ABSTRACT

Zhang, Jianlei. Particle Learning and Gated Recurrent Neural Network for Online Tool Wear Diagnosis and Prognosis (Under the direction of Dr. Starly, Binil).

Automated tool condition monitoring is critical in intelligent manufacturing to improve both productivity and sustainability of manufacturing operations. Estimation of tool wear in real-time for critical machining operations can improve part quality and reduce scrap rates. The motivation of this work is to study two approaches, which aim to provide an online diagnosis and prognosis of machine tool wear conditions using indirect measurements. This work covers four aspects within the approach: 1) Diagnosis of the tool wear itself; 2) Prognosis estimates of the tool wear ahead in time; 3) Mode of collecting indirect measurements from the process, and finally, 4) A method by which we continuously update in real-time tool wear estimates during a machining operation.

The first proposed approach is a probabilistic method based on Particle Learning by building a linear system transition function whose parameters are updated by online in-process observations of the machining process. By applying Particle Learning (PL), the method helps to avoid developing the closed form formulation for a specific tool wear model. It increases the robustness of the algorithm and reduces the time complexity of the computation.

Our first approach assumes linearity and a Markovian process, which may not always hold for broader applications. Our second approach is based on Recurrent Neural Networks (RNN) for the online diagnosis and prognosis for cutting tool wear. It avoids the need to build an analytical model for specific tool wear model, and aims to capture the long term dependencies.

Capturing both long-term and short term memories through Gated Recurrent Units distinguishes our work from other RNNs developed by the community. Without increasing the complexity of the Neural Networks, our approach can realize multi-step ahead tool wear prediction and forecasting Remaining Useful Life (RUL). Both methods were tested experimentally to validate the diagnosis (online estimation), arbitrary multiple-step ahead prediction and Remaining Useful Life capability of our approach.

Particle Learning and Gated Recurrent Neural Network for Online Tool Wear Diagnosis and Prognosis

by
Zhang, JianLei

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Industrial and Systems Engineering

Raleigh, North Carolina

2017

APPROVED BY:

Starly, Binil
Chair of Advisory Committee

Cohen, Paul H.

Lee, Yuan-Shin

Shen, Xipeng

Medhin, Negash

DEDICATION

To Socrates

To Nicolas Copernicus

To Giordano Bruno

To Galileo Galilei

BIOGRAPHY

Zhang, JianLei completed his Bachelor Degree in Engineering at JMSU, and Master's Degree in Mechanical Engineering at University of Shanghai for Science and Technology. He worked as an Electrical Control Engineer at Shanghai General Motors. Presently, he is pursuing his Doctoral degree in Industrial and Systems Engineering at North Carolina State University, Raleigh, NC. His further career will focus on Data Analytics related problems in Engineering and global challenges areas, by applying Statistical, Optimization, Machine learning, Neural Networks, and other data driven approaches. He has hobbies in badminton, jogging, ancient Greek philosophy, documentary, anthropology, history, futuristic study and sci-fi movies.

TABLE OF CONTENTS

LIST OF TABLES

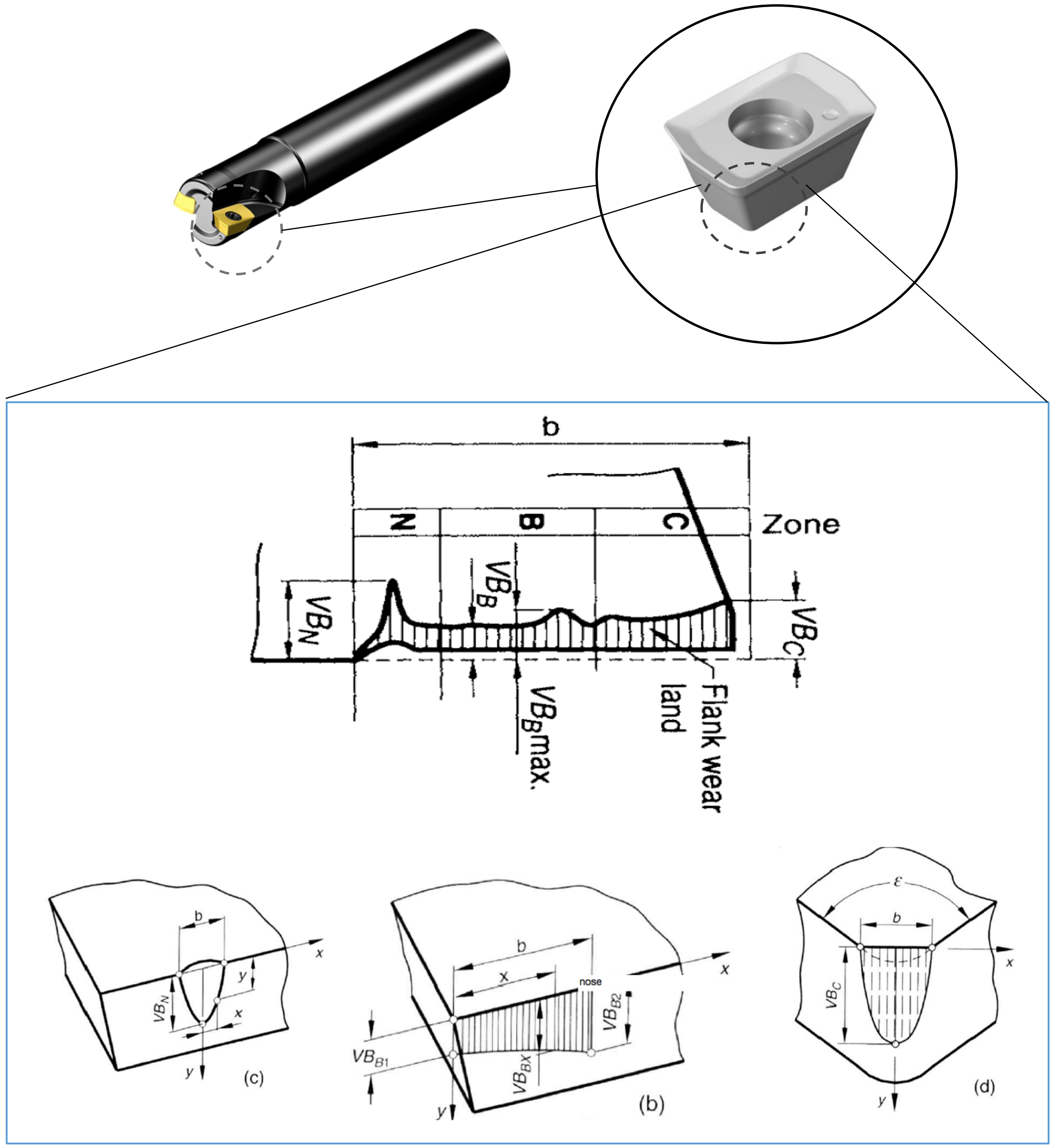
LIST OF FIGURES

Chapter 1 Introduction

1.1 Background

Maximizing the use of cutting tool inserts and effectively utilizing them in automated machining based manufacturing process is critical to reducing machine downtime and improving part quality. Early removal of tooling inserts before they are completely worn out can lead to unnecessary machine downtime and increased costs. Extended use of tooling inserts beyond its lifetime can also induce unexpected machine downtime, undesirable surface deterioration, dimensional inaccuracies and even be detrimental to overall machine health through increased vibrations (Lee, J. H., et al, 1996). Optimizing the tooling insert exchange time interval will require methods to effectively predict in real-time the tool wear estimates based on machining conditions. Current industrial practice of switching out tooling inserts are based on vendor recommended tool lifetime use and machine operator experience. Offline methods of tool wear estimation are by direct observation of tool wear amounts by removing the tool insert and observing them in optical microscopes or through direct vision systems (Kurada, S., et al, 1997). While accurate, this manual method of observation is not practical in high-speed automated manufacturing and ineffective during the adaptive machining of critical features of an in-process workpiece (Si, X. S., et al, 2013).

Tool wear type can be categorized by the zone on which it is located on the cutting tool. There are three zones for measurement of the cutting edge, zone C, B, and N. In Figure 1.1, VB_C is the Nose Wear, VB_B is the flank wear, and VB_N is notch wear (Astakhov, 2006). Flank Wear is most commonly used to indicate the tool life (Kuttolamadom, et al., 2012). Usually, the standard criteria of the end of the tool life, for uniform average depth VB_B, is about 0.3mm, or for a localized wear depth, VB_Bmax, to be about 0.6 mm.



Cutting Tool – Image Courtesy of Sandvik Coromant and Astakhov (2006)
Figure 1.1 Tool Wear Types

Various cutting tool vendors such as Sandvik Coromant (ToolGuide™) or Kennametal (Kennametal 2013 Master Catalog) publish recommended tool wear rates and times based on machining conditions. These recommendations are specific to each cutting tool type, the type of coating applied on the insert and the workpiece material type. Most of the recommended tool life times are based on cutting time intervals rather than actual tool wear estimation. Experienced machinists also are trained to hear the 'tune' of metal cutting to determine if the tool has chipped or worn out before its recommended change out interval. Other observations such as a bad surface finish on the in-process part material can be used to assess tool wear states.

There are several mechanisms causing the tool wear, including abrasion, adhesion, diffusion, chemical reactions, and plastic deformation. The tool wear rate will get accelerated if the cutting speeds are high via most mechanisms. In addition, with elevated temperature, the two mechanisms, diffusion and chemical reaction will speed up, and thus the tool wear rate will rise rapidly leading to premature failure. Moreover, other machining conditions, including feed rate, depth of cut, and hardness of work material also affect the mechanisms of the tool wear (Groover, 2010). Over decades of research, cutting tool inserts have advanced to increase both its lifetime and its ability to cut through hard to machine alloys. The physics-base tool wear model is constructed to consider all the factors into the model. But, the physics-based models require detailed and complete knowledge of the system behavior, which is not ready available. Moreover, the parameter values in the physics-based models need to be determined through experimentation. These experimentally determined values help to understand the underlying mechanism of tool wear but cannot be used generally across all machining conditions. The values will need to be determined for each specific cutting condition and combination of the cutting tool and workpiece material (Gao, et al., 2015). Also, these parameters values change between the individual cutting tool and are often affected by the real time environment factors, such as temperature, humidity and more importantly the variability seen in the work piece quality and cutting tool quality, which cannot be predetermined. This variability in these factors are hardly predictable leading to variations in actual tool wear conditions.

The other alternative for estimating tool wear is with data driven approaches by collecting in-process sensor data to estimate tool wear conditions and life state. Based on whether the analytics models exist in the data-driven methods, they can get categorized into two groups. One category of methods is based on statistical approaches which combines with the underlying analytics model, and another one is artificial intelligence approaches, which are often the model-free methods without the defined analytics model. The data driven approaches can model the system which are analytically difficult to be modeled. The data driven approaches can incorporate the variety of indirect measurements into the model estimation, so that it can update the model parameter values on the fly and reduce the overhead time on the direct measurement.

Among the online methods of tool wear measurement, direct measurement using touch probe or optical sensors within the machining environment can be used as a diagnosis tool to estimate tool wear amounts to appropriately adjust machining process plans. However, the direct measurement interrupts the machining process, leading to an increase in the total machining time. Second, such tool wear probes can only be used to measure after every operation plan and are rarely if ever used continuously after every feature step. Third, they are not prevalent in every machine and at best are integrated in high speed precision machining operations.

Sensors used for indirect measurement of tool wear can be classified as either intrusive sensors or non-intrusive sensors. Intrusive sensors are the sensors that need to be incorporated into the machining environment. They are typically installed on the tool holder, or on the workpiece or the spindle itself. While they offer very high detailed measurement signals, they are limited in practical machining environments due to the often harsh environments within the machining testbed. These harsh environments can include the use of cooling fluid, high temperature, excessive vibration, chips flying off the cutting surface, all of which can limit the use of intrusive sensors in production environments. Such sensors are also affected by the position of the sensor with respect to the workpiece and can affect estimated amount of tool wear. Sensors such as force dynamometers have known to be good indicators of tool wear but they are impractical for use in production environments (Haber, et al., 2004).

In order to monitor the machining process, the intrusive sensors include force sensor (https://www.kistler.com/us/en/), acoustic emission sensor (http://www.physicalacoustics.com/), and vibration sensors (https://www.kistler.com/us/en/). Although indirect forms of measurements, they have been utilized in real-time tool wear estimation and Remaining Useful Life (RUL) prediction of tool inserts. However, the placement of these sensors in hazardous environments can limit its use in practical industrial settings. Noise in the signals can also significantly affect the quality of the tool wear estimates and often need to be adjusted based on machining conditions. Such adjustments and re-calibration becomes often impractical in production environments.

The primary reason for direct or indirect measurements of cutting conditions is to assess machining state and monitoring tool wear conditions as they significantly influence the quality of the part produced. Depending on monitoring the current or the future tool wear state, the monitoring approach can fall under two categories: Diagnosis and Prognosis. Diagnosis, according to (Sikorska, 2011, ISO13381), means Fault Detection, which is for detecting and reporting an abnormal operating condition; Fault Isolation, which is for determining which component (subsystem, system) is failing or has failed; Fault Identification, which is for Estimating the nature and extent of the fault. In the scenario of tool wear monitoring, Diagnosis here refers to the online estimation of the extent of the tool wear via direct measurements.

Prognosis, applied on the tool wear, for level 1 and level 2 (Sikorska, et al., 2011, ISO13381), includes the one-step ahead tool wear prediction, multi-step ahead tool wear prediction and Remaining useful life (RUL) prediction. Here RUL refers to identifying the lead time to failure of the cutting tool insert. The ability to conduct prognosis (Sikorska, et al., 2011; Ahmadzadeh, F., et al, 2014) of tool wear states in real-time and during the process, represents a promising direction in optimizing the use of cutting tool inserts. The highest form of the prognosis (Sikorska, et al., 2011) is the *post-action prognostics*. It is to identify potential actions to retard, halt, or prevent fault progression. For the tool wear estimation case, it is to correlate the one-step ahead, multi-step ahead tool wear and RUL prediction associated with different machining parameters. If multi-step ahead tool wear amounts can be estimated, corrective action can be initiated to adjust machining conditions to extend tool insert lifetime until after the machining sequence is completed.

1.2 Research Motivation

Estimation of tool wear has been carried out over a decade with various strategies and methods proposed. However, none of these methods have been fully implemented in practical machining conditions. Current industrial practice involves removing the worn out cutting tool based on recommendations made by the cutting tool vendor. In other cases, cutting tools are removed based on the experience of the machine operator or in certain conditions when the 'noise' from the cutting operation does not follow the normal sound wave patterns. Physical inspection of the parts, based on surface finish obtained after cutting operation can be a good estimate on the state of tool. Such practices, while effective cannot be used in highly automated 24/7 production based environments.

Most accurate estimates of tool wear are through direct measurement of tool wear states or through expensive indirect measurements made through force dynamometers. This research tries to address the question - can we estimate and predict tool wear by analyzing data that comes from non-intrusive sensors such as hall-effect sensors which measure power consumption to the spindle or at best be dependent on minimally intrusive sensors such as three-axis vibration sensors.

This research explores both statistical and data driven approaches to diagnosis and prognosis of tool wear processes and be as practically relevant as possible within highly automated production environments. Current research work either applies a complex model for tool wear processes such that its solution becomes too intractable to solve in real-time manufacturing settings or they may be too simple to account for the dynamics of the process. So that the more reasonable model will be promoted for precise modeling with reasonable assumptions and tractable computation in time complexity. A solution for a comprehensive model which properly incorporates the real time indirect measurement to update the diagnosis and prognosis, including one-step ahead, multi-step ahead tool wear prediction, and RUL prediction is needed.

1.3 Research Goal and Objectives

The goal of this work is to present a new approach to monitoring and predicting tool wear in machining processes. Specifically, this work will focus on developing two approaches to conduct online tool wear diagnosis and prognosis via indirect measurements. For diagnosis, the developed method intends to update estimation of current tool wear from online and historical measurements. For prognosis, the developed methods should predict the tool wear during the next run or two runs ahead in terms of feature cutting operation. Both diagnosis and prognosis will be conducted using online and indirect forms of measurement, such as spindle power consumption or through vibration based minimally intrusive sensors.

The following research objectives aim to help achieve these goals.

Research Objective 1: Develop the Particle Learning approach to realize online updating of tool wear prognosis and diagnosis through indirect measurement, with the assumption that the system transition function and observation function are linear and static, and the system state is subject to a Markovian process. We approach this problem by separating the training process for the system transition function and the observation function from each other.

Research Objective 2: Develop the Recurrent Neural Network (RNN) approach to realize the online updating of tool wear diagnosis and prognosis through indirect measurements, by deeming that the system transition function and observation function are nonlinear. We attempt to capture the dynamics of the tool wear process without assuming the Markov process properties of the system state. Moreover, a strategy for combining the training process for system transition function and observation function will be developed.

1.4 Dissertation Outline

In Chapter 1, it introduces the background of our research. It discusses the different types of tool wear and significance of monitoring the tool wear. Then, the motivation, the goal and objectives are related. In Chapter 2, the current and past researches about tool wear diagnosis and prognosis with indirect or direct measurement with online and offline updating get reviewed. These methods generally follow into these categories, including the physics-based model, stochastic model, statistical model, and data driven model. In Chapter 3, at beginning, the stochastic tool wear modeling approaches are briefly reviewed, followed by the detailed explanation of the Particle Learning (PL) methods with linear system models. Further, how the online diagnosis and prognosis are implemented here get explained. Moreover, the experiments and conditions at which the data collected is described. After the data is collected the proposed PL method for data analysis is implemented. The results show the capabilities of the proposed methods on the online tool wear estimation and prediction, and RUL prediction.

In Chapter 4, it begins with brief introduction on the the application of Neural Network method in tool wear estimation and prediction. Then, it introduces the sequence to sequence modeling algorithms, which is similar to our tool wear modeling, and then gives two examples on how RNN are applied in another field. Then, there is the detailed explain on modeling the tool wear process with system transition and observation functions by taking advantage of the simple RNN, gated RNN. The advantage of the gated RNN than the simple RNN is discussed. Thenceforth, it shows the schema on how to realize the online tool wear estimation and prediction, as well as the RUL prediction by the system model via (gated) RNN. Lastly, the methods get validated by the experiment data, and the result on the online estimation, prediction and RUL prediction are showed. The result shows the performance of different types of RNN. In Chapter 5, it summaries the work in this dissertation, and presents the uniqueness and contributions of the work. Also, the orientation of the future work get discussed.

Chapter 2 Literature Review

2.1 Overview of Research Area

Several decades of work have been carried out with respect to machine tool wear. Based on work done so far, methods developed for tool wear and other degradation processes for machines can be classified into four categories: Physical model, Statistical model, Stochastic model, and Data-driven model (Figure 2.1).

Physical model, including Taylor's model, Failure model, proportional hazard model and nonlinear dynamics, specifies applying the mathematical representation of the physical manner of the deterioration process for a machine. The challenge of such models are that it needs comprehensive and instant information of the physical entity. However, such information is not available in practical setting, so that such models are limited to application in well-known and understood scenarios.

Statistical Model, including autoregressive moving average model, Bayesian updating model, Kalman filter, and particle filter, apply a general purpose statistical model to model the degradation process. The statistical model itself has certain assumptions. For example, the Kalman filter is a Gaussian-Markov process based algorithm and assumes both process and measurement are zero-mean white stochastic processes, which is not applicable to model the tool wear process. Particle filter can be applicable to no zero-mean process, and it is easier and robust in computation but its assumption is based on which underlying system transition model and observation model is used.

Stochastic Model, including hidden Markov model, Wiener process, and random walk, assumes state transition subjects to certain kind of stochastic process, including Markov process, Wiener process, Gamma process, or random walk. After settling the distribution properties, the prognosis such as RUL prediction can be realized given the machine or its components current state. Because it is based on the probability, the results on the RUL will be obtained as the confidence intervals. However, the assumption on certain distribution limits the application and may not be adaptive to a variety of situations.

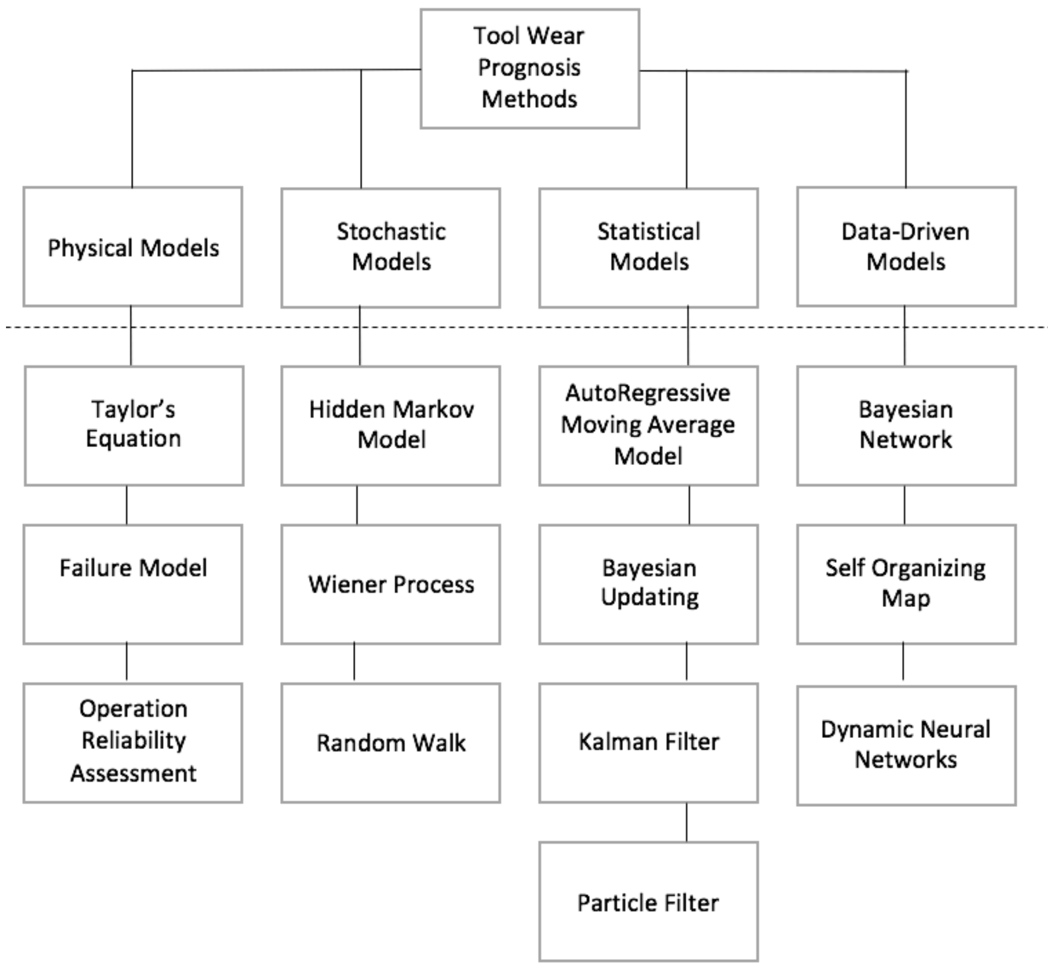


Figure 2.1 Categories of Approaches for Tool Wear Prognosis

Data Driven Model, including Bayesian Network, self-organizing map, and dynamic neural network, has least set of assumptions made about the underlying process. These methods can model the complex system with highly nonlinear properties and does not need and produce the analytic model. It does not require prior information about the system. However, these model need large amount of initial data to train. Since these methods are not probabilistic based, their prediction is not in the form of a confidence interval.

In the following sections, each category method is broken down and current state of art in terms of the work carried out by the scientific community in terms of diagnosis and prognosis of tool wear has been described. As mentioned, each method has its advantages and limitations.

2.2 Physics-based Models

2.2.1 Taylor's Equation

The earliest work in tool wear assessment has been based on the physics of the tool wear process under different materials and machining conditions. As stated in the review paper by (Ahmadzadeh, 2014), physics-based models include Taylor's model and Tool wear rate model. They can be highly accurate if the machining conditions remain fairly consistent across multiple systems. It does require fewer data to build the model than other data-driven systems. However, coefficients in model are empirically determined and only work for particular combinations of tools and workpieces (Gao, et al. 2015). Material variations and inevitable varying machining conditions can result in inaccurate wear estimates. Due to the stochastic nature of the machining environment, physics based models become too complex to capture the physical process. They provide a good overall estimation of the underlying process but cannot be used for individual tool degradation estimates in production based environments. The most common approach to estimate tool wear is through Taylor's tool life equation.

$vT^n = C.$ (2.1)

where, v is the cutting speed, T is the tool life, and n and C are the parameters depends on feed, depth of cut, work material, tooling, and tool life criterion used (Groover, 2010). In order to incorporate more machining parameters and material hardness within the model, the extended version of the Taylor's equation was developed as follows:

$v\, T^n f^m d^p H^q = KT_{ref}^{\ n} f_{ref}^{\ m} d_{ref}^{\ p} H_{ref}^{\ q}.$ (2.2)

where f is the feed rate, d is depth of cut, H is the hardness. f_{ref}, d_{ref}, and H_{ref} are the reference values for feed rate, depth of cut, and hardness. However, m, p, and q are the exponents determined experimentally (Groover, 2010). Specifically, in the paper of (Ginta, et al, 2009), their work is to establish the tool life models based on cutting speed, depth of cut, and feed rate.

