Research Article

Prognostics and Health Management of an Automated Machining Process

Cheng He,1,2 Jiaming Li,3 and George Vachtsevanos3

1 School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China
2 School of Intelligent Manufacturing and Control Engineering, Shanghai Second Polytechnic University, Shanghai 201209, China
3 School of Electrical & Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0250, USA

Correspondence should be addressed to George Vachtsevanos; giv@gatech.edu

Received 21 September 2015; Accepted 4 November 2015

Academic Editor: Uchechukwu E. Vincent

Copyright © 2015 Cheng He et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine failure modes are presenting a major burden to the operator, the plant, and the enterprise causing significant downtime, labor cost, and reduced revenue. New technologies are emerging over the past years to monitor the machine’s performance, detect and isolate incipient failures or faults, and take appropriate actions to mitigate such detrimental events. This paper addresses the development and application of novel Prognostics and Health Management (PHM) technologies to a prototype machining process (a screw-tightening machine). The enabling technologies are built upon a series of tasks starting with failure analysis, testing, and data processing aimed to extract useful features or condition indicators from raw data, a symbolic regression modeling framework, and a Bayesian estimation method called particle filtering to predict the feature state estimate accurately. The detection scheme declares the fault of a machine critical component with user specified accuracy or confidence and given false alarm rate while the prediction algorithm estimates accurately the remaining useful life of the failing component. Simulation results support the efficacy of the approach and match well the experimental data.

1. Introduction

Prognosis and Health Management (PHM) has emerged over recent years as significant technologies that are making an impact on both military and commercial maintenance practices. Automated machining processes are employed extensively in manufacturing. Screw-tightening machines are critical assets of an automated machining process. This study focuses on such a machining process with novel features for automatic screw tightening in crucial manufacturing, assembly, and other operations. Figure 1 shows a picture of the equipment. Screw tightening is usually carried out manually resulting in wasted time, operator mistakes, and inconsistent applied torques. The automatic assembly line requires, therefore, for improved performance an automatic screw-tightening machine.

The machine consists mainly of seven parts: feeder, screw falling device, screw separating device, screwdriver, guiding device, 3-axis motion platform, and a clamping device. The function of the feeder is to arrange the screws in a line. The operator places thousands of screws in the center of the feeder. The screws are arranged in a row along the track of the feeder via a vibration mechanism and arranged in a line at the end of the track. The screws move along the track of the feeder and, via gravity, they drop into the pipe one by one and arranged in a line in the pipe. A screw separating device picks one screw at a time and sent it to the head of the screwdriver. The 3-axis motion platform moves the screw hole of the parts behind the head of the screwdriver. Then, the screwdriver moves and tightens the screw.

2. Failure Analysis

Machine critical failure modes include the following: the screws may pile up on the trail of the feeder, thus preventing the screw separating device from separating the screws and arranging them one at a time. The 3-axis motion platform does not move to the precise location. The screwdriver is unable to tighten the screw. Among several other failure
modes, the 3-axis motive platform not moving to its precise location is the most serious one. In this case, the screw cannot be tightened in the screw hole and the parts cannot be connected correctly. The major reason for the 3-axis motive platform not being able to move precisely is the presence of a fault condition in bearings. Bearing fault/failure is deemed to be critical, severe, and frequent and must be addressed if the automated process will perform properly and expeditiously.

The overall architecture for the bearing health management system is shown in Figure 2. We will describe in the next section the function of each module of the architecture.

3. The Test Platform

In order to obtain test data, under normal and faulty conditions, for the targeted bearings a test platform was designed and built. Accelerated life testing is performed on two rolling element bearings. The experimental data was processed and appropriate features or condition indicators (CIs) were extracted. The processed data combined with a suitable model of the fault growth and a novel estimation algorithm, called particle filtering, are used to implement the diagnostic and prognostic routines.

3.1. Test Platform Design. The test platform consists of a single-axis servomotor-driven module and a data acquisition and analysis system. Figure 3 is the single-axis servomotor-driven module. Figure 4 is the data acquisition system.

