Weihong Guo and Shenghan Guo
Rutgers University

Hui Wang
Ford Motor Company

Xiao Yu
Optimal CAE

Annette Januszczak and Saumuy Suriano
Ford Motor Company

Abstract

The wide applications of automatic sensing devices and data acquisition systems in automotive manufacturing have resulted in a data-rich environment, which demands new data mining methodologies for effective data fusion and information integration to support decision making. This paper presents a new methodology for developing a diagnostic system using manufacturing system data for high-value assets in automotive manufacturing. The proposed method extends the basic attributes control charts with the following key elements: optimal feature subset selection considering multiple features and correlation structure, balancing the type I and type II errors in decision making, on-line process monitoring using adaptive modeling with control charts, and diagnostic performance assessment using shift and trend detection. The performance of the developed diagnostic system can be continuously improved as the knowledge of machine faults is automatically accumulated during production. An example of the analysis at one Ford production plant is provided to demonstrate the implementation of this methodology.

Keywords

Feature selection, attributes control chart, Bayesian inference, Markov chain Monte Carlo (MCMC) algorithm, process monitoring, manufacturing system

Introduction

Among manufacturing industries, the automotive industry is the technological trend setter - it keeps evolving on fast pace and is now at the forefront of data analytics. Different kinds of information sources and sensor signals are increasingly used for quality control,

prediction, decision making, and insight in the automotive industry. This creates a tremendous capability for researchers and engineers to access valuable information from various enterprise entities to improve their designs and operations. As the quantity of data has dramatically increased, thanks to the wide applications of low-cost and smart sensing devices along with fast and advanced computer systems, the challenges to fully realizing the value of data have increased and researchers are eager to meet those challenges.

There are a large number of assets and resources in an automotive manufacturing plant, among which are the high-value assets that this paper will focus on. High-value assets are those that are either expensive or mission-critical. Take the gantry system for example. A gantry is commonly used for the pick and place process for the automated transfer of a part from a feed conveyor, a buffer, or an upstream process, to a downstream process, for example, one of the multiple parallel Computer Numeric Control (CNC) systems for a machining operation. The gantry is mission critical in the sense that down time in the gantry will cause the CNCs to idle while waiting for parts, which will result in lost production time and serious impact on the entire production line. Normal production is interrupted from the start of the failure to the moment the gantry is returned to production. The total maintenance time of an asset includes notification time, while the asset waits for the engineer to come, and the time for the actual diagnosis, fix, and test. While the latter varies from a quick reset to a major repair, it is essentially inevitable. The time spent waiting, however, has been recognized as one of the more serious wastes in lean production systems [1]. If we could extract useful information from manufacturing data and gain more insights about the failures, it's very likely that we would be able to get a better

understanding of the manufacturing operation, to do prognostics, to allocate maintenance resources [2-3] in a more efficient manner, and, eventually, to reduce lost production time and improve productivity.

This motivates our research to develop a data-driven diagnostic system that facilitates effective monitoring of the manufacturing process, early detection of system anomalies, quick diagnosis of fault root causes, and intelligent maintenance planning and resource allocation. The huge amounts of recorded manufacturing data enable us to study and explore data-driven methods to efficiently and effectively monitor the process and provide prognostics and diagnostics in manufacturing analytics. This research assumes that features or quality characteristics have already been defined and extracted based on the physical and mathematical understanding of the manufacturing process. Hence, feature extraction will not be discussed in this paper. Instead, this study focuses on what can be learned from those features - finding out the underlying distribution of a feature, determining monitoring thresholds, balancing false positives and false negatives, handling time series data with multiple features and/or correlated features, updating the monitoring strategy continuously as new observations become available, and integrating multiple decision rules for process monitoring and change detection.

Data-driven diagnosis and prognosis have been widely accepted in industrial practice [4, 5, 6, 7] including semiconductor manufacturing [8-9], the steel industry [10, 11, 12], and the chemical industry [13-14]. Many studies in this area, however, focus on multivariate statistics for extracting variables from complex raw data. The concept and application of statistical control charts have been widely studied [15], while (i) balancing false positives and false negatives and (ii) integrating control charts with feature selection and adaptive modeling are usually not handled in the areas of traditional control charts design.

The objective of this paper is to develop a diagnostic system using manufacturing data for high-value assets in automotive manufacturing. The proposed method extends the basic attributes control chart with the following key elements: optimal feature subset selection considering multiple features and correlation structure, balancing the type I and type II errors in decision making, on-line process monitoring using adaptive modeling with control chart, and diagnostic performance assessment using process shift and trend detection. The performance of the developed diagnostic system can be continuously improved as the knowledge of machine faults is automatically accumulated during production.

The remainder of this paper is organized as follows. An overview of the proposed methodology is given in the next section, followed by detailed research procedures for the methodology development. After that, an example using data from one Ford production plant is provided to demonstrate the implementation of the proposed method. Finally, the last section concludes the paper and discusses future work.

Methodology Overview

The generic procedures of the proposed methodology are shown in Figure 1. The methodology starts with the basic control chart once a training dataset is collected. The width of control limits is determined by considering the tradeoff between the required type I error rate and expected type II error rate. To improve the effectiveness and efficiency

of diagnostic system, multiple observations that are collected in time order of the production will be used in monitoring, and thus the next step is feature subset selection. As more and more knowledge about the process faults can be accumulated during the use of the process monitoring system, that knowledge is then used to further improve the diagnostic performance. For this purpose, adaptive modeling of the feature and adaptive control limits will be addressed.
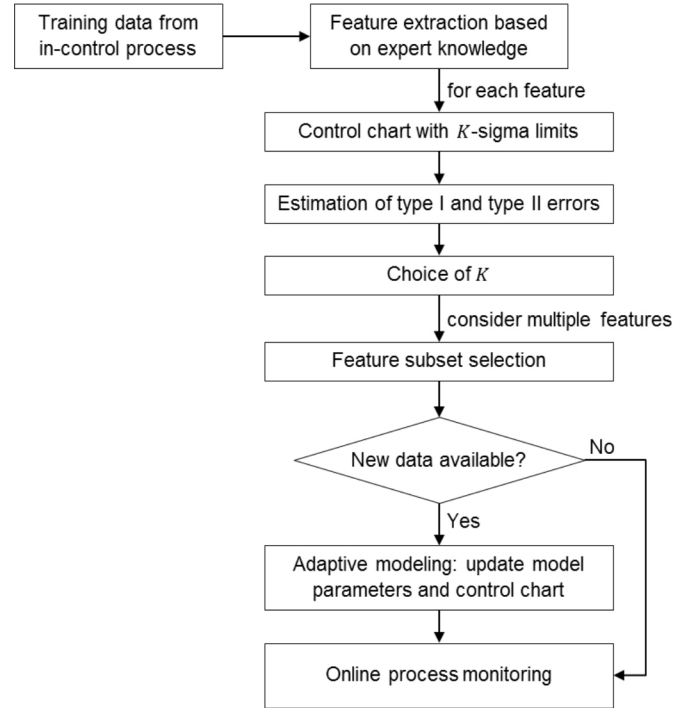

Figure 1. The framework of diagnostic system development with adaptive learning

In order to develop this new methodology, the following issues will be discussed in the paper: (1) Determining the underlying probability distribution of a feature using the goodness-of-fit test, based on which the basic control charts are designed; (2) Determining control limits, which balances the type I and type II errors; (3) Optimal feature subset selection by minimizing a weighted summation of the type I and type II errors associated with the features considered; two selection strategies are proposed; (4) Adaptive modeling of the feature distribution using Bayesian inferences; (5) Online process monitoring with adaptive control limits and sensitizing rules. The detailed procedures and algorithms will be discussed in the next section.

