Huang, N. Nguyen, and K. Krishnakumar, "Design of neural network-based estimator for tool wear modeling in hard turning," *Journal of Intelligent Manufacturing*, vol. 19, no. 4, pp. 383-396, 2008.

[8] Y. Huang, Y.K. Chou,and S.Y. Liang, “CBN tool wear in hard turning: A survey on research progresses,” *International Journal of Advanced Manufacturing Technology*, vol. 35, no. 56, pp. 443453, 2007.

[9] A. K. S. Jardine, D. Lin, 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.

[10] A. Heng, S. Zhanga, A. C. C. Tana, J. Mathew, “Rotating machinery prognostics: State of the art, challenges and opportunities,” *Mechanical Systems and Signal Processing*, vol. 23, no. 3, pp. 724-739, 2009.

[11] J. H. Suh, S. R.T. Kumara, S. P. Mysore, “Machinery Fault Diagnosis and Prognosis: Application of Advanced Signal Processing Techniques,” CIRP Annals - Manufacturing Technology, vol. 48, no. 1, pp. 317-320, 1999.

[12] D.P. Filev, R.B. Chinnam, F. Tseng, P. Baruah, “An Industrial Strength Novelty Detection Framework for Autonomous Equipment Monitoring and Diagnostics,” *IEEE Trans. Ind. Informat.*, vol.6, no.4, pp.767-779, Nov. 2010.

[13] R. Gong, S. H. Huang, T. Chen, “Robust and efficient rule extraction through data summarization and its application in welding fault diagnosis, *IEEE Trans. Ind. Informat.*, vol. 4, no. 3, pp. 198206, Aug. 2008.

[14] S. Simani, “Identification and fault diagnosis of a simulated model of an industrial gas turbine,” *IEEE Trans. Ind. Informat.* , vol. 1, no. 3, pp. 202–216, Aug. 2005.

[15] C. Cempel, “Simple condition forecasting techniques in vibroacoustical diagnostics,” *Mechanical Systems and Signal Processing*, vol. 1, no. 1, pp. 75-82, Jan. 1987.

[16] S. Achiche, M. Balazinski, L. Baron, K. Jemielniak,“Tool wear monitoring using genetically-generated fuzzy knowledge bases”, *Engineering Applications of Artificial Intelligence*, vol. 15, p. 303-314, 2008.

[17] A. Iqbal, N. He, N.U. Dar, L. Li, " Comparison of fuzzy expert system based strategies of offline and online estimation of flank wear in hard milling process", *Expert Systems with Applications*, vol. 33, p. 61-66, 2009.

[18] O. Massol, X. Li, R. Gouriveau, J.H. Zhou, O.P. Gan,"An exTS based neuro-fuzzy algorithm for prognostics and tool condition monitoring," Control Automation Robotics & Vision (ICARCV), 2010 11th International Conference on , pp.1329-1334, 7-10 Dec. 2010.

[19] L. Rutkowski, A. Przybyl, K. Cpalka, "Novel Online Speed Profile Generation for Industrial Machine Tool Based on Flexible Neuro-Fuzzy Approximation ," *IEEE Trans. on Industrial Electronics*, vol. 59, no. 2, pp. 1238 - 1247 , Feb 2012.

[20] X. Li, A. Djordjevich, P.K. Venuvinod,"Current-Sensor-Based Feed Cutting Force Intelligent Estimation and Tool Wear Condition Monitoring", *IEEE Trans. on Industrial Electronics*, vol. 47, no. 3, June 2000.

[21] X. Li, H.-X. Li, X.-P. Guan, and R. Du,"Current-Sensor-Based Feed Cutting Force Intelligent Estimation and Tool Wear Condition Monitoring", *IEEE Trans. on Systems, Man, and Cybernetics-Part C: Applications and Reviews*, vol. 34, no. 4, Nov 2004.

[22] F. Cus, U. Zuperl,"Real-Time Cutting Tool Condition Monitoring in Milling", *Journal of Mechanical Engineering*, vol. 57, P. 142-150, 2011.

[23] P. Palanisamy, I. Rajendran, S. Shanmugasundaram, "Prediction of tool wear using regression and ANN models in end-milling operation," *Int. J. Adv. Manuf. Technol.*, vol. 37, no. 1/2, pp. 2941, Apr. 2008.