$T= CV^k a^m f_z^{\ l}$ (2.3)

where, T is the predicted tool life, V, a, and f are cutting speed, depth of cut, and feed rate. Then the model gets linearized by the logarithmic transformation, shown as following:

$\ln T = \ln C + k\ln V + m\ln a + l\ln f$ (2.4)

equivalently, the above equation can be written as:

$y = \beta_0 x_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$ (2.5)

Their experiment executed on 5 different cutting speeds, 5 different depths of cut, 5 different feed rates, from lowest to highest. The result of their model show that the model is statistically significant. However, the depth of the cut has insignificant effect on the tool life. With the doubled cutting speed, depth of cut, and feed rate, the reduction of the tool life by 70%, 27%, and 37%, respectively. The variance analysis shows that the interactions of the variables are statistically insignificant.

The physical model on tool wear also include other methods. Usui, et al. (1997) promoted a practical way by applying finite element method (FEM) into tool wear estimation. The tool wear rate model, which is deemed as belonging to crack growth model (Gao, et al., 2015). Beyond certain criterion, the tool wear was modeled as the Power law, more specific is Paris' law (Paris, et al., 1961). There are several improvement and extension work based on the Paris' law, including the research by Pugno, et al (2006) and Aslantas, et al (2004). The model parameters need to be determined by large number of experiment, but the determined values only affective to the specific machining conditions and workpiece and cutting tool material combination. Another limitation of the physics based model methods is that it cannot estimate the Remaining Useful Life (RUL) under every possible cutting condition. The tool wear process may be varied during the cutting process and the model is susceptible to variations often seen in the material workpiece, the cutting tool itself, the state of the machine and other process variables not accounted for in the model. Such models cannot capture such variation and therefore the error in tool wear estimates is much larger than some of the online monitoring methods utilized.

2.2.2 Operation Reliability Assessment

The operation reliability assessment based methods construct a relationship between the indirect measurement and the reliability of the cutting tool. In (Cai, et al., 2012, Ding, et al., 2011), proportional hazards model (PHM), proportional covariate model (PCM), and Cox model are the methods to realize estimate the reliability of the tool via the online indirect measurement, the features of the sensor signal. Both PHM and PCM enable the incorporation of the condition monitoring into the assessment of the risk of the cutting tool. It assumes the multiplicative relationship between the failure probability of the cutting tool and the indirect measurements from the features of the sensor signals. As shown by Allison (1995), PHM can be expressed as the following):

$$h_i(t)=\lambda_0(t)\exp[\beta_1 x_{i1} + \ldots + \beta_k x_{ik}(t)] \quad (2.6)$$

The hazard for individual i at time t is the product of two factors: A baseline hazard function $\lambda_0(t)$ that is left unspecified, except that it cannot be negative. It is the base value, which can be estimated by the training model with real data. The second factor is a linear function of a set of k fixed covariates, which is then exponentiated. The coefficient β_1, β_2, …, β_k can be also get estimated. The expected Remaining Useful Life is then determined to be as

(RUL) = 1/h(t), assume h(t) remain constant.

The experiment by Ding, et al.'s (2011) paper uses the 0.6 mm as the criterion of the flank wear. Their work demonstrated the use of vibration signals from which features were extracted - root mean square, peak, kurtosis, and crest factor. The result shows that the root mean square and peak are strongly correlated with the tool degradation processes, so that these two features get incorporated into the proportional hazard model. Their work established the linkage between the time domain features and the reliability statistics. For example, if we get the feature value, we can infer how much hazard the tool will fail.

The benefits to use the survival analysis, as stated by (Allison, 1995), is the basic idea to divide the time scale into intervals, assume that the hazard is constant within each interval but can vary across intervals. Another benefit, is the ability to incorporate time dependent covariates, so it can incorporate indirect measurements. However, these models suffer from limitations in their theoretical grounds (Sikorska, et al., 2011): from the statistical view, the models suffer from several disabilities. The ability to estimate RUL for prognosis of varied faults in complex systems is uncertain. The assumption that failures are independent and identically distributed is rarely justified. It is difficult to define a comprehensive set of covariates that is applicable to all failure model experienced. Additionally, proportional hazards model (Ding, et al., 2011), proportional covariate model (Cai, et al., 2012), and Cox model have limitations as proportional hazard model. From practical view, most often these models have been built using intrusive indirect sensors such as vibration and/or force sensors that detect vibration of cutting and cutting forces respectively. This dependence on these sensors can limits its usefulness in production environments.

2.3 Stochastic Models

2.3.1 Hidden Markov Model

In the hidden Markov Model (HMM) (Rabiner, 1989), there are two combined stochastic processes. The underlying one, which is not observable, subject to the finite state homogeneous Markov chain. The underlying one impacts another process which is observable.

Given the sequence of the observed process, its probability gets derived, then the most probably underlying state sequence gets identified. Also, the parameters in the model get updated at the maximum likelihood of the observed sequence (Gao, et al., 2015). The RUL can get predicted with updated model. Specifically, for the HMM in the application of the tool wear, Baruah, et al., (2005), and Vallejo Jr, et al., (2005) use the hidden Markov Model (HMM) for tool wear diagnosis and prognosis. Briefly the HMM process assumes that the system states transition is subject to the Markovian process.

$$P(X_t|X_{t-1}, O_{t-1}, \ldots, X_1, O_1) = P(X_t|X_{t-1}). \qquad (2.9)$$

X_t is hidden variables representing real system state, and O_t is observed variables. It also assumes the state transition duration is subject to exponential distribution. In their application case, the thrust-force and torque dynamometer are applied for the observed variables. Another assumption is that the current observations only depend on their current states:

$$P(O_t|X_T, O_T, X_{T-1}, O_{T-1}, \ldots, X_1, O_1) = P(X_t|X_{t-1}). \qquad (2.10)$$

In order to use the model, it should be trained with time series paired data, for example, like $(O_t, X_t), t = 1, 2, \ldots, T$. For training process, it uses Forward-backward algorithm to estimate $P(O|\lambda)$ and use Baum-Welch algorithm (EM for HMMs) to estimate model parameters $\lambda^* = \arg\max_\lambda p(O|\lambda)$. After the model is trained, it can realize online diagnosis and prognosis. The prognosis includes the one step ahead tool wear prediction, multi-step prediction, and RUL prediction (Baruah, et al. 2005).

In the (Baruah, et al., 2005) paper, they define 4 health states, good, medium, bad, and worst, for the drill bit. The result shows for diagnosis, the accuracy for the classification of the health state of the drill bit reach 96.9% for the test data. And also it shows that it can be effective to predict the health state of the drill bit. However, the assumption about the state transition duration subject to exponential distribution is often not the case, so the accuracy of the model gets decreased. Further, for the application in tool wear prediction it fails to incorporate machining conditions in their model. Also they use the intrusive sensors, thrust-force and torque dynameters, which are not ideal in practical applications.

2.3.2 Other HMM

Variable Duration HMM (VD-HMM) relaxes the assumption that the system states transition subject to Markov process. In (Yue, et al., 2010, Yue, 2012) papers, it uses the Gaussian and Weibull distribution as state duration. So, by using these two distributions, it increases the model accuracy by avoiding the assumption by the HMM. Same as the HMM, variable HMM fails to incorporate multiple cutting conditions and may not include all possible variability in cutting conditions. Additionally, model fit and prediction is obtained through intrusive sensors such as thrust-force and torque dynameters.

The Hidden Semi-Markov Model (HSMM) also does not assume the state transition duration subject to exponential distribution, it can be modeled as a more flexible distribution. Specifically, for HSMM developed in (Geramifard, et al., 2011), at each time step, only two options exist: 1. Remain at same state as previous time step; 2. Proceed to next state subject to the duration distribution. Therefore, state transition matrix only has binary element: 0 or 1. However, like other HMM models, the model fails to incorporate multiple cutting conditions. And in their applications, intrusive sensors are used, thrust-force and torque dynameters, which is also not desirable. Overall, these HMMs have the common assumption that the system state must subject to the Markov process, and the states of the indirect observation and direct observation should be discrete. Overall, while these methods have shown some success, they are highly dependent on data from all possible cutting conditions and on very intrusive indirect measurement sensors. This can limit the practicality of use of such models. Nevertheless, among the stochastic models in use for tool wear prognosis, methods based on the Hidden Markov Model appears to be the most well studied method.

2.4 Statistical Models

2.4.1 Bayesian Updating with Taylor Equation

The Taylor Equation is proposed as deterministic equations; however, the values of the parameters in the model is inherently uncertain, which subject to change with internal and external time varied factors. The Bayesian updating approach is applied to update the estimation of the probability distribution of these uncertain values in the Taylor equation, with the experimental results to the cutting tool. Further, with the updated belief of the values, the tool life can be predicted. As stated in (Karandikar, et al, 2014a, 2014b), the steps to realize the Bayesian updating with Taylor Equation includes: 1st - Record tool life at different cutting conditions; 2nd - Update the parameter estimation; and then to predict the tool life. The parameters are the exponent n, and the constant C. The criterion of the tool wear is set to 0.3mm. The prior distribution of the parameters, n and C, are determined by the literature or through experiments. For every new data set collected by experiments, the distribution of these parameters are updated. Bayesian updating of the distribution of the parameters of the Taylor equation and then computed to predict tool life via the probability function (Karandikar, et al, 2014a). In part 2 (Karandikar, et al, 2014b), utilized the extended form of the Taylor equation, which includes the cutting speed, feed rate into the equation. Their work utilized the Metropolis-Hastings algorithm of Markov Chain Monte Carlo approach to update an estimate of the parameters of their probability density distributions. Therefore, the tool life is constantly predicted via the updated values.

The advantage of this method is that it can be generalized to apply to different machining conditions. However, there are several limitations for the method. The methods developed by the authors are subject to direct measurements of the tool wear at intermittent intervals. Therefore, the applicability of this approach by using indirect measurement is not clear and most likely may not work. It cannot realize the real-time online updating of the parameters in the Taylor equation. The parameters can only be updated when one cutting tool gets fully worn out.

2.4.2 Bayesian Updating with Growth Curves

Rather than use the Taylor's equation, growth curves can be used in combination with Bayesian updating to realize tool wear estimates. Similarly, the parameters in the growth curve model can get updated by the experiment results. The model with estimated parameters can estimate the RUL of the tool. As stated in (Karandikar, et al. 2012, 2013), the root mean square of the power and flank wear gets measured after every pass as the target value in the growth curve model. The procedure of applying the Bayesian updating with growth curves is briefly mentioned. First, the process begins with generating a large number of the growth curves via the second order formula:

$$p_{rms,\,i} = a\,t^2 + b\,t + c \qquad (2.7)$$

Second, the posterior probabilities of the tool wear growth curves gets updated by the likelihood:

P(path = p_{rms} growth curve| test result) $\propto$ P(test result|path = true p_{rms} growth curve)×P(path = p_{rms} growth curve) (2.8)

here, likelihood $L= \exp[-(p-p_m)^2/k]$.

The advantage of the method is that it incorporates the indirect measurement into the model, and with that, the remaining useful life (RUL) can also be estimated. Different percentile of the RUL can be considered according to the user requirements. Karandikar, et al (2012), for example, set 99% is a conservative percentile. The higher the percentile, the more conservative the tool get used. Another positive aspect is that the method can incorporate data from non-intrusive sensors, such as the hall effect sensor detecting power. However, it has the limitations, in that, generating the growth curves is based on rule of thumb. The methods do not provide the ways to realize the tool wear estimation and prediction. The methods only give the most possible growth curve of the indirect measurement of the tool wear, which are only the feature values of the sensor signal. Because there is no direct measurement of the tool wear incorporate in the model, the estimation of the tool wear by the online indirect measurement is not viable. No to mention the prediction of the tool wear. By set a criterion of the indirect measurement, the growth curve can estimate the RUL of the cutting tool.

2.4.3 Kalman Filter

Kalman filter (KF) provides a mean for continuous learning. It explicitly formulates the state space model and observation function with linearity assumption. It recursively utilizes the up-to-date information contained in the training data. Thus, the estimation of the state can be computed recursively. KF is capable of updating the current belief and then make prediction about the state given the current and historical indirect measurement. For its application in the tool wear, in the paper (Niaki, et al. 2015), they formulate following equations as the state transition equations.

$$\begin{bmatrix} \mathrm{VB(k)} \\ \mathrm{VB'(k)} \end{bmatrix} = \begin{bmatrix} 1 & MR \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mathrm{VB(k-1)} \\ \mathrm{VB'(k-1)} \end{bmatrix} + \begin{bmatrix} w_1(k) & 0 \\ 0 & w_2(k) \end{bmatrix},$$

$$\begin{bmatrix} w_1(k) \\ w_2(k) \end{bmatrix} \sim \mathrm{N}\left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} Q_1 & 0 \\ 0 & Q_2 \end{bmatrix}\right]. \qquad (2.11)$$

Where, VB is the tool flank wear. For the experiment, the spindle power and the flank wear is measured for every pass, and k is the k^{th} pass, w is noise, and the noise of the VB and its derivative VB' are assumed independent with each other. MR is the volume of the material get removed for every pass. Their result shows the maximum error is at 18%, which is quite large, and even larger than the 15%, which is often used as the reference error for experiment. The model assumes the tool wear transition for every pass is subject to linear function. However, it may not be the case that it is linear, and the linearity may not capture the complex dynamic of the tool wear. Furthermore, the Kalman filter method itself has several deficiencies in calculations. As stated by (Hartikainen, et al. 2011), the limitation of the Kalman filter includes: 1) The error propagation should be well approximated by a linear or a quadratic function, otherwise, poor performance will result; 2) The Jacobian matrices or Hessian matrices should exist, it is often not the case, such as the state belongs to jump linear model. 3) Even when Jacobian matrices or Hessian matrices exist, it is still challenging to calculate them. Finally, the Kalman filter has been superseded by the other methods such as the Particle Filter methods which produces far super results (Niaki, et al. 2015a, 2015b).

2.4.4 Particle Filter

Particle filter (PF) is an alternative to the KF as one of the Bayesian updating approaches. PF relaxes the assumption on the linearity and Gaussian distribution for the noise. Similar to KF and other Bayesian updating approaches, it can realize diagnosis and prognosis of the system state, given current and past indirect observation of the system. PF is more flexible in the form of its underlying system transition model, which can be linear with zero-mean, without zero-mean or nonlinear. Also, with heuristic like computation process, by mimicking the particles in the sampling and propagation calculation, the algorithm becomes more robust, flexible, and easier to calculate than other Bayesian inference approaches.

2.4.5 Particle Filter with Nonlinear Transition Function

Considering the flexibility of the PF, a nonlinear underlying system transition model is implemented. In the papers (Wang, et al. 2015a, Wang, et al. 2015b, Wang, et al. 2013), their methods use the state space model, while implementing a nonlinear state space model. After collecting the sensor signals from the accelerometer and force sensors, through the procedure of feature extraction and selection, it builds the relationship via the autoregressive and support vector regression for system observation function. Through a non-linear system Transition function, which incorporates with Taylor' equation, as follows

$$\mathrm{x}_k = C_{k-1} x_{k-1}^{m_{k-1}} dt + x_{k-1} \qquad (2.12)$$

Particle filter was applied here to update the estimation of the parameters, makes a prediction on the amount of the tool wear, and at the end of every run, the RUL is estimated. With the help of the autoregressive and support vector regression, the prediction error rate gets decreased. This promoted method realized the indirect measurement, combined with the robustness of the Particle filter. Parameters in the Taylor equation get estimated and updated every time step. It can realize diagnosis and prognosis of the tool wear, including RUL estimation for every pass. However, nonlinear system transition function increases the calculation time complexity. There is also the issue that the system works because of very high resolution force dynamometers which aid in the estimation process.

2.4.6 Particle Filter Finite Difference as System Transition Function

Moreover, when considering the robustness and lower time complexity in computation, the linear model is also realized by the PF as mean for calculation. In the paper of (Niaki, et al. 2015a, 2015b), Root mean square (RMS) of the power signal of spindle motor gets collected, and used as the indirect measurement to the tool wear. Kalman filter and particle filter are applied in this paper. It applied state space model as following to realize tool wear prediction via online measurement:

For system transition function, it assumes that the tool wear process is subject to the first order finite difference

$VB(k) = VB(k-1) + VB'(k)\Delta t$

$VB'(k) = VB'(k-1)$ (2.13)

Moreover, assume the linear relationship between power and VB, it gets

$$P = K_1 + K_2\,VB \qquad (2.14)$$

where, P is the spindle power, VB is the flank wear, and k is the k-th pass. Their result shows that compared with the Kalman filter, the estimation error by the particle filter gets decreased up to 33%. For their application in tool wear diagnosis and prognosis, the nonintrusive sensors get applied, the power sensor with the capabilities of incorporating indirect measurement via particle filter model. Also, particle filter is robust, which do not suffer the calculation problem as Kalman filter. However, the assumption of first order finite difference of the tool wear limits its application, the tool wears process do not follow the first order finite difference.

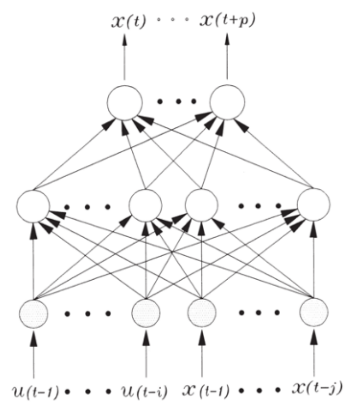
2.5 Model-free methods

As model-free methods, Neural networks (NN) were widely used for modeling and predicting tool wear because they can learn complex non-linear patterns without the need for an underlying data relationship. They have strong self-learning and self-adaptive capabilities. Most previous work do not consider the sequential nature of the sensor data or are reliant on a combination of intrusive sensors that limits its practical use. Feature extraction and selection are most commonly performed, which therefore require some form of human labor involved in deciphering the sensor signals.

2.5.1 Feed Forward Neural Networks

Artificial Neural Network (ANN) uses the mathematical model to simulate neuron architecture typically seen in the nervous systems of animals. The nodes as neurons are connected with associated weights to form a network, and the weights are trained with training data. The weights establish the relationship between the nodes which denote the variables, including the indirect and direct observations. ANN is capable of approximating the continuous nonlinear function desirably without understanding the object in physical perspective, so that no analytic formula is needed. Feed-Forward Neural Network (FFNN) is a version of the ANN with all inputs and outputs in the training data originating from same time step. Specifically, for the application in tool wear, as mentioned by Liu, et al. (1999), a Multilayer Feed-Forward Neural Network (MLFF NN) is designed. There are three layers, the input layer with input variables, the hidden layer and output layer which is the tool wear. The online indirect measurement in their work is the cutting force ratio. Other inputs are the cutting conditions, including feed rate, depth of cut, and spindle speed. The weights between input layer and hidden layer, between hidden layer and the output layer are identified by the iterative learning process and Back-propagation (BP) methods. With a trained Neural Network, it can realize real time online estimation of tool wear. Although it is flexible enough to include the machining conditions and the indirect measurement to the network, yet it can only realize the online diagnosis. Feed Forward Neural Networks only consider the static relationship between the indirect measurement and the tool wear, which is why it is a considered a static model for estimation of tool wear.

Moreover, there are numerous neural network methods that have been developed over the past decades. It can generally be categorized into two groups, one is static neural network and the other is the dynamic neural network. The static one does not have the feedback loop, which take the output or hidden nodes values into the input nodes at delayed time steps, or time delayed input. This group includes feed forward neural network, with single layer perceptions, or multilayer perceptions (Haykin, 2009). The dynamic one often have the input or output at different time step, so that this type neural network can model the dynamics of the state or input variables. This group include time delay neural networks (Hush, et al., 1993), recurrent neural networks, like Nonlinear Autoregressive Exogenous (NARX) (Lee, 1996, 1999).



(Lee, 1996, 1999)

Figure 2.2 Nonlinear Autoregressive Exogenous (NARX)

Wang, et al.'s (2008) work proposed the Fully forward connected neural network (FFCNN), which do not need to specify the number of layers for the hidden nodes, comparing to the Multilayer perceptron (MLP), and the optimized FFCNN by pruning some weights among the network. Their training processes are accomplished by the extended Kalman filter (EKF), rather than the back propagation (BP) algorithm. The inputs to the model include cutting velocity, feed rate, cutting force ratio, and depth of cut (constant), and outputs are the binary expression of the tool wear. The result shows that the optimized FFCNN trained with EKF has a faster converge rate, more accurate and a robust estimation than MLP trained with BP. However, the methods still did not consider the time dependencies of the inputs and outputs to model, and cannot make prediction of the future tool wear.

Ozel, et al.'s (2005) work applied the feed forward neural network (FFNN) to estimate the surface roughness, and tool flank wear on variety of cutting conditions. Bayesian regularization with Levenberg–Marquardt method is used in training to obtain good generalization capability and help in determine the number of neurons in hidden layer. The inputs to the model are tool edge geometry, Rockwell-C hardness, cutting speed, feed rate, cutting length, cutting force, and the outputs are VB and Ra. Surface roughness (Ra) and flank wear (VB).

2.5.2 Dynamic Neural Networks

Dynamic neural networks are ANNs which involve the construction of the temporal information which can be the inputs and outputs from different time steps. It can simulate dynamic system with nodes representing the system state and observations at different time steps. Nonlinear Autoregressive Exogenous (NARX) is a kind of recurrent neural network (RNN) which construct a nonlinear relationship between present target value and the input including present and past inputs and delayed outputs. In the papers of (Lee, 1996, 1999), Nonlinear Autoregressive Exogenous (NARX) Model is applied. The inputs are u(t) cutting force ratio, and x(t-1), …, x(t - j) is flank wear with time lag j. The Output is flank wear x(t), …, x(t+p). It can realize one-step and multi-step ahead tool wear prediction. In the application of the model, it considers the different cutting conditions; however, the model does need the historical direct measurements of the tool wear to realize its functions, which is not realistic in practical usage, because continuous direct measurements are not always available. Further, although they do consider applying the model in different machining conditions, yet they have not studied that in detail.

It is worth mentioning that in the work by Bukkapatnam, et al., 1999, 2000, they incorporate the cutting conditions as inputs, including the cutting speed V, feed rate f and depth of cut b. The output is the tool flank wear. Recurrent neural networks are applied in their context. The output with a time unit lag acts as the input of the network. The force sensor and accelerometers are used to collect the signals, and then the features are the input nodes in the model. The advantage of this method is that it incorporates the cutting condition as the inputs for the neural networks. Moreover, after the neural network gets trained, it can only rely on indirect measurement, without the need for the direct measurement. However, the methods are limited in that it can only realize one step ahead tool wear prediction, it cannot realize multi-step and RUL prediction. In practical usage, intrusive sensors—force sensor, and accelerometers, are used to make this prediction. Recently, Niaki, et al (2015, 2016) uses NARX Model to do the tool wear diagnosis and prognosis. They use features of the power signal extracted by the wavelet methods as inputs to the network, and output of the network is flank wear o(t). Their work uses the nonintrusive sensor, hall sensor, which is better than other intrusive sensors, like the dynamometers, force sensor and acoustic emission sensors. However, the limitations of the method are that they are only able to predict one step ahead tool wear prediction, and it cannot realize RUL prediction.

Another type of dynamic neural network which can also consider the time series are also used, that is Time Delay Neural Networks (TDNN). It is straightforward way to model the dynamic system. By feeding into a tapped delay line of finite extent the input sequence, then feeding into a static neural networks the taps from the delay line, the TDNNs gets constructed (Hush, et al., 1993). In the application for the tool wear, Ghasempoor, et al (1998, 1999) uses such network by combining the state space model:

$\dot{w}$ = g(w, cutting conditions) (State space transition function)

F = h(w, cutting conditions) (System observation function) (2.15)

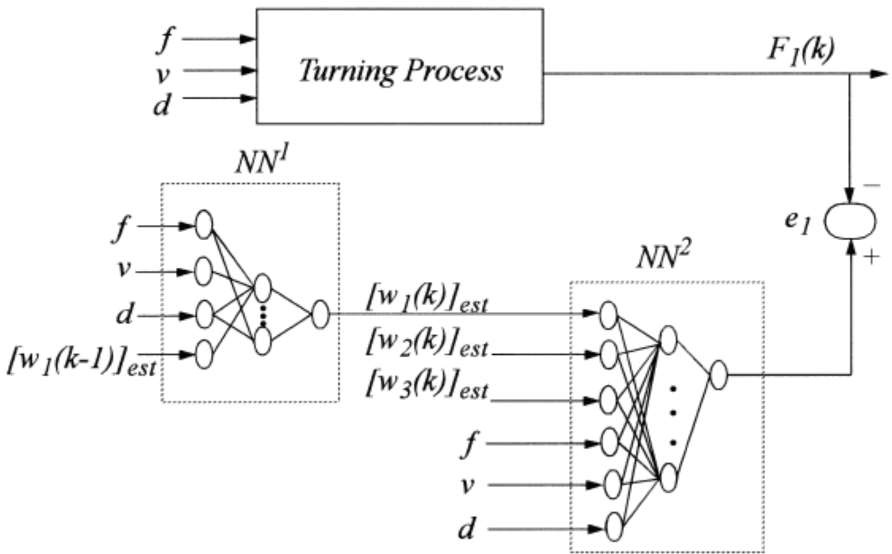


Figure 2.3 TDNN for tool wear estimation

This method includes machining conditions as inputs, includes feed rate f, cutting speed v, and depth of cut d. In Ghasempoor's (1998, 1999) work, two Neural Networks get trained, one for State space transition function, and another for System observation function. The method incorporated the cutting condition as the inputs to the neural network. It can only rely on indirect measurement, if it gets used online in the machining process. Further, with trained network for System observation function, the network for State space transition function can get trained with online indirect measurement. However, it can only realize one-step ahead tool wear prediction, but it cannot realize prognosis and RUL prediction. Also, the intrusive sensors, force sensor, get used.