Methodology Development

The Basic Attributes Control Chart for Fault Occurrences

Many quality characteristics in manufacturing can be conveniently represented as attributes, for example, the proportion of warped automobile engine connecting rods in a day's production, the number of nonconformities in a subassembly area, and the number of maintenance requests that require a second call to complete the repair. In this study, we focus on attribute features such as the occurrence of a machine fault or warning or repair within a certain period of time. Some examples of these attributes are the number of assembly errors in a day or in a rolling 24-hour period, the number of low oil level warnings in a week, and the number of repairs in a month. When the

manufacturing process is stable and in statistical control, only chance causes of variation will be present and thus the quality characteristics are more likely to fall within a certain range [15]. If the process is operating in the presence of assignable causes, the process will be unstable, out of control, and requires actions; the quality characteristics are more likely to fall outside the control limits.

Take the occurrence of a machine fault in a high-value asset in a rolling 24-hour period as an example. In most occasions this fault can be fixed by a quick reset which takes no longer than a few seconds. However, the down times accumulate and thus too many resets would lead to lost production time. In the attributes control chart, we would like to detect high occurrences of this fault, while assuming the process is in control if the occurrences appear to be random and are inside the control limits. Let $X_1, X_2, \ldots, X_n$ denote the occurrences observed in each 24-hour window when the process is under normal operating conditions. When the process is in control, it is usually assumed that the occurrence of nonconformities in samples of constant size is well modeled by the Poisson distribution. Without loss of generality, we model the data with a Poisson distribution and estimate the Poisson parameter λ using the Maximum Likelihood Estimation (MLE) [16]:

$$\hat{\lambda} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \ ,$$

(1)

where $\bar{X}$ is the sample mean. We then apply the chi-square goodness-of-fit test [16] to test if the observed data $X_1, X_2, \ldots, X_n$ indeed come from the Poisson distribution with parameter $\hat{\lambda}$, denoted as Poi($\hat{\lambda}$).

H_0:   The sample data $X_1, X_2, \ldots, X_n$ follow Poi($\hat{\lambda}$) distribution.

H_a: The sample data $X_1, X_2, \ldots, X_n$ do not follow Poi($\hat{\lambda}$) distribution.

The p-value from the chi-square goodness-of-fit test is then examined to decide whether to reject the null hypothesis. If the p-value is less than or equal to a chosen significance level, for example, 0.05, the test suggests that the observed data is inconsistent with the null hypothesis and thus H_0 should be rejected. The control chart for fault occurrences will have control limits determined by the percentiles of the correct underlying distribution. These percentiles could be obtained from a histogram if a large sample (at least 100 observations) is available, or from a probability distribution fit to the data.

If $p > 0.05$, the null hypothesis will not be rejected. It is then statistically reasonable to assume that $X_1, X_2, \ldots, X_n$ follow the Poi($\hat{\lambda}$) distribution. A control chart for fault occurrences can be constructed in a similar procedure as the control chart for nonconformities, also known as the c chart. The attributes control chart with K-sigma limits is defined as follows:

$$\text{UCL} = \hat{\lambda} + K\sqrt{\hat{\lambda}}$$
$$\text{Center line} = \hat{\lambda}$$
$$\text{LCL} = \hat{\lambda} - K\sqrt{\hat{\lambda}}$$

(2)

where UCL is the upper control limit, LCL is the lower control limit, $K\sqrt{\hat{\lambda}}$ is the distance from the center line to control limits. Should the LCL calculation in (2) yields a negative value, set LCL = 0.

Choice of Control Limits: Balancing Type I and Type II Errors

One of the critical decisions in designing the control chart for fault occurrences is specifying the control limits. As expressed in Eq. (2), the value K determines how far the control limits are from the center line. The choice of control limits will then determine the decision errors in the control chart monitoring system. The type I error occurs when the control chart shows a point falling beyond the control limits or points forming non-random patterns, when no assignable cause of variation is present; the type II error occurs when the control chart fails to detect an out-of-control process.

In automobile manufacturing, type I errors result in unnecessary maintenance efforts, manual inspections, and loss of production time. If a monitoring system has too many false positives, the plant engineers will tend to ignore the alarms, distrust the monitoring decisions, and eventually abandon the monitoring system. The risk of a type I error can be decreased if we select a larger K and move the control limits farther from the center line. However, widening the control limits will increase the risk of a type II error. Type II errors would result in misdetections of process shifts and delays in actions, potentially storing up trouble for the future. If we use a smaller K and move the control limits closer to the center line, the risk of type II error can be decreased, while the risk of type I error will be increased. Therefore, finding the optimal K that balances the type I and type II errors becomes a critical issue.

Assume that the fault occurrence x follows a Poisson distribution with parameter λ_0 when the process is in control and a Poisson distribution with parameter λ_1 when the process is out of control. The type I error from the control chart defined in Eq. (2) is expressed as:

$$\alpha = P(x \geq \text{UCL}|x \sim \text{Poi}(\lambda_0)) + P(x \leq \text{LCL}|x \sim \text{Poi}(\lambda_0)),$$

(3)

where λ_0 can be estimated by Eq. (1). The type II error is expressed as:

$$\beta = P(\text{LCL} < x < \text{UCL}|x \sim \text{Poi}(\lambda_1)),$$

(4)

where $\lambda_1 = \lambda_0 + \delta$ and δ is the magnitude of the process mean shift.

In order to find the optimal K that balances the type I and type II errors, we use Monte Carlo simulation with N iterations to estimate the type I and type II errors. Specifically, in the jth iteration, $j = 1, 2, \ldots, N$, we generate a sample of n independent and identically Poi(λ_0) distributed random values representing the fault occurrences in a 24-hour window when the process is in control, that is, $X_{ij}^0 \sim \text{Poi}(\lambda_0), i = 1, 2, \ldots, n$; we also generate n independent and identically Poi(λ_1) distributed random values representing the fault occurrences in a 24-hour window when the process is out of control, that is, $X_{ij}^1 \sim \text{Poi}(\lambda_1), i = 1, 2, \ldots, n$. The estimated type I error from the jth iteration is

$$\hat{\alpha}_j = \frac{1}{n}\sum_{i=1}^{n}\left[1 - I(\text{LCL} < X_{ij}^0 < \text{UCL}|X_{ij}^0 \sim \text{Poi}(\lambda_0))\right],$$

(5)

where $I(\cdot)$ represents an indicator function. The estimated type II error from the jth iteration is

$$\hat{\beta}_j = \frac{1}{n}\sum_{i=1}^{n} I(\text{LCL} < X_{ij}^1 < \text{UCL} | X_{ij}^1 \sim \text{Poi}(\lambda_1)).$$

(6)