3.1.1. Single-Axis Servomotor-Driven Module. The single-axis servomotor-driven module includes a 400 w servo motor, a coupling device, two bearings, two bearing-bases, a ball screw, and a suitable electrical control system. The servomotor is connected to the ball screw via the coupling. Both ends of the ball screw are installed in the bearings. The bearings are fixed in the bearing-housing.

3.1.2. Data Acquisition System. The data acquisition system is composed of a 16-channel dynamic signal data acquisition instrument (model: DH5902), a charge adapter, and two vibration sensors (model: PCB356A32). The vibration sensors are installed in a rack with a magnetic base. They acquire the vibration signal and send it to the 16-channel dynamic signal data acquisition instrument via the charge adapter. The Industrial Control Computer (IPC) receives the data from the 16-channel dynamic signal data acquisition instrument via wireless transmission.

3.1.3. Experimental Parameters. The servomotor has three speeds: 500 r/min, 1000 r/min, and 1500 r/min. The sampling frequency is 2.5 Kz, and the sampling period is 10 seconds. Two bearings were tested. Each bearing is tested along the Z direction resulting in two test points for each speed and each test point providing three entries; that is, there are 6 data entries for each speed setting.

3.2. Fault Analysis. Rolling bearings are typically damaged due to various causes, such as improper assembly, poor lubrication, and moisture. Corrosion and overload may also lead to premature bearing damage. If lubrication and maintenance
procedures are not performed regularly over time, fatigue spall and bearing wear may occur. The main failure modes are fatigue spall, wear, plastic deformation, corrosion, fracture, gluing, cage damage, and so forth.

3.2.1. Fault Conditions. Among all the bearing faults, abrasion is one of the most common. Abrasion implies that material is removed from the metal surface. The characteristic feature for abrasion is a shallow trench, with a bright surface. It is produced in rolling contact surface or guide surface. With abrasion present, the bearing clearance increases.

3.2.2. Root-Cause Analysis. Bearing abrasion is due primarily to two causes: insufficient bearing lubrication and small particle penetration into the bearing resulting in material removal from the metallic surface via sliding friction. Surface wear is the result of the relative motion between rolling track and rolling body when dust and other objects enter the bearing surface resulting in increased bearing surface roughness and reduced motion accuracy. The motion accuracy of the machine is also reduced increasing the machine’s vibration and noise. For precision bearings, the amount of wear determines the life of the bearing. On the other side, there is also a kind of wear with slight vibration.

3.3. Fault Injection. In testing, faults are injected periodically until the bearing reaches a failure condition and fault data are recorded. The fault injection consists of the following arrangement: a hole is drilled on the side of the bearing base and a screw is tightened in the hole step by step. The screw generates friction with the bearing surface resulting in an abrasion fault mode.

4. Feature Extraction and Selection

As a bearing defect evolves, due to poor lubrication for example, the fault mode excites a specific frequency associated with the particular type of defect. The amplitude and time duration of the defect frequency are generally good indicators of defect severity. The increase of defect size usually increases the ratio of impulse force to operational noise. The increase of defect length also increases the impulse force duration. If the defect area is large along the raceway turning direction, harmonics of the frequency will also imply an indication of the severity of the defect present. Acquired vibration data are used to extract appropriate features.

4.1. Feature Extraction. The vibration signal is converted from the time to the frequency domain. Analysis shows that the signal changes rapidly between 1700 Hz and 1900 Hz with features extracted in this area. Several vibration signals from bearings with different fault conditions are illustrated in Figure 5. For each data set we cut the period with signal as shown in Figure 5, the section between these two lines.

Frequency domain studies indicate when a bearing defect exists with the defect exhibiting a signature in the frequency spectra of the vibration signals. The signal is digitized in a frequency region around the resonances of the structure and such features as the energy, amplitude, and so forth are extracted. The frequency spectra are shown in Figure 6.

We choose four frequency domain features, as is shown in Table 1.

Time domain features include mean value ($f_9$) and standard deviation ($f_{10}$). We calculate the correlation coefficient for each feature. Table 2 shows the results.

