1

Introduction

1.1 Motivational Illustrations

Consider the following scenarios:

Scenario A

You are a plant operator, and a gas analyser reading triggers an alarm for a low level of a vital reaction component, but from experience you know that this gas analyser is prone to error. The difficulty is, however, that if the vital reaction component is truly scarce, its scarcity could cause plugging and corrosion downstream that could cost over $120 million in plant downtime and repairs, but if the reagent is not low, shutting down the plant would result in $30 million in downtime. Now, imagine that you have a diagnosis system that has recorded several events like this in the past, using information from both upstream and downstream, is able to generate a list of possible causes of this alarm reading, and displays the probability of each scenario. The diagnosis system indicates that the most possible cause is a scenario that happened three years ago, when the vital reagent concentration truly dropped, and by quickly taking action to bypass the downstream section of the plant a $120-million incident was successfully avoided. Finally, imagine that you are the manager of this plant and discover that after implementing this diagnosis system, the incidents of unscheduled downtime are reduced by 60% and that incidents of false alarms are reduced by 80%.

Scenario B

You are the head of a maintenance team of another section of the plant with over 40 controllers and 30 actuators. Oscillation has been detected in this plant, where any of these controllers or actuators could be the cause. Because these oscillations can push the system into risky operating regions, caution must be exercised to keep the plant in a safer region, but at the cost of poorer product quality. Now, imagine you have a diagnosis tool that has data recorded from previous incidents, their troubleshooting solutions, and the probabilities of each incident.

Process Control System Fault Diagnosis: A Bayesian Approach, First Edition. Ruben Gonzalez, Fei Qi and Biao Huang.

With this tool, we see that the most probable cause (at 45%) was fixed by replacing the stem packing on Valve 23, and that the second most probable cause (at 22%) was a tank level controller that in the past was sometimes overtuned by poor application of tuning software. By looking at records, you find out that a young engineer recently used tuning software to re-tune the level controller. Because of this information, and because changing the valve packing costs more, you re-tune the controller during scheduled maintenance, and at startup find that the oscillations are gone and you can now safely move the system to a point that produces better product quality. Now that the problem has been solved, you update the diagnosis tool with the historical data to improve the tool's future diagnostic performance. Now imagine, that as the head engineer of this plant, you find out that 30% of the most experienced people on your maintenance team are retiring this year, but because the diagnostic system has documented a large amount of their experience, new operators are better equipped to figure out where the problems in the system truly are.

Overview

These stories paint a picture of why there has been so much research interest in fault and control loop diagnosis systems in the process control community. The strong demand for better safety practices, decreased downtime, and fewer costly incidents (coupled with the increasing availability of computational power) all fuel this active area of research. Traditionally, a major area of interest has been in detection algorithms (or *monitors* as they will be called in this book) that focus on the behaviour of the system component. The end goal of implementing a monitor is to create an alarm that would sound if the target behaviour is observed. As more and more alarms are developed, it becomes increasingly probable that a single problem source will set off a large number of alarms, resulting in an alarm flood. Such scenarios in industry have caused many managers to develop alarm management protocols within their organizations. Scenarios such as those presented in scenarios A and B can be realized and in some instances have already been realized by research emphasizing the best use of information obtained from monitors and historical troubleshooting results.

1.2 Previous Work

1.2.1 Diagnosis Techniques

The principal objective in this book is to diagnose the operational mode of the process, where the mode consists of the operational state of all components within the process. For example, if a system comprises a controller, a sensor and a valve, the mode would contain information about the controller (e.g. well tuned or poorly tuned), the sensor (e.g. biased or unbiased) and the valve (e.g. normal or sticky). As such, the main problem presented in this book falls within the scope of *fault detection and diagnosis*.

Fault detection and diagnosis has a vast (and often times overwhelming) amount of literature devoted to it for two important reasons:

1. The problem of fault detection and diagnosis is a legitimately difficult problem due to the sheer size and complexity of most practical systems.

2. There is great demand for fault detection and diagnosis as it is estimated that poor fault management has cost the United States alone more than $20 billion annually as of 2003 (Nimmo 2003).

In a three-part publication, Venkatasubramanian et al. (2003b) review the major contributions to this area and classify them under the following broad families: quantitative model-driven approaches (Venkatasubramanian et al. 2003b), qualitative model-driven approaches (Venkatasubramanian et al. 2003a), and process data-driven approaches (Venkatasubramanian et al. 2003c). Each type of approach has been shown to have certain challenges. Quantitative model-driven approaches require very accurate models that cover a wide array of operating conditions; such models can be very difficult to obtain. Qualitative model-driven approaches require attention to detail when developing heuristics, or else one runs the risk of a spurious result. Process data-driven approaches have been shown to be quite powerful in terms of detection, but most techniques tend to yield results that make fault isolation difficult to perform. In this book, particular interest is taken in the quantitative model-driven and the process data-driven approaches.