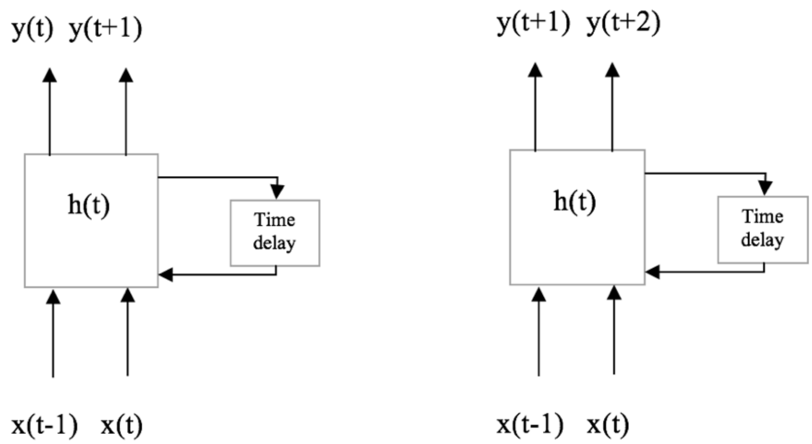


Figure 2.4 RNN with one-step ahead (left) and two-step ahead (right) prediction

The general RNN can realize one-step ahead and two-step ahead prediction, but its horizon on prediction is limited, so that it cannot be extended to the RUL prediction. Kamarthi, et al., (1995) and Luetzig, et al., (1997) composes RBF networks and a RNN. The radial basis function networks (RBFN) fuse and map the sensor signal to the tool wear, so they get the estimated tool wear. The RNN will provide the final flank wear estimation, given the predefined certain number of time steps delayed tool wear estimation provided by the RBF networks and the predefined certain number time step delayed output of tool wear provided by RNN. The method considers the time dependencies of the inputs, the sensor signal, and the past estimated tool wear, which is better than the FFNN. However, how many time steps of temporal dependencies is predefined in the model, rather than determined by the input value in our RNN method. Moreover, such method only give out the online estimation of the tool wear, but there is not prediction of the future tool wear, not to mention the multiple time step ahead prediction and RUL prediction.

2.5 Chapter Summary

The Taylor's tool life equation, without modification, can only use the direct measurement. It needs a large number of experiments to resolve the values of the parameters. As a result, the application is limited by the well known parameters. Moreover, owing to the changing internal and external factors, the estimation of the tool life cannot get re-estimated. As long as the constants do not change, the tool wear estimated will be the same. Nevertheless, the Bayesian updating with Taylor equation (Karandikar, et al, 2014a, 2014b) can update the estimation of the values of parameters. The methods can only reevaluate the parameters after the end of the tool life, which cannot realize the online updating of the values. The works (Karandikar, et al. 2012, 2013) using the Bayesian updating with the growth curves as the underlying assumption can incorporate the indirect measurement into the model; however, the assumptions about the growth curves generation do not have strong theoretical foundation, and their generation is based on the rule of thumb.

After introducing the Cox's model into the tool life estimation, which PHM, PCM are evolved from, the indirect measurements can get incorporated into the model, and these measurements can be time dependent. However, the assumptions that the failures are independent and identically distributed are seldom justified. There is no diagnosis and prognosis including one-step, multi-step prediction of the tool wear within these models.

Further, there are several variations of the HMM, including the Variable Duration HMM, and HSMM. These models can incorporate the direct and indirect observations into the model and realize the diagnosis and prognosis including one step prediction, multi-step prediction and RUL prediction. However, the assumption about the Markov process and discrete states should be justifiable. In fact, although the direct measurement can discretize into discrete state, the indirect measurement is better to keep as the continuous variables in order to capture the subtle changes. Furthermore, KF get introduce into the tool wear problem and it can take the continuous direct and indirect measurements into the model. However, it possesses the assumption that linear relationship for the indirect and direct measurements, and the transition of direct measurement with the time evolving. It also suffers from the calculation problems, if the assumptions do not hold. Such properties lead to its limited application and undesirable results. Further, the particle filter methods has overridden some of the inherent problems that the KF method has. It can take the underlying model with nonlinear relationship, and have better performance in computation. Currently, Niaki, et al (2015a, 2015b) apply the finite difference function as system transition function. Although some workpiece material can be well defined by finite difference function, yet some workpiece reject such underlying formula. Wang, et al (2015a) and Wang, et al (2015b) promoted nonlinear transition function, which can estimate the coefficients in the Taylor's equation, but it does not help in diagnosis and prognosis. The calculation of these coefficients cause large amount calculation which is not necessary. Our work promotes another underlying model which is capable of modeling wider range of workpiece material than Niaki, et al (2015a, 2015b) and less computation consumption than Wang, et al (2015a) and Wang, et al (2015b).

Further, analytic formula must be provided for PF regardless if it is linear or nonlinear, which leads the formula specific for certain condition such as the combination of the cutting tool and workpiece material and the formula must be provided by the research. ANNs model provide a way to parry the problem. It is model free approach and model relationship as a black box. FFNN is static one which can only model the static relationship between the observation and tool wear. Dynamic NN, including the NARX, TDNN, and any form of Recurrent Neural Networks (RNNs), can model the dynamic tool wear process. Also, the RNN relax the assumption about the Markov process in the HMMs. This dissertation focuses on further studying the RNN and its framework of training and testing practices that will be effective in producing an RNN that is suitable for the diagnosis and prognosis of tool wear. Further, we will propose the study about combine the training and test process in RNN model, and give study on the training and testing strategy. Initial data will still be required to train the RNN.

Chapter 3 Particle Learning in Machine Tool Wear Diagnosis and Prognosis

(The content of this chapter has been previously published in Zhang, J., et al., 2017a)

3.1 Introduction

Various models have been proposed by researchers for more than four decades on model selection for tool wear estimation. Among the earliest were physics based models relating cutting forces or power consumed to mechanically deriving equations to estimate tool wear (Gao, R., et al, 2015; Yen, Y. C., et al, 2004). These static models however depended on empirical models that characterize tool wear through multiple experimental conditions and then producing a set of constant coefficients which are plugged into the mechanistic models. These mechanistic models overlook the mechanism of the progression of tool wear and cannot be adaptive during the in-process machining operations (Liu, Q., et al, 1999). These static type models are useful in tool wear diagnosis, in that it does not take full advantage of prior information, which ultimately leads to inaccuracy when differential machining conditions are present. Methods such as of Support Vector Machines (SVM) (Widodo, A., et al, 2007), Neural Networks (Liu, Q., et al, 1999; Sick, B., 2002) and Self-Organizing Maps (SOM) (Owsley, L. et al, 1997) in tool wear diagnosis, all of which are machine learning methods based on frequency or time-frequency domain analysis methods have also been developed.

By contrast, dynamic models often incorporate the progression of tool wear through a state-space model, incorporating historical observations and online observations of the system to improve accuracy of prediction (Niaki, F. A., et al, 2015a, 2015b, 2015c, 2016; Wang, J., et al, 2015; Hanachi, H., et al, 2016; Lopes, H. F., et al, 2011; Hartikainen, J., et al, 2011). Models of these type include Bayesian updating, Kalman filter, particle filter, and dynamic neural networks (Carvalho, C., et al, 2010). The state space model provides certain benefits in that it needs less data to train the model, because it gets modeled with certain knowledge and assumptions of the tool wear degradation process.

The objective of this study is to improve tool wear estimation and RUL prediction using low cost minimally intrusive sensing in machining operations. The focus of this paper will be on method development using Particle Learning (PL) and implementation of the approach to estimate tool wear and RUL of uncoated tooling inserts for steel based alloys by online measurement of vibration signals. The organization of this work is as follows: first we provide a theoretical background of Particle Learning method, experimental setup and selected cutting conditions for experiments. We finally report on the ability of the method to provide one-step and two step look ahead predictions of tool wear state for two experimental runs.

3.2 Stochastic Modeling Methods for Tool Wear

For a dynamic system, such as machine tool wear prediction, the amount and type of wear itself is difficult to be directly observed during the machining process. It is either measured offline (eg. Optical microscope, laser probe) or indirectly through online sensing techniques (vibration, spindle power, acoustic emission or cutting forces). In the latter case, it is possible to write a system transition function with time, t, by assuming the Markovian state transition, wherein, the current state is related to the previous state. The explicit system transition function is

$$x_{t+1} = f(x_t) + \omega_t \qquad (3.1)$$

where ω_t is caused by uncertainty of the state model with a known probability density of $f(x_t|x_{t-1})$. This transition function can be further described using a Bayesian updating framework (Si, X. S., 2013), wherein direct or indirect measurements can be used to predict the state variable, in this case, amount of tool wear:

$x_{t+1} \sim p(x_{t+1}|x_t, \theta)$ (3.2)

The tool wear processes may not be a linear degradation model or an exponential like degradation model. The Bayesian updating with closed form here is to estimate the parameters in the models. Si, *et al*, (2013) applied the linear degradation model as given by:

$X(t) = \varphi + \theta t + \sigma B(t)$ (3.3)

where φ is the initial degradation, and assumed $\varphi = 0$, θ and σ are the drift and diffusion parameters, and B(t) denotes the standard Brownian Motion. The authors developed the closed form of the Bayesian updating for the parameters of both the linear degradation model and exponential like degradation model. However, for specific model, the closed form for Bayesian updating do not exist or hard to develop. Further, even if developed, building a one-step ahead prediction on the closed form equation of high time complexity and will require significant computation time.

Therefore, in order to avoid developing the closed form equations for Bayesian based updating, Particle filter based methods are adopted in the Bayesian updating calculation. Niaki, *et al*. (2015a) and Wang, *et al*. (2015) have applied the particle filter method upon the developed system transition functions. Niaki, *et al*. (2015a, 2016) assumes that the tool wear rate does not change with time, thus the process is subject to the first order finite difference, so that:

$VB(k) = VB(k-1) + VB'(k-1) \times MR \times \Delta t + w_1(k)$

$VB'(k) = VB'(k-1) + w_2(k)$ (3.4)

Where k is the k-th cutting pass, VB is flank wear of tool, MR is material removal rate, $w_1(k)$ and $w_2(k)$ subject to Normal distribution, and are independent of each other. In their experiment, they apply above equation at fixed machining conditions, and the time interval of every pass is equal, meaning that MR×Δt is a constant. Wang, *et al*. (2015a, 2015b) incorporates tool wear growth behavior evolving with time in their system transition function:

$$x_k = C_{k-1} x_{k-1}^{m_{k-1}} dt + x_{k-1} \quad (3.5)$$

where x is the tool wear, m_{k-1} and C_{k-1} are the parameters and subject to certain distribution function. Their model incorporates the physical model, the power law, specifically Paris's law, into their system transition functions, and realized online updating for estimating the parameters of the model. Since their initial experimentation was conducted at the same machining conditions, there is limited benefit to estimating the parameters of the state transition function through a non-linear function than a linear approximation of the system transition function. Additionally, nonlinear system transition function increases the time complexity and need for computational resources. This can also limit its applicability to real-time updating of the remaining useful life of the tool insert.

In summary for the current research, some of them apply the direct measurement, some cannot realize online updating, some have high time complexity, some apply the more intrusive measurement, and some do not figure out the RUL prediction. There is no comprehensive solution for the above problems. In order to compensate the shortages in the current research, we should implement a set of solutions to deal with them. First, incorporate indirect measurement into model, so that reduce the overhead time for the machine. Second, utilize the less nonintrusive sensor, so that reduce the interference to the machining process at best. Third, it should provide such robust and low calculation cost algorithm on online estimating and predicting the tool wear and RUL. Also, such method should be applicable to certain range machining conditions and combinations of the cutting tool and workpiece.

3.3 Technical Approach

In our work, we consider the property that the cutting tool inserts wears out gradually (Lee, J. H., et al., 1996) and by taking advantage of the benefit of the linear system transition function, we relax the assumption about finite difference to a first order linear equation. This then leads to:

$$x_t|x_{t-1}, \theta \sim N(\alpha+\beta x_{t-1}, \tau^2) \qquad (3.6)$$

Where x_t is the tool wear at the t-th cutting pass, θ is composed of (α, β, τ), the parameters of the transition function (Lopes, H. F., et al., 2011). In the real system state, tool wear cannot be observed after every cutting state, hence this needs to be estimated from indirect measurements, which must be incorporated into the state space model. Generally, system observation function is given by

$$y_t = g(x_t) + e_t \qquad (3.7)$$

where, y is the indirect measurement, e_t is the error term, subject to certain distribution. Under the constant machining parameters, including depth of cut, spindle speed, and feed rate, the tool wear can be determined by the linear regression with root mean square (RMS) of the vibration signal.

$VB = K_1 + K_2 . RMS_{vibration}$ (3.8)

Therefore, regression analysis is applied for constructing the relationship between RMS of the vibration signal and the real tool wear measurement (VB).

3.3.1 Particle Learning

To realize online updating of tool wear diagnosis and prognosis through indirect measurements, techniques using Kalman Filter and Particle Filtering can be used to fulfill this task (Niaki, F. A., et al., 2015a, 2015b). In order to effectively use the Kalman Filter, the error propagation should be well approximated by a linear or a quadratic function; second, the Jacobian matrices or Hessian matrices should exist; third, even when they do exist, it is still challenging to calculate Jacobian matrices or Hessian matrices and Kalman filter is not robust to jump linear problem (Hartikainen, J., et al., 2011). As a result, *Particle filter* is superior to Kalman filter (Niaki, F. A., et al., 2015a) when estimating parameters in a state space model. Moreover, Particle Learning (PL) (Carvalho, C., et al., 2010) is a special kind of Particle Filter method which provides system state filtering, sequential parameter learning, and smoothing for state space models. Specifically, in order to realize online updating with indirect measurement from past and current states, generally, the *Bayes' theorem* can be applied to build the system transition function as:

$$p(x_t|y^t) = \frac{p(y_t|x_t)p(x_t|y^{t-1})}{p(y_t|y^{t-1})} \qquad (3.9)$$

Within the system transition function, there are a few parameters in the system transition model which must be updated after every cutting pass.

$p(x_t, \theta|y^t)$, where $y^t = (y_1, \ldots, y_t)$. (3.10)

Further, in order to make a prediction of the system state, state prediction probability distribution can be generally written as:

$p(x_{t+1}|y^t) = \int p(x_{t+1}|x_t)p(x_t|y^t)dx_t$ (3.11)

Where $y^t = (y_1, \ldots, y_t)$.

To solve eqn 3.11, the integral w.r.t. x_t must be solved to obtain the marginal distribution. To implement Bayes' theorem and the integration with respect to x_t are both analytically intractable and/or computationally costly. In this respect, the particle filter method introduces the Dirac measure into the calculation to help solve for the parameters. Detailed explanation of the Particle Learning methods is provided in the work by Carvalho, *et al*. (2010) and Lopes, *et al*. (2011). Briefly, the Dirac measure is given by:

$\delta x(A) = 1_A(x) = \begin{cases} 0, & x \notin A \\ 1, & x \in A \end{cases}$ (3.12)

Therefore, for a probability distribution, it can be discretized with the Dirac measure as *particles*:

$p^N(x) = \{x_t^{(i)}\}_{i=1}^N = \frac{1}{N}\sum_{i=1}^N \delta_{(x)^{(i)}}$ (3.13)

After adding the parameters in the model, the *Bayes' theorem* for updating system state can be modified as

$p(x_t, \theta|\, y_{t+1}) \propto p(y_{t+1}|x_t, \theta)\, p(x_t, \theta\, |y^t)$ (3.14)

With the aid of the Dirac measure, (eqn 3.11) can be substitute by the calculation given by:

$p^N(x_t, \theta|y_{t+1})$ = Sample with weights $\propto p(y_{t+1}|x_t, \theta)$ from the $p^N(x_t, \theta\, |y^t)$ (3.15)

On the other hand, Marginal distribution for Prediction

$p(x_{t+1}|y^t) = \int p(x_{t+1}|x_t)p(x_t|y^t)dx_t$ (3.16) can be calculated as sampling process

$p^N(x_{t+1}|y^t) = p(x_{t+1}|x_t^{(i)})$ (3.17)

More detailed, by applying *Resample-Propagate Rule*

Bayes' theorem $p(z_t|y^{t+1}) \propto p(y_{t+1}|z_t)p(z_t|y^t)$ (3.18)

get updated with:

$p(z_t|y^{t+1})$ = sample $p^N(z_t|y^{t+1})$ with weights $\propto p(y_{t+1}|z_t)$ (3.19)

and the marginal distribution is given by:

$p(z_{t+1}|y^{t+1}) = \int p(s_{t+1}|x_{t+1}, s_t, y_{t+1})p(x_{t+1}|z_t, y_{t+1})p(z_t|y^{t+1})dx_{t+1}\, dz_t$ (3.20)

Get estimated by sample from:

First, draw $s_{t+1}^{(i)} = p\left(s_{t+1}\middle|z_t^{(i)},\ y_{t+1}\right)$; (3.21a)

Then, draw $x_{t+1}^{(i)}=p\left(x_{t+1}\middle|z_t^{(i)}, s_{t+1}^{(i)}, y_{t+1}\right)$ (3.21b)

Where $z_t = (x_t, s_t, \theta)$, s_t is sufficient statistics, which is the auxiliary variable used as to save and transfer the information between time steps. By adding the conditional sufficient statistics as an auxiliary variable to realize the sequential parameter learning for θ, the general particle filter is upgraded to a particle learning algorithm (Carvalho, C., et al., 2010).

In conclusion, the particle filter and particle learning, which introduce the Dirac measure, make intractable calculation in eqn (3.9) and (3.11) into tractable one, and greatly reduces the computational cost for the Bayes' theorem and Marginal distribution. More specifically, for our case about the tool wear, the *first order dynamic linear model* gets applied as system transition function. So, the system transition function is now given as:

$x_t|x_{t-1}, \theta \sim N(\alpha+\beta x_{t-1}, \tau^2)$ (3.22)

And the system observation function can be written as

$y_t|x_t, \theta \sim N(x_t, \sigma^2)$ (3.23)

where $\theta = (\alpha, \beta, \tau^2, \sigma^2)$ are termed as the hyperparameters.

The initial values of the hyperparameters are

$p(\theta) = p(\sigma^2)\, p(\tau^2) p(\alpha, \beta|\tau^2)$ (3.24)

Where $\sigma^2 \sim IG(n_0/2, n_0\sigma_0^2/2)$;

$\tau^2 \sim IG(v_0/2, v_0\tau_0^2/2)$;

$(\alpha, \beta) \sim N(b_0, \tau^2 B_0)$

The initial state $x_s \sim Norm\ (m_0, sC_0)$. With the particle learning algorithm, the parameters in the system transition function and hence the system state variable - the tool wear, can be updated for every cutting pass.

3.3.2 Diagnosis and Prognosis

Therefore, our proposed method is to realize online tool wear diagnosis, by taking advantage of the indirect measurement from past and current states: $p(x_t, \theta|y^t)$, where $y^t = (y_1, \ldots, y_t)$. Generally, it applies *Bayes' theorem* in equation (3.14) which can be solved by PL with equation (3.19). Second, one-step ahead tool wear prediction can be done with $p(x_{t+1}, \theta\, |y^t, \theta)$, where $y^t = (y_1, \ldots, y_t)$. And the parameter θ gets updated at the same time. With the aid of the PL, it can be solved by eqn (3.17) and (3.21). Third, for multi-step ahead prediction, because the indirect measurement at time t+1 is not available, the two-step look ahead tool wear cannot get predicted directly. By taking the one-step ahead tool wear prediction into the new iteration, as input $p^N(x)$ in eqn (3.25) and then the new $p^N(x)$ for the next step is obtained from eqn

(3.26), the two-step ahead prediction get estimated. Here, the assumption is the standard deviation τ^2 of the normal distribution in the system transition function is constant after time t.

$p^N(\mu) = p^N(\alpha)+ p^N(\beta)p^N(x)$ (3.25)

where μ is the mean for the normal distribution.

$p^N(x)$ = resample N particles from $Norm(p^N(\mu), \tau^2)$ (3.26)

For two-step ahead tool wear prediction, eqn (3.25) and (3.26) are run once, and three-step ahead run twice, and so on.

Lastly, for Remaining Useful Life (RUL) prediction, the criterion of the tool wear is set as 0.3 mm, for certain level conservation, 90 percentile of the tool wear is used as the reference. By looping through eqns (3.25) and (3.26), until the certain percentile tool wear reaches the tool wear criterion, the number of loops can be used to predict the RUL, which in our case, will be the number of cutting passes to reach the set tool wear criteria.

3.4 Experimental setup

There are two experiment cases for tool wear experiment. First case is the experiment conditions in which nose wear is obtained, by keeping the axial depth of cut to be shallow. In such shallow cutting, the nose of the cutting edge is worn during the cutting. Second case is the experiment conditions which cause complete flank wear, which is the axial depth of cut is deep enough, so that the flank of the cutting edge is worn out during the cutting.

In order to evaluate the Particle Learning method on the proposed state space model, experiments were conducted to obtain actual tool wear measurements using a HAAS VF2 under dry milling machining conditions.

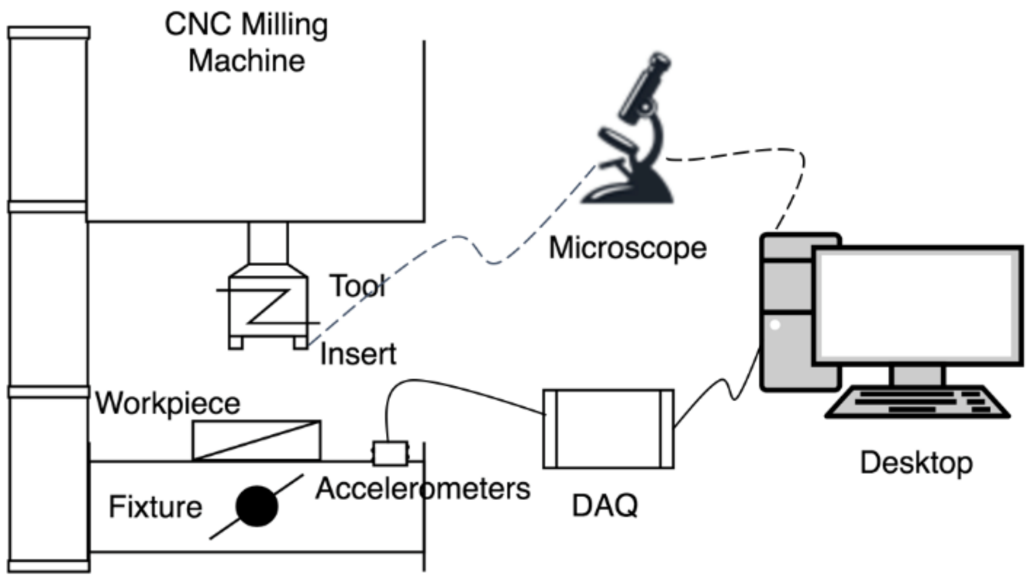


Figure 3.1 Schematic diagram of experimental setup

A two-flute indexable milling tool with diameter 12.7 mm, Sandvik Coromill 390 (RA390-013O13-07L) installed with one uncoated insert, Sandvik Coromill 390 Insert (390R-070204E-NL) was used in the cutting process. The workpiece material utilized is Steel 4142 (cold rolled, 40-45HRC). Each single cutting pass was 37.0mm long, with the spindle speed at 500rpm, leading to a surface speed of 19.81m/min and chip load at cutting to be 0.05mm. The radial depth of cut is 6.5mm and axial depth of cut at 0.5mm. A 3-axis vibration sensor (Kistler accelerometers 8762A10) was installed on the fixture attached to the workpiece. Sensor data was sampled at 1652Khz and its RMS signal formed the online indirect measurement. After every single pass of cut along the length of the workpiece, offline direct measurements were taken to measure actual tool wear. Table 3.1 provides actual flank wear measurements and its associated vibration index – the RMS of the vibration signals. The machining process was repeated for five experimental runs.