With N iterations in Monte Carlo simulation, we obtain an unbiased estimate for α as

$$\hat{\alpha} = \frac{1}{N}\sum_{j=1}^{N} \hat{\alpha}_j$$

(7) and an unbiased estimate for β as

$$\hat{\beta} = \frac{1}{N}\sum_{j=1}^{N} \hat{\beta}_j.$$

(8)

We illustrate how $\hat{\alpha}$ varies with K and how $\hat{\beta}$ varies with K and δ in the operating-characteristic (OC) curves for the attributes control chart for fault occurrences. Based on the OC curves, we select an interval for K where $\hat{\alpha}$ is smaller than or equal to a chosen threshold α_0. The endpoints of the interval are

$$K_L = \min\{K | \hat{\alpha} \leq \alpha_0\},$$

$$K_U = \max\{K | \hat{\alpha} \leq \alpha_0\}.$$

(9)

The optimal K will then be selected as the value that minimizes $\hat{\beta}$:

$$K^* = \arg\min_{K \in [K_L, K_U]} \hat{\beta}\ .$$

(10)

Feature Subset Selection

One of the standard assumptions in conventional control charts is that the data generated by the process when it is in control are independently distributed. Conventional control charts may not work well if the quality characteristic exhibits even low levels of correlation over time. Unfortunately, this independence assumption is not even approximately satisfied in many manufacturing processes. It is also well acknowledged that all manufacturing processes are driven by inertial elements, and when the interval between samples becomes small relative to these forces, the observations on the process will be correlated over time.

While the underlying correlation structure may be too complicated to be modeled, we explore the possibility of monitoring multiple observations that are taken in time order of the production to improve the effectiveness and efficiency of the diagnostic system. We would like to use the features observed prior to time t to predict if the process will be in control at time t. An example of the feature set is the fault occurrences observed in the l 24-hour windows prior to the decision time t. In order to remove redundant or irrelevant features, reduce overfitting, and simplify the monitoring system, we need to select an optimal feature subset that gives the best predictive power.

Let t denote the time when a prediction on process status needs to be made. Consider the features from the W windows prior to t. We propose two methods in selecting the optimal feature subset from the W features.

Method 1
Selecting w consecutive windows out of the W windows

In this method, we want to select the most recent w consecutive windows out of W such that the feature subset will minimize the cost function in Eq. (11), in which w is to be determined. This cost function is a weighted summation of the type I and type II errors associated with the features considered:

$$\min_{w \in \mathcal{W}} c_1 \alpha_w + c_2 \beta_w,$$

(11)

where $\mathcal{W} = \{1, 2, \dots, W\}$ is the set of all possible values for w, $c = (c_1, c_2)$ is the cost vector, $c_1 + c_2 = 1$. The cost vector is used to balance the type I and type II errors since they may not be equally important in real applications. We recommend a large value of c_1 if false alarms are more expensive and a large c_2 if misdetections are more expensive. α_w and β_w denote the type I and type II errors calculated for the most recent w consecutive windows, respectively. They can be computed using conditional probabilities as:

$$\hat{\alpha}_w = P(\cup_{l=1}^{w} f_l \geq \text{UCL} \mid \text{process in control})$$

$$+ P(\cup_{l=1}^{w} f_l \leq \text{LCL} \mid \text{process in control}) , \text{ and}$$

(12)

$$\hat{\beta}_w = P(\cap_{l=1}^{w} \text{LCL} < f_l < \text{UCL} \mid \text{process out of control}) ,$$

(13)

where f_l is the feature from window l. The rationale behind our interest in the most recent w consecutive windows is straightforward. The closer the features are to the decision time t, the more predictive power they may have. If the observations on the process are less correlated, then the cost function will be a non-decreasing function of w. As w increases, we consider more prior windows and thus the accuracy of monitoring decisions will degenerate. In that case, the optimal feature for prediction is exactly the feature from the most recent one window prior to t.

Method 2
Selecting w individual windows out of the W windows

In this method we want to select the best combination of w individual windows out of W such that the feature subset will minimize the cost function in Eq. (14), in which an index set of size w is to be determined. Similar to Method 1, this cost function is also a weighted summation of the type I and type II errors associated with the features considered. The number of w-combination from set $\mathcal{W} = \{1, 2, \dots, W\}$ of W elements is denoted as $\binom{W}{w}$, which is calculated by $\frac{W!}{w!(W-w)!}$. Let Ω_w denote the set of w-combination indices. The size of Ω is $\binom{W}{w}$. Let ϖ denote an element in Ω_w, $\varpi \in \Omega_w$. ϖ then represents the ϖth w-combination to form w individual windows out of the W windows. The cost function is defined as follows:

$$\min_{\varpi \in \Omega_w} c_1 \alpha_\varpi + c_2 \beta_\varpi,$$

(14)

where α_ϖ and β_ϖ denote the type I and type II errors calculated for the ϖth w-combination from the W windows prior to decision time t, respectively. Denote the index set of w selected features by $\mathcal{F}$, then they can be computed using conditional probabilities as:

$$\hat{\alpha}_w = P(\cup_{l \in \mathcal{F}}^w f_l \geq UCL \mid \text{process in control})$$

$$+P(\cup_{l \in \mathcal{F}}^w f_l \leq LCL \mid \text{process in control}) \text{, and}$$

$$\text{(15)}$$

$$\hat{\beta}_w = P(LCL < \cap_{l \in \mathcal{F}}^w f_l < UCL \mid \text{process out of control}) \text{,}$$

$$\text{(16)}$$

where f_l is the feature from window l. The optimal feature subset can be found by greedy algorithms.

Adaptive Modeling

As the knowledge of process faults is automatically accumulated during production, we are interested in developing an adaptive model to continuously improve the performance of the diagnostic system. The classic Markov Chain Monte Carlo (MCMC) algorithm [17] is extended for this purpose. Specifically, we adopt the Metropolis-Hastings (MH) algorithm [17-18] in data modeling and continuously update our estimate of the parameter(s) of the underlying distribution as new data become available.

Take the fault occurrence in a 24-hour window as an example. The initial estimate of the Poisson distribution parameter $\hat{\lambda}_0$ is calculated based on a training dataset, using Eq. (1). For a typical MCMC algorithm, we assume an acceptance rate of

$$\rho(X_t, X_{t-1}) = \min\left\{\frac{\pi(X_t; \lambda)}{\pi(X_{t-1}; \lambda)} \frac{q(X_{t-1}|X_t; \lambda)}{q(X_t|X_{t-1}; \lambda)}, 1\right\},$$

$$\text{(17)}$$

where X_t denotes the observations available at time t, whose underlying probability distribution depends on a parameter λ, and $\pi(X_t; \lambda)$ is the stationary distribution of the process. In the MH algorithm, we assume a symmetric proposal density $q(\cdot)$, i.e.,

$$X_t = X_{t-1} + \varepsilon \text{,}$$

$$\text{(18)}$$

where ε is a noise following a Normal distribution $N(0, \sigma_\varepsilon^2)$. This representation is reasonable since we assume that the random variable X_t will reach a steady probability distribution in the long run. Such an assumption leads to the following equality:

$$q(X_{t-1}|X_t; \lambda) = q(X_t|X_{t-1}; \lambda).$$

$$\text{(19)}$$

The acceptance rate can be then written as

$$\rho(X_t, X_{t-1}) = \min\left\{\frac{\pi(X_t; \lambda)}{\pi(X_{t-1}; \lambda)}, 1\right\}.$$

$$\text{(20)}$$

It is noted that the acceptance rate is a ratio of two probabilities, which are both very small and can cause problems in computation. In order to improve computation efficiency, we adopt the following calculation

$$\rho(X_t, X_{t-1}) = \min\left\{\exp\{\log(\pi(X_t; \lambda)) - \log(\pi(X_{t-1}; \lambda))\}, 1\right\}$$

$$\text{(21)}$$

instead of the original formulation in Eq. (20).