Quantitative Model-driven Approaches

Quantitative model-driven approaches focus on constructing the models of a process and using these models to diagnose different problems within a process (Lerner 2002) (Romessis and Mathioudakis 2006). These techniques bear some resemblance to some of the monitoring techniques described in Section 1.2.2 applied to specific elements in a control loop. Many different types of model-driven techniques exist, and have been broken down according to Frank (1990) as follows:

1. *The parity space approach* looks at analytical redundancy in equations that govern the system (Desai and Ray 1981).
2. *The dedicated observer and innovations approach* filters residual errors from the Parity Space Approach using an observer (Jones 1973).
3. *The Fault Detection Filter Approach* augments the State Space models with fault-related variables (Clark et al. 1975; Willsky 1976)
4. *The Parameter Identification Approach* is traditionally performed offline (Frank 1990). Here, modeling techniques are used to estimate the model parameters, and the parameters themselves are used to indicate faults.

A popular subclass of these techniques is deterministic fault diagnosis methods. One popular method in this subclass is the parity space approach (Desai and Ray 1981), which set up parity equations having analytical redundancy to look at error directions that could correspond to faults. Another popular method is the observer-based approach (Garcia and Frank 1997), which uses an observer to compare differences in the predicted and observed states.

Stochastic techniques, in contrast to deterministic techniques, use fault-related parameters as augmented states; these methods enjoy the advantage of being less sensitive to process noise (Hagenblad et al. 2004), being able to determine the size and precise cause of the fault, but are very difficult to implement in large-scale systems and often require some excitement

(Frank 1996). Including physical fault parameters in the state often requires a nonlinear form of the Kalman filter (such as the extended Kalman filer (EKF), unscented Kalman filter (UKF) or particle filter) because these fault-related parameters often have nonlinear relationships with respect to the states. Such techniques were pioneered by Isermann (Isermann and Freyermuth 1991), (Isermann 1993) with other important contributions coming from Rault et al. (1984). The motivation for including fault parameters in the state is the stochastic Kalman filter's ability to estimate state distributions. By including fault parameters in the state, fault parameter distributions are automatically estimated in parallel with the state. Examples of this technique include that of Gonzalez et al. (2012), which made use of continuous augmented bias states, while Lerner et al. (2000) made use of discrete augmented fault states.

Process Data-driven Approaches

A popular class of techniques for process monitoring are data-driven modeling methods, where one of the more popular techniques is principal component analysis (PCA) (Ge and Song 2010). These techniques create black-box models assuming that the data can be explained using a linear combination of independent Gaussian latent variables (Tipping and Bishop 1998); a transformation method is used to calculate values of these independent Gaussian variables, and abnormal operation is detected by performing a significance test. The relationship between abnormal latent variables and the real system variables is then used to help the user determine what the possible causes of abnormality could be. There have also been modifications of the PCA model to include multiple Gaussian models (Ge and Song 2010; Tipping and Bishop 1999) where the best local model is used to calculate the underlying latent variables used for testing.

All PCA models assume that the underlying variables are Gaussian, but more recent methods (Lee et al. 2006) do away with this assumption by first using independent components analysis (ICA) to calculate values of independent latent variables (which are not assumed to be Gaussian under ICA) and then using a kernel density estimation to evaluate the probability density of that value. Low probability densities indicate that the process is behaving abnormally. Even more recent work (Gonzalez et al. 2015) uses Bayesian networks instead of PCA/ICA to break down the system into manageable pieces; this allows the user to define variables of interest for monitoring and determine the causal structures used to help narrow down causes. Abnormality is detected if key process variables take on improbable values or if groups of key process variables take on improbable patterns. Results from this approach are generally easier to interpret than PCA/ICA-based methods.

Bayesian Data-driven Approaches

This book focuses on using the Bayeisan data-driven approach, which is distinct from other fault detection and diagnosis methods, mainly for the reason that the Bayesian approach is a *higher-level diagnosis method* (Pernestal 2007; Qi 2011). This type of approach is not meant to compete with previously mentioned fault detection and diagnosis methods; instead, the Bayesian approach provides a unifying framework to simultaneously use many of these methods at once. As such, it can take input from many different fault detection and diagnosis techniques in order to make a final decision. In this book, other diagnosis methods and even instruments themselves are treated as input sources and are referred to as *monitors*; this term is chosen mainly because previous work (Qi 2011) focused heavily on using input from control loop monitoring techniques (described in Section 1.2.2).

For Bayesian diagnosis, data from monitors must be collected for every scenario that one would wish to diagnose. In this book, such scenarios are referred to as operational *modes*. When new monitor information arrives, the new information is compared to historical data in order to determine which historical mode best fits the new information. The Bayesian diagnosis technique ranks each of the modes based on posterior probability, which is calculated using Bayes' theorem (Bayes 1764/1958):

$$p(M|E) = \frac{p(E|M)p(M)}{p(E)}$$

$$P(E) = \sum_i p(E|m_i)p(m_i)$$

where

- $p(M|E)$ is the posterior probability, or probability of the mode M given evidence E
- $p(E|M)$ is the likelihood of the evidence E given the historical mode M
- $p(M)$ is the prior probability of the historical mode M
- $p(E)$ is the probability of the evidence E (which is a normalizing constant).