3.5 Results and Discussion

3.5.1 Flank (Nose) Wear Monitoring Experiment

The initial values of the hyperparameters for the experiment settings are set at α_0 = -0.1, β_0 = 1, τ_0^2 = 0.0001, σ_0^2 =0.0001, m_0 = 0.0, C_0 = 1.0, n_0 = 1.0, v_0 = 1.00, $b_0 = [\alpha_0, \beta_0]$, $B_0 = [1/\tau_0^2, 1/\tau_0^2]$, $sC_0 = \sqrt{C_0}$, $sB_0 = \sqrt{B_0}$, and the number of particles in PL is 10,000. Given indirect measurement y^t, with the aid of the Particle Learning, we realized the online updating of the model parameters, $\theta = (\alpha, \beta, \tau^2, \sigma^2)$, the estimation of the tool wear at current cut, next cut, and next two cuts. For the replication 1 of the cutting experiments, the final estimation of the parameter, with 95% confidence level, $\alpha \in$ (-0.0412 -0.0155 0.0141), and $\beta \in$ (0.9702 1.1188 1.2530). In the scenario of cutting with a single uncoated insert on Steel 4142, the proposed method was able to achieve real-time online tool wear estimation using the vibration signal as the indirect measurement. With the updated estimations of the model parameters, with aid of PL, by eqn (3.21), the online diagnosis to the tool wear is obtained. Also, for the tool wear prediction, with the aid of the PL, one step ahead prediction can be solved by equation (3.17). And for two-step ahead tool wear prediction, eqn (3.25) and (3.26) run once to obtain it.

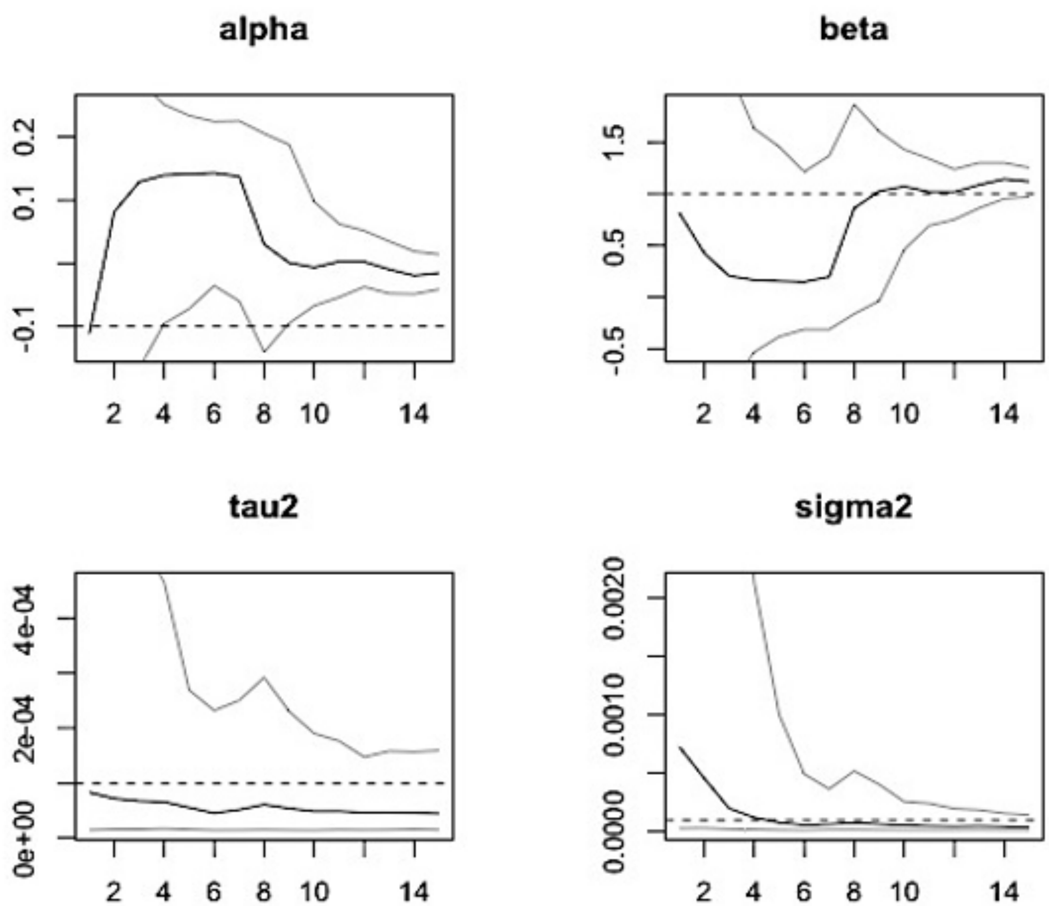


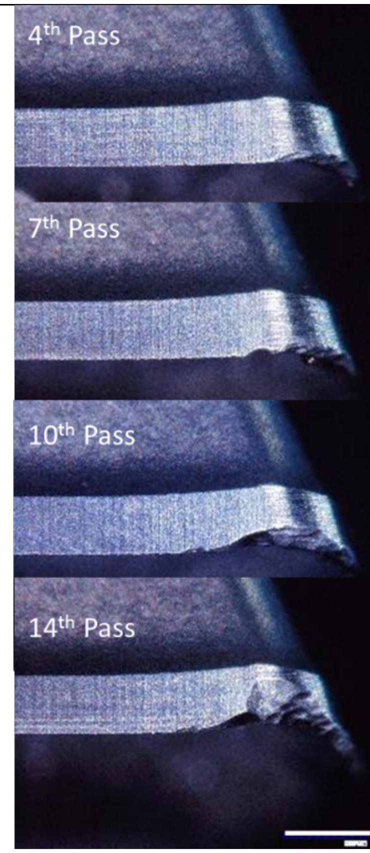
Figure 3.2 Online Updating of Model Hyperparameters of Replication 1

The x-axis represents the number of cuts; y-axis is the value of parameters. The dark line is the sample mean of the prediction and the light grey lines are the confidence limits. The dashed straight lines indicate initial values.

Table 3.1 Nose Wear Measurement and Vibration Index

Images on right show tool wear micrographs. Scale bar 500μm.

Replication 1			Replication 2		
Test	Flank Wear	Vibration Index	Test	Flank Wear	Vibration Index
#	(μm)	$\times 10^{-3}$ (g)	#	(μm)	$\times 10^{-3}$ (g)
1.1	159	11.6336	2.1	85	11.5215
1.2	175	13.2887	2.2	94	12.2370
1.3	175	13.7673	2.3	98	11.7758
1.4	179	13.9541	2.4	99	11.9476
1.5	179	13.6441	2.5	106	12.2356
1.6	180	13.5912	2.6	110	12.9962
1.7	180	15.9381	2.7	115	13.7390
1.8	184	18.6700	2.8	142	13.6430
1.9	191	18.7381	2.9	183	17.5361
1.10	199	20.0806	2.10	186	18.0902
1.11	214	20.2440	2.11	287	20.2915
1.12	220	20.9609	2.12	305	22.0541
1.13	248	24.5650	2.13	330	25.8142
1.14	274	27.4840			
1.15	299	28.4834			



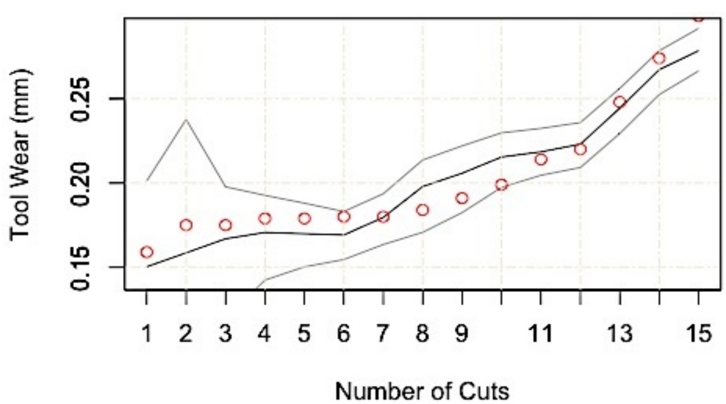


Figure 3.3 Online Tool Wear Diagnosis with Indirect Measurement in Replication

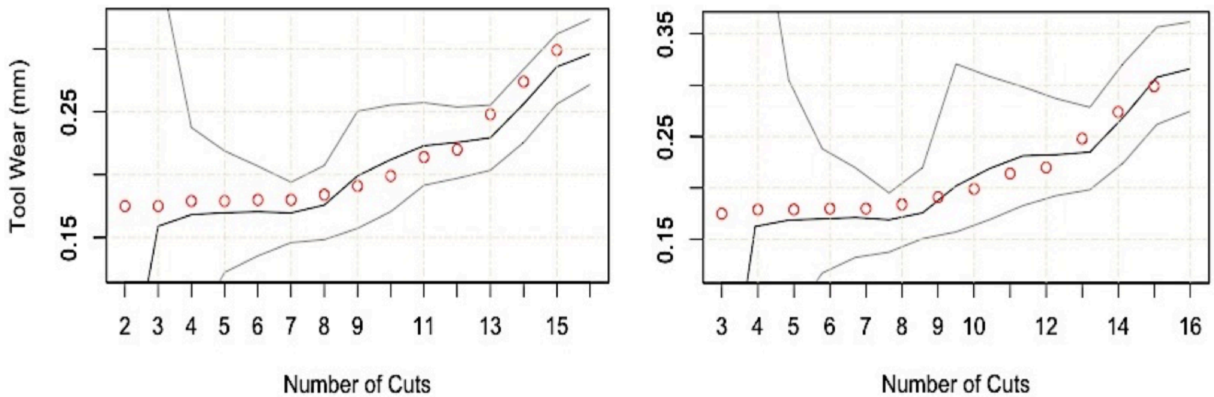


Figure 3.4 A. Online Tool Wear One-Step Ahead Prediction of Replication 1;

B. Online Tool Wear Two-Step Ahead Prediction of Replication 1.

Light grey lines indicate 95% confidence interval. The dark line indicates the sample mean of the prediction and the light grey lines are the confidence limits. The red circles are actual tool wear measurements.

We repeated the same experimental setup with similar machining conditions to estimate the hyperparameters and the RUL. The final estimation of the parameter, with 95% confidence level, $\alpha \in$ (-0.0277, -0.0111, 0.0052), and $\beta \in$ (1.1201, 1.2254 1.3491). Despite similar machining conditions, these values are different from the experimental trial 1, indicating the importance of online monitoring. It is worthy to mention that here the confidence interval of β with 95% confidence level reject the hypothesis that β is always equal to 1 as given by the first order finite difference method used in the state space model developed by Niaki, *et al*. (2015a)

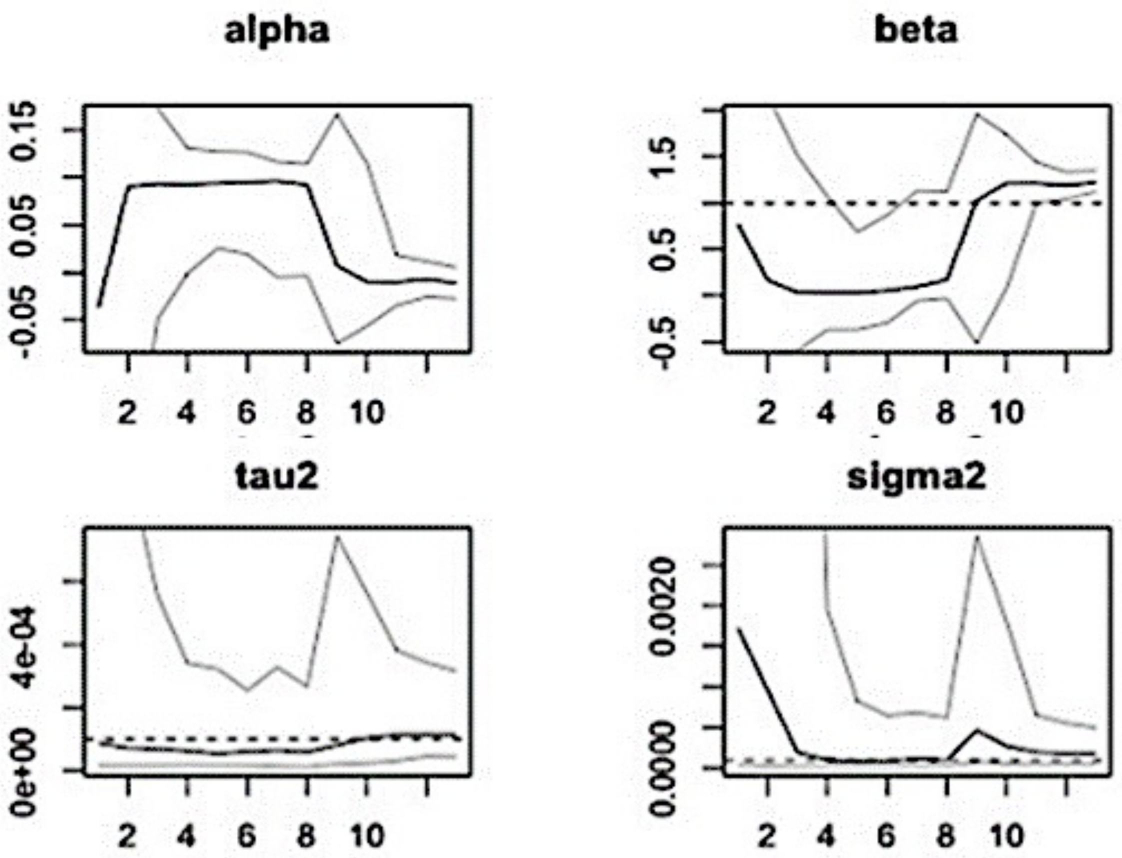


Figure 3.5 Online Updating of Model Hyperparameters of Replication 2

Furthermore, the RUL prediction for the tool insert is calculated via the online indirect measurement. Here, the RUL of 90 percentile tool wear prediction is used as comparison to the criteria of the tool wear 0.300 mm. When the RUL prediction curve reaches 0, which means it has reached the tool wear criteria, then the insert will most likely wear out in the next cut and the operator can decide if this tool needs to be used in the next cut or not. Current industrial practice of determining when the tool is worn out depends on a set tool lifetime as recommended by the cutting tool vendor. Material inconsistencies, unpredictability in machine states and quality of the cutting inserts leads to tool wear amounts to vary significantly. Therefore, online indirect measurements of the process state can be used to estimate and predict tool wear amounts. These online indirect measurements can range from measuring spindle power, vibration, acoustic emission and cutting forces during the machining operation. Desired solution features must be non-intrusive sensing, relatively inexpensive to implement and must enable fast calculations within the context of the machining operation times.

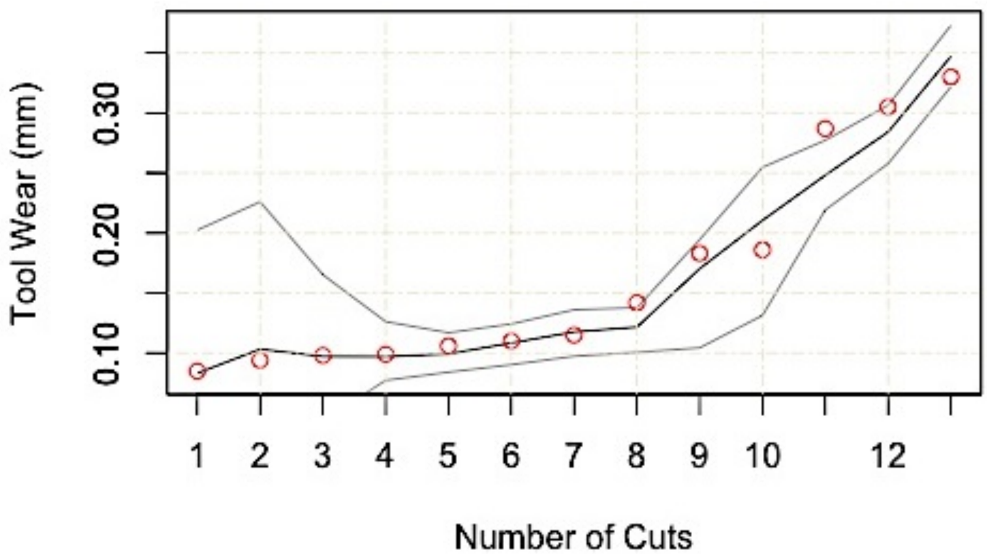
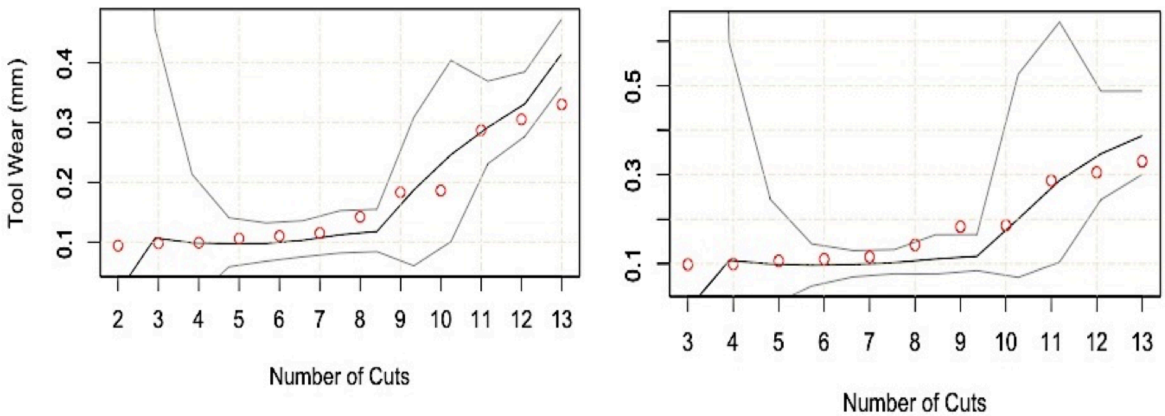


Figure 3.6 Online Tool Wear Diagnosis with Indirect Measurement in Replication 2

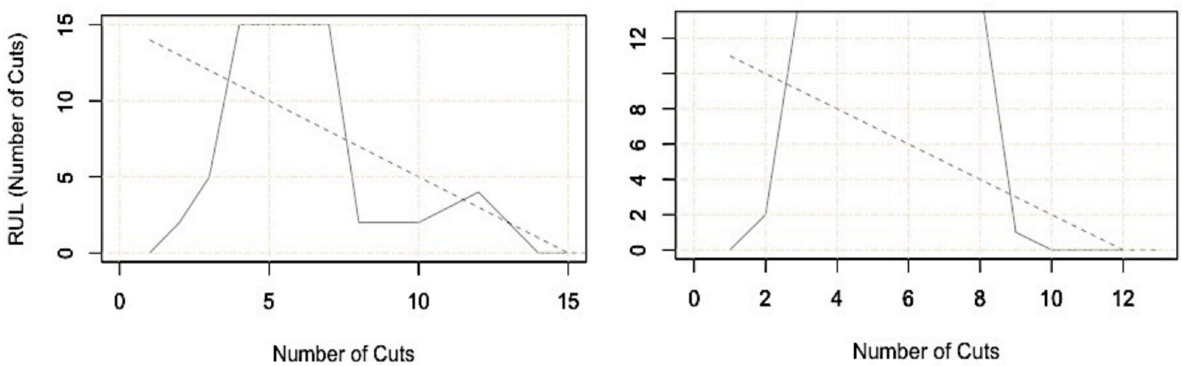
The linear system transition function adequately captures the dynamics behavior of the tool wear processes, which significantly reduces the time complexity of the nonlinear system transition function. By taking place the first order finite difference, the proposed model relaxes the assumption by the first order finite difference and thus become more flexible to capture the dynamics of the tool wear process.



A. Online Tool Wear One Step Ahead Prediction of Replication 2;

B. Online Tool Wear Two Step Ahead Prediction of Replication 2.

Figure 3.7 Online Tool Wear One-Step & Two-Step Ahead Prediction of Replication 2



The dashed lines indicate actual tool wear remaining useful life based on wear criterion.
A: Online RUL Prediction of Replication 1
B: Online RUL Prediction of Replication 2.

Figure 3.8 Online RUL Prediction of Replication 1 & 2

The PL method itself does not incorporate the machining parameters into the model. The system transition function can account for gradual tool wear but will not be able to predict abrupt tool wear such as a chip breakage or rapid degeneration of the cutting tool insert beyond normal wear. It assumes that the model parameters are static in the proposed method, although future work will need to look at whether the model parameters are indeed static or time varying during the life of a cutting tool insert.

After the experiment, we envision the following implementation approach in production settings. The company that sells the cutting tools will conduct prior experimental runs to build the system transition and observation functions based on various cut-ting tool insert materials and workpieces. Such industrial vendors will conduct detailed experimental runs, including direct observations based on their recommendation for machining parameters to be used. Once the models are trained and functions are developed, their customers rely on the indirect measurements from either vibration signals or from power readings generated to make an assessment on tool wear estimate. If the equipment is outfit-ted with an accurate tool wear probe, these measurements can be further improve the model based on the existing manufacturing conditions. While we have not shown spindle power readings in this study, this can be further extension of our work to be as non-intrusive as possible. The limitation of our approach is that further refinement in models may have to be conducted when customer use the same cutting tool for various materials. In such hybrid use cases, the development model approach may not be sufficient and needs to be augmented. The other limitation is that we assume that tool wear occurs gradually and our model cannot account for sudden fracture of cutting tool surfaces.

3.5.2 Flank Wear Monitoring Experiment

For the flank wear experiment, the results are as following Figure 3.9 :

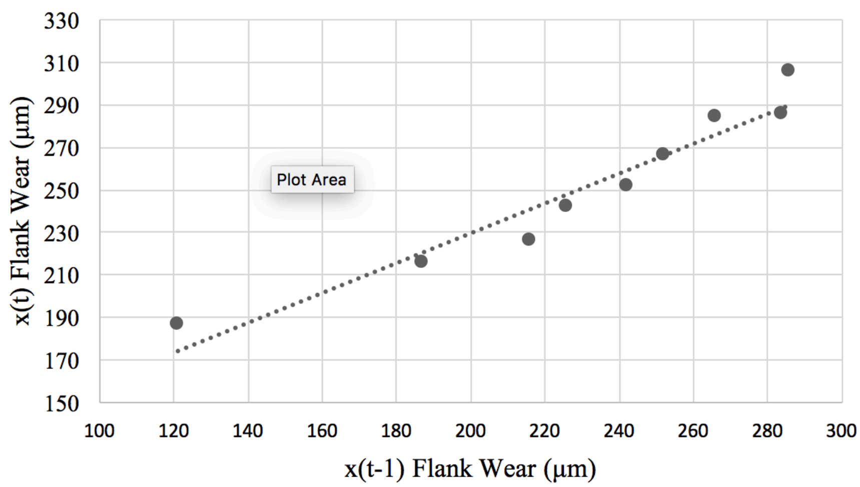


Figure 3.9 Plot the System Transitions of Tool Wear

The system transition function $x(t)=0.7012x(t-1)+89.62$ has $R^2 = 0.92942$, which indicate good linear regression between x(t) and x(t-1).

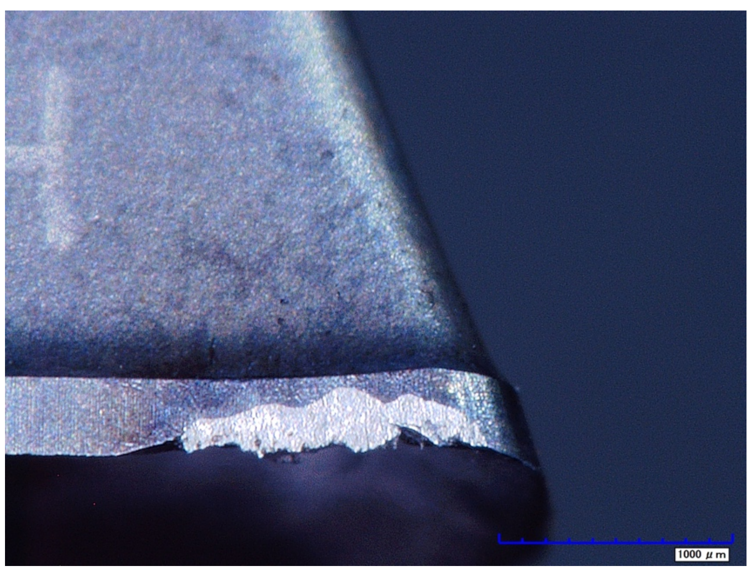


Figure 3.10 Full flank tool wear

However, it is highly nonlinear between the RMS of the vibration and flank wear on this machining condition in this experiment. Due to our model requires the good linear relationship between the indirect measurement and the direct measurement, it is impossible to use the RMS of vibration signal for the proposed model. Moreover, from the data analysis, there is no obvious correlation between the RMS of the hall sensor and the tool wear, therefore, the RMS of the hall sensor cannot be used to reflect the tool wear.