[24] Sr. D.E. Dimla, P.M. Lister, "On-line metal cutting tool condition monitoring. II: tool-state classification using multi-layer perceptron neural networks," *International Journal of Machine Tools and Manufacture*, vol. 40, no. 5, pp. 769-781, 2000.

[25] K. Raimond, "Minimizing Error in Tool Wear Estimation using Artificial Neural Network," Industrial Technology, 2006. ICIT 2006. IEEE International Conference on, pp. 2049–2054, 15-17 Dec. 2006.

[26] P. Tse, and D. Atherton, "Prediction of machine deterioration using vibration based fault trends and recurrent neural networks," *Transactions of the ASME: Journal of Vibration and Acoustics*, no. 121, pp. 355-362, 1999.

[27] R. C. M. Yam, P. W. Tse, L. Li, P. Tu, "Intelligent predictive decision support system for CBM," *Int. J. Adv. Manuf. Technol.*, vol. 17, pp. 383-391, 2001.

[28] W. Q. Wang, M. F. Golnaraghi, F. Ismail, "Prognosis of machine health condition using neuro-fuzzy systems," *Mechanical Systems and Signal Processing*, vol. 18, pp. 813-831, 2004.

[29] F. Cus, U. Zuperl, M. Milfelner, "Dynamic neural network approach for tool cutting force modelling of end milling operations," International Journal of General Systems, vol. 35, no. 5, pp. 603-618, 2006.

[30] Y. Shao, K. Nezu, "Prognosis of remaining bearing life using neural networks, Proceedings of the Institution of Mechanical Engineers," Part I : Journal of Systems and Control Engineering, pp. 217-230, 2000.

[31] S.S.H. Zaidi, S. Aviyente, M. Salman, Kwang-Kuen Shin, E.G. Strangas, "Prognosis of Gear Failures in DC Starter Motors Using Hidden Markov Models ," *IEEE Trans. on Industrial Electronics*, vol. 58, no. 5, pp. 1695 - 1706 , May 2011.

[32] 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, pp. 2248-2266, 2007.

[33] K. P. Zhua, G. S. Hong, and Y. S. Wong, "A Comparative Study of Feature Selection for Hidden Markov Model-Based Micro-Milling Tool Wear Monitoring," *Machining Science and Technology*, vol. 12, no. 3, pp. 348-369, 2008.

[34] L. Atlas, M. Ostendorf, and G.D. Bernard, ”Hidden Markov models for monitoring machining tool-wear,” Acoustics, Speech, and Signal Processing. ICASSP ’00. Proceedings. 2000 IEEE International Conference on , vol. 6, pp. 3887-3890, 2000.

[35] P. Baruah, R.B. Chinnam, “ HMMs for diagnostics and prognostics in machining processes,” *International Journal of Production Research*, vol. 43, pp. 12751293, 2005.

[36] L. Wang, M.G. Mehrabi, E. Kannatey-Asibu, “Hidden Markov model-based tool wear monitoring in machining,” *ASME Journal of Manufacturing Science and Engineering*, vol. 124, pp. 651658, 2002.

[37] H.M. Ertunc, K.A. Loparo, H. Ocak, “Tool wear condition monitoring in drilling operations using hidden Markov models (HMMs),” *International Journal of Machine Tools and Manufacture*, vol. 41, pp. 13631384, 2001.

[38] M. Orchard, B. Wu, and G. Vachtsevanous, “A particle filter framework for failure prognosis,” in Proc. of the World Tribology Congress, Washington, DC, 2005.

[39] R.L. Deauff, and M.E.H. Zaïm, *Rotating Electrical Machines*, 1st ed. John Wiley & Sons Inc., 2010.

[40] S. Huang,X. Li, O.P. Gan, “Tool Wear Estimation using Support Vector Machines in Ball-nose End Milling,” In Proceedings of Annual Conference of the Prognostics and Health Management Society, Portland, USA 2010.

[41] T.W. Liao,G. Hua, J. Qu, and P.J. Blau, “Grinding wheel condition monitoring with hidden Markov model-based clustering methods,” *Machining science and technology*, vol. 10, no. 4, pp. 511-538, 2006.