We give the pseudo code for the MH algorithm in Table 1. Let $D(t)$ denote the available samples at time t and $n(t)$ denote the sample size. Let $\hat{\lambda}_{old}$ denote the previous estimate of the parameter from one iteration and $\hat{\lambda}_{new}$ as the new estimate. In addition, let $L(D(t))$ represent the log-likelihood function of the feature.

Table 1. Pseudo code for the application of Metropolis-Hastings algorithm in data modeling for the fault occurrences in 24-hour windows

---

initial parameter $\hat{\lambda}_0$

initial log-likelihood function at t_0:

$L(D(t_0)) = \exp\left\{\sum_{j=1}^{n(t_0)} x_j \log\hat{\lambda}_0 - n(t_0)\hat{\lambda}_0 - \sum_{j=1}^{n(t_0)} \log(x_j!)\right\}$ (22)

for $i = 1, ..., \text{total sample size}$ **do**

  Calculate $\hat{\lambda}_{new}$ using Eq. (1)

  $L(D(t_{i+1}))$
  $= \exp\left\{\sum_{j=1}^{n(t_{i+1})} x_j \log\hat{\lambda}_{new} - n(t_{i+1})\hat{\lambda}_{new} - \sum_{j=1}^{n(t_{i+1})} \log(x_j!)\right\}$ (23)

  $L(D(t_i))$
  $= \exp\left\{\sum_{j=1}^{n(t_i)} x_j \log\hat{\lambda}_{old} - n(t_i)\hat{\lambda}_{old} - \sum_{j=1}^{n(t_i)} \log(x_j!)\right\}$ (24)

  $\rho_i = \min\left\{\exp\left(L(D(t_{i+1})) - L(D(t_i))\right), 1\right\}$ (25)

  Generate a random number $U \sim \text{Unif}(0, 1)$

  if $U < \rho_i$ **then**

    Set $\hat{\lambda}_{old} = \hat{\lambda}_{new}$

  else

    Keep $\hat{\lambda}_{old}$

  end if

end for

The utilization of the MH algorithm provides us with a close estimation of the distribution parameter given sufficient data. For an in-control process, after enough MH iterations, we should be able to reach the true distribution and the point estimates of the distribution parameter would be almost identical in iterations after the true distribution is reached.

Online Process Monitoring

Instead of using the basic attributes control chart for process monitoring, we propose to use the adaptive control chart for attributes, especially fault occurrences. While we still use Eq. (2) for

calculating the control limits, the Poisson parameter $\hat{\lambda}$ is continuously updated using the MH algorithm as new observations of process faults accumulate during production.

To improve the detection of nonrandom patterns on the adaptive control chart, we apply multiple decision rules as suggested by the Western Electric Handbook [15, 19]. Specifically, the process is concluded to be out of control if either

1. One or more points outside of the control limits,
2. Two of three consecutive points outside the two-sigma warning limits but still inside the control limits,
3. Four of five consecutive points beyond the one-sigma limits, or
4. A run of eight consecutive points on one side of the center line.

Note that Rule 1 can be applied whenever a new observation is available, while Rules 2, 3, and 4 require more than one point on the control chart. With the adaptive control chart, we apply only Rule 1 when we have no more than 3 new samples, apply only Rules 1 and 2 when there are no more than 4 new samples, apply Rules 1, 2, and 3 when there are no more than 8 new samples, and we apply all four rules only when there are at least 9 new samples.

Example: A Data-Driven Diagnostic System for Powertrain Manufacturing

Data Description

We choose a gantry system from one Ford transmission plant for the study. The collected dataset includes faults, warnings, and repairs of the gantry from April 2014 to August 2016. The dataset contains 32 types of machine faults; their occurrences in the 28-month period vary from one to a few hundreds. For the purpose of demonstrating our proposed methodology, we present the data and analysis results for one of the critical faults, coded as the "F1" fault.

There is a total of 302 occurrences of the F1 fault in the 28-month data collection period. Figure 2 shows these faults in the time order that they occurred. The vertical axis represents the repair time for each F1 occurrence in seconds. It can be seen from Figure 2 that most repairs are very quick, but the down times accumulate and result in lost production time. Thus, the feature of our focus is the occurrence of the F1 fault in a rolling 24-hour window, as illustrated in Figure 3.
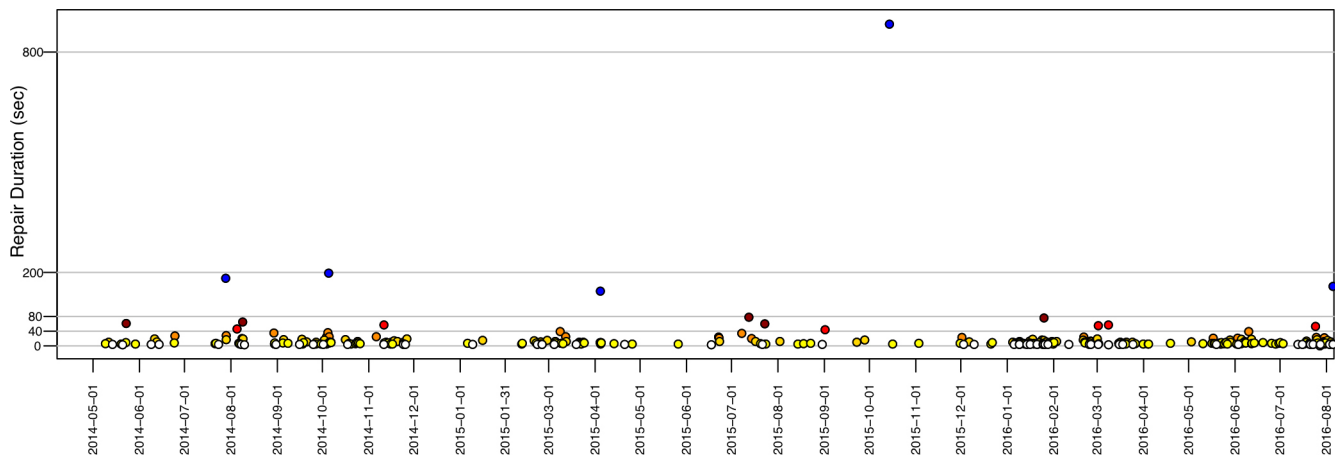