4.2. Feature Fusion. Within an automated health management system, there are many areas where fusion technologies play a contributing role. At the lowest level, data fusion can be used to combine information from a multisensory data array to validate signals and create features. At a higher level, fusion may be used to combine features in intelligent ways so as to obtain the best possible diagnosis information. Finally, knowledge fusion is used to incorporate experience-based information such as legacy failure rates or physical model predictions with signal-based information.
Table 2: Correlation coefficient.

<table>
<thead>
<tr>
<th>Feature</th>
<th>$f_1$</th>
<th>$f_2$</th>
<th>$f_3$</th>
<th>$f_4$</th>
<th>$f_5$</th>
<th>$f_6$</th>
<th>$f_7$</th>
<th>$f_8$</th>
<th>$f_9$</th>
<th>$f_{10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>0.7972</td>
<td>0.3797</td>
<td>0.7867</td>
<td>0.5902</td>
<td>0.2729</td>
<td>0.0327</td>
<td>0.0097</td>
<td>0.2957</td>
<td>0.2368</td>
<td>0.9347</td>
</tr>
</tbody>
</table>

Three features with the highest correlation coefficient are picked up: $f_1$, $f_3$, $f_{10}$.

Table 3: Correlation coefficient.

<table>
<thead>
<tr>
<th>Feature</th>
<th>$f_1$</th>
<th>$f_3$</th>
<th>$f_{10}$</th>
<th>$f_{combined}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>0.79</td>
<td>0.78</td>
<td>0.93</td>
<td>0.95</td>
</tr>
</tbody>
</table>

The combined feature has the best highest coefficient and performs better than any individual feature.

Thus, we combine the features and get the combined feature:

$$F_{combined} = f_1^{0.29} \cdot f_3^{0.18} \cdot f_{10}^{2.01}.$$  \hspace{1cm} (1)

The combined correlation coefficient is 0.95. The combined feature performance is shown in Figure 7.

4.3. Performance Comparison. We apply correlation coefficient as the performance metric and come up with Table 3.

5. Modeling

Reliable, high-fidelity fault growth models form the foundation for accurate and robust detection and failure prediction. A suitable modeling framework assists in the development, testing, and evaluation of detection and prediction algorithms. It may be employed to generate data for data-driven methods to diagnostics/prognostics and test and validate routines for data processing tool development, among others. The flexibility provided by a simulation platform, housing appropriate detection and progression models, is a unique attribute in the study of how fault processes are initiated and propagating so that corrective action can be taken before a catastrophic event occurs. The objective of the modeling effort is to develop, test, and evaluate novel fault initiation and progression models that will assist in the design and implementation of “smart” sensors and sensing modalities for critical machine systems.

5.1. Modeling-Symbolic Regression. Fault detection and prediction algorithms rely on data, a model of the degradation process and an estimation method that, given the current
state of the system, predicts its evolution over the next time step. Such models are typically based on first principles while others are built on the basis of data. We exploit in this effort a modeling framework called symbolic regression. It is a type of regression analysis that searches the space of mathematical expressions to find the model that best fits a given dataset, in terms of both accuracy and simplicity. No particular model is provided as a starting point to the algorithm. Instead, initial expressions are formed by randomly combining mathematical building blocks such as mathematical operators, analytic functions, constants, and state variables. (Usually, a subset of these primitives will be specified by the operator, but that is not a requirement of the technique.) New equations are then formed by combining previous ones, using genetic programming. In linear regression, the dependent variable is a linear combination of the parameters (but need not be linear in the independent variables). Nonlinear symbolic regression and other regression techniques incorporating uncertainty in the independent variables. Nonlinear symbolic regression is a useful tool to implement the symbolic regression routine. Exploiting the test data, Eureqa searches to find the model (both model structure and model parameters) that best fit a given data set. We arrive at the solution by using a swarm of points called particles and a set of weights associated with them representing the discrete probability masses. Particles can be generated and recursively updated easily given a nonlinear process model (which describes the evolution in time of the system under analysis), a measurement model, a set of available measurements \( z_{1:k} = (z_1, \ldots, z_k) \), and an initial estimation for the state PDF \( p(x_0) \), as shown in the following equations:

\[
x_k = f_k(x_{k-1}, \omega_k) \quad \leftrightarrow \quad p(x_k | x_{k-1}),
\]

\[
z_k = h_k(x_k, v_k) \quad \leftrightarrow \quad p(z_k | x_k).
\]