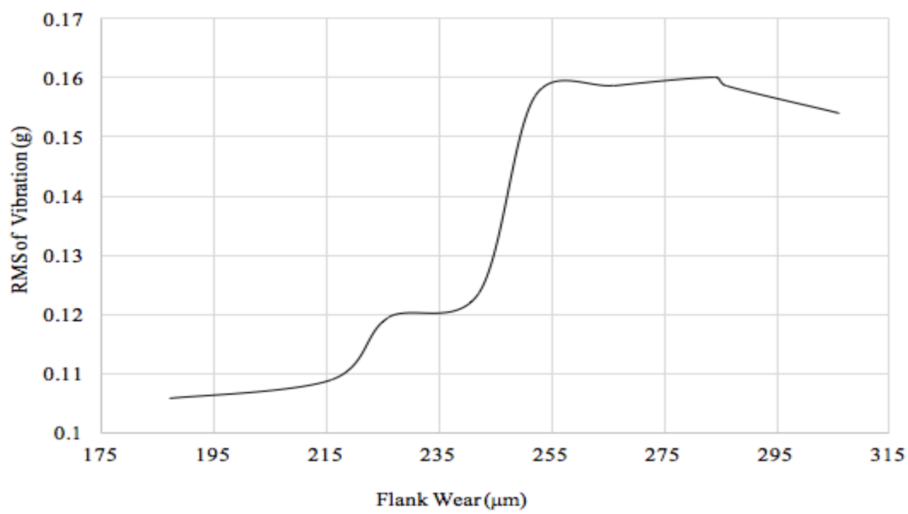


Figure 3.11 Indirect Measurement Versus Direct Measurement

3.6 Conclusion

The linear system transition function adequately captures the dynamic behavior of the tool wear processes, which significantly reduces the time complexity of the nonlinear system transition function. By taking the first order finite difference, the proposed model relaxes the assumption by the first order finite difference and thus become more flexible to capture the dynamics of the tool wear process. The PL method itself does not incorporate the machining parameters into the model. The system transition function can account for gradual tool wear but will not be able to predict abrupt tool wear such as a chip breakage or rapid degeneration of the cutting tool insert beyond normal wear. It assumes that the model parameters are static in the proposed method, although future work will need to look at whether the model parameters are indeed static or time varying during the life of a cutting tool insert. The PL with the proposed linear system transition described in this study utilizes indirect measurement data to account of the uncertainty of measurement and the inherent stochastic property in all tool wear processes. The contribution of our study is that we have proposed the use of a first order autoregressive relation as the system transition function for the tool wear. This reduce time complexity and robustness of the solution without having to take in nonlinear system transition functions. In addition, the model does not assume a constant finite order difference in tool wear amounts. The specific results of this initial study indicate that:

(a) The linear system transition function proposed in the work show its capability of capture the dynamics behavior of the tool wear processes. This significantly reduces the time complexity of solving the solution and making it practical for use in real-time machining; and

(b) The PL combining linear regression for the indirect measurement show the capabilities of the online diagnosis and prognosis for the tool wear. It is able to capture the linear transition of the tool wear rate.

Future studies will investigate applicability of the method to harder to machine metal alloys and coated insert tool types. In addition, different machining conditions and suitability of other sensor signals, such as spindle power can increase the practical relevance and robustness of the solution.

Chapter 4 Recurrent Neural Networks in Machine Tool Wear Diagnosis and Prognosis

4.1 Introduction

Data-driven approaches to predicting machine condition has seen rapid advancements due to low cost sensor technology and reduced computational cost. In machining based operations, a critical asset to obtaining superior part quality are the cutting tool inserts. These inserts wear out and are constantly changed in a production based environment. The continuous estimation of the tool wear classifications either by abrasion, erosion or tool breakage is necessary to improve the reliability of the metal-cutting manufacturing processes. Tool breakage can be detected by robust power monitoring techniques, leading to emergency shut-off of the machine. However, tool wear due to abrasion and erosion are hard to estimate in continuous production environments. At present, production environments resort to offline detection through the use of in-machine laser probes which measure wear amounts on the cutting tool surface or a scheduled replacement of cutting tools based on regular machining time intervals. Non-optimized changeover of cutting tool inserts can lead to lost production efficiency and reduced part quality. Detection and prediction of tool wear amounts in real-time is necessary to improve discrete manufacturing systems operations.

Diagnosing and prediction of tool wear research has been carried out for decades. This problem is particularly difficult due to modeling difficulties involved in a non-linear time-variant process. To cope with this problem, prediction techniques have been applied to estimate future tool wear amounts but often resort to intrusive sensing techniques which are limited in practical industrial implementation. Sensor signals are dependent on a number of factors including machine state, machining conditions and position of the sensor. Adding to the complexity is the apparent variability in the cutting tool, inconsistencies in the material properties of the workpiece and the larger number of machining conditions that must be accounted for.

Several methods have been proposed in the literature for tool wear estimation and prognosis through both linear and non-linear prediction. The most common non-linear prediction methods involve neural networks. Various work has shown the power of NN in terms of predicting tool wear amounts which can meet accuracy requirements. They do however require large data to collect over periods of time under various machining conditions. Improvements made to regular feed-forward neural network, recurrent neural networks were proposed to account for cyclic connections over time. Each successive time-step in the machining states are stored in the internal state of the network to provide a sort of temporal memory. However, conventional RNN did suffer from training artificial neural networks with gradient based learning and back-propagation through gradient vanishing or exploding issues (Hochreiter, et al., 1997). New variations in RNN, such as utilizing multi-level hierarchy, long short-term memory and residual networks have been developed to address this problem in deep learning.

Long Short-Term Memory (LSTM) is a RNN architecture that is excellent at remembering both short-term and long term dependencies. Originally designed by Hocheiter (1997) this method was developed to address the issue of the vanishing and/or exploring gradient problem when number of layers in the neural network was increased. These RNNs with LSTMs have been very successfully used in speech recognition, language modeling, network traffic regulation, visual recognition and many others.

The objective of this study is to improve tool wear estimation and RUL prediction by utilizing an LSTM based RNN framework, which makes more effective use of model parameters to train system observation and transition models of tool wear process. We utilize minimally intrusive vibration sensors to help train the RNN model. We train and compare our LSTM models at various numbers of parameters and configurations. The solution approach can be utilized to assess other critical assets of a manufacturing machine. The organization of this work is as follows: first, we provide a brief overview of related work in tool wear and in general recurrent neural networks. The LSTM RNN architecture and equations are laid out as the basis of our work. We detail our experimental setup and selected cutting conditions for tool wear diagnosis and prognosis. We finally report on the ability of the method to provide one-step and two-step look ahead predictions of tool wear state.

4.2 Related Works

4.2.1 Sequence Modeling Techniques Overview

The tool wear process is a sequence of time steps. Further, there are two variables in the sequence, one is direct measurement which is the actual tool wear, another is the indirect measurement- the features of the sensor signal. There are several sequence modeling techniques having dependent and independent variables in different areas, including the N-Grams in language modeling, Linear Dynamics System, Hidden Markov Models, and so on.

A brief introduction is given below to discuss the sequence modeling techniques and how these methods are possible candidate for the defined tool wear problem.

N-Grams: The N-Grams are widely used in Natural language processing (NLP). The bilingual units express as the tuples (t, s). With the text from two corresponding languages, the model is trained. After training, with input from one text from one language, the model can generate another text in another language. Specifically, for the machine translation, the tuple N-gram Model was proposed by (Marino, et al., 2006).

$$p(T, S) \approx \prod_{k=1}^{K} p((t, s)_k \,|(t, s)_{k-1}, (t, s)_{k-2}, \ldots, (t, s)_{k-n+1}) \qquad (5.1)$$

where t refer to the target, s to source, and $(t, s)_k$ to the k-th tuple of a given bilingual sentence pair, n is the number of time steps considered in the model.

Disadvantage of N-Grams models: Generally, N-Grams model is vulnerable to curse of dimensionality. To explain further, there are V^n possible n-grams, and thus more parameters need to be estimated (here, V is the dimension of the variable). Secondly, owing that it assumes the n time steps historical information, the long range dependencies are not captured. Third, N-Grams models are specifically designed for the linguistic model; whose data type is the discrete labelling data. For our application in sensor data, they are continuous streaming data. Although the continuous data can be discretized into label data. any two state values in the N-Grams model are at same distance from each other, which leads to loss of information when using the model. Last, there is no proposed strategy to deal with the missed target labels and mixing the training and inference, compared to other methods, which will be introduced later.

For **Linear Dynamical System** (LDS), it applies the linear system transition function and observation function to model dynamic process by the indirect observations. Owing to the assumption, it can only model the simple linear dynamic processaes. Such models include Kalman filter and our proposed methods, the Particle Learning on linear system transition function.

Hidden Markov Models (HMM). It considers the sequence in HMM as a stochastic process. There are two combined stochastic processes (Rabiner, 1989). The underlying one, which is not observable, subject to the finite state homogeneous Markov chain. The underlying one impacts another process which is observable. Given the sequence of the observed process, the underlying process probability gets estimated, and then the most probable underlying state sequence gets identified. Also, the parameters in the model gets updated at the maximum likelihood of the observed sequence (Gao, et al., 2015). The RUL can also be predicted with the updated HMM models. However, there are several assumptions in the HMM which limits its applicability. First, the Markov assumption as described earlier does not hold in all cases. Second, the assumption on the statistically independence of the current observation and the previous observation is not always true. For illustration, the sequence of the observations or indirect measurement of the model:

$Y = y_1, y_2, \ldots, y_T$;

With the second assumption for HMM with parameter λ,

$p\{Y|x_1, x_2, \ldots, x_T, \lambda\} = \prod_{t=1}^{T} p(y_t|x_t, \lambda)$. (5.2)

Such assumption fails to capture the dynamics of the indirect measurement, similar to the PL method.

Recurrent Neural Networks (RNN) are also designed to learn the sequential or time-varying patterns. RNNs is a neural network framework with feedback obtained from the past output. Compared to HMMs, N-Grams, LDS models, RNN does not have the following assumptions: 1) The assumption about the analytic formula and the linearity of the system. RNN can have nonlinear system transition function; 2) The assumption about the Markov process is that there is a linear relation between the indirect measurement and direct measurement is not made; 3) The assumption on the statistically independence of the current observation and the previous observation is also removed. Moreover, it is flexible for RNN to incorporate any input variables into the model, such as the hidden system state can be continuous or discrete for RNN. There are still assumptions made in the RNN approach (Goodfellow, I.J, et al. 2016). It assumes that the system transition functions are same for every time step, so that the values of parameters are the same for different time steps. Equivalently, the conditional probability distribution over the variables at time step t+1 given the variables at time t is stationary, which means that the relationship between previous time step and the next time step does not depend on time t.

There are two ways to express the variables: one way is to use the continuous value for the variable. Another way is to use the one-hot vector to express one variable. If the variable is the continuous value, the assumption in RNN will model tool wear as homogeneous transition process for the hidden layer given the indirect measurement as input. In comparison, if the tool wear and/or the indirect measurement are modeled as one-hot vectors, it carries more information for modeling the tool wear process given the indirect measurement, and thus need more data to train the model. One more problem is that the distance between the tool wear values and the indirect measurement are same in the one-hot vector, that could lead to loss of information and thus may need more data as compensation to train the model.

4.2.2 RNN Applications

RNN can model the dependencies of the elements of input and/or output sequence, and can capture of long-range dependencies. (Lipton, et al., 2015). Here, two applications will be introduced in two areas. One is in the system identification and another is in translation for languages.

RNN Application 1: System Identification

There are two corresponding system state sequence (x_1, x_2, …, x_T) and output sequence (y_1, y_2, …, y_T) out of the system. The NNs are applied to identify the systems.

Two neural network to identify the system for the State-Space Model. One RNN is for modeling system transition function:

$x(t+1) = f(W_a x(t), W_b u(t))$

One RNN is for modeling system observation function:

$y(t)= h(x(t))$

Where, u(t): input; x(t): output of the hidden layer; $f(\cdot, \cdot)$, h() are nonlinear activation functions; W_a, W_b, and C are matrixes with weight as element.

For System Transition Function:

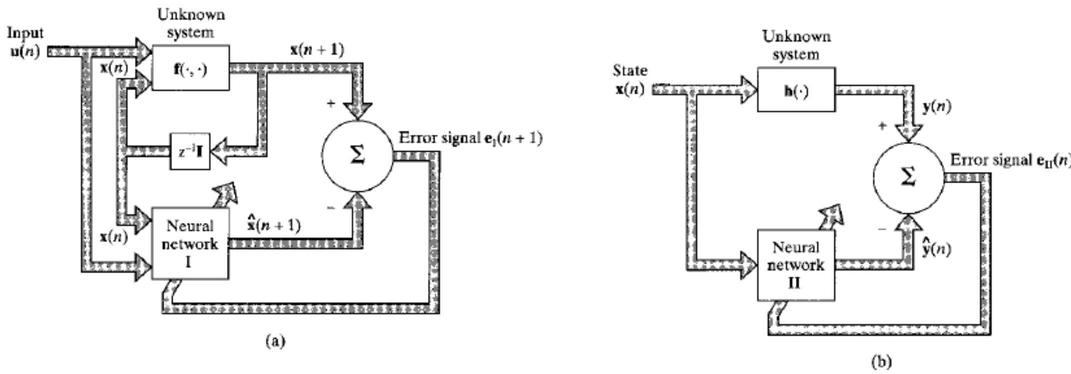
$e_I(t+1) = x(t+1) - \hat{x}(t+1)$

Here, e_I is the estimation error for system transition function; x is the real system state; and $\hat{x}$ is the estimated system state. e_I is used to update the synaptic weights of network I.

For System Observation Function:

$e_{II}(t+1) = x(t+1) - \hat{x}(t+1)$

Similarly, e_{II} is used to update the synaptic weights of network II.



Network I Network II

(a) for system transition and (b) for System Observation
(Haykin, 1998, 2008; Narendra, et al., 1990)
Figure 4.1 State-Space Model for System Identification

Traditional RNN, including the vanilla RNN and Elman RNN cannot capture long term time dependencies, due to the gradient vanishing problem (Goodfellow, et al. 2016).

RNN Application 2: Natural Language Translation using LSTM

For the RNN in the application of the translation from one language to another language, it assumes each language form a sequence, and another language form another sequence.

Specifically, $x_1, \ldots, x_T$ is an input sequence in one language; $y_1, \ldots, y_{T'}$ is its corresponding output sequence in another language.

For translation, its goal to estimate $p(y_1, \ldots, y_{T'} | x_1, \ldots, x_T)$, and then get most probable sequence as the output sequence in another language. Specifically, in the work of Sutskever, et al. (2014), it realized the translation by following steps.

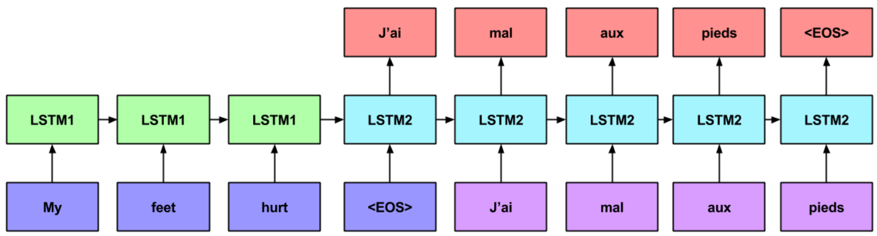
Input is one hot vector in English and output is also one hot vector in French. The model applied two different deep LSTMs. One for input sequence, and another for output sequence.

During the training process, the algorithm maximize the log probability of a correct translation T given source sentence S:

$1/|S| \sum_{(T,\,S)\in S} \log p(T|S)$

After training, make the translation by finding most likely translation according to the LSTM model:

$\hat{T} = \arg\max_{T} p(T|S)$



(Sutskever, et al., 2014)

Figure 4.2 LSTM translation diagram

Specifically, it obtained the fixed dimensional representation *v* of the input sequence, given by last hidden state of the LSTM, then compute the probability of $(y_1, \ldots, y_{T'})$

$p(y_1, \ldots, y_{T'}| x_1, \ldots, x_T) = \prod_{t=1}^{T'} p(y_t|v,\ y_1,\ \ldots,\ y_{t-1})$

Here, v keeps all the input sequence information. More detail refer to Sutskever, et al. (2014).

4.2.3 Neural Network for Tool Wear

A comprehensive review about Neural Network for tool wear has been conducted at Chapter 2.4. Feed Forward Neural Networks, include MLFF NN (Liu, et al., 1999), FFCNN (Wang, et al., 2008; Ozel, et al., 2005), are static and do not consider the time dependencies of the tool wear. Dynamic Neural Networks, include NARX (Lee, 1996, 1999; Niaki, et al., 2015, 2016), TDNN (Hush, et al., 1993; Ghasempoor, et al., 1998, 1999), and RBFN combined with RNN (Kamarthi, et al., 1995; Luetzig, et al., 1997), although consider the time dependencies, yet did not figure out the prediction of the tool wear, not to mention the multiple time step ahead prediction and RUL prediction.

In order to overcome the limitations of simple RNN, the gated recurrent neural networks were proposed in tool wear monitoring and prediction. The simple RNN can only realize one-step ahead and two-step ahead prediction, but its horizon on prediction is limited, meaning that it is limited in its ability to be extended to RUL prediction. The conventional RNNs cannot capture long term-dependencies, resulting in the Neural Networks not able to capture previous long-term history into current estimation. This is due to the well described problem of the vanishing or exploding gradient as the number of layers time step in the NN increases. Therefore, gated recurrent neural networks, including LSTM and GRU were developed to avoid this problem.

The motivation for this work is that a model-free approach which did not need an underlying analytical model would be a better approach to capture long term dependences, prediction in an arbitrary horizon and remaining useful life prediction for a non-linear time variant process. GRU has simpler structure than the LSTM, and less parameters to be trained, and asserted as having similar effect on capturing long term temporal dependencies.

The advantages of our approach over existing ANN and RNN approaches are:

• LSTM based RNNs capture time dependencies and make prediction over convention NNs that are based on Feed Forward Neural Networks.

• The proposed RNN approach does not have the limitation of limited prediction horizon period, which is the major drawback of time-delayed neural networks.

• Time delayed based RNNs, although can achieve time dependencies, the method cannot realized arbitrary time step-ahead and RUL prediction.

Our approach of including a model System Transition and System Observation function separately can naturally realize diagnosis and prognosis within the LSTM based RNN architecture. However, the assumption about the Markov process and discrete states should be justifiable. In fact, although the direct measurement can be discretized into discrete state, the indirect measurement is better to keep as the continuous variables in order to capture the subtle changes. Furthermore, KF get introduce into the tool wear problem and it can take the continuous direct and indirect measurements into the model. However, it possesses the assumption that linear relationship for the indirect and direct measurements, and the transition of direct measurement with the time evolving. It also suffers from the calculation problems, if the assumptions do not hold. Such properties lead to its limited application and undesirable results. Further, the particle filter methods has overridden some of the inherent problems that the KF method has. It can take the underlying model with nonlinear relationship, and have better performance in computation. Currently, Niaki, et al (2015a) apply the finite difference function as system transition function. Although some workpiece material can be well defined by finite difference function, yet some workpiece reject such underlying formula. Wang, et al (2015a) and Wang, et al (2015b) promoted nonlinear transition function, which can estimate the coefficients in the Taylor's equation, but it does not help in diagnosis and prognosis. The calculation of these coefficients cause large amount calculation which is not necessary. Our work promotes another underlying model which is capable of modeling wider range of workpiece material than Niaki, et al (2015a) and less computation consumption than Wang, et al (2015a) and Wang, et al (2015b).

Further, analytic formula must be provided for PF regardless if it is linear or nonlinear, which leads the formula specific for certain condition such as the combination of the cutting tool and workpiece material and the formula must be provided by the research. ANNs model provide a way to parry the problem. It is model free approach and model relationship as a black box. FFNN is static one which can only model the static relationship between the observation and tool wear. Dynamic NN, including the NARX, TDNN, and any form of Recurrent Neural Networks (RNNs), can model the dynamic tool wear process. Also, the RNN relax the assumption about the Markov process in the HMMs. This chapter focuses on further studying the RNN and its framework of training and testing practices that will be effective in producing an RNN that is suitable for the diagnosis and prognosis of tool wear. Further, we will propose the study about combine the training and test process in RNN model, and give study on the training and testing strategy. Initial data will still be required to train the RNN.

There are several assumptions in the Particle Learning method that may not always be valid in certain cutting conditions. First, is the assumption about the linearity of the system transition function. Although it is usually the case, for certain cutting conditions the system transition function may not always be subject to linearity. If the system transition function is nonlinear, certain explicit analytic formula should be given for the PL methods (Wang, et al., 2015a, 2015b). Second, is the assumption about the Markov process. The current system state only depends on the last system state. Such an assumption may not capture the entire dynamics of the system; Third, is the assumption about the linear relation between the indirect measurement and direct measurement.

The objective of this study is to improve tool wear estimation and RUL prediction using low cost minimally intrusive sensing in machining operations. The focus of this paper will be on method development using RNN and implementation of the approach to estimate tool wear and RUL of uncoated tooling inserts for steel based alloys by online measurement of vibration signals. The organization of this work is as follows: first, we provide a theoretical background of the RNN method, followed by the experimental setup and selected cutting conditions for experiments. We finally report on the ability of the method to provide one-step and two-step look ahead predictions of tool wear state.

4.3 The RNN State Space Model

The RNN State Space Model approach is a model-free method based on Recurrent Neural Networks by formulating a temporal dependent relationship between in-process indirect measurement, in-process real machine state and the relationships between the real machine states themselves in the temporal domain. Due to its model-free nature, RNN can model the non-linear relationship between their inputs and outputs. Therefore, no assumption pertaining to the Markov model and linearity can be avoided. Long term dependencies cannot be modeled by conventional RNN approaches. Here we adapt the Long Short Term Memory (LSTM) and Grated Recurrent Units (GRU) to model long term dependencies for multi-step prognosis of tool wear processes. We build a system observation function and a system transition function with RNN, LSTM and GRU. We split experimental data collected to a training set, validation and a test set. We train the system observation function and the system transition function separately to help tune its individual parameters and to implement independent training protocols. More importantly, the independent training also allows these functions to go through varying training iterations. A comparison will be made between RNN, LSTM and GRU for the system observation and system transition function respectively, and then the best RNN model will be selected for one of these functions. With the trained models, an online machine state estimation is realized by the indirect sensor measurement. Building on the idea of generative models (Graves, 2013), an arbitrary time step ahead prediction can be updated in real-time, as well as predict Remaining Useful Life (RUL).

4.3.1 System Transition Function

For any dynamic mechanical systems, the systems state may have temporal dependencies. More specifically, the current system state is affected by its past system states. The System Transition function can be generically written as follows:

$y_{t+1} \sim p(y_{t+1}|y^t; \theta)$, where $y^t = (y_1, y_2, \ldots, y_t)$

Here, y^t is the direct measurement of the system state. When the system transition function is modeled as an RNN, the function can be rewritten as:

$y_{t+1} \sim p(y_{t+1}|y^t, h_t^{trans}; \theta)$

where h_t^{trans} is the hidden nodes for the system transition function. The current hidden node h_t^{trans} is a function of the past hidden node h_{t-1}^{trans} and current input yt. so the via the hidden node, current hidden node bear all the current and pass information to the inputs.

$h_t^{trans} = f(h_{t-1}^{trans}, y_t; \theta)$

so that $h_t^{trans} = g_t^{trans}(y_1, y_2, \ldots, y_t; \theta)$, which the function g_t^{trans} takes the whole pass sequence (y_1, y_2, …, y_t), so that the system state at time t+1 can be modeled given all the pass system state information. Note that the variable here is a continuous value and its distribution is assumed as unimodal.

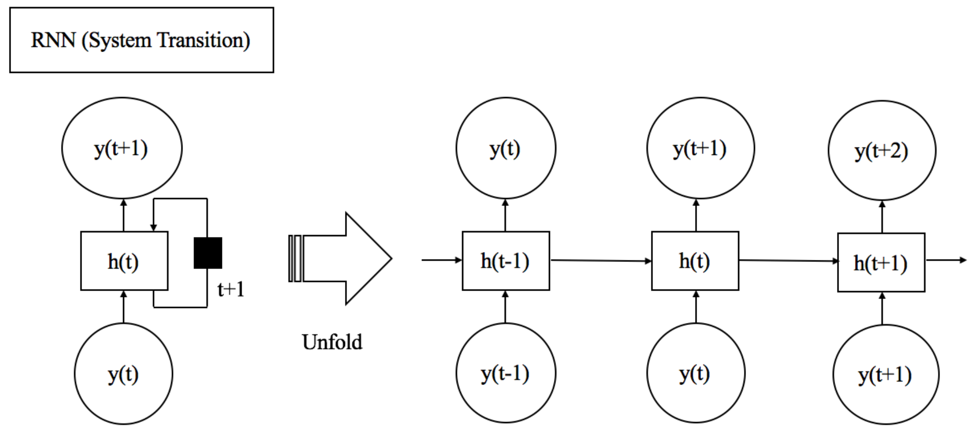


Figure 4.3 RNN for System Transition

4.3.2 System Observation Function

The direct measurement of the real-system state is often not practically feasible to be collected in production environments, such as actual tool wear amounts on cutting tool inserts. The system state can be inferred from indirect measurements, such as one or more sensor signals. Also, in most cases, there is not a good function that can build a relationship between the direct measurement of the system state and its associated indirect measurement. Also any developed relationship must consider the temporal effects. Specifically, with availability of any past direct measurements of the system state, the determination of current system state should refer to all past indirect measurements. Therefore the system observation function can be written as:

$y_t \sim p(y_t|x_t, \theta)$, where $x^t = (x_1, x_2, \ldots, x_t)$

here, x_t is the indirect measurement at time step t; and y_t is the direct measurement of the system state.