In the Bayesian diagnosis technique, the historical data and mode classifications are used to construct the likelihood $p(E|M)$, and prior probabilities of modes are assigned to $p(M)$ using expert knowledge. While collecting data for historical modes may be a challenge, the Bayesian method at least allows us to collect data in a way that is not necessarily representative of the true mode occurrence rate. For example, if mode 1 occurs 90% of the time, then representative sampling would require that 90% of the data come from mode 1. Bayesian methods (which use prior probabilities to cover mode representation) allow us to collect an arbitrary amount of data for each mode, giving us a lot more flexibility in data collection than other methods.

1.2.2 Monitoring Techniques

Much of this work focuses on monitoring and diagnosing control-loops (a schematic for a typical control loop is given in Figure 1.1); for this area of research, there exists abundant

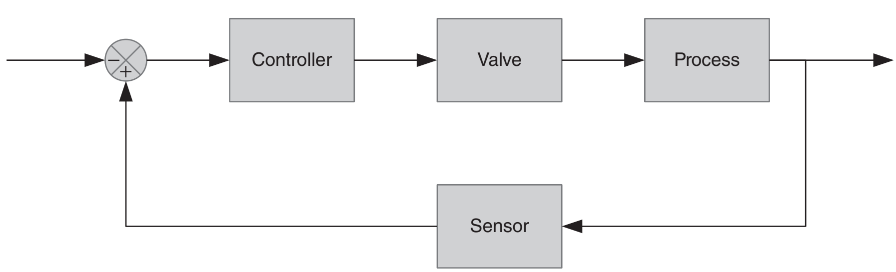

Figure 1.1 Typical control loop

Table 1.1 List of monitors for each system

Simulated	Bench-scale	Industrial-scale
Control performance Valve stiction Process model	Sensor bias Process operation	Raw sensor readings

literature on assessing the performance of the entire loop as well as diagnosing problems within the loop's core components. These methods (defined as monitors in this book) can be directly used to create alarms or notification statuses which alert operators and engineers about risky or inefficient operation.

Monitors tend to focus on one or more of the main components in a control system: for example, the controller, the actuator (often a valve), the process and the sensor. The following monitors will be considered in this book as examples but the diagnosis approach as proposed in this book can be applied to other monitors as well.

- **Control performance monitors** are intended to monitor the performance of the controller, but are often affected by other parts of the control loop.
- **Sensor bias monitors** focus on sensor performance.
- **Valve stiction monitors** focus on valve performance, but can sometimes be affected by other sources of oscillation.
- **Process model monitors** evaluate the correctness of the process model, which has utility in diagnosing controller performance and process performance. Deviation from the model can indicate a change in the system operation, and perhaps even a fault. In addition, because control tuning is performed with a model in mind, changes in the model may indicate that the current controller configuration is not suitable for current operation.
- **Process operation monitors** tend to fall under the category of fault detection, and aim to diagnose abnormalities and faults within a process.

The methods in this book are tested on three particular testbed systems: a simulated system, a bench scale system and an industrial scale system. Each type of monitor has been used in at least one of the testbed systems; a summary of monitors for each system is presented in Table 1.1. The simulated system makes use of three monitors (control performance monitors, valve stiction monitors and process model monitors) while the bench scale system makes use of the two remaining monitor types (a process operation monitor and two sensor bias monitors). The industrial-scale system uses no monitors, but instead directly uses data from the various sensors within the facility.

Control Performance Monitoring

Control performance assessment is concerned with the analysis of available control loop performance against certain benchmarks, while control performance monitoring is concerned with monitoring control performance change with respect to certain references. Due to their similarity, the two terminologies have often been used interchangeably and it is commonly accepted that they can represent each other. Research in this areas was pioneered by Harris et al. 1999 for proposing the minimum variance control (MVC) benchmark. Huang et al. (1995) developed a filtering and correlation (FCOR) algorithm to estimate the MVC benchmark that can be easily extended to multivariate systems. A state space framework for the MVC benchmark was proposed by McNabb and Qin (2005). The MVC index was extended to multivariate systems by Harris et al. (1996) and Huang et al. (1997); the latter tackled MIMO MVC benchmark by introducing the unitary interactor matrix. The MVC benchmark provides a readily computable and physically significant bound on control performance.

The theoretical variance lower bound of MVC may not be achievable for most practical controllers. More realistic performance indices are needed. Ko and Edgar (1998) discussed a PID benchmark. An approach was presented by Qin (1998), stating that MVC can be achievable for a PID controller when the process time delay is either small or large, but not medium. Huang and Shah (1999) proposed the linear quadratic Gaussian (LQG) regulator benchmark as an alternative to the MVC benchmark, based on the process model. Model-based approaches also exist for benchmarking model predictive control (MPC) systems (see Shah et al. (2001) and Gao et al. (2003)).