[42] A.A. Kassim, Z. Mian, and M.A. Mannan, “Tool condition classification using Hidden Markov Model based on fractal analysis of machined surface textures,” *Machine Vision and Applications*, vol. 17, no. 5, pp. 327-336, 2006.

[43] M. Wang, and J. Wang, "CHMM for tool condition monitoring and remaining useful life prediction," *The International Journal of Advanced Manufacturing Technology*, vol. 59, no. 5, pp. 463-471, 2012.

[44] T. Boutros, and M. Liang, "Detection and diagnosis of bearing and cutting tool faults using hidden Markov models," *Mechanical Systems and Signal Processing*, vol. 25, no. 6, pp. 2102-2124, 2011.

[45] O. Cetin, M. Ostendorf, and G.D. Bernard, "Multirate coupled hidden Markov models and their application to machining tool-wear classification," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2885-2896, 2007.

[46] L.T. Wang, M.G. Mehrabi, E. Kannatey-Asibu, "Hidden Markov model-based tool wear monitoring in turning," *Journal of Manufacturing Science*, vol. 124, no. 3, pp. 651658, 2002.

[47] S. Lee, L. Li, and J. Ni, "Online Degradation Assessment and Adaptive Fault Detection Using Modified Hidden Markov Model," *Journal of Manufacturing Science and Engineering*, vol. 132, no. 2, 2010.

[48] Z. Kunpeng, H.-G. Soon, W.-Y. San, "Multiscale Singularity Analysis of Cutting Forces for Micromilling Tool-Wear Monitoring ," *IEEE Trans. on Industrial Electronics*, vol. 58, no. 6, pp. 2512 - 2521 , June 2011.

[49] T. Dawson, "Machining hardened steel with polycrystalline cubic boron nitride cutting tools," Ph.D. Thesis, Georgia Institute of Technology, Atlanta, GA, 2002.

[50] S. Nandi, H.A. Toliyat, X. Li, "Condition Monitoring and Fault Diagnosis of Electrical Motors-A Review," *IEEE Trans on Energy Convers*, Vol. 20, no. 4, pp. 719-729, 2005.

[51] M.L. Adams,*Rotating Machinery Vibration: From Analysis to Troubleshooting*. Marcel Dekker, New York, 2001.

[52] D.E. Bently, C.T. Hatch, B. Grissom, *Fundamentals of Rotating Machinery Diagnostics*. Bently Pressurized Bearing Press, Minden, Nev, USA, 2002.

[53] S.A. McInerny, Y. Dai, "Basic vibration signal processing for bearing fault detection," *IEEE Trans Educ*, vol.46, no.1, pp. 149-156, 2003.

[54] C. Bianchini, F. Immovilli, M. Cocconcelli, R. Rubini, A. Bellini, "Fault Detection of Linear Bearings in Brushless AC Linear Motors by Vibration Analysis," *IEEE Trans Ind Electron*, vol.58, no.5, pp. 1684-1694, 2011.

[55] P. Jayaswal, S.N. Verma, A.K. Wadhwani, "Development of EBP-Artificial neural network expert system for rolling element bearing fault diagnosis," *Journal of Vibration and Control*, vol. 17, pp. 1131-1148, 2011.

[56] J.R. Stack, T.G. Habetler, R.G. Harley, "An amplitude modulation detector for fault diagnosis in rolling element bearing," *IEEE Trans Ind Electron*, vol. 51, no. 5, pp. 1097-1102, 2004.

[57] J.R. Stack, T.G. Habetler, R.G. Harley, "Fault-signature modeling and detection of inner-race bearing faults," *IEEE Trans Ind Appl*, vol. 42, no. 1, pp. 61-68, 2006.

[58] H. Ocak, K.A. Loparo, F.M. Discenzo, "Online tracking of bearing wear using wavelet packet decomposition and probabilistic modeling: A method for bearing prognostics," *J Sound Vib*, vol. 302, no. 4-5, pp. 951-961, 2007.

[59] F. Immovilli, A. Bellini, R. Rubini, C. Tassoni, "Diagnosis of Bearing Faults in Induction Machines by Vibration or Current Signals: A Critical Comparison," *IEEE Trans Ind Appl*, vol.46, no.4, pp. 1350-1359, 2010.

[60] M.H. Benbouzid,"A Review of Induction Motors Signature Analysis as a Medium for Faults Detection", *IEEE Trans. on Industrial Electronics*, vol. 47, no. 5, pp. 984-993, Oct 2000.