Figure 2. F1 faults and their repair durations in the data collection period
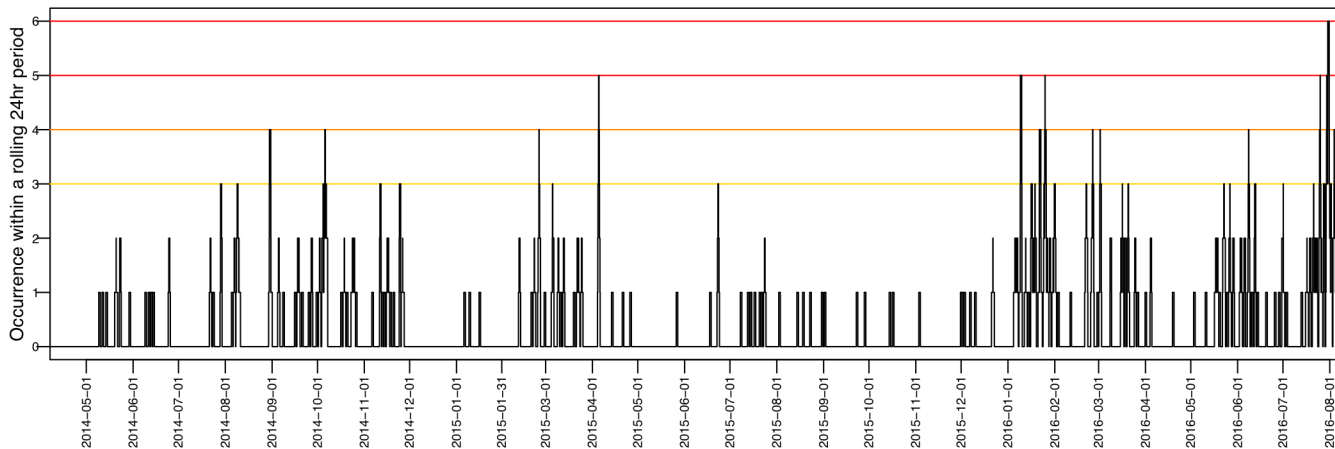

Figure 3. Occurrences of the F1 fault in rolling 24-hour windows

The rolling windows are incremented based on the start time and end time of each F1 fault and the time that is 24 hours after a F1 fault ends. In order to avoid repetitive counts for an occurrence, we consider a 24-hour window instead of a rolling 24-hour window. The entire dataset S is then partitioned into 857 subsamples based on 24-hour intervals, with each subsample producing one count of the F1 fault occurrence, $S = \{X_1, X_2,\ldots, X_{857}\}$. Since dataset S consists of observations from both the in-control and out-of-control processes, we need to remove the out-of-control subsamples before fitting a Poisson distribution to the data. Based on expert knowledge from Powertrain manufacturing engineers, we decided to set a threshold at 2, which is approximately the 97.5th percentile of the values in S, to partition S into two groups S^0 and S^1. Let $S^0 = \{X_i| X_i \leq 2; i = 1,2, \ldots,857\}$ and $S^1 = \{X_i| X_i > 2; i = 1,2, \ldots,857\}$ denote the subsamples from the in-control and out-of-control processes, respectively. With in-control data S^0, we use the Maximum Likelihood Estimation (MLE) shown in Eq. (1) to estimate the Poisson parameter and get $\hat{\lambda}$ = 0.2792363. Figure 4 shows the histogram of the values in S^0 with the fitted Poi($\hat{\lambda}$ = 0.2792363) distribution.
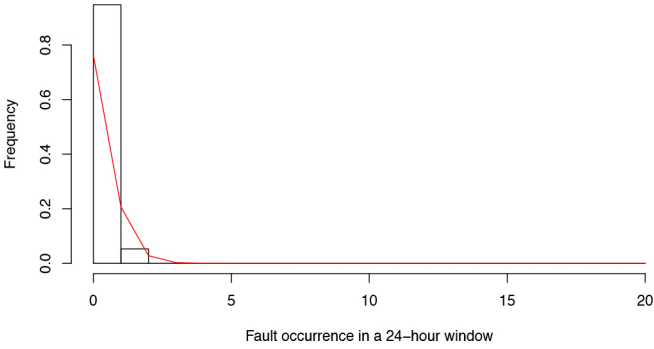
Figure 4. Histogram of S^0 with the Poi($\hat{\lambda}$ = 0.2792363) distribution fit

To test if the observed data in S^0 indeed come from the Poi($\hat{\lambda}$ = 0.2792363) distribution, the chi-square goodness-of-fit test calculates the test statistic as $\chi^2 = 40$ and degree of freedom as 38. The p-value of the goodness-of-fit test is 0.3814, which is greater than 0.05. Therefore, it is statistically reasonable to state that the data in S^0 follow the Poi($\hat{\lambda}$) distribution with $\hat{\lambda}$ = 0.2792363. Since S^0 is in-control data, this $\hat{\lambda}$ gives the baseline parameter λ_0 = 0.2792363.

Analysis Results

To find the optimal K for control limits, we apply Monte Carlo simulation with $N = 100$ iterations to estimate the type I and type II errors. In each iteration, $n = 10,000$ independent and identically Poi(λ_0) distributed random values are generated to represent the fault occurrences in a 24-hour window when the process is in control, where λ_0 = 0.2792363. An unbiased estimate of the type I error rate, $\hat{\alpha}$, is calculated from simulation and shown in Figure 5. It can be seen from Figure 5 that $\hat{\alpha}$ is non-increasing as K increases. When K is small and close to 0, the control limits are close to the center line, resulting in the worst type I error rate at approximately 0.244. The commonly used 3-sigma control limits give $\hat{\alpha}$ = 0.033 when $K = 3$; the 2-sigma limits also give $\hat{\alpha}$ = 0.033 when $K = 2$. Therefore, we recommend the choice of K to be as least 2.
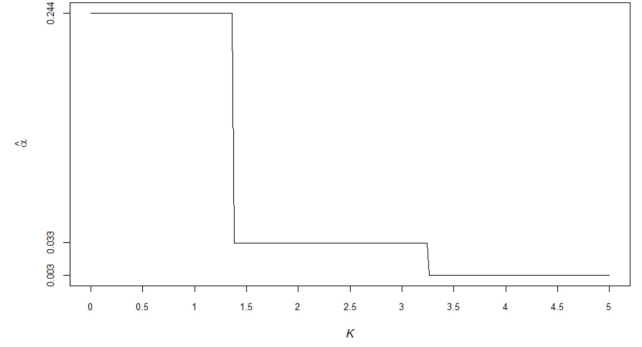
Figure 5. Type I error rate for the control chart for fault occurrences with K-sigma limits

To estimate the type II error rate, $n = 10,000$ independent and identically Poi(λ_1) distributed random values are generated to represent the fault occurrences in a 24-hour window when the process is out of control, where $\lambda_1 = \lambda_0 + \delta$ and δ is the magnitude of the process mean shift. Figure 6 plots the unbiased estimate of the type II error rate, $\hat{\beta}$, against δ for $K = 2$, 3, and 5. The worst type II error rate is approximately 0.243 when $\delta \in [0.36, 0.70]$. When $|\delta|$ gets larger, $\hat{\beta}$ decreases since larger process shifts are easier to detect. For a large shift $\delta \geq 2.72$, $\hat{\beta}$ approximates 0. It is also worth noting that $\hat{\beta}$ doesn't show significant differences when using different values of K. Based on simulation results, we come to the conclusion that $K = 2$, 3, and 5 are all good candidates for K. Taking $K = 3$ would then be preferred since it coincides with the widely used 3-sigma control limits in statistical quality control and thus easier to implement in practice.
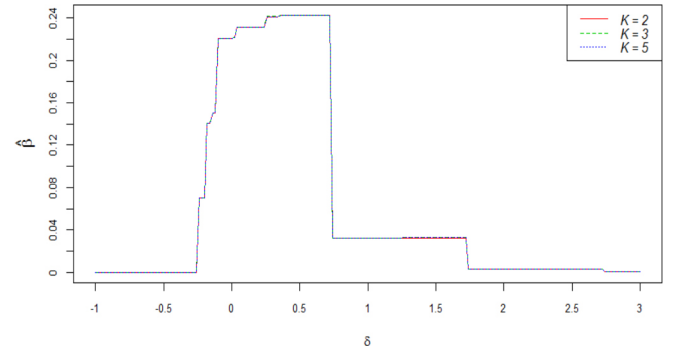
Figure 6. Operating-characteristic curves for the control chart for fault occurrences with K-sigma limits, where the process mean shifts by δ