With RNN, the system observation function can be rewritten as:

$y_t \sim \ p(y_t|x_t, h_t^{obs}, \theta)$

where h_t^{obs} is the hidden nodes for the observation function with superscript letter 'obs' denoting it.

$h_t^{obs} = f(h_{t-1}^{obs}, x_t; \theta)$

so that $h_t^{obs} = g_t^{obs}(x_1, x_2, \ldots, x_t; \theta)$,

which the function g_t^{obs} takes the whole pass sequence $(x_1, x_2, \ldots, x_t)$

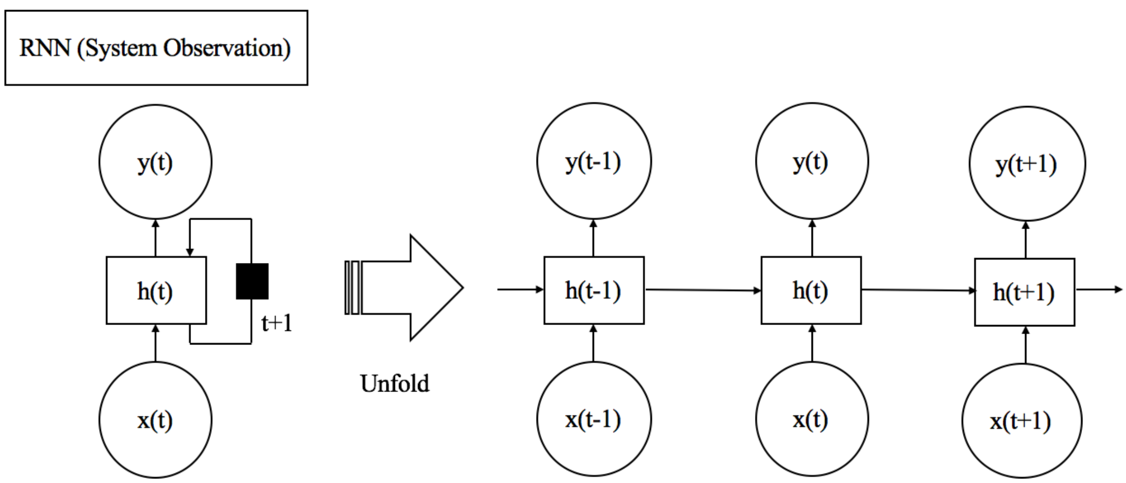


Figure 4.4 RNN for System Observation

Here, x_t is the indirect measurement, the sensor signal, at time step t ; y_t is the direct measurement, the machine state, at time step t.

4.3.3 RNN Cells

Elman RNN

The Elman RNN Cell reserve the past information via the hidden nodes h_t, so that the input information xt from past time step can be referred by the current and future output y_t by the hidden nodes h_t.

$h_t = \sigma_h(W_h x_t + U_h h_{t-1} + b_h)$

$y_t = \sigma_y(W_y h_t + b_y)$

However, for the general RNN, owing to the gradient vanishing effect during training, the long term dependencies may not get capture properly. So that for design of the RNN cells, in order to avoid the gradient vanishing issue, Long Short Term Memory (LSTM) (Hochreiter, et al., 1997), or Gated Recurrent Units (GRU) (Chung, et al., 2014) will be used as the RNN model. LSTM and GRU can tackle the vanishing gradient issue so that it is more strength at capture the long-term dependencies. Also, GRUs has same mechanism as LSTM, but less parameter need to be tuned.

Long Short Term Memory (LSTM) and Gated Recurrent Units (GRU)

For design of the RNN cells, in order to avoid the gradient vanishing issue, Long Short Term Memory (LSTM) (Hochreiter, et al., 1997), or Gated Recurrent Units (GRUs) (Chung, et al., 2014) will be used as the RNN model. LSTM can tackle the vanishing gradient issue so that it is more strength at capture the long-term dependencies. Also, GRUs has same mechanism as LSTM, but less parameter to tune, so need less training data. Based on their similar performance (Goodfellow, et al., 2016), Gated Recurrent Units (GRUs) need less data to get trained, so that GRUs may be preferred here.

There will be a comparison between the existing methods, which are capable of all the features in the prognosis, and diagnosis. The Mean Square Error (MSE) will be applied as a criterion for the different model. Also, the number of whole lives of tool will also be counted as a criterion.

Long Short Term Memory (LSTM) Cell

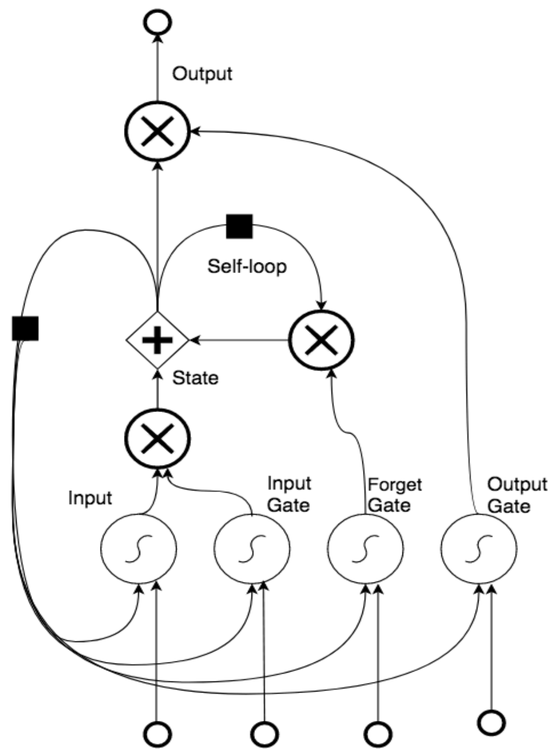


Figure 4.5 LSTM cell block diagram

LSTM cell is advanced version of the RNN cell, which are excellent at remembering values for either long or short time periods. It consists of repeating memory cell units, each of which consists of interacting elements – the cell state, and three gates that protect the cell state

- input gate, a forget gate and an output gate (see Figure 4.5). The gates control the flow of information by updating the cell state, which transfers from one cell to the next. The input gate controls which values are updated by input activations and the output gate controls the flow of activations into the rest of the chain. The forget gate implements self-loops, meaning that the gradient from previous cells can flow from the long temporal past, thereby adaptively forgetting or including the distant memory. Additionally, with gated self-loops controlled by the weights, the time scale of integration can also be controlled by the input gates (Goodfellow, et al. 2016). At each time step t, the hidden state, h_t is updated by x_t, the current data at the time step, the hidden state at previous time h_{t-1}, input gate i, forget gate f, output gate o and the internal memory cell c.

Input gates:

$i = \sigma_i(x_t U^i + h_{t-1} W^i)$

Forget gates:

$f = \sigma_f(x_t U^f + h_{t-1} W^f)$

Output gates:

$o = \sigma_o(x_t U_o + h_{t-1} W_o)$

$g = \tanh(x_t U_g + h_{t-1} W_g)$, where g is candidate hidden state which is used to calculate the real hidden state.

$c_t = c_{t-1} \cdot f + g \cdot i$

$h_t = \tanh(c_t) \cdot o$

Here, W and U are the weights matrices

As seen from the above equations, all the gates can be determined by the input x_t and hidden nodes h_{t-1}, with weights that can be tuned during the training process.

Finally, the output calculation, the output y_t is written as

$y_t = \sigma_y(h_t W^y + b_y)$

Where, σ is activation function, x_t is the input at time step t; and y_t is the output at time step t. and b is bias.

The input gate, output gate and forget gate like "throttle" controlling the amplitude of the input, output, and self-loop, respectively.

Gated Recurrent Units (GRU) Cell

GRU cell is a special kind of gating mechanism for the LSTM cell of an RNN. It has fewer parameters and simpler structure than the LSTM cell. A simplified version of LSTM cell, the GRU cell was proposed by (Cho, et al., 2014), which has less gates and thus less weights to be tuned. There are only two kinds of gates for GRU: reset gate, r, and update gate, z. The reset gate determines how to combine new input with that of the previous memory in its cells. On the other hand, the update gate decides the portion of its previous memory does it retain or throw out. The equations for the GRU cell are similar to the LSTM:

$z = \sigma_z(x_t U_z + s_{t-1} W_z)$

$r = \sigma_r(x_t U_r + s_{t-1} W_r)$

$h = \tanh(x_t U_h + s_{t-1} W_h)$

$s_t = (1-z)*h + z*s_{t-1}$

Since there are only 2 gates, the important difference between a GRU versus an LSTM is that GRU does not possess an internal memory (c_t). The reset gate is directly applied to the previous hidden state (h_{t-1}). Thus it can train faster since there are only few parameters in the U and W matrices and hence maybe useful when not much training data is available.

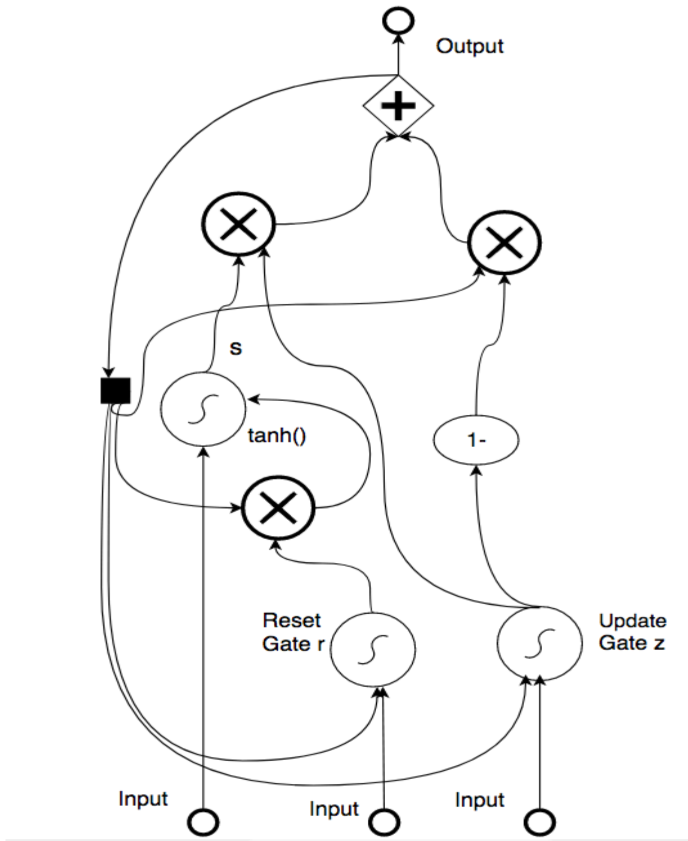


Figure 4.6 GRU cell block diagram

4.3.4 Diagnosis and Prognosis with RNN

After the three kinds of RNN model of System Transition and System Observation is trained, they can be further applied for diagnosis and prognosis of the system state, by inputting the indirect measurement in real time. Our RNN state space model method is to realize online diagnosis, by taking advantage of the indirect measurement from past and current states: $p(y_t, \theta|x_t)$, where $y^t = (y_1, \ldots, y_t)$. With the generative recurrent neural networks (Graves, 2013), the output nodes and the hidden nodes is produced by the RNN of the System transition, and then input into the input node and hidden node into the model itself. Repeat the process, and number of the repetition is number of time step horizon in the prediction. Lastly, for Remaining Useful Life (RUL) prediction, the criterion of the degradation is set to certain value, for example 0.3 mm for the tool wear. By apply the generative system transition RNN, until the predicted degradation variable reaches the criterion, the number of time step it passes through can be the predicted RUL, which in our case, will be the number of time steps it will take to reach the wear criteria.

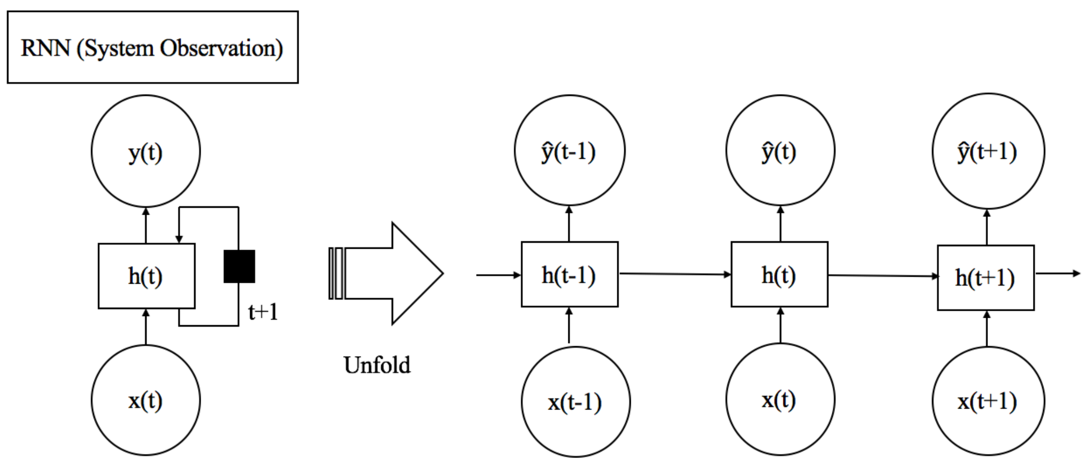


Figure 4.7 RNN System Observation for online estimation

The above steps will be investigated through developing both the modeling framework and testing with experimental conditions to test the validity of the approach. Appropriate training strategies will be conducted within the context of monitoring tool wear processes.

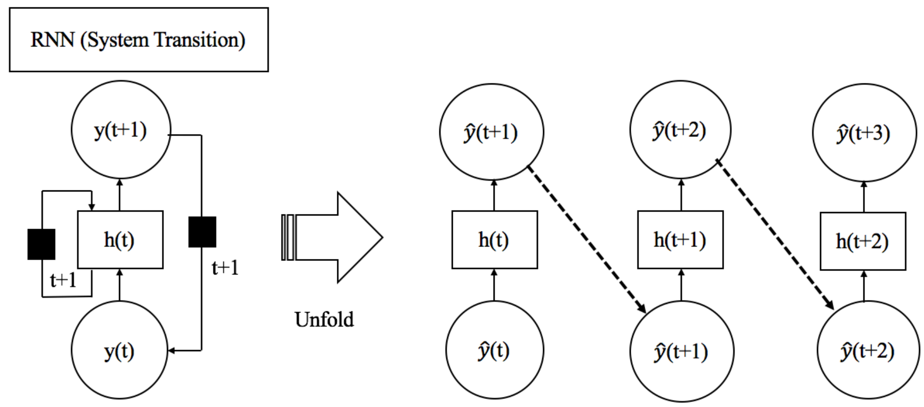


Figure 4.8 RNN System Transition for Prediction

4.4 Experiment Design

In order to evaluate our RNN state space model method, experiments were conducted to obtain actual tool insert wear measurements using the HAAS VF2 under dry milling machining conditions. A two-flute indexable milling tool with diameter 12.7 mm, Sandvik Coromill 390 (RA390-013O13-07L) installed with one uncoated insert, Sandvik Coromill 390 Insert (390R-070204M-PM) was used in the cutting process. The workpiece material utilized is Steel 4142 (cold rolled, 40-45HRC,). Each single cutting pass was 37.0 mm long, with the spindle speed at 500rpm, leading to a surface speed of 19.81m/min and chip load at cutting to be 0.05mm. The radial depth of cut is 6.5 mm and axial depth at 0.4 mm. Feed rate is 50 mm/min. A 3-axis vibration sensor (Kistler accelerometers 8762A10) was installed on the fixture attached to the workpiece. Sensor data was sampled at 1652Khz and its RMS signal formed the online indirect measurement. After every single pass of cut along the length of the workpiece, offline direct measurements were taken to measure actual tool wear. The machining process was repeated for two experimental runs. The RMS of the vibrations signals for the two repeated runs are shown in Figure 4.10 . The vibration signal curves do not monotonically increase with the tool wear. The signal curves show that it is critical to account for the time dependency for any diagnosis and prognosis of cutting tool wear.

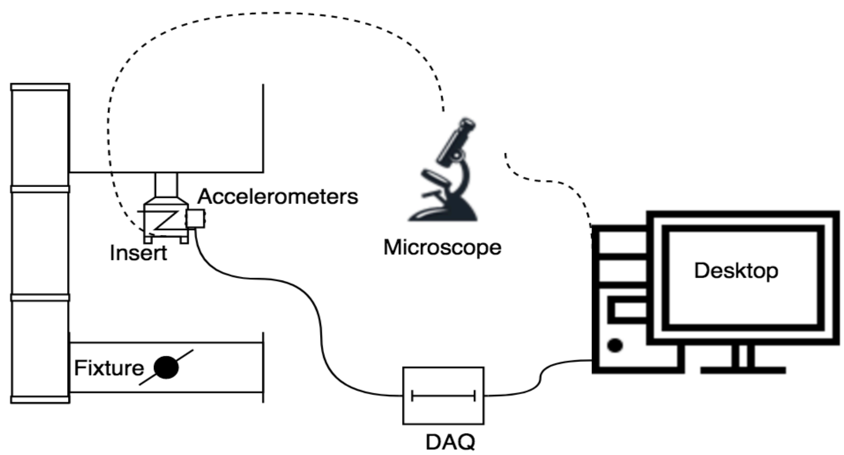


Figure 4.9 Schematic diagram of experimental setup

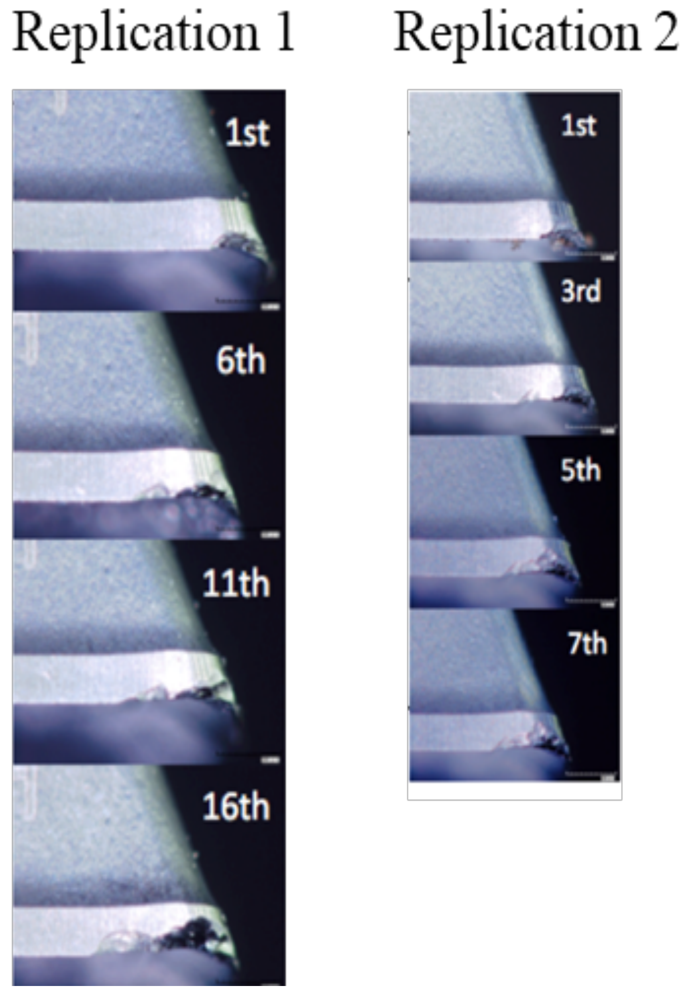


Figure 4.10 Tool Wear Images

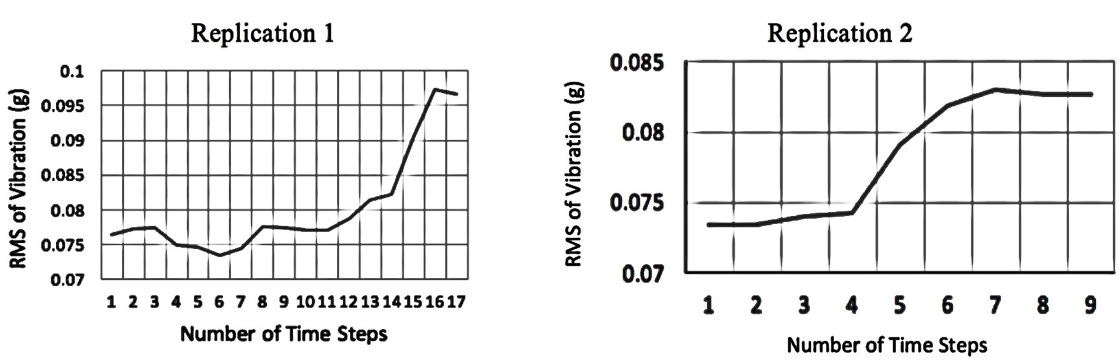


Figure 4.11 Vibration Index for two replications

The plotting of the RMS of vibration of replication 1 and replication 2 shows that the vibration signal curves do not monotonically increase with the tool wear. So that if the relationship between RMS of the vibration and tool wear is constructed without the factor of time dependency considered, the model may not be accurate.

4.5 Model Training

Early stopping is applied to loop through the training set a number of iteration for the three kinds of RNN model for the system transition and observation functions, respectively. The MSE of validation set get evaluated for each iteration. The whole dataset get divided into three parts, the first one is the training set (7 sequences), the second one is the validation set (6 sequences), and the last one is the test set (2 sequences).

For each time step, the weights in the RNN model get updated by the BPTT. For each sequences, the values of hidden nodes is reset to zeroes. The RNN for System Transition and System Observation are trained independently. The activation function for Elman RNN cell is tanh().

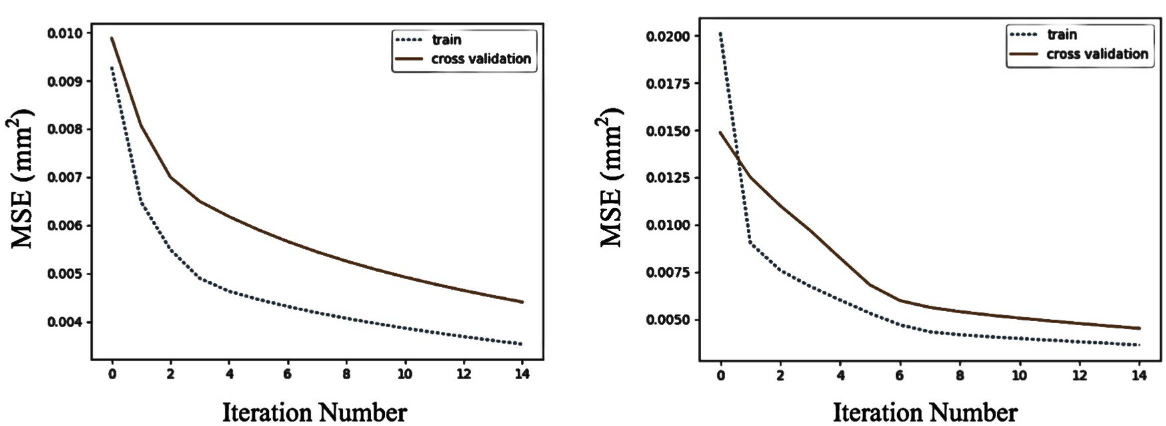


Figure 4.12 Average (Avg.) MSE on Training set and Validation set

As shown in the figure, the reduction rate of the average MSE decrease with the number of the iterations. The iteration number is set as 14, given the computation capacity and average MSE is good enough for estimation. Comparing of MSE on System Transition Function, the results show that average MSE of the LSTM is lower than the other two, RNN and GRU. And the Standard deviation (Std.) also show the stability of the LSTM is have relative good performance. It is also the case for the MSE on System Observation Function. See Table 4.1 .