[61] A. Bellini , F. Immovilli , R. Rubini, and C. Tassoni, "Diagnosis of bearing faults in induction machines by vibration or current signals: A critical comparison", Conf. Rec. IEEE IAS Annu. Meeting, pp.1-8, 2008.

[62] A. Bellini, F. Filippetti, C. Tassoni, G.-A. Capolino, "Advances in Diagnostic Techniques for Induction Machines", *IEEE Trans. on Industrial Electronics*, vol. 55, no. 12, pp. 4109-4126, Dec 2008.

[63] J. Cusid, L. Romeral, J.A. Ortega, J.A. Rosero, A.G. Espinosa, "Fault Detection in Induction Machines Using Power Spectral Density in Wavelet Decomposition", *IEEE Trans. on Industrial Electronics*, vol. 55, no. 2, pp. 633-643, Feb 2008.

[64] A. Bouzida, O. Touhami, R. Ibtiouen, A. Belouchrani, M. Fadel, A. Rezzoug, "Fault Diagnosis in Industrial Induction Machines Through Discrete Wavelet Transform ," *IEEE Trans. on Industrial Electronics*, vol. 58, no. 9, pp. 4385-4395 , Sep 2011.

[65] M. Blodt, P. Granjon, B. Raison, G. Rostaing, "Models for Bearing Damage Detection in Induction Motors Using Stator Current Monitoring," *IEEE Trans. on Industrial Electronics*, vol. 55, no. 4, pp. 1813-1822, April 2008.

[66] M. Djeddi, P. Granjon, B. Leprettre, "Bearing Fault Diagnosis in Induction Machine Based on Current Analysis Using High-Resolution Technique", IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives, SDEMPED 2007, pp. 23-28, 6-8 Sep 2007.

[67] M. Pineda-Sanchez, M. Riera-Guasp, J. Roger-Folch, J.A. Antonino-Daviu, J. Perez-Cruz, R. Puche-Panadero, "Diagnosis of Induction Motor Faults in Time-Varying Conditions Using the Polynomial-Phase Transform of the Current ," *IEEE Trans. on Industrial Electronics*, vol. 58, no. 4, pp. 1428-1439 , April 2011.

[68] L. Eren, M.J. Devaney, "Bearing damage detection via wavelet packet decomposition of the stator current", *IEEE Trans. on Instrumentation and Measurement*, vol. 53, no. 2, pp. 431-436, Apr 2004.

[69] P.S. Barendse, P. Pillay, "The Detection of Unbalanced Faults in Inverter-Fed Induction Machines", IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives, SDEMPED 2007, pp. 46-51. 2007.

[70] Enzo C.C. Lau, H.W. Ngan, “Detection of Motor Bearing Outer Raceway Defect by Wavelet Packet Transformed Motor Current Signature Analysis”, *IEEE Trans. on Instrumentation and Measurement*, vol. 59, no. 10, pp. 2683-2690, Oct 2010.

[71] I.P. Georgakopoulos, E.D. Mitronikas, A.N. Safacas, ”Detection of Induction Motor Faults in Inverter Drives Using Inverter Input Current Analysis ”, *IEEE Trans. on Industrial Electronics*, vol. 58, no. 9, pp. 4365-4373, Sep 2011.

[72] I. Tsoumas, E. Mitronikas, G. Georgoulas, A. Safacas, “A comparative study of induction motor current signature analysis techniques for mechanical faults detection”, IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives, SDEMPED 2005, pp. 1-6, Sep 2005.

[73] R. Yan, Robert X. Gao, “Energy-Based Feature Extraction for Defect Diagnosis in Rotary Machines”, *IEEE Trans. on Instrumentation and Measurement*, vol. 58, no. 9, pp. 3130-3139, Sep 2009.

[74] W. Zhou; T.G. Habetler, R.G. Harley, “Stator Current-Based Bearing Fault Detection Techniques: A General Review”, IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives, SDEMPED 2007, pp.7-10, Sep 2007.

[75] Z. Li, Z. Wu, Y. He, C. Fulei, “Hidden Markov model-based fault diagnostics method in speed-up and speed-down process for rotating machinery,” *Mech Syst Signal Process*, Vol. 19, no. 2, pp. 329-339, 2005.