The benchmarks discussed above mainly focus on stochastic performance. However, these benchmarks can also be related to deterministic performances, such as overshoot, decay ratio, settling time, etc. Ko and Edgar (2000) modified the MVC index to include setpoint variations in the inner loop of cascade control. The influence of setpoint changes on the MVC index was discussed by Seppala et al. (2002), who proposed a method to decompose the control error into two components: one that resulted from setpoint changes, and another from a setpoint detrended signal. Thornhill et al. (2003) examined the reasons why performance during setpoint change differs from the performance during operation at a constant setpoint. The extension of the MVC index to the varying setpoint case has also been discussed by McNabb and Qin (2005).

Some other methods have also been proposed for control performance assessment. Kendra and Cinar (1997) applied frequency analysis to evaluate control performance. An r statistic was introduced by Venkataramanan et al. (1997) that detects deviations from setpoint, regardless of the output noise. Li et al. (2003) proposed a relative performance monitor, which compares the performance of a control loop to that of a reference model.

A number of commercial control performance assessment software packages are available on the market, such as the Intune software tools by Control Soft, LoopScout by Honeywell Hi-Spec Solutions, Performance Surveyor by DuPont, etc. (Jelali 2006). Various successful industrial applications have also been reported (Hoo et al. 2003; Jelali 2006).

In this book, the FCOR algorithm (Huang et al. 1995) is used to calculate the MVC benchmark and serves as the control performance monitor for the simulated system.

Valve Stiction Monitors

The undesirable behaviour of control valves is the biggest single contributor to poor control loop performance (Jelali and Huang 2009). According to Jelali and Huang (2009), 20–30% of control loop oscillations are induced by valve nonlinearities, including stiction, deadband, hysteresis, etc. Among these problems, stiction is the most common one in the process industry (Kano et al. 2004). Oscillation in control loops increases the variability of process variables, which in turn affects product quality, increases energy consumption, and accelerates equipment wear. Detecting valve stiction in a timely manner will bring significant economic benefits, and thus there is a strong incentive for valve stiction detection research. A comprehensive review and comparison of valve stiction detection methods can be found in Jelali and Huang (2009).

Singhal and Salsbury (2005) proposed a stiction detection methodology by calculating the ratio of the areas before and after the oscillation peaks of the PV signal. A method for diagnosing valve stiction was developed based on observations of control loop signal patterns by Yamashita (2006). The method determines typical patterns from valve input and valve output/process variables in the control loop, and thus does not allow detection of stiction which shows up in different patterns. Scali and Ghelardoni (2008) improved the work of Yamashita (2006) to allow different possible stiction patterns to be considered. Choudhury et al. (2007) proposed a controller gain change method, which is based on the change in the oscillation frequency due to changes in the controller gain to detect valve stiction. Yu and Qin (2008) showed that this method can fail to detect the presence of the sticky valve in interacting multi-input multi-output systems. A strategy based on the magnitude of relative change in oscillation frequency due to changes in controller gain is proposed to overcome the limitations of the existing method.

Despite the various work regarding stiction detection, valve stiction quantification remains a challenging problem. Choudhury et al. (2008) proposed a method to quantify stiction using the ellipse fitting method. The PV vs. OP plot is fitted to an ellipse and the amount of stiction is estimated as the maximum length of the ellipse in the OP direction. Chitralekha et al. (2010) treated the problem of estimating the valve position as an unknown input estimation problem. The unknown input is estimated by means of an input estimator based on the Kalman filter. Jelali (2008) presented a global optimization based method to quantify valve stiction. A similar method was also proposed by Srinivasan et al. (2005). The approach is based on identification of a Hammerstein model consisting of a sticky valve and a linear process. The stiction parameters and the model parameters are estimated simultaneously with a global grid search optimization method. Jelali and Huang (2009) presented a closed-loop stiction quantification approach using routine operating data. A suitable model structure of valve stiction is chosen prior to conducting valve stiction detection and quantification. Given the stiction model structure, a feasible search domain of stiction model parameters is defined, and a constrained optimization problem is solved in order to determine the stiction model parameters.

The aforementioned stiction qualification methods all assume that the process is linear. Nallasivama et al. (2010) proposed a method to qualify the stiction for closed-loop nonlinear systems. The key idea used in the approach is based on the identification of extra information available in process output, PV, compared to the controller output, OP. Stiction phenomenon leads to many harmonic components compared to the Fourier transform of the Volterra system, which allows stiction detection in nonlinear loops.

In this book, a simple stiction monitoring algorithm is used for the simulated system based on fitting the valve's input–output relationship to an ellipse (Choudhury et al. 2008). If stiction is absent, the data should be easily fitted by an ellipse. However, if the fit is poor, it is likely that stiction is present.