Table 4.1 Compare RNN, LSTM, and GRU for System Transition and Observation

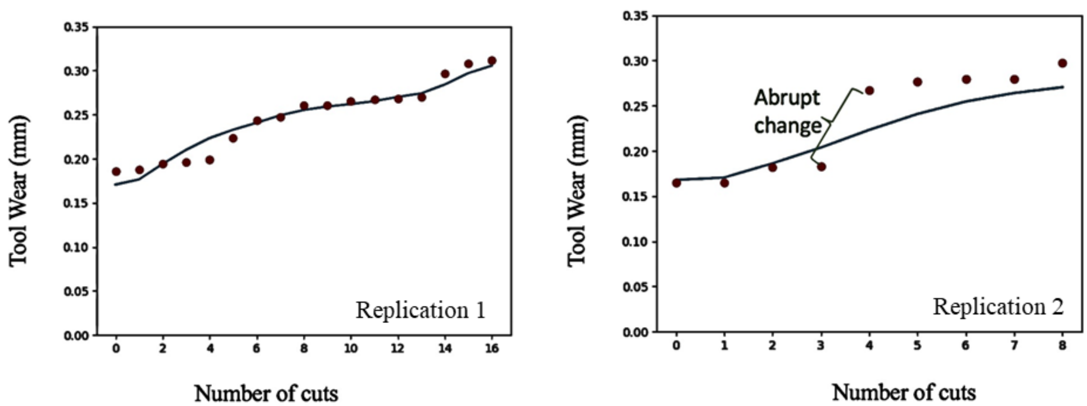
RNN Cell Type	**System Transition**		**System Observation**	
	Avg. of MSE (mm^2)	**Std. of MSE (mm^2)**	**Avg. of MSE (mm^2)**	**Std. of MSE (mm^2)**
Elman RNN	0.0027180	0.0007639	0.0038187	0.0016212
LSTM	0.0015628	0.0007708	0.0037855	0.0003046
GRU	0.0030236	0.0010090	0.0038729	0.0010628

The MSE of applied trained LSTM of the System Transition function is 0.0020405mm, and the the MSE of applied trained LSTM of the System Observation function is 0.0038631mm. (0.002804779, hidden nodes = 20, iterations = 30). Because there are more parameters in the LSTM cell than the other two, Elman RNN and GRU, it could more flexible to fit the data and more adaptive to construct the complexity of the time dependencies, although more computation resource could be applied to update the parameters in the model.

4.6 Results

Online Diagnosis

By applying the LSTM for system observation, given the indirect measurement of the current and past time step, the RMS of the vibration signal, the tool wear at current time step get estimated.

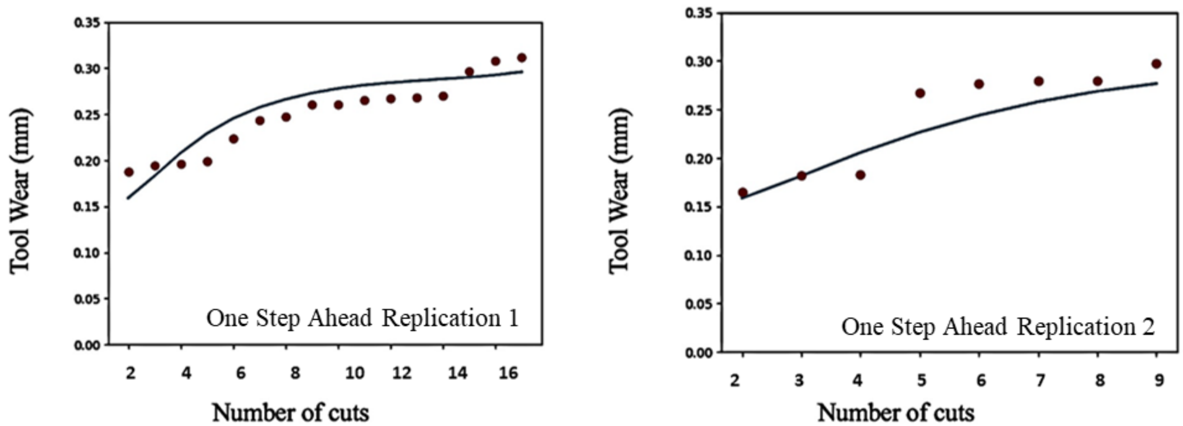


The dots indicate the actual measured tool wear; the solid lines are the estimated tool wear.

Figure 4.13 Online Tool Wear Diagnosis with indirect measurement.

Online One-Step Ahead Prediction

Input the estimated tool wear at current time step into the input node of the LSTM for System transition function. It assumes that the tool wear at time step 0 equal to 0.0 mm ($y_{t=0}$ =0.0 mm). The predicted tool wear at next time step t+1 is the value from the output node of the RNN for System transition, given the hidden node and estimated tool wear at current time step t, which means the tool wear at next time step get predicted given the current and past tool wear estimation by the system observation function.

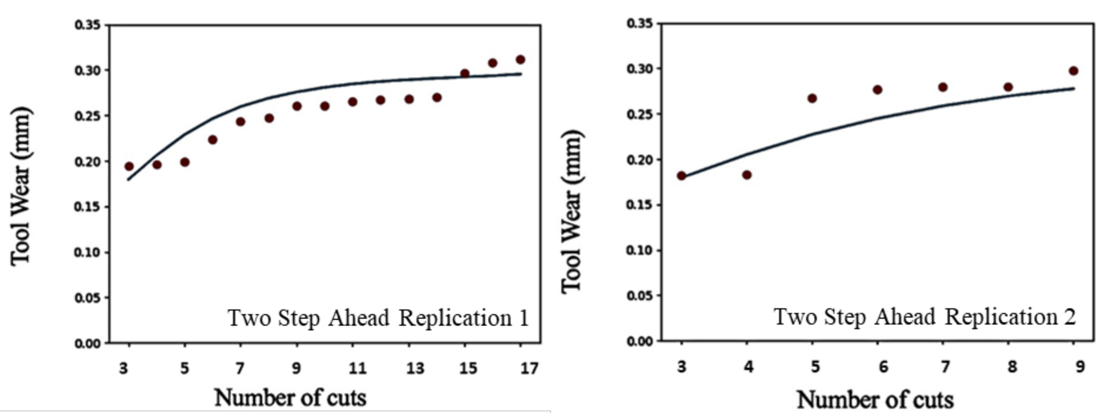


The dots indicate the actual tool wear; the solid lines are the predicted tool wear.

Figure 4.14 Online Tool Wear One-Step Ahead Prediction with indirect measurement.

Online Two-Step Ahead Prediction

Similar to online one-step ahead prediction, the tool wear at two time step ahead get predicted given the next, current and past tool wear estimation by the system observation function and system transition function.



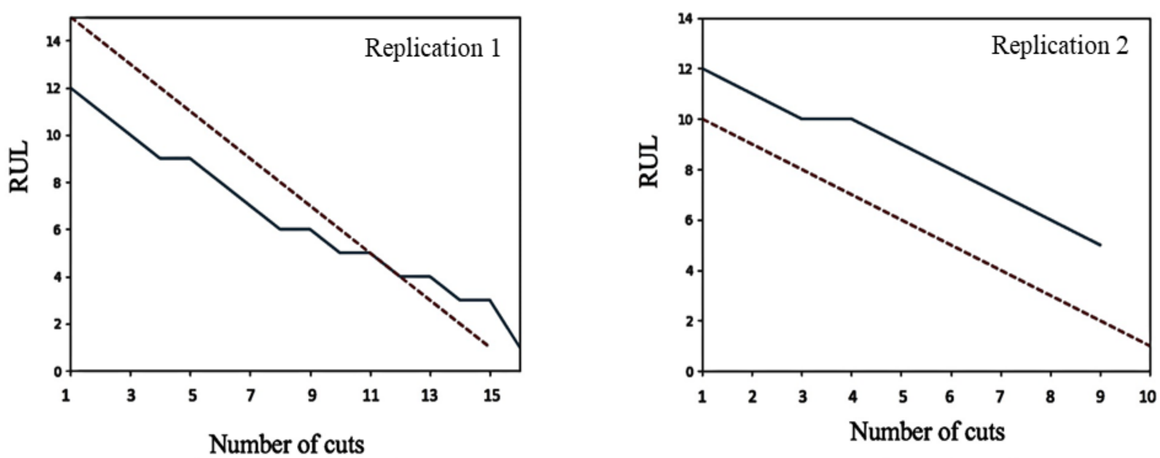
The red dot indicate the actual tool wear; the solid lines are the predicted tool wear

Figure 4.15 Online Tool Wear Two-Step Ahead Prediction with indirect measurement.

The RNN state space model method is robust to certain abrupt change of the system state, the tool wear. Arbitrary number of time step ahead prediction can be realized by our method. Similar to online two-step ahead prediction, the tool wear at three or multiple time step ahead get predicted given the two step ahead or two step ahead to multiple time minus one step ahead, current and past tool wear estimation by the system observation function and system transition function.

Remaining Useful Life (RUL) Prediction

At each time step, by looping the RNN for system transition, which means input the values of hidden and output nodes of RNN for system transition of the last time step into the hidden nodes and input nodes of the next time step, and repeat the process, until the value of the output nodes reach the criterion for the tool wear, 0.3 mm.



The dashed lines indicate the actual RUL; the solid lines are the predicted RUL

Figure 4.16 Online RUL Prediction.

Both of the RUL predictions start at 12 cuts. For the replication 1, there are several time step show the RUL remain with the proceeding number of cuts. By contrast, the replication 2, there is few time step show the RUL remains with the proceeding number of cuts.

The plotting illustrates that the RUL is able to get updated with the online indirect measurement, the RMS of the vibration signal.

Discussion

LSTM and GRU can capture the long term time dependencies as model for the system transition and observation functions, compared to the simple RNN, Elman RNN. By taking advantage of the trained RNN system observation model, the tool wear can be estimated according to the current indirect measurement, and past indirect measurement. Also, for tool wear prediction, one time step ahead tool wear prediction can be realized by input into RNN system transition function the estimated current and past tool wear by the RNN system observation function. For two step ahead tool wear prediction, the one step ahead tool wear prediction and all the estimated tool wear by the RNN system observation function act as inputs into the RNN system transition model, thus, the two step ahead tool wear prediction get realize. Similarly, the multiple time step ahead prediction can get estimated. Thank to the property of the generative RNN, our model can predict arbitrary number of time step ahead tool wear degradation in the future.

Two replications of tool were tested to demonstrate the how our RNN state space model model work. However, for the replication 1 and replication 2, despite their same cutting conditions, the tool wear in diagnosis and prognosis and RUL are different with each other, and the value get changed and updated on the fly, which confirms that the necessary of the diagnosis and prognosis methods adaptive online. Furthermore, the RUL prediction also get prediction the process until the tool wear reach the criterion. The process is that by iteratively inputting certain time step ahead tool wear predictions by the RNN system transition function and all the tool wear estimation by the RNN system observation function into the RNN system transition function to get next time step prediction. The criterion of the tool wear here is set as 0.3 mm.

Therefore, the tool wear diagnosis and prognosis can be estimated online via the indirect measurement. Here, the indirect measurement is the vibration signal to measure the vibration generated during the cutting process. As a black box methods, the Neural Network model fail to explicit show the formula of the system observation and system transition function as the Particle Learning method proposed by Zhang, et al, (2017a) and any analytic models, increase the computation cost, and the challenge to decide the complexity of the network, number of layers and number of hidden nodes, given the reward is the flexibility of the fitting any form of relationship. Additionally, the system transition function also need to be trained before applying into the prognosis process, unlike the Particle Learning method (Zhang, et al,. 2017a) that the system transition function will be determined online given the prior parameter value distribution. The method can have desirable diagnosis and prognosis result if the tool wear gradually occurs, and it can tolerate certain abrupt tool wear change. For different material and different machining conditions, the new data should be collected to train the bran new model.

It can make prediction which is superior than the feed forward neural network. Second, our RNN state space model is better than the time-delay neural network, which can only predict limited time step ahead machine state. Thirdly, for the RNN with only time-delayed form can capture the time dependencies which is superior than the time delayed neural network, but it cannot realize the arbitrary number time step ahead prediction, not to mention the RUL prediction. Fourthly, some papers use the RUL as target variable and sensor signal to model the RNN. However, the RUL does not reflect any physical characterization, so that there is no direct physical connection between the sensor signal and RUL. Also, these method cannot realize machine state diagnosis and prognosis. Compared to these methods, model the system transition function and system observation function with the RNN render the online estimation considering the time dependencies become natural. And taking the schema of the generative RNN, arbitrary time step ahead prediction get realized by combining the system transition function and system observation function modeled by RNNs. Further, by applying advanced form RNN cell, the LSTM, GRU, the long term time dependencies can get captured.

The neural network model is more flexible than the analytic model, which do not need physical formula within the model. Our method, models the system transition and system observation with long term time dependency RNN, LSTM and GRU, through which we realize estimate the system state and predict the arbitrary future horizon of the system state, by given the current and past indirect measurement. The application of such method can be extended into any physical machine, such as bearings, gear, engine, ball screw, grind wheel, nano-machining, bio-reactor, in which only indirect measurements are accessible in real time and need the estimate the system state and predict the system state. Additionally, for the degradation problem, with set up criterion, the RUL can be predicted via our method.

There several possible ways to improve the accuracy of the prediction. First one is to loop through more number of iterations of the training set to train the model; second one is to use the more hidden nodes or more layers of hidden nodes within the neural network model.

Possibly, the dropout technique can be used for the multilayer RNN; third one is to collect more data to train the model. With more data to train the model, and more data for the validation, the model get more robust and have more stable performance in diagnosis and prognosis. Moreover, one can also use the n-fold or leave-one-out cross validation to get stable evaluation of the MSE of both the RNN for system transition and RNN for system observation. Incorporating more time domain and frequency domain feature of the signal into the inputs of the model, so that feature engineering work should be implemented.

4.7 Conclusion

In conclusion, RNN models the system transition and observation functions, so that both of functions realized the online diagnosis and prognosis by inputting the indirect measurement. The LSTM and GRU, which can capture the long term dependencies, take place the RNN to model those two functions. After the two RNN model trained with the indirect measurement paired with direct measurement, and direct measurement paired with its one time step delay, respectively. The validation set was used to evaluate the estimation accuracy by the MSE. Two replications of sequence of the indirect measurement, as test set, is used by our model, to achieve the online estimation, one and two time step prediction, and RUL prediction. The results were compared with the real tool wear and the one at one and two time step ahead. And RUL prediction was also compared with the real RUL. The validation set show the acceptable MSE. The two replication test sequences were as the demonstration to the diagnosis and prognosis. After compared with the real values, the results show the good performance and strong possibility for practical usage. First, as model-free methods, the Neural Network requires no analytic formula in modeling. Secondly, the LSTM cell modeled system transition and system observation functions, which capture the long term temporal dependencies, has better performance than the simple RNN to model them. The arbitrary horizon of prediction can be realized taking advantage of the modeled system transition function and the generative model, and thus the RUL prediction. It is very possible to extend our methods for diagnosis and prognosis of system state by indirect measurement for other physical entities, such as gear, bearings, ball screw, surface roughness, etc.

Chapter 5 Conclusions, Contributions, and Future Work

This chapter summarizes the research work in this dissertation, and also underlines the distinctions of this work and possible directions of future work.

5.1 Summary of the Work

In this study, two new online tool wear diagnosis and prognosis methods via the indirect measurement is developed. The first one is the method development using Particle Learning (PL) and implementation of the approach to estimate tool wear and RUL of uncoated tooling inserts for steel based alloys by online measurement of vibration signals. The second approach utilizes Recurrent Neural Networks (RNN) and its advanced form, gated RNN, with system model, which is also applied to the cutting tool wear diagnosis and prognosis.

For the first method, it needs to construct the a relationship, named system observation function, between the indirect measurement and direct measurement. Then, apply the PL realize the updating the tool wear online estimation, and forecast. The updating process get realized given the online tool wear estimation by the system observation function. The prognosis, including multi-step ahead prediction, and RUL prediction, get realized by feeding the one step ahead prediction into the PL model iteratively. The diagnosis and prognosis get updated when the new indirect measurement get collected at on-going time step. For the first method, the PL, as a semi-heuristic algorithm, make the calculation process become possible and realizable.

The second method, the RNN and gated RNN with system model, is to improve tool wear estimation and RUL prediction by utilizing an LSTM based RNN framework, which makes more effective use of model parameters to train system observation and transition models of tool wear processes. We utilize minimally intrusive vibration sensors to collect the signal online and help train the RNN model. There are two RNNs to model the system observation function, and system transition function, a.k.a. system model, respectively. RNN modeled system observation function formulate the relationship between the indirect measurements and direct measurement and RNN help it consider the time dependencies. Similarly, RNN modeled system transition function formulate the relationship between the direct measurement at adjacent time steps, and also the RNN help it consider the time dependencies. With trained RNN for the system model, the tool wear online diagnosis and prognosis can get estimated and updated via the indirect measurement. The RNN modeled system observation function realized the online diagnosis, and RNN modeled system transition using the generative schema can realize multi-step ahead prediction and RUL prediction. Gated RNN, acted as RNN, try to capture the long temporal dependencies between the indirect measurement and direct measurement itself.

5.1.1 Overall Limitations

Both analytical model and model-free methods are proposed in this dissertation, and their attributes, including strengths and shortcomings are discussed here. Taylor's Equation, as traditional analytic model to prediction RUL for tool wear, can only use the direct measurement to predict RUL. The values of parameters need to be determined by experiment, and cannot get updated during the cutting process. PHM, PCM, Cox Model integrate the indirect measurement into the model, and determine the hazard rate by it. The models can predict RUL and also consider the historical indirect measurements, which means the temporal dependencies; however, it cannot give out the tool wear online estimation and forecasting. Bayesian Updating with Taylor Equation and Bayesian Updating with Growth Curves can update the predefined analytic tool wear formula; however, they cannot estimate the real tool wear. HMM, HSMM, VD-HMM, are capable of using the indirect measurement to realize the online diagnosis and prognosis; however, because it assumes the process are Markov process, so that only one pass time step information will be considered for diagnosis and prognosis, and the long-term temporal dependencies of the indirect measurement or direct measurement itself cannot realized.

The current research on Kalman Filter have not figured out RUL prediction by Kalman filter and Particle filter. The analytic models are needed for KF and PF. Similar to HMM, KF and PF only consider the one pass time step information. PL can update the estimation of the tool wear and the parameters in the analytic model, and also with the aid of sufficient statistics, the long temporal past information could be considered in the updating process for the system transition function; however, for the system observation function, there is only linear regression is applied and thus not time dependencies get considered here. The Neural Network methods do not need analytic formula for modeling the tool wear. FFNN as static NN, does not consider any time dependencies. General RNN can model the indirect measurement and direct measurement, and consider the time dependencies. But for the forecasting, the increased horizon need the increasing the complexity of the RNN, more time delayed inputs will be needed for more horizon of forecasting, so that the arbitrary forecasting is hard to realize. Modeling the system model with RNN can capture the temporal dependencies for both system observation and transition functions. Suffer from the deficiency of the simple RNN, the long temporal time dependencies cannot be captured properly. With the help of the gated RNN, the problem get solved. The system model can realize the diagnosis and prognosis via online indirect measurement.

Table 5.1 Comparison of the methods for tool wear modeling

Methods/ Features	Indirect Meas.	Diag.	Limited horizon predict	Arbitrary horizon predict	RUL predict	Model-free	Limited time depend	Long time depend
Taylor's Eqn.	✗	✗	✗	✗	✓	✗	✗	✗
PHM, PCM, Cox Model	✓	✗	✗	✗	✓	✗	✓	✓
Bayesian Updating w/ Taylor Eqn.	✗	✗	✗	✗	✓	✗	✓	✓
Bayesian Updating w/ Growth Curves	✓	✗	✗	✗	✓	✗	✓	✓
HMM, HSMM, VD-HMM.	✓	✓	✓	✓	✓	✓	✓	✗
Kalman Filter (KF)	✓	✓	✓	✓	✗	✗	✓	✗
Particle Filter (PF)	✓	✓	✓	✓	✓	✗	✓	✗
Particle Learning (PL)	✓	✓	✓	✓	✓	✗	✓	✓ / ✗
FFNN	✓	✓	✗	✗	✗	✓	✗	✗
General RNN	✓	✓	✓	✗	✗	✓	✓	✗
RNN w/ syst. model	✓	✓	✓	✓	✓	✓	✓	✗
Gated RNN w/ syst. model	✓	✓	✓	✓	✓	✓	✓	✓

The proposed PL and gated RNN with system model have more advantages than others in tool wear modeling. Although theoretically, our approaches are better than others in several aspects, yet, they are not perfect during the engineering application. The limitation of our approach is that further refinement in models may have to be conducted when customer uses the same cutting tool for various materials. In such hybrid use cases, the development model approach may not be sufficient and needs to be augmented with more training data. The other limitation is that we assume that tool wear occurs gradually and our model cannot account for sudden fracture of cutting tool surfaces. However there are other approaches to detect sudden breakage of cutting tool surfaces, particularly using real-time power consumption monitoring devices. Many such devices from Sandvik-Coromant have been recently installed in high-throughput production. Our approach is better suited to detect gradual cutting tool wear.

5.2 Uniqueness and Contributions of the Work

The following highlights distinguishes this work from other work conducted by the community.

- The proposed method can realize the online diagnosis and prognosis via indirect measurement and does not completely rely on direct measurements or intrusive indirect measurements. Indirect and non-intrusive measurements minimizes the need to build in intrusive sensors within a production environment. Often time, such intrusive measurements are practically impossible. This is either due to harsh environments or physical obstruction to the placement of the sensor. Moreover, the online response or monitoring proposed more challenge to collect the direct measurement. It is often the case that the indirect measurement from sensor signal can be collected in real time with lower cost and a more convenient way;

• Proposed the use of a first order autoregressive relation as the system transition function for the tool wear, which reduce time complexity and capture the non-constant tool wear rate. First order autoregressive as system transition function can be used to fit the the first derivative and the second derivative of a curve. Also with updating parameters in the First order autoregressive as system transition function, it is widely capable of fitting to any form of curve. With strong capacity of curve fitting, the time complexity to updating the parameters and tool wear estimation is much lower than the nonlinear function as system transition function;

• The PL, as a semi-heuristic algorithm, realizes calculation for an intractable formula. Formulating a Bayesian updating process is easy for the practitioner; however, to solve them is a big challenge. It is often the case it is difficult to formulate the closed form solution or even if the closed form solutions get solved, it would require significant more time to solve them. Therefore, with the aid of the semi-heuristic algorithm, the situation can be am meliorated to get the approximated analytical solution;

• RNNs capture time dependencies and make prediction over conventional NNs and TDNN. The TDNN cannot increase the forecast horizon without increase the complexity of the network, aka the input nodes and the hidden nodes added correspondingly. RNN can model time series data naturally; This work applied the gated RNN to capture the long term temporal dependencies of the indirect measurement, and direct measurement across time steps. The gated RNN, the LSTM and GRU, are designed to conquer the gradient vanishing and explosion in the BPTT, which is the denunciation of simple RNN;

• The two proposed methods model the tool wear, the direct measurement, and sensor signal, the indirect measurement as system model, the system transition function and system observation function. With these functions in the system model, it become possible to apply generative model to realize the prognosis by system transition function, including the multi-step ahead prediction and RUL prediction. So that the method can make arbitrary horizon of prediction horizon and RUL prediction, which is incapable within time-delayed (Recurrent) Neural Networks;

• The solution approaches are easily extendable to assess other degrading physical entity. The PL with first order autoregressive function and the gated RNN with system model are capable of large applications. Especially, they are applicable to the degradation process of a time series based physical environment. Furthermore, the condition based maintenance is core of industry 4.0 and intelligent manufacturing, the proposed two methods is applicable and expected to render significant contribution to this era.

5.3 Future Work

There are three possible directions to extend the current research conducted in this dissertation. New advancements made in deep learning network which is essentially multi-layered neural networks have opened up the possibility of training such models to cater to the highly dynamic manufacturing environment. Coupled with advancements in manufacturing hardware and compute technologies, smarter machines which self-adapt and adjust based on observed real-time scenario may become a reality.

5.3.1 Broaden the Applicability of the Methods

The basis of this work can be broadened beyond just cutting tool wear diagnosis and prognosis. Extend the methods into the application of the different materials of the workpiece with different hardness. It is possible to include the hardness as input in our model. As a consequence, experiment data should be collected to certify the modified method. Also, future studies can investigate applicability of the method to harder to machine metal alloys and coated insert tool types. In addition, different machining conditions and suitability of other sensor signals, such as spindle power can increase the practical relevance and robustness of the solution.

Specifically, different workpiece materials should be considered. The properties of the materials, including the hardness, toughness, surface roughness, yield strength. For the steel, the percentage of the carbon can also considered. Moreover, the shape of the workpiece, the stiffness of workpiece, fixture, the machine tool, the tool holder, the cutting tool etc. can affect the vibration signal. Therefore, the data collected for a specific machine is only useful to that machine tool. Moreover, different cutting tool, coated or uncoated, different material of the tool, the dimension, indexable or not, number of cutter, etc. can affect the vibration signal. Therefore, the data collected for specific combination of the tool, workpiece, and machine tool are only valid for that one. Also, the room and local temperature, use coolant or not also be factors. It needs to notice whether the machining condition change when the trained model use to make diagnosis and prognosis.

Furthermore, precision of the machine tools to control the depth of cut, feed rate, spindle speed, etc. Especially, the depth of cut could have large impact on the vibration signal. Therefore, Design of Experiment (DoE) may need to be conducted to check the sensitivities of each factor for the tool wear process. Further, it is also possible to consider some factors to be incorporated into the model. For example, the machine conditions can be as input nodes in the RNN model. Lastly, more sensor, and more features of the signal can be extracted for improving the accuracy of the methods to counter the inaccuracy and uncertainty cause by the complex impacting factors. These sensor could include power sensor, vibration sensor, acoustic emission sensor, temperature sensor, etc. More feature can include the average, standard deviation, Kurtosis, Standard variation, Skewness, Kurtosis, Maximum peak value, Absolute mean value, RMS, Crest factor, Shape factor, Impulse factor, Clearance factor, Kullback–Leibler divergence, Entropy, complexity, correlation dimension, etc.