As in every Bayesian estimation problem, the estimation process can be achieved into two steps, namely, prediction and filtering. On the one hand, prediction uses both knowledge of the previous state estimation and the process model to generate the a priori state PDF estimation for the next time instant, as is shown in the following expression:

\[
p(x_k | z_{1:k-1}) = \int p(x_k | x_{k-1}) p(x_{k-1} | z_{1:k-1}) \, dx_{k-1}.
\]

A sufficient data set is obtained by using the following formula to enrich the data:

\[
\text{Feature} = g(x[k]) + n(0, S),
\]

where \( \delta(x) \) is the model fitted by regression, \( S \) is the standard deviation, \( N \) is the number of samples, and \( x \) is a variable:

\[
S = \sqrt{\frac{\sum_{i=1}^{N} [1/i - \delta(x_i)]^2}{N-1}}.
\]

6. The Particle Filtering Framework for Fault Diagnosis and Failure Prognosis

Particle filtering is an emerging and powerful methodology for sequential signal processing based on the concepts of Bayesian theory and Sequential Importance Sampling (SIS). Particle filtering is very suitable for nonlinear systems or in the presence of non-Gaussian process/observation noise. For this approach, both diagnosis and prognosis rely upon estimating the current value of a fault/degradation dimension, as well as other important parameters, and use a set of observations (or measurements) for this purpose. This research team has pioneered the introduction of particle filtering techniques into fault diagnosis and failure prognosis [1, 2]. The team has also demonstrated how these techniques can be combined with traditional artificial intelligence methods in a synergistic way [3]. The success of this novel model-based approach has been demonstrated in a number of diverse application domains from rotorcraft critical components to electrical systems, environmental control systems, and high power amplifiers [4]. Figure 11 depicts the flow of the component modules comprising the health management architecture.

The underlying principle of the methodology is the approximating of the conditional state probability distribution \( p(z_k|x_k) \) using a swarm of points called particles and a set of weights associated with them representing the discrete probability masses. Particles can be generated and recursively updated easily given a nonlinear process model (which describes the evolution in time of the system under analysis), a measurement model, a set of available measurements \( z_{1:k} = (z_1, \ldots, z_k) \), and an initial estimation for the state PDF \( p(x_0) \), as shown in the following equations:

\[
\text{Feature} = 8.11 \times 10^3 + 0.262x^3.
\]

\[
x \text{ implies cycle times; here we use } x \text{ as time, } t.
\]

Figure 10(a) shows the result from the application of Eureqa. Figure 10(b) shows a plot of \( (8.11 \times 10^3 + 0.262x^3) \) and the error as the process proceeds. Figure 10(c) shows the accuracy versus complexity plot. By trading off accuracy for complexity, the best model from the list generated by Eureqa is \( (8.11 \times 10^3 + 0.262x^3) \).
On the other hand, the filtering step considers the current observation $z_t$ and the a priori state PDF to generate the a posteriori state PDF by using Bayes’ formula:

$$p(x_k | z_{1:k}) = \frac{p(z_k | x_k) p(x_k | z_{1:k-1})}{p(z_k | z_{1:k-1})}. \quad (7)$$

The actual distributions then would be approximated by a set of samples and the corresponding normalized importance weights $\tilde{w}_k = \tilde{w}_k(x_{0:k}^i)$ for the $i$th sample:

$$p(x_k | z_{1:k}) \approx \sum_{i=1}^{N} \tilde{w}_k(x_{0:k}^i) \delta(x_{0:k} - x_{0:k}^i), \quad (8)$$

where the update for the importance weights is given by

$$w_k = w_{k-1} \frac{p(z_k | x_k) p(x_k | z_{1:k-1})}{p(z_k | x_{0:k-1}, z_{1:k-1})}. \quad (9)$$