Model–plant Mismatch Monitors

A large volume of work has been published for open-loop model validation. However, the literature has been relatively sparse on studies concerned with on-line model validation using closed-loop data.

Huang (2001) developed a method for the analysis of detection algorithms in the frequency domain under closed-loop conditions. The divergence algorithm is extended to the model validation for the general Box–Jenkins model under closed-loop conditions through the frequency domain approach. Based on the two-model divergence method, Jiang et al. (2009) developed two closed-loop model validation algorithms, which are only sensitive to plant model changes. Of the two algorithms, one is sensitive to changes in both plant and disturbance dynamics, while the other is only sensitive to changes in plant dynamics, regardless of changes in disturbance dynamics and additive process faults, such as sensor bias.

Badwe et al. (2009) proposed a model mismatch detection method based on the analysis of partial correlations between the model residuals and the manipulated variables. The more significant this correlation is, the higher is the possibility that there exists model mismatch. Badwe et al. (2010) further extended their earlier work by analysing the impact of model mismatch on the control performance.

Selvanathan and Tangirala (2010) introduced a plant-model ratio (PMR) as a measure to quantify the model–plant mismatch in the frequency domain. The PMR provides a mapping between its signatures and changes in process models, and thus the changes in model gain, time constant and time delay can be identified. Although it is claimed that the PMR can be estimated from closed-loop operating data, a significant underlying assumption is that the setpoint contains at least a pulse change. This assumption, however, can be restrictive in practice.

In this book the output error (OE) model method is used for model error monitoring. This algorithm focuses on multi-input, single-output (MISO) systems, even though the simulated process is a MIMO (multiple input, multiple-output) system; however, a MIMO system can be easily constructed using several MISO systems.

Bias Monitors

Sensor bias can also be a problem in control loops, as sensors are the main reference for control action. A common method for detecting sensor bias in process industries is the use of data reconciliation and gross error detection (Mah and Tamhane 1982). Most data-reconciliation and gross error detection methods have been proposed for offline implementation (Ozyurt and Pike 2004); recently, Qin and Li. (2001) and Gonzalez et al. (2012) developed on-line versions.

In this book, bias monitors for the bench-scale process focus on the flow meter output versus pump speed. This type of monitor is effective for positive-displacement pumps such as those found in the bench-scale process.

Process Operation Monitors

Process operation monitoring is a broad area of research, mainly because of the large variety of processes that can be monitored and the large number of operation phenomena that can be targeted (such as faults/breakdowns, abnormal/suboptimal operation and violation of operating limits). Literature in this area falls under fault detection and diagnosis literature, which is reviewed in Section 1.2.1.

In this book, for the bench-scale system, a quantitative model-driven technique is used based on the Kalman filter; here, the state is augmented in order to include two fault-related parameters (representing leaks). Under ideal conditions, the parameters have values of zero (no leak), but as leaks are introduced, the parameter values change to values significantly greater than 0.

1.3 Book Outline

This book is broken down into two major parts, *Fundamentals* and *Application*, and each of the major contributions is generally represented in both parts. The fundamentals section focuses on theoretical development and justification of the proposed techniques, while the application section focuses on succinctly conveying all information required to apply these techniques. Since both parts are meant to be stand-alone, there may be some slight overlap between them, namely the parts in the fundamentals section that are directly relevant to applications.

A number of techniques exist in this book that many readers may not be familiar with, namely Bayesian diagnosis, Dempster–Shafer theory, kernel density estimation and bootstrapping. A tutorial is provided which covers fundamental aspects of all four of these techniques.

1.3.1 Problem Overview and Illustrative Example

The main objective of this work is to diagnose the process operating mode (which contains information about the state of each process component of interest, such as sticky valves, biased sensors, inaccurate process models etc.). Before diagnosing modes, we collect historical data from monitors for each mode; this historical data is used to diagnose the mode when new evidence becomes available online. Because it is assumed that corresponding modes are available with the historical data, this book takes a *supervised learning approach* when applying historical data.

In order to easily illustrate the challenges associated with Bayesian diagnosis, we start from a toy problem where the modes consist of two different coins, one with a bias toward heads (probability of heads $= 0.6$) and one that is fair (probability of heads $= 0.5$). The probability estimates are obtained through historical data of coin flips. For the diagnosis problem, a coin is randomly selected and we wish to use evidence of coin flipping to determine which coin was selected. The evidence is provided by two people flipping the same coin once.

1.3.2 Overview of Proposed Work

This book aims to address various challenging issues with respect to Bayesian diagnosis. A visual map of these solutions is given in Figure 1.2, where shaded boxes indicate problems, and white boxes indicate solutions proposed by this book; dotted lines indicate a combination of multiple solutions.