In feature subset selection, we consider the features from $W = 20$ windows prior to a decision time t. The feature selection results from Method 1 are shown in Figure 7. The three curves in Figure 7 correspond to three different cost vectors $c_1 = (0.5, 0.5)$, $c_2 = (0.8, 0.2)$, and $c_3 = (0.95, 0.05)$. It is shown in Figure 7 that the cost function is an increasing function of w, which is consistent with the intuition that as we consider more windows, the accuracy of monitoring decisions will degenerate. Since we want to minimize the weighted summation of the type I and type II errors, the optimal w for Method 1 would be $w^* = 1$, i.e., we select the most recent one window out of the 20 windows prior to each decision time.
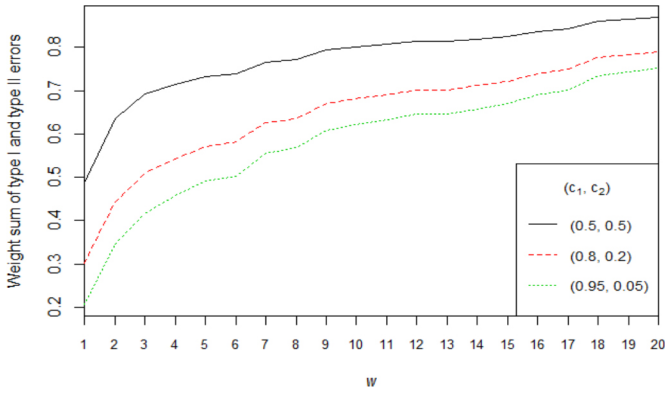
Figure 7. Feature subset selection Method 1: cost function (weighted summation of type I and type II errors) varies with w

In Method 2 of feature selection, we aim to select w individual windows out of the $W = 20$ windows prior to a decision time. When selecting w individual windows out of W, there is a total of $\binom{W}{w}$ possible combinations. Take $w = 3$, $w = 5$, and $w = 8$ for example. We have 1,140 combinations of individual windows when $w = 3$, 15,504 combinations when $w = 5$, and 125,979 combinations when $w = 8$. The feature selection results with the three different cost vectors are shown in Figure 8, where the horizontal axis represents the w-combination feature indices.

Specifically, Figure 8(a) shows that at $w = 3$, the smallest weighted summation of the type I and type II errors with cost vector c_1 is approximately 0.57, which is achieved at $\{8, 14, 17\}$ when combining the 8th, 14th, and 17th windows out of the twenty 24-hour windows prior to a decision time. Figure 8(b) shows that at $w = 3$, the minimums with cost vectors c_2 and c_3 are 0.36 and 0.25, respectively, both achieved at $\{1, 4, 8\}$ when combining the 1st, 4th, and 8th windows out of the 20 windows prior to the decision time.

Similarly, Figure 8(c) shows that at $w = 5$, the smallest weighted summation of the type I and type II errors with cost vector c_1 is approximately 0.64, which is achieved at $\{8, 10, 14, 15, 17\}$ when combining the 8th, 10th, 14th, 15th, and 17th windows out of the twenty 24-hour windows prior to a decision time. Figure 8(d) shows that at $w = 5$, the minimums with cost vectors c_2 and c_3 are both achieved at $\{1, 4, 6, 7, 8\}$ when combining the 1st, 4th, and 6th through 8th windows out of the 20 windows prior to the decision time. At $w = 8$, minimums with cost vectors c_1, c_2, and c_3 are achieved at the same combination $\{6, 7, 8, 10, 13, 15, 16, 17\}$, as shown in Figure 8(e).

In adaptive modeling, we use the Metropolis-Hastings (MH) algorithm in MCMC simulation to continuously update the estimate of the Poisson distribution parameter λ. Whenever a new F1 fault occurs, a new observation of the fault occurrence is recorded and our estimate of λ will be updated accordingly. Figure 9 shows how the estimated λ (solid line) approaches the true λ (dashed line) as new data become available during production. The true λ plotted in Figure 9 is the $\lambda_0 = 0.2792363$ estimated by MLE as stated earlier in this section.
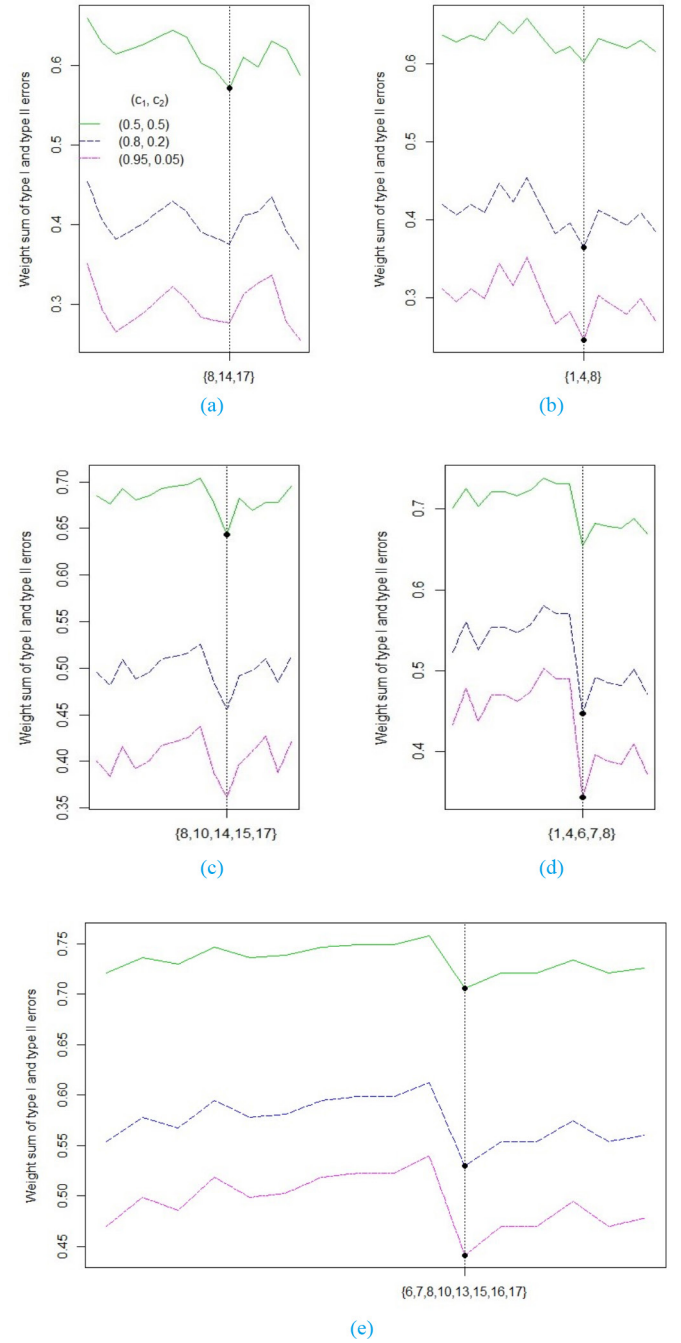