[76] Y. Xu, M. Ge, “Hidden Markov model-based process monitoring system,” *J Intell Manuf*, vol. 15, no. 3, pp. 337-350, 2004.

[77] D. Ye, Q. Ding, Z. Wu, “New method for faults diagnosis of rotating machinery based on 2-dimension hidden Markov model,” in: Proceedings of the International Symposium on Precision Mechanical Measurement, vol. 4, pp. 391-395, Hefei, China, 2002.

[78] Q. Miao, V. Makis, "Condition monitoring and classification of rotating machinery using wavelets and hidden Markov models," *Mech Syst Signal Process*, Vol. 21, no. 2, pp. 840-855, 2007.

[79] F.V. Nelwamondo, T. Marwala, U. Mahola, "Early classifications of bearing faults using hidden Markov models, Gaussian mixture models, Mel-frequency Cepstral Coefficients and fractals," *Int J Innov Comput Info Control*, vol. 6, no. 2, pp. 1281-1299, 2006.

[80] H. Ocak, K.A. Loparo, "A new bearing fault detection and diagnosis scheme based on hidden Markov modeling of vibration signals," IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'01), vol. 5, pp. 3141-3144, 2001.

[81] M.R. Maurya, R. Rengaswamy, V. Venkatasubramanian, "Fault diagnosis using dynamic trend analysis: A review and recent developments," *Eng Appl Artif Intell*, vol. 20, no. 2, pp. 133-146, 2007.

[82] S.S.H. Zaidi, S. Aviyente, M. Salman, K. Shin, E.G. Strangas, "Prognosis of Gear Failures in DC Starter Motors Using Hidden Markov Models," *IEEE Trans Ind Electron*, vol. 58, no. 5, pp. 1695-1706, 2011.

[83] T. Dietterich, "Machine learning for sequential data: A review," *Structural, syntactic, and statistical pattern recognition*, pp. 227-246, 2002.

[84] B. LeBaron, W.B. Arthur, and R. Palmer, "Time series properties of an artificial stock market," *Journal of Economic Dynamics and Control*, vol. 23, no. 9, pp. 1487-1516, 1999.

[85] R. Mitkov, *The Oxford handbook of computational linguistics.* Oxford: Oxford University Press, 2003.

[86] R.C. Daniels, , C.M. Caramanis, and R.W. Heath, "Adaptation in convolutionally coded MIMO-OFDM wireless systems through supervised learning and SNR or-

dering," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 1, pp. 114-126, 2010.

[87] A. Mateos, J. Dopazo, R. Jansen, Y. Tu, M. Gerstein, and G. Stolovitzky, "Systematic learning of gene functional classes from DNA array expression data by using multilayer perceptrons," *Genome Research*, vol. 12, no. 11, pp. 1703-1715, 2002.

[88] A. Schlapbach, and H. Bunke, "Off-line handwriting identification using HMM based recognizers," In Proceedings of the 17th International Conference on Pattern Recognition (ICPR 2004), pp. 654-658, August, 2004.

[89] P. Dollar, Z. Tu, and S. Belongie, "Supervised learning of edges and object boundaries," In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2006), pp. 1964-1971, 2006.

[90] M.B. Waldron, and S. Kim, "Isolated ASL sign recognition system for deaf persons," *IEEE Transactions on Rehabilitation Engineering*, vol. 3, no. 3, pp. 261-271, 1995.

[91] V. Venkatasubramanian, R. Rengaswamy, S.N. Kavuri, and K. Yin, "A review of process fault detection and diagnosis: Part III: Process history based methods," *Computers & chemical engineering*, vol. 27, no. 3, pp. 327-346, 2003.

[92] A.G. Rehorn, J. Jiang, and P.E. Orban, "State-of-the-art methods and results in tool condition monitoring: a review," *The International Journal of Advanced Manufacturing Technology*, vol. 26, no. 7, pp. 693-710, 2005.

[93] O. Geramifard, J. -X. Xu, C. K. Pang, J. -H. Zhou, and X. Li, "Data-Driven Approaches in Health Condition Monitoring - A Comparative Study," in Proceedings of the 8th IEEE ICCA, FrA7.3, pp. 1618-1622, Xiamen, China, June 9-11, 2010.