**Fault Diagnosis.** The proposed particle-filter-based diagnosis framework aims to accomplish the tasks of fault detection and identification, under general assumptions of non-Gaussian noise structures and nonlinearities in process dynamic models, using a reduced particle population to represent the state pdf [5]. A compromise between model-based and data-driven techniques is accomplished by the use of a particle filter-based module built upon the nonlinear dynamic state model:

$$x_d(t+1) = f_b(x_d(t), n(t)), \quad (10)$$

$$x_c(t+1) = f_f(x_c(t), x_d(t), w(t)), \quad (11)$$

Features $(t) = h_t(x_d(t), x_c(t), v(t))$, where $f_b$, $f_f$ and $h_t$ are nonlinear mappings, $x_d(t)$ is a collection of Boolean states associated with the presence of a particular operating condition in the system (normal operation, fault type #1, #2, etc.), $x_c(t)$ is a set of continuous-valued states that describe the evolution of the system given those operating conditions, and $w(t)$ and $v(t)$ are non-Gaussian distributions that characterize the process and feature noise signals, respectively. At any given instant of time, this framework provides an estimate of the probability masses associated with each fault mode, as well as a pdf estimate for meaningful physical variables in the system. Once this information is available within the diagnostic module, it is conveniently processed to generate proper fault alarms and to inform about the statistical confidence of the detection routine. Customer specifications are translated into acceptable margins for types I and II errors in the detection routine. The algorithm itself will indicate when type II error (false negatives) has decreased to the desired level. Typical results of the diagnostic algorithm are shown in Figure 12.

**Failure Prognosis.** The evolution in time of the fault dimension may be described through the state equation:

$$x_1(t + 1) = x_1(t) + x_2(t) \cdot F(x(t), t, U), w_1(t),$$

$$x_3(t + 1) = x_2(t) + w_2(t), \quad (12)$$

where $x_1(t)$ is a state representing the fault dimension under analysis, $x_2(t)$ is a state associated with an unknown model parameter, $U$ are external inputs to the system (load profile, etc.), $F(x(t), t, U)$ is a general time-varying nonlinear function, and $w_1(t)$ and $w_2(t)$ are white noises (not necessarily Gaussian) [5]. The nonlinear function may represent a model based on first principles, a neural network, or even a fuzzy system. Long term predictions are generated on the basis of the current state pdf estimate, and using kernel functions to reconstruct the state pdf estimate for future time instants,
<table>
<thead>
<tr>
<th>Size</th>
<th>Fit</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.000</td>
<td>Feature = 8.89e3</td>
</tr>
<tr>
<td>5</td>
<td>0.748</td>
<td>Feature = 6.54e3 + 198x</td>
</tr>
<tr>
<td>3</td>
<td>0.998</td>
<td>Feature = 8.8e3 + x</td>
</tr>
<tr>
<td>6</td>
<td>0.721</td>
<td>Feature = ( \frac{4.46e5}{62.8} - x )</td>
</tr>
<tr>
<td>7</td>
<td>0.707</td>
<td>Feature = 7.81e3 + 6.4x²</td>
</tr>
<tr>
<td>9</td>
<td>0.676</td>
<td>Feature = 8.11e3 + 0.262x³</td>
</tr>
<tr>
<td>11</td>
<td>0.653</td>
<td>Feature = 8.2e3 + 0.0108x⁴</td>
</tr>
<tr>
<td>13</td>
<td>0.649</td>
<td>Feature = 8.27e3 + 0.000415x⁵</td>
</tr>
<tr>
<td>14</td>
<td>0.584</td>
<td>Feature = 7.01e3 + 159x - 1.15e3 sin(175x)</td>
</tr>
<tr>
<td>16</td>
<td>0.449</td>
<td>Feature = 7.13e3 + 139x + 84.4x cos(144x)</td>
</tr>
<tr>
<td>18</td>
<td>0.399</td>
<td>Feature = 7.15e3 + 131x + 5.54x² cos(144x)</td>
</tr>
<tr>
<td>20</td>
<td>0.390</td>
<td>Feature = 7.98e3 + 4.37x² + 5.25x² cos(144x)</td>
</tr>
<tr>
<td>22</td>
<td>0.390</td>
<td>Feature = 7.97e3 + x + 4.35x² + 5.27x² cos(144x)</td>
</tr>
</tbody>
</table>