Figure 8. Feature subset selection Method 2: cost function (weighted summation of type I and type II errors) varies with different feature combinations. (a) $w = 3$ and c_1, (b) $w = 3$ and c_2 and c_3, (c) $w = 5$ and c_1, (d) $w = 5$ and c_2 and c_3, (e) $w = 8$ and c_1, c_2, and c_3
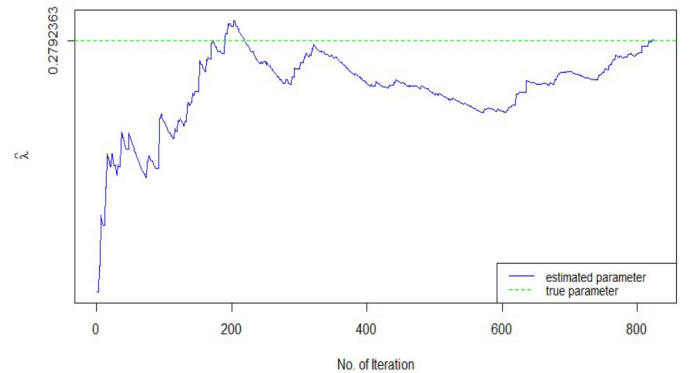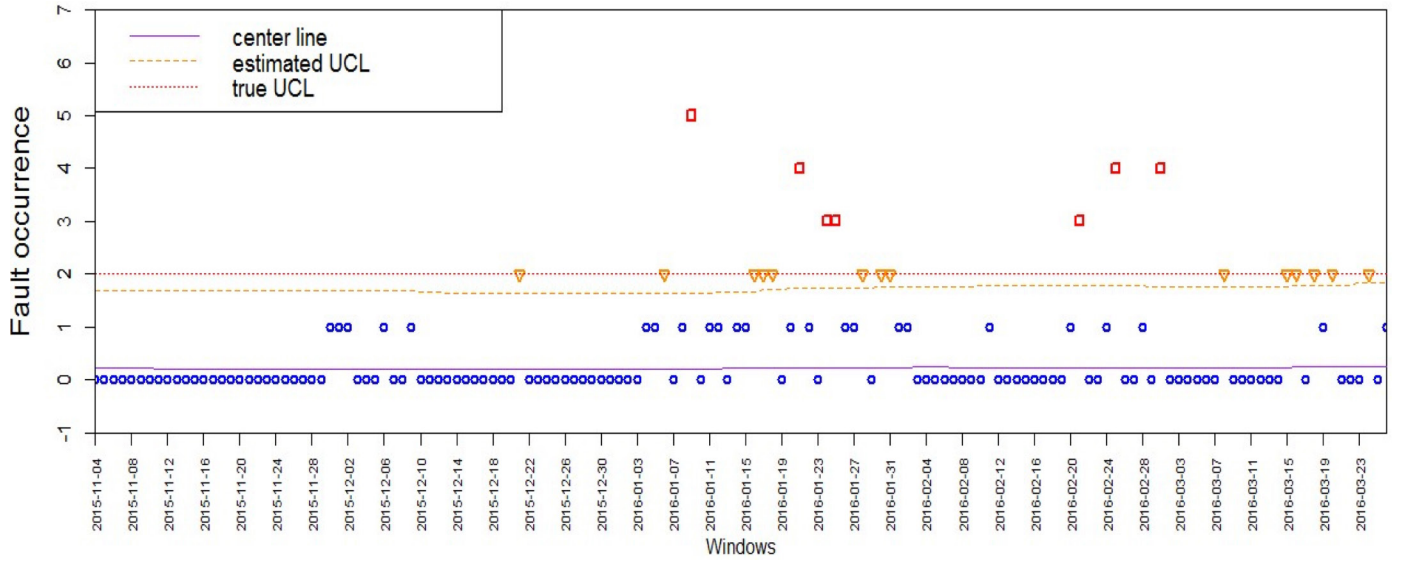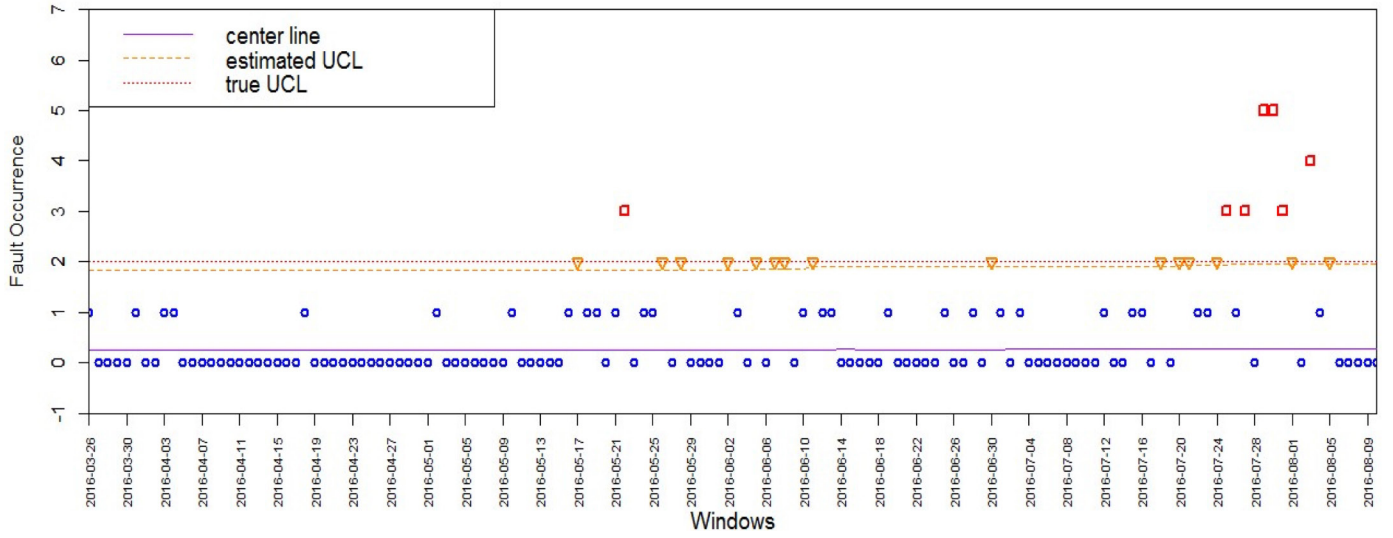

Figure 9. Adaptive modeling of the Poisson parameter λ

Figure 10. Adaptive control chart for the occurrence of the F1 fault in a 24-hour window: (a) from 2015-11-04 to 2016-03-26, (b) from 2016-03-26 to 2016-08-10

An adaptive control chart for the occurrence of the F1 fault in a rolling 24-hour window is constructed for online process monitoring and change detection. The center line and control limits are continuously updated using adaptive modeling as new observations of process faults accumulate during production. We present two segments of the entire control chart in Figure 10. The lower control limit (LCL) is 0 and hence not plotted in the chart. The true UCL is obtained based on $\lambda_0 = 0.2792363$; estimated UCL is calculated based on the $\hat{\lambda}$ obtained from adaptive modeling.