[94] Z. Xing, P. Jian, and K. Eamonn, "A brief survey on sequence classification," *ACM SIGKDD Explorations Newsletter*, vol. 12, no. 1, pp. 40-48, 2010.

[95] L.R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc IEEE*, vol. 77, no. 2, pp.257-286, 1989.

[96] K. P. Murphy, "Dynamic Bayesian Networks: Representation, Inference and Learning," PhD, University of California, Berkeley, USA, 2002.

[97] Z. Ghahramani, "Learning Dynamic Bayesian Networks," Lecture Notes in Computer Science, no. 1387, pp. 168-197, 1998.

[98] R. O. Duda, P. E. Hart and D. G. Stork, *Pattern Classification*, Wiley-Interscience, pp. 654, 2001.

[99] J. Bilmes, "A gentle tutorial on the EM algorithm and its application to parameter estimation for gaussian mixture and hidden markov models," Technical report, University of California, Berkeley, 1997.

[100] C. A. Bouman, M. Shapiro, G. W. Cook, C. B. Atkins, and H. Cheng, "Cluster: An unsupervised algorithm for modeling gaussian mixtures," In http://dynamo.ecn.purdue.edu/b̃ouman/software/cluster, Oct. 1998.

[101] A. D. McQuarrie, and C.L. Tsai, *Regression and time series model selection*, Vol. 43. Singapore: World Scientific, 1998.

[102] S. Haykin, *Neural Networks: a Comprehensive Foundation*. Prentice Hall, pp. 842, 1998.

[103] F. D. Foresee, and M.T. Hagan, "Gauss-Newton approximation to Bayesian learning," Neural Networks, 1997., International Conference on , vol. 3, pp. 1930-1935, 9-12 Jun. 1997.

[104] D. J. C. MacKay, "Bayesian interpolation," *Neural Computation*, vol. 4, pp. 415-447, 1992.

[105] J. Sansom, and P. Thomson, "Fitting hidden semi-Markov models to breakpoint rainfall data," *Journal of Applied Probability*, vol. 38, pp. 142-157, 2001.

[106] S.Z. Yu, "Hidden Semi Markov Models," *Artificial Intelligence*, vol. 174, no. 2, pp. 215-243, 2010.

[107] S.Z. Yu and H. Kobayashi, "An Efficient Forward-Backward Algorithm for an Explicit Duration Hidden Markov Model," *IEEE Signal Processing Letters*, Vol. 10, no. 1, pp. 11-14, 2003.

[108] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer, 2003.

[109] M.G. Genton, ed. *Skew-elliptical distributions and their applications: a journey beyond normality*. Chapman & Hall/CRC, 2004.

[110] O. Geramifard, J. -X. Xu, J. -H. Zhou, and X. Li, "Continuous Health Assessment Using a Single Hidden Markov Model", Control, Automation, Robatics and Vision, 2010. ICARCV 2010. 11th International Conference on, 7-10 Dec. 2010. 1990 - 2001, May 2011.

[111] O. Geramifard, J.X. Xu, J.H. Zhou, X. Li, "A Physically Segmented Hidden Markov Model Approach for Continuous Tool Condition Monitoring: Diagnostics and Prognostics ," *IEEE Trans. Ind. Informat.* , vol. 8, no. 4, pp. 964 - 973, Nov. 2012.

[112] K. B. Korb and A. E. Nicholson, *Bayesian Artificial Intelligence*, ser. Chapman & Hall/CRC Computer Science & Data Analysis, 2nd ed. Boca Raton, FL: CRC Press, 2011.

[113] D.Y. Kim and M. Jeon, "Spatio-Temporal Auxiliary Particle Filtering with L1-Norm-Based Appearance Model Learning for Robust Visual Tracking," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 511-522, Feb. 2013.

[114] L. Xu, S. Kwong, H. Wang, Y. Zhang, D. Zhao, and W. Gao, "A Universal Rate Control Scheme for Video Transcoding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 4, pp. 489-501, Apr. 2012.

[115] A. Choudhury, G. Medioni, "A Framework for Robust Online Video Contrast Enhancement Using Modularity Optimization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 9, pp. 1266-1279, Sept. 2012.

[116] D.W. Mount, *Bioinformatics: Sequence and Genome Analysis*. Cold Spring Harbor Press, 2001.