Autodependent Modes

For industrial processes, mode changes tend to be quite rare, which means that the mode at time t is highly dependent on the mode at time $t-1$. Taking this type of dependency into account can significantly increase the precision of the diagnosis results, as consecutive pieces

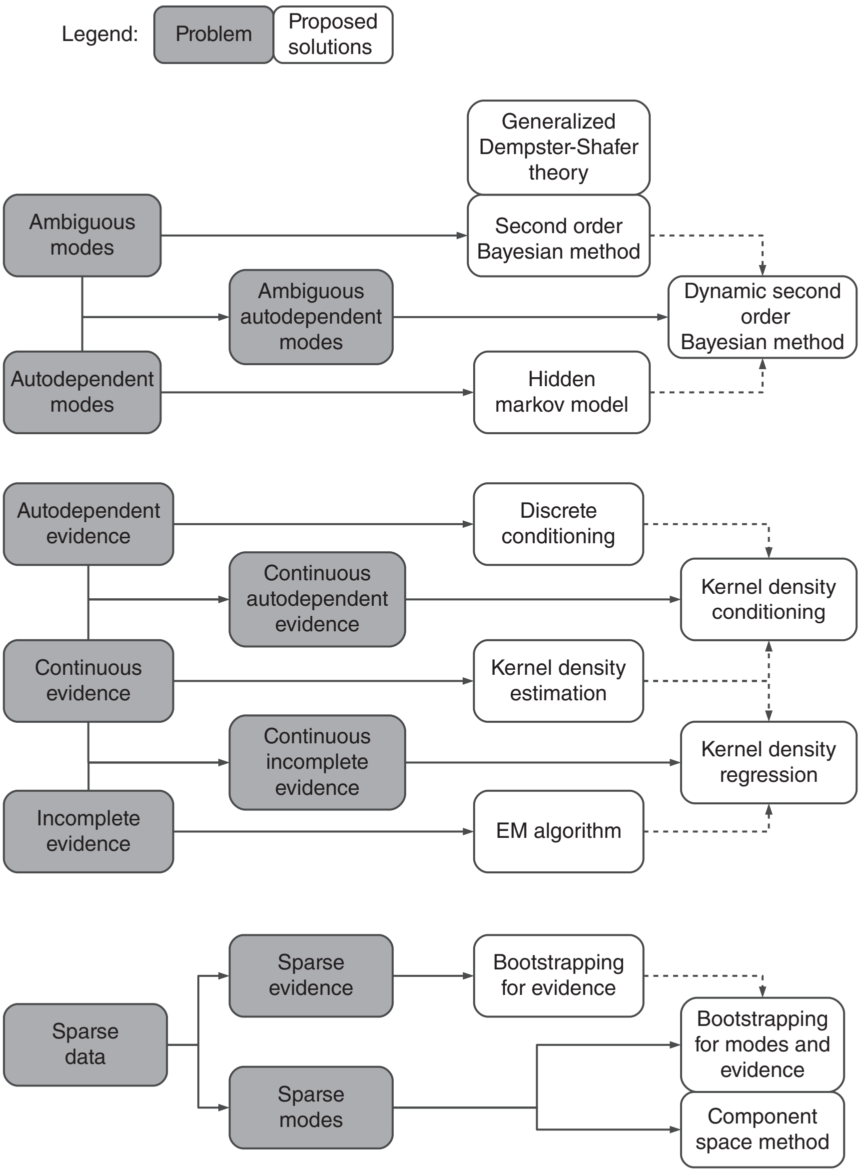

Figure 1.2 Overview of proposed solutions

of evidence contain more information than individual pieces. This type of dependency has been addressed in Qi and Huang (2010b).

Returning to our coin-flipping example, consider the case where after each pair of flips there is a probability of the coin being switched. If that probability is low, the 'mode' has a strong time-wise autodependence. This means that consecutive pieces of evidence contain even more information about the mode than single pieces of evidence themselves.

Autodependent Evidence

Monitor readings often use data from previous time steps in order to calculate a result. If monitor readings are not sampled slowly enough, the evidence will be autodependent. Taking the autodependence of evidence into account was addressed in Qi and Huang (2011a).

Autodependent evidence can also be applied to our coin-flipping example. If the second coin flipper obtained heads at $t-1$, and the first coin flipper at time t placed the coin on his thumb heads-side-up (with tails being similarly treated) then results would exhibit time-wise dependence.

Incomplete Discrete Evidence

It is not uncommon in process industries that historical records are unavailable during certain time intervals. Since sensors are also used for monitoring, the corresponding monitor will also be unavailable, rendering a data point incomplete (as some elements are missing). Simply discarding incomplete data points will result in a loss of information so Qi and Huang (2010a) proposed using Bayesian methods to recover the useful information from these incomplete data points.

Using our coin-flipping example, consider the case where the evidence from the two people flipping coins is dependent. For example, after the first person flips a coin, whatever side faces up will be placed face up on the second person's thumb. Now the historical data contains the results of both people flipping coins. In some circumstances, the result from one of the two flips will be missing. Because the coin flips are dependent, Qi and Huang (2010a) adopt Bayesian methods to use the information present to make up for the missing information.