We notice that the estimated UCL approaches the true UCL as new data become available during production. The estimated UCL eventually converges to the true UCL when sufficient samples are available. We observed approximately 72 to 75 alarms from the entire control chart, representing possible process shifts. All these shifts were detected by Rule 4, *a run of eight consecutive points on one side of the center line*. Note that the number of alarms varies slightly when we run the MH algorithm multiple times due to the randomness in the algorithm.

Conclusion

In this paper, a new methodology is developed for developing a diagnostic system using manufacturing system data for high-value assets in automotive manufacturing. Assuming features or quality characteristics have already been defined and extracted based on expert knowledge, this research focuses on utilizing data and features for process monitoring, prognostics, and diagnostics. The result from this research will accomplish in-process and continuous diagnostic performance enhancement and improvement. A real case study of a gantry system at one Ford production plant is provided to demonstrate the effectiveness of the proposed methodology.

The proposed methodology brings in additional benefits to current manufacturing system monitoring. It is purely data-driven such that the utilization of data from routine production is maximized, which is rather desirable in the recent data-intensive manufacturing environment. More importantly, with its adaptive nature, the methodology is very sensitive

to quality changes in operation and can accommodate shifts automatically. This enables prompt diagnosis for the manufacturing system and potentially leads to lower costs in process monitoring. The results from the case study have shown good applicability of the method in an industrial environment, and demonstrated its effectiveness in adaptive modeling as well as shifts detection.

The developed data-driven diagnostic system is generic, which is applicable to nearly any high-value asset used in an industrial environment that has data collection capabilities. The proposed methodology can also be applied to the monitoring of many mission-critical processes. There are a few issues to be studied further in the methodology development. Examples of those open issues include: (1) the impact of subgroup size on the diagnostic performance, (2) the potential of using more sensitizing rules in adaptive control charts, and (3) the burn-in period for the Metropolis-Hastings algorithm and the efficiency of MCMC for online monitoring.

References

1. Askin, R.G. and Goldberg, J.B., *Design and Analysis of Lean Production Systems*. John Wiley & Sons, 2007.

2. Gu, X., Lee, S., Liang, X., Garcellano, M., Diederichs, M., and Ni, J., "Hidden Maintenance Opportunities in Discrete and Complex Production Lines." *Expert Systems with Applications*, 40(11): 4353-4361, 2013, doi: 10.1016/j.eswa.2013.01.016.

3. Guo, W., Jin, J., and Hu, S.J., "Allocation of Maintenance Resources in Mixed Model Assembly Systems," *Journal of Manufacturing Systems*, 32(3): 473-479, 2013, doi:10.1016/j.jmsy.2012.12.006.

4. Yin, S., Ding, S. X., Xie, X., and Luo, H., "A Review on Basic Data-Driven Approaches for Industrial Process Monitoring," *IEEE Transactions on Industrial Electronics*, 61(11): 6418-6428, 2014, doi: 10.1109/TIE.2014.2301773.

5. Qin, S.J., "Survey on Data-driven Industrial Process Monitoring and Diagnosis,", *Annual Reviews in Control*, 36(2): 220-234, 2012, doi: 10.1016/j.arcontrol.2012.09.004.

6. Gao, Z. and Dai, X., "From Model, Signal to Knowledge: A Data-driven Perspective of Fault Detection and Diagnosis," *IEEE Transactions on Industrial Informatics*, 9(4): 2226-2238, 2013, doi: 10.1109/TII.2013.2243743.

7. Chiang, L., Russell, E., and Braatz, R., *Fault Detection and Diagnosis in Industrial Systems*. London, U.K.: Springer-Verlag, 2001.

8. Yue, H., Qin, S., Markle, R., Nauert, C., and Gatto, M., "Fault Detection of Plasma Etchers using Optical Emission Spectra," *IEEE Transactions on Semiconductor Manufacturing*, 13(3): 374-385, 2000, doi: 10.1109/66.857948.

9. Cherry, G. and Qin, S. J., "Multiblock Principal Component Analysis based on a Combined Index for Semiconductor Fault Detection and Diagnosis," *IEEE Transactions on Semiconductor Manufacturing*, 19(2): 159-172, 2006, doi:10.1109/TSM.2006.873524.

10. Miletic, I., Quinn, S., Dudzic, M., Vaculik, V., and Champagne, M., "An Industrial Perspective in Implementing On-line Applications of Multivariate Statistics," *Journal of Process Control*, 14(8): 821-836, 2004, doi:10.1016/j.jprocont.2004.02.001.

11. Zhang, Y. and Dudzic, M. S., "Online Monitoring of Steel Casting Processes using Multivariate Statistical Technologies: From Continuous to Transitional Operations," *Journal of Process Control*, 16(8): 819-829, 2006, doi:10.1016/j.jprocont.2006.03.005.

12. Kano, M. and Nakagawa, Y., "Data-based Process Monitoring, Process Control and Quality Improvement: Recent Developments and Applications in Steel Industry," *Computers and Chemical Engineering*, 32(1-2): 12-24, 2008, doi:10.1016/j.compchemeng.2007.07.005.

13. Kosanovich, K., Dahl, K., and Piovoso, M., "Improved Process Understanding using Multiway Principal Component Analysis," *Industrial & Engineering Chemistry Research*, 35(1): 138-146, 1996, doi: 10.1021/ie9502594.

14. Russell, E., Chiang, L., and Braatz, R., *Data-Driven Methods for Fault Detection and Diagnosis in Chemical Processes*. London, U.K.: Springer-Verlag, 2000.

15. Montgomery, D.C., *Introduction to Statistical Quality Control*. Vol. 7. New York: Wiley, 2009.

16. Walpole, R. E., Myers, R. H., Myers, S. L., and Ye, K., *Probability and Statistics for Engineers and Scientists*. Vol. 5. New York: Macmillan, 1993.

17. Robert, C. and George C., *Monte Carlo Statistical Methods*. Springer Science & Business Media, 2013.

18. Hastings, W. K., "Monte Carlo Sampling Methods using Markov Chains and Their Applications," *Biometrika*, 57(1): 97-109, 1970, doi: 10.1093/biomet/57.1.97.

19. Western Electric Company, *Statistical Quality Control Handbook*, Indianapolis, Indiana: Western Electric Co., 1956.

Contact Information

Dr. Hui John Wang
Ford Global Data Insight & Analytics
Ford Motor Company, 15403 Commerce Dr. S, Dearborn, MI 48120
hwang168@ford.com
Phone: (313) 322-3032

Dr. Weihong (Grace) Guo
Department of Industrial and Systems Engineering, Rutgers - The State University of New Jersey
96 Frelinghuysen Rd, Piscataway, NJ 08854
wg152@rutgers.edu
Phone: (848) 445-8556

Acknowledgments