[117] O. Geramifard, J.X. Xu, J.H. Zhou, and X. Li, and O. P. Gan, "Feature selection for tool wear monitoring: A comparative study," In Proc. of 7th IEEE Conference on Ind. Electron. and Appl. (ICiEA 2012), Singapore, 18-20 July, 2012.

[118] J.R. Stack, T.G. Habetler, R.G. Harley, "Effects of machine speed on the development and detection of rolling element bearing faults," *IEEE Power Electron Lett*, vol.1, no.1, pp.19-21, 2003.

[119] E. Keogh, C.A. Ratanamahatana, "Exact indexing of dynamic time warping," *Knowl Inf Syst*, vol. 7, no. 3, pp. 358-368, 2005.

[120] T.C. Fu, "A review on time series data mining," *Eng Appl Artif Intell*, vol. 24, no. 1, pp. 164-181, 2011.

[121] R. Casimir, E. Boutleux, G. Clerc, A. Yahoui, "The use of features selection and nearest neighbors rule for faults diagnostic in induction motors," *Eng Appl Artif Intell*, vol. 19, no. 2, pp. 169-177, 2006.

List of Publications

Articles in Refereed Journals

1. O. Geramifard, J.X. Xu, J.H. Zhou, and X. Li, "A Physically Segmented Hidden Markov Model Approach for Continuous Tool Condition Monitoring: Diagnostics and Prognostics," *IEEE Trans. Industrial Informatics*, vol. 8, no. 4, pp. 964-973, Nov. 2012.

2. O. Geramifard, J.X. Xu, and S.K. Panda, "Fault Detection and Diagnosis in Synchronous Motors Using Hidden Markov Model-Based Semi-Nonparametric Approach," *Eng. Appl. Artif. Intel*, vol. 26, no. 8, pp. 1919-1929, Sep 2013.

3. O. Geramifard, J.X. Xu, J.H. Zhou, and X. Li, "Multi-Modal Hidden Markov Model-based Approach for Tool Wear Monitoring," *IEEE Trans. on Industrial Electronics*. (accepted for publication)

Chapters in Edited Books

1. O. Geramifard, J.X. Xu, and J.H. Zhou. "A Temporal Probabilistic Approach for Continuous Tool Condition Monitoring." chapter in *Diagnostics and Prognostics of Engineering Systems: Methods and Techniques*, ch. 11, IGI Global, 2012. 205-228.

Papers in Refereed Conference Proceedings

1. O. Geramifard, J.X. Xu, J.H. Zhou, X. Li, and O.P. Gan, "Feature Selection for Continuous Tool Condition Monitoring: A Comparative Study," In Proc. of 8th IEEE Conference on Industrial Electronics and Applications (ICiEA 2012), Singapore, 18-20 July, 2012.

2. O. Geramifard, J.-X. Xu, and W.Y. Chen, "An HMM-based Semi-Nonparametric Approach for Fault Diagnostics in Rotary Electric Motors," In Proc. of 21st IEEE International Symposium on Industrial Electronics (ISIE 2012), Hangzhou, China, 28-31 May 2012.

3. O. Geramifard, J.X. Xu, J.H. Zhou, and X. Li, "Continuous Health Condition Monitoring: A Single Hidden Semi-Markov Model Approach" In Proc. of IEEE International Conference on Prognostics and Health Management (ICPHM 2011), Denver, USA, June 20-23, 2011.

4. O. Geramifard, J.X. Xu, T. Sicong, J.H. Zhou, and X. Li, "A multi-modal hidden Markov model based approach for continuous health assessment in machinery systems," In Proc. of 37th Annual Conference on IEEE Industrial Electronics Society (IECON 2011), Melbourne, Australia, 7-10 Nov. 2011.

5. O. Geramifard, J.X. Xu, C.K. Pang, J.H. Zhou, and X. Li, "Data-driven approaches in health condition monitoring - A comparative study," In Proc. of 8th IEEE International Conference on Control and Automation (ICCA 2010), Xiamen, China, 9-11 June 2010.

6. O. Geramifard, J.X. Xu, J.H. Zhou, and X. Li, "Continuous health assessment using a single hidden Markov model," In Proc. of 11th International Conference on Control Automation Robotics & Vision (ICARCV 2010), Singapore, 7-10 Dec. 2010.