**Figure 10:** Symbolic regression using Eureka.

**Figure 11:** The health management architecture.
7. Results

Using the machinery data, we determine its health status and hence its current fault/failure condition. The main modules of an integrated approach to fault diagnosis and failure prognosis are shown in Figure 13. It includes experiment design, fault classification, data collection, signal process, feature extraction, algorithm development, and schedules required maintenance.

The resulting predicted state pdf contains critical information about the evolution of the fault dimension over time. This information is condensed through the computation of statistics (expectations, 90% confidence intervals), the Time-of-Failure (ToF), or the Remaining Useful Life (RUL) of the faulty system. Figure 16 depicts a typical prognostic configuration.

Data from simulation as well as the experiments were used to implement and evaluate the fault diagnosis and failure prognosis routines. Performance metrics were applied at all levels of the architecture (data analysis/feature extraction, diagnosis, and prognosis).

7.1. Fault Predict. Using Eureka software and the model, we fitted a curve shown in Figure 14. The points represent 27 features extracted in 27 cycles. The red line is the model fitted by the feature. According to the deviation of the data itself, the model was enriched and data were generated with noise injected into the model. The data is shown in Figure 15. The red line is the model and the blue curve is the generated data.

Figures 16 and 17 show the results obtained when the proposed approach is applied to the problem of fault detection/prediction due to degradation of the bearing lubrication.
In Figure 19, the feature value is depicted as a function of time. It exhibits the fault growth (green line) and the set threshold, that is, the fault value at a predetermined level (red line). One hundred points in each time window are selected. The blue histogram represents the normal pdf. The blue line is selected as the threshold. 5% of the normal pdf is on the right of the threshold implying that type I error is 5%. The red histogram represents the current pdf. 6% of the current pdf is on the left of the threshold. So, type II error is 6%. Thus, the corresponding confidence in the declaration of the fault condition is 90% meaning that after 998 cycles, there is 94% possibility that the test platform is experiencing a fault.

7.2. Failure Prognosis. For prediction purposes the model is tuned first, as shown in Figure 18. The distance of the bearing run in miles corresponds to time in this case. 5000 particles are used for resampling purposes. The red line corresponds to the measurement data. The blue dot line represents the estimated data.

In Figure 19, the red "line" is the threshold value. The threshold or hazard zone is actually a probability density function (pdf) set by the user and representing the latter's conception from past experiences of the fault dimension a failure event is imminent and must be corrected. Multiple hazard zones may be specified. The histogram along the time/mileage axis (approximate pdf) represents the time or miles traveled to failure. Statistics, such as the mean and standard deviation, are computed to provide the user with useful information, that is, when action should be taken to repair the failing component (earlier or later than the mean time depending on the risk the user is willing to take under the prevailing circumstances). Appropriate performance metrics are defined and employed relating to the prediction accuracy [1].

8. Conclusions

Prognosis and health management is an important subject worthy of study for automated machining processes. This paper introduces a framework for the process, fault type, and assessment and testing of a critical component from an automated machining process based on novel particle filtering algorithms. Fault diagnosis has been shown to result in detection and isolation of a faulty machine component with user specified accuracy (confidence) and given false alarm rate. The prognostic approach, based on streaming data, an appropriate feature vector, a data-driven symbolic regression model, and a Bayesian estimation process-particle filtering, has been demonstrate to provide accurate results for the test case under consideration. As a result, the maintainer is provided with useful information as to when corrective action is required considering risk issues. The application example employs real fault/failure data derived from a seeded fault test in a bearing of the automated machining process. It provides excellent insight about the effect of model inaccuracies and customer specifications in the algorithm performance.
Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References