Ambiguous Modes from a Bayesian Perspective

Qi and Huang (2010a) addressed the issue where some elements of historical evidence records are missing, causing the evidence to be incomplete. However, just as evidence requires input from multiple monitors, the mode requires information from multiple components. If any information about the components is missing, a number of different modes will be possible, causing the mode to be *ambiguous*.

For example, if some of the historical data from coin-flipping exercises contained no information on which coin was flipped (biased or fair) then either coin could have produced the results (heads or tails) and the corresponding mode (or coin in this case) is *ambiguous*. Since the conditioning variable is unknown, the probability cannot be calculated in a straightforward manner.

Ambiguous Modes from a Dempster–Shafer Perspective

Demspter–Shafer theory (Dempster 1968; Shafer 1976) has been deemed by many to be a generalization of Bayesian diagnosis that is able to handle ambiguity. However, it is shown in this book that Dempster–Shafer theory does not adequately formulate the problem of ambiguous modes in the historical data when likelihoods $p(E|M)$ are used. Some modifications are required in order to properly fit the data-driven diagnosis problem into a Dempster–Shafer framework.

Using Continuous Evidence

Previously, it was assumed that information used by the diagnosis method was discrete (our coin-flipping exercise yields discrete evidence). In reality, however, most monitors yield an output that is continuous (e.g. a monitor that monitors changes in compressor pressure). In order to reduce the amount of information lost through discretizing continuous values, this book proposes the use of *kernel density estimation* (as proposed in Gonzalez and Huang (2014)) to make use of the continuous data directly.

Continuous evidence can also suffer from missing data, but because kernel density estimation is non-parametric, the expectation-maximization (EM) algorithm is not directly applicable. The most common method used to deal with missing evidence in a kernel density estimation framework is kernel density regression (a method used to calculate the expected value of the missing data). The completed data is then included in the data set used for kernel density estimation. Due to the simplicity of this solution, the problem of missing continuous evidence is included in the continuous evidence chapter.

Sparse Evidence given a Mode

If a process has a large number of components, the number of possible modes will be very large. In such cases, it is quite possible for data from a particular mode to be quite sparse. Qi and Huang (2011b) recommended the use of bootstrapping as a method to generate additional data and get a better representation of the monitor distribution.

For the coin-flipping example, consider the case where one of the coins (such as the biased one) does not have a large amount of historical data. Bootstrapping is a method that was suggested in Qi and Huang (2011b) to resolve this issue by simulating more coin flips by randomly drawing from previous results recorded in the historical samples.

Sparse or Missing Modes in the Data

As the number of components in a system increases, the possible modes will increase exponentially. For systems with a large number of components, it is likely that data for a significant number of modes will be missing entirely.

For example, in our coin-flipping exercise, even though we only have two coins (e.g. modes), we might not have any historical data from one of the coins. This book will present techniques on what one can do if certain modes of interest are absent from the historical data.

Dynamic Application of Continuous Evidence and Ambiguous Modes

When accounting for ambiguous modes, the solution for addressing mode autodependence will be affected. Similarly, when accounting for continuous evidence, the solution for autodependent evidence will be affected. In Part One, which deals with fundamentals, the solution to autodependent modes is addressed in Chapter 6, which discusses ambiguous modes. Likewise, the solution to autodependent evidence is addressed in Chapter 8. However, in Part Two, which deals with application, the solutions to both autodependent modes and evidence are dealt with in one chapter (Chapter 18).

References

Badwe A, Gudi R, Patwardhan R, Shah S and Patwardhan S 2009 Detection of model-plant mismatch in mpc applications. *Journal of Process Control* **19**, 1305–1313.

Badwe A, Patwardhan R, Shah S, Patwardhan S and Gudi R 2010 Quantifying the impact of model-plant mismatch on controller performance. *Journal of Process Control* **20**, 408–425.

Bayes T 1764/1958 An essay towards solving a problem in the doctrine of chances. *Biometrika* **45**, 296–315.

Chitralekha SB, Shah SL and Prakash J 2010 Detection and quantification of valve stiction by the method of unknown input estimation. *Journal of Process Control* **20**, 206–216.

Choudhury MAAS, Kariwala V, Thornhill NF, Douke H, Shah SL, Takada H and Forbes JF 2007 Detection and diagnostics of plant-wide oscillations. *Canadian Journal of Chemical Engineering* **85**, 208–219.

Choudhury M, Jain M and Shah S 2008 Stiction – definition, modelling, detection and quantification. *Journal of Process Control* **18**(3-4), 232–243.

Clark RN, Fosth DC and Walton VM 1975 Detection instrument malfunctions in control systems. *IEEE Trans. Aerospace Electron. Syst* **AES-II**, 465–473.

Dempster A 1968 A generalization of Bayesian inference. *Journal of the Royal Statistical Society. Series B* **30**(2), 205–247.

Desai M and Ray A 1981 A fault detection and isolation methodology. *Proc. 20th Conf. on Decision and Control*, pp. 1363–1369.

Frank P 1990 Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy: a survey and some new results. *Automatica* **26**(3), 459–474.

Frank P 1996 Analytical and qualitative model-based fault diagnosis – a survey and some new results. *European Journal of Control* **2**, 6–28.

Gao J, Patwardhan R, Akamatsu K, Hashimoto Y, Emoto G and Shah SL 2003 Performance evaluation of two industrial MPC controllers. *Control Engineering Practice* **11**, 1371–1387.

Garcia E and Frank P 1997 Deterministic nonlinear observer-based approaches to fault diagnosis: a survey. *Control Engineering Practice* **5**(5), 663–670.

Ge Z and Song Z 2010 Mixture Bayesian regularization method of PPCA for multimode process monitoring. *Process Systems Engineering* **56**(11), 2838–2849.

Gonzalez R and Huang B 2014 Control-loop diagnosis using continuous evidence through kernel density estimation. *Journal of Process Control* **24**(5), 640–651.

Gonzalez R, Huang B, Xu F and Espejo A 2012 Dynamic Bayesian approach to gross error detection and compensation with application toward an oil sands process. *Chemical Engineering Science* **67**(1), 44–56.

Gonzalez R, Huang B and Lau E 2015 Process monitoring using kernel density estimation and Bayesian networking with an industrial case study. *ISA Transactions* **58**, 330–347.

Hagenblad A, Gustafsson F and Klein I 2004 A comparison of two methods for stochstic fault detection: the parity space approach and principle component analysis. *13th IFAC Symposium on System Identification (SYSID 2003)*, Rotterdam, The Netherlands, 27–29 August 2003.

Harris T, Boudreau F and MacGregor JF 1996 Performance assessment using multivariable feedback controllers. *Automatica* **32**, 1505–1518.

Harris T, Seppala C and Desborough L 1999 A review of performance monitoring and assessment techniques for univariate and multivariate control systems. *Journal of Process Control* **9**(1), 1–17.

Hoo K, Piovoso M, Sheneller P and Rowan D 2003 Process and controller performance monitoring: overview with industrial applications. *International Journal of Adaptive Control and Signal Processing* **17**, 635–662.

Huang B 2001 On-line closed-loop model validation and detection of abrupt parameter changes. *Journal of Process Control* **11**, 699–715.

Huang B and Shah S 1999 *Performance Assessment of Control Loops*. Springer-Verlag.

Huang B, Shah S and Kwok K 1995 On-line control performance monitoring of MIMO processes. *Proceedings of American Control Conference*, 21–23 June 1995, Seattle, USA.

Huang B, Shah S and Kwok E 1997 Good, bad or optimal? Performance assessment of multivariable process. *Automatica* **33**, 1175–1183.

Isermann R 1993 Fault diagnosis of machines via parameter estimation and knowledge processing. *Automatica* **29**, 815–836.

Isermann R and Freyermuth B 1991 Process fault diagnosis based on process model knowledge, parts i (principles) and ii (case study experiments). *ASME Journal of Dynamic Systems, Measurement Control* **113**, 620–633.

Jelali M 2006 An overview of control performance assessment technology and industrial applications. *Control Engineering Practice* **14**(5), 441–466.

Jelali M 2008 Estimation of valve stiction in control loops using separable least-squares and global search algorithms. *Journal of Process Control* **18**, 632–642.

Jelali M and Huang B (eds) 2009 *Detection and Diagnosis of Stiction in Control Loops: State of the Art and Advanced Methods*. Springer.

Jiang H, Huang B and Shah S 2009 Closed-loop model validation based on the two-model divergence method. *Journal of Process Control* **19**, 644–655.

Jones H 1973 Failure detection in linear systems. PhD thesis. MIT, Cambridge, MA.

Kano M, Maruta H, Kugemoto H and Shimizu K 2004 Practical model and detection algorithm for valve stiction. *Proceedings of the IFAC Symposium on Dynamics and Control of Process Systems*, Boston, USA.

Kendra S and Cinar A 1997 Controller performance assessment by frequency domain techniques. *Journal of Process Control* **7**, 181–194.

Ko B and Edgar T 1998 Assessment of achievable PI control performance for linear processes with dead time. *Proceedings of American Control Conference*, 24–26 June 1998, Philadelphia, Pennsylvania, USA.

Ko B and Edgar TF 2000 Performance assessment of cascade control loops. *AIChE Journal* **46**, 281–291.

Lee JM, Lee IB and Qin S 2006 Fault detection and diagnosis based on modified independent component analysis. *AIChE Journal* **52**, 3501–3514.

Lerner UN 2002 *Hybrid Bayesian networks for reasoning about complex systems*. PhD thesis. Stanford University.

Lerner U, Parr R