5.3.2 Adapt the Model to New Machining Scenarios

It would be desirable if the model can cater or is modified to situations that arise in practical implementation scenarios:

Another real scenario that it is convenient in the production surrounding that only measurement the direct measurement of the tool wear after certain number of cuts, which is refer by the specification from the manufacturer. Because in the production setting, the only good chance to measure the tool wear directly is when the tool is expected to be worn out. It would solve the the problem that the material properties could change slightly and slowly across different batches, and therefore, the values of the parameters in the model should be updated accordingly. Specifically, for our RNN model, The RNN model is trained first in the current setting, and diagnosis and prognosis get conducted via the online indirect measurement. The values of the parameters in the model is never updated after training. Thanks to the flexibility of the RNN methods, the modification of the model for the new scenarios can be done for RNN methods. The conceived way to realize the updating in the new scenario is possibly deployed as following steps:

First, the both direct and indirect measurement collect for each time step for several number of whole time series of the tool wear processes, which the number of time step is refer to the specification from manufacturer. Then, the training step is conducted as before.

Then, only measure the tool wear at certain number of time step which specified by manufacturer. Then the possible model as solution is : Combine Backward RNN and Forward RNN to estimate the tool wear across the whole wearing process of the cutting tool. For the forward RNN, it starts with $y0 = 0$, and generate the tool wear estimation by generative model in forward direction. On another hand, for the backward RNN, it start with yT, where T is the suggested number of cuts by the manufacturer, equals to the measured tool wear, and generate the estimated tool wear in backward direction. Afterward, combine these estimation from forward and backward RNNs. Applied the estimated tool wear in time series and the indirect sensor signal measured online, to retrain the RNN model. Then, the values of the model is updated.

5.3.3 Expend the Methods to Other Degrading Process

One of the significant goal of industrial 4.0. is condition based maintenance for the machining and fabricating equipment, aiming at reduce the cost of overhaul, damage to the machine or equipment, and work piece. With the advances computational capacity, cloud computation architecture, it becomes more executable in online decision making via computation with real time data via the algorithms from the methods.

The two methods is easy expandable to other physical degradation processes, as long as long as the degradation process of the physical entity can be measured indirectly in real time, and the direct measurement can be measured offline at certain time interval. Also, it is possible to use different kinds of sensors, such as vibration, power, acoustic emission, etc., and multiple features extracted from the signals, include time domain feature and frequency domain features. At last, if the RUL need to be predicted, the criterion to the direct measurement should be set up.

Specifically, there are several processes in smart manufacturing and intelligent machine can be modeled by both or one of my proposed methods. For example, the surface roughness, which is similar to the tool wear that it is also gradually increase during the machine process owing to the tool wear. Surface roughness is a direct measurement or index for quality of a workpiece. To our best knowledge, and because our methods is pretty new and creative, there is no exact method that has been applied in the surface roughness problem. In practice, the measurement of the surface roughness need to be conducted, which although accessible, yet time consuming for such measurement and thus online measurement is challenge. We can model the surface roughness as the offline direct measurement and sensor signal, the vibration during the machine as the online indirect measurement, and use both the methods.

Another case is the wearing process of gears in the gear box. Certainly, it is not possible to measure the degree of wear in real time. The real degree of wear for the gear can be measured after a period, the gearbox get uninstalled, taken off the main axle and inspected under the microscope and then installed back. This process is time consuming and labor intensive, so that online measurement is hard to achieve. So that our methods can also get applied for the gear wear. Next one is the bearings, for the balls of bearing, the direct measurement of the wear of the balls is less than accessible, so that in order to reduce cost incurred by such measurement, the periods between inspection could be longer. The frequency of inspection will determine the accuracy for the tool wear estimation and forecast and thus need to be determined experimentally.

REFERENCES

1. Ahmadzadeh, F., and Lundberg, J., 2014. Remaining useful life estimation: review. *International Journal of System Assurance Engineering and Management*, 5(4), 461-474.
2. Allison, P. D. Survival analysis using the SAS system: A practical guide. 1995. Cary, NC: SAS Institute.
3. Aslantas K, Tasgetiren S (2004) A Study of Spur Gear Pitting formation and Life Prediction. Wear 257:1167–1175.
4. Astakhov, V. P., 2006, Tribology of Metal Cutting (Tribology and Interface Engineering Series), Elsevier, New York.
5. Baruah, Pundarikaksha, and Ratna B. Chinnam*. "HMMs for diagnostics and prognostics in machining processes." International Journal of Production Research 43.6 (2005): 1275-1293.
6. Bukkapatnam, S. T. S., S. R. T. Kumara, and A. Lakhtakia. "Analysis of acoustic emission signals in machining." Journal of manufacturing science and engineering 121.4 (1999): 568-576.
7. Bukkapatnam, Satish TS, Soundar RT Kumara, and Akhlesh Lakhtakia. "Fractal estimation of flank wear in turning." Journal of dynamic systems, measurement, and control 122.1 (2000): 89-94.
8. Cai, G., Chen, X., Li, B., Chen, B. and He, Z., 2012. Operation reliability assessment for cutting tools by applying a proportional covariate model to condition monitoring information. Sensors, 12(10), pp.12964-12987.

9. Carvalho, C., Johannes, M. S., Lopes, H. F., and Polson, N., 2010. Particle learning and smoothing. Statistical Science, 25(1), 88-106.

10. Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. and Bengio, Y., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.

11. Chung, J., Kastner, K., Dinh, L., Goel, K., Courville, A.C. and Bengio, Y., 2015. A recurrent latent variable model for sequential data. In Advances in neural information processing systems (pp. 2980-2988).

12. Ding, Feng, and He, Zhengjia. "Cutting tool wear monitoring for reliability analysis using proportional hazards model." The International Journal of Advanced Manufacturing Technology 57.5-8 (2011): 565-574.

13. Gao, R., Wang, L., Teti, R., Dornfeld, D., Kumara, S., Mori, M., and Helu, M., 2015. Cloud-enabled prognosis for manufacturing. CIRP Annals-Manufacturing Technology, 64(2), 749-772.

14. Geramifard, O., Xu, J.X., Zhou, J.H. and Li, X., 2011, June. Continuous health condition monitoring: a single hidden semi-Markov model approach. In Prognostics and Health Management (PHM), 2011 IEEE Conference on (pp. 1-10). IEEE.

15. Ghasempoor, A., T. N. Moore, and J. Jeswiet. “On-line wear estimation using neural networks.” Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture 212.2 (1998): 105-112.

16. Ghasempoor, A., J. Jeswiet, and T. N. Moore. "Real time implementation of on-line tool condition monitoring in turning." International Journal of Machine Tools and Manufacture 39.12 (1999): 1883-1902.

17. Ginta, T.L., Amin, A.K.M.N., Radzi, H.C.D.M. and Lajis, M.A., 2009. Tool life prediction by response surface methodology in end milling titanium alloy Ti-6Al-4V using uncoated WC-Co inserts. European Journal of Scientific Research, 28(4), pp.533-541.

18. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y., 2014. Generative adversarial nets. In Advances in Neural Information Processing Systems (pp. 2672-2680).

19. Goodfellow, I.J., Bengio, Y., and Courville, A., 2016. Deep learning (Adaptive Computation and Machine Learning series). MIT Press book.

20. Graves, A., 2013. Generating sequences with recurrent neural networks. arXiv preprint arXiv:1308.0850.

21. Groover, Mikell P. Fundamentals of modern manufacturing: materials processes, and systems. John Wiley & Sons, 2010.

22. Haber, R.E., Jiménez, J.E., Peres, C.R. and Alique, J.R., 2004. An investigation of tool-wear monitoring in a high-speed machining process. Sensors and Actuators A: Physical, 116(3), pp.539-545.

23. Hanachi, H., Liu, J., Banerjee, A., and Chen, Y., 2016. Sequential state estimation of nonlinear/non-Gaussian systems with stochastic input for turbine degradation estimation. Mechanical Systems and Signal Processing, 72, 32-45.

24. Hartikainen, J., Solin, A., and Särkkä, S., 2011. Optimal Filtering with Kalman Filters and Smoothers a Manual for the Matlab toolbox EKF/UKF Version 1.3.

25. Haykin, S., 1998. Neural Networks, and Comprehensive Foundation, Pearson/Prentice-Hall.

26. Haykin, S., 2008. Neural Networks and Learning Machines, Pearson/Prentice-Hall.

27. Hochreiter, S. and Schmidhuber, J., 1997. Long short-term memory. Neural computation, 9(8), pp.1735-1780.

28. Hush, D.R. and Horne, B.G., 1993. Progress in supervised neural networks. IEEE signal processing magazine, 10(1), pp.8-39.

29. Julier, S. J. and Uhlmann, J. K. (2004b). Unscented filtering and nonlinear estimation. Proceedings of the IEEE, 92(3):401–422.

30. Kamarthi, S.V., Kumara, S.R.T. and Cohen, P.H., 1995. Wavelet representation of acoustic emission in turning process. Intelligent Engineering Systems Through Artificial Neural Networks, 5, pp.861-866.

31. Karandikar, Jaydeep M., Ali E. Abbas, and Tony L. Schmitz. "Tool life prediction using Bayesian updating. Part 1: Milling tool life model using a discrete grid method." Precision Engineering 38.1 (2014a): 9-17.

32. Karandikar, Jaydeep M., Ali E. Abbas, and Tony L. Schmitz. "Tool life prediction using Bayesian updating. Part 2: Turning tool life using a Markov Chain Monte Carlo approach." Precision Engineering 38.1 (2014b): 18-27.

33. Karandikar, Jaydeep M., Ali Abbas, and Tony L. Schmitz. "Remaining useful tool life predictions in turning using Bayesian inference." Int. J. Prognostics Health Manag 4.2 (2012): 25-36.

34. Karandikar, Jaydeep M., Ali E. Abbas, and Tony L. Schmitz. "Tool life prediction using random walk Bayesian updating." Machining Science and Technology 17.3 (2013): 410-442.

35. Kormushev, Petar, and Darwin G. Caldwell. "Direct policy search reinforcement learning based on particle filtering." Proceedings of the 10th European Workshop on Reinforcement Learning. 2012.

36. Kurada, S., and Bradley, C., 1997. A machine vision system for tool wear assessment. Tribology International, 30(4), 295-304.

37. Kuttolamadom, M.A., Mears, M.L. and Kurfess, T.R., 2012. On the Volumetric Assessment of Tool Wear in Machining Inserts With Complex Geometries—Part 1: Need, Methodology, and Standardization. Journal of Manufacturing Science and Engineering, 134(5), p.051002.

38. Lee, J. H., Kim, D. E. and Lee, S. J., 1996. Application of neural networks to flank wear prediction. Mechanical Systems and Signal Processing 10.3: 265-276.

39. Lee, J. H., and Lee, S. J. "One-step-ahead prediction of flank wear using cutting force." International Journal of Machine Tools and Manufacture 39.11 (1999): 1747-1760.

40. Lipton, Z.C., Berkowitz, J. and Elkan, C., 2015. A critical review of recurrent neural networks for sequence learning. arXiv preprint arXiv:1506.00019.

41. Liu, Q., and Altintas, Y., 1999. On-line monitoring of flank wear in turning with multilayered feed-forward neural network. International Journal of Machine Tools and Manufacture, 39(12), 1945-1959.

42. Lopes, H. F., and Tsay, R. S., 2011. Particle filters and Bayesian inference in financial econometrics. Journal of Forecasting, 30(1), 168-209.

43. Luetzig, G., Sanchez-Castillo, M. and Langari, R., 1997, June. On tool wear estimation through neural networks. In Neural Networks, 1997., International Conference on (Vol. 4, pp. 2359-2363). IEEE.

44. Mak, M.W., Ku, K.W. and Lu, Y.L., 1999. On the improvement of the real time recurrent learning algorithm for recurrent neural networks. Neurocomputing, 24(1), pp.13-36.

45. Malhi, A., Yan, R. and Gao, R.X., 2011. Prognosis of defect propagation based on recurrent neural networks. IEEE Transactions on Instrumentation and Measurement, 60(3), pp.703-711.

46. Marino, J.B., Banchs, R.E., Crego, J.M., de Gispert, A., Lambert, P., Fonollosa, J.A. and Costa-Jussà, M.R., 2006. N-gram-based machine translation. Computational Linguistics, 32(4), pp.527-549.

47. Motorcu, A.R., Kus, A., Arslan, R., Tekin, Y. and Ezentas, R., 2013. Evaluation of tool life-tool wear in milling of Inconel 718 superalloy and the investigation of effects of cutting parameters on surface roughness with Taguchi method. Techn Gaz, 20, pp.765-774.

48. Narendra, K.S. and Parthasarathy, K., 1990. Identification and control of dynamical systems using neural networks. IEEE Transactions on neural networks, 1(1), pp.4-27.

49. Nemeth, Christopher, Paul Fearnhead, and Lyudmila Mihaylova. "Sequential Monte Carlo methods for state and parameter estimation in abruptly changing environments." Signal Processing, IEEE Transactions on 62.5 (2015): 1245-1255.

50. Niaki, Farbod Akhavan, Durul Ulutan, and Laine Mears. "Parameter Inference Under Uncertainty in End-Milling γ′-Strengthened Difficult-to-Machine Alloy." Journal of Manufacturing Science and Engineering 138.6 (2016): 061014.

51. Niaki, Farbod Akhavan, Durul Ulutan, and Laine Mears. "Stochastic tool wear assessment in milling difficult to machine alloys." International Journal of Mechatronics and Manufacturing Systems 8.3-4 (2015a): 134-159.

52. Niaki, Farbod Akhavan, Durul Ulutan, and Laine Mears. "In-Process Tool Flank Wear Estimation in Machining Gamma-Prime Strengthened Alloys Using Kalman Filter." Procedia Manufacturing 1 (2015b): 696-707.

53. Niaki, Farbod Akhavan, Durul Ulutan, and Laine Mears. "Wavelet based sensor fusion for tool condition monitoring of hard to machine materials." Multisensor Fusion and Integration for Intelligent Systems (MFI), 2015 IEEE International Conference on. IEEE, (2015c).

54. Niaki, F.A., Feng, L., Ulutan, D. and Mears, L., 2016. A wavelet-based data-driven modelling for tool wear assessment of difficult to machine materials. International Journal of Mechatronics and Manufacturing Systems, 9(2), pp.97-121.

55. Owsley, L. M., Atlas, L. E., and Bernard, G. D., 1997. Self-organizing feature maps and hidden Markov models for machine-tool monitoring. IEEE Transactions on Signal Processing, 45(11), 2787-2798.

56. Özel, T. and Karpat, Y., 2005. Predictive modeling of surface roughness and tool wear in hard turning using regression and neural networks. International Journal of Machine Tools and Manufacture, 45(4), pp.467-479.

57. Paris, PC, Gomez, MP, Anderson, WE (1961). A Rational Analytic Theory of Fatigue. The Trend in Engineering 13:9–14.

58. Pugno N, Ciavarella M, Cornetti P, Carpinteri A (2006) A Generalized Paris' Law for Fatigue Crack Growth. Journal of the Mechanics and Physics of Solids. 54:1333–1349.

59. Rabiner, L.R., 1989. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, 77(2), pp.257-286.

60. Ranzato, M'A., Chopra, S., Michael Auli, M. and Wojciech Zaremba, W., "Sequence level training with recurrent neural networks." ICLR, San Juan, Puerto Rico (2016).

61. Rezende, D.J., Mohamed, S. and Wierstra, D., 2014. Stochastic backpropagation and approximate inference in deep generative models. arXiv preprint arXiv:1401.4082.

62. Si, X. S., Wang, W., Chen, M. Y., Hu, C. H., and Zhou, D. H., 2013. A degradation path-dependent approach for remaining useful life estimation with an exact and closed-form solution. European Journal of Operational Research, 226(1), 53-66.

63. Sick, B., 2002. On-line and indirect tool wear monitoring in turning with artificial neural networks: a review of more than a decade of research. Mechanical Systems and Signal Processing, 16(4), 487-546.

64. Sikorska, J. Z., Hodkiewicz, M., and Ma, L., 2011. Prognostic modelling options for remaining useful life estimation by industry. Mechanical Systems and Signal Processing, 25(5), 1803-1836.

65. Sutskever, I., Vinyals, O. and Le, Q.V., 2014. Sequence to sequence learning with neural networks. In Advances in neural information processing systems (pp. 3104-3112).

66. Tse, P.W. and Atherton, D.P., 1999. Prediction of machine deterioration using vibration based fault trends and recurrent neural networks. Journal of vibration and acoustics, 121(3), pp.355-362.

67. Tsung, F.S. and Cottrell, G.W., 1995. Phase-space learning. Advances in Neural Information Processing Systems, pp.481-488.

68. Usui E., Hirota A., Masuko M., (1997) Analytical Prediction of Three-Dimensional Cutting Process—Part 1: Basic Cutting Model and Energy Approach. Journal of Engineering for Industry 100(2):222–228.

69. Vallejo Jr, A.G., Nolazco-Flores, J.A., Morales-Menéndez, R., Sucar, L.E. and Rodríguez, C.A., 2005, November. Tool-wear monitoring based on continuous hidden Markov models. In Iberoamerican Congress on Pattern Recognition (pp. 880-890). Springer Berlin Heidelberg.

70. Venkatraman, A., Hebert, M. and Bagnell, J.A., 2015, January. Improving Multi-Step Prediction of Learned Time Series Models. In AAAI (pp. 3024-3030).

71. Wang, X., Wang, W., Huang, Y., Nguyen, N. and Krishnakumar, K., 2008. Design of neural network-based estimator for tool wear modeling in hard turning. Journal of intelligent manufacturing, 19(4), pp.383-396.

72. Wang, J., Wang, P. and Gao, R. X. "Tool life prediction for sustainable manufacturing." (2013).

73. Wang, Peng, and Robert X. Gao. "Adaptive resampling-based particle filtering for tool life prediction." Journal of Manufacturing Systems 37 (2015a): 528-534.

74. Wang, Jinjiang, Wang, Peng, and Robert X. Gao. "Enhanced particle filter for tool wear prediction." Journal of Manufacturing Systems 36 (2015b): 35-45.

75. Wang, Lihui, and Robert X. Gao, eds. Condition monitoring and control for intelligent manufacturing. Springer Science & Business Media, 2006.

76. Widodo, A., and Yang, B. S., 2007. Support vector machine in machine condition monitoring and fault diagnosis. Mechanical systems and signal processing, 21(6), 2560-2574.

77. Williams, R.J. and Zipser, D., 1989. A learning algorithm for continually running fully recurrent neural networks. Neural computation, 1(2), pp.270-280.

78. Yen, Y. C., Söhner, J., Lilly, B., and Altan, T., 2004. Estimation of tool wear in orthogonal cutting using the finite element analysis. Journal of Materials Processing Technology, 146(1), 82-91.

79. Yue, Wu, G. S. Hong, and Y. S. Wong. "HMM with explicit state duration for prognostics in face milling." 2010 IEEE Conference on Robotics, Automation and Mechatronics. IEEE, 2010.

80. Yue, Wu. Prognostics for Tool Condition Monitoring Based on Long-Term and Short-Term Prognostic Approaches. Diss. 2012.

81. Zhang, J., Starly, B., Cai, Y., Cohen, P.H. and Lee, Y.S., 2017a. Particle learning in online tool wear diagnosis and prognosis. Journal of Manufacturing Processes.

82. Zhang, J., Starly, B., 2017b. Recurrent Neural Networks with Long Term Temporal Dependencies in Machine Tool Wear Diagnosis and Prognosis. Accepted for review by Engineering Applications of Artificial Intelligence.

Appendix (Jianlei Zhang Resume)

Area of Expertise

Data Mining/Machine Learning, (Recurrent) Neural Networks/Deep Learning, Time Series Analysis, Statistics, Linear/Nonlinear/Heuristic Optimization, Design of Algorithm, Database Management, Logistics, Simulation

Education

Ph.D. student, Industrial & Systems Engg.(ISE), NC State Univ., 3.52/4.0, Aug. 2013 -Now

Ph.D. Program, ISE, Univ. of Oklahoma, 4.0/4.0, Aug. 2012 – Aug. 2013

Master Degree in MechE, (Thesis), Univ. of Shanghai for Sci & Techy, 3.5/4.0, Sept. 2011

Skills

• Expert in Python (Torch, Pandas, NLTK, Scikit-Learn, Matplotlib), R, MATLAB, SQL, Macro, VBA

• Proficient in SAS, SPSS, Tableau, Simio, CPLEX, Minitab, C

• Familiar with Hadoop, Apache Pig, LabView, C++

Working Experience

Electric Control Engineer, Shanghai General Motors. Oct. 2011 – Aug. 2012, Shanghai, China

Project of Retooling Machining Center

• Retool and debug the GROB machining center with FANUC CNC systems

• Exam its electrical control cabinet and electrical control diagram

Research Assistant, Shanghai Electric Group. Sept. 2010 – Jun. 2011, Shanghai, China

Judging the Degradation of the Ball Screw

• Collecting the servomotor's electric signal at different spindle rotational speed by LabView

• Design the digital low-pass filter, and apply wavelet multi-resolution signal decomposition to detect the wear and tear of the ball screw

Publications

1. Zhang, J., et al., Recurrent Neural Networks with Long Term Temporal Dependencies in Machine Tool Wear Diagnosis and Prognosis. Accepted for review by Engineering Applications of Artificial Intelligence

• Applied Elman RNN, LSTM, GRU as system model for online estimation and forecast

2. Zhang, J et al., Particle Learning in Online Tool Wear Diagnosis and Prognosis. JMP (2017)

3. Nordberg, R. C., Zhang, J., et al., (2016). ECI Spectroscopy Can Monitor Age-Grouped hASC Variability During Osteogenic Differentiation. StemCellsTM.

• Worked on the Signal Analysis, Statistics analysis, and Data Visualization

4. Zhang, J., Li, H., Principal Component Analysis in Grinding Chatter Detection. (2011)

5. Zhang, J., Li, H., The Study on the Methods of Degradation Assessment of the Feed Drive System of Machine Tools, Aug. 2011. Master Thesis.

• Designed digital filter, adaptive filter to filter out noise and applied wavelet analysis to extract features

Projects

Harvest System Optimization for Bioenergy Production

• Create a model for inventory, harvest system, and logistics for biomass material production and determine the optimal method for harvest

• Create a model for the lowest stock to minimize the inventory cost

• Decide the optimal number of machines, considering the production rate of machine, storage lost and cost

• Solve the optimal location for the refinery plant, considering the farm area and location

Time Series Analysis Projects

• Utilize ARMA model and GARCH model to model Walmart stock

• Model the CPI and GDP by Vector Autoregressive (VAR) model, compute the forecast errors by AIC and BIC.

• Assess the Granger-causes between two variables

Inventory Management Analysis for a Newspaper Company

• Build a simulation model to calculate the optimal stock for two agents, both supplier and retailer

• Propose a new buyback policy to solve the problem to non-cooperative games of two agents

Recommendation System for Yelp Business Rating Prediction

• Predict business rating by Singular Value Decomposition (SVD)

• Apply five-fold-cross validation and root mean square error (RMSE) discover SVD is better than other

Big Data Project on Wikipedia Words Count and PageRank Algorithm

• Words count and PageRank for 41 GB Wikipedia by Hadoop, Pig Latin, and python writing the map-reduce functions

Old Navy In-Store Inquiry and Order Systems

• Developed the multi-column and multi-selection menu and search engine interface by VBA

• Associated the interface with Macro and SQL to the database system for customer and inventory

Computer Vision for Butterfly Classification

• Extract various features of butterflies according to statistics

• Utilize Fourier coefficients to extract shape attributes and to do the clustering

Simio Aerospace Manufacturing Problem

• Modeled the assembly line with tasks for each workcell, and assigned the workers

• Simulated the model and achieved the line balance by shifting the tasks between workcells

• Adjusted the shift number and overtime work when the new plane get introduced into the same assembly line

Professional Activities

• NSF Student Travel Award and NAMRC Conference Presenter, June 2017, Los Angeles, CA

• Presenter at Institute of Industrial Engineers IIE, June 2015, Nashville, TN, USA

• Production Intern, FAW Automotive, Aug. 2008. Changchun, Jilin, China

• Audit research seminar: Smart Manufacturing--Research implications for quality assurance and productivity, presented by Dr. Satish T. S. Bukkapatnam from Texas A&M University

• Completed the Machine Learning course, taught by Dr. Andrew Ng, from Stanford University and by Pedro Domingos from Univ. of Washington on Coursera.

• Participate in Meetup - Discuss artificial intelligence, big data, and the IBM Watson project

Interests

Badminton, Basketball, Jogging, Ancient Greek Philosophy, Documentary, Anthropology, History, Future Study, Sci-fi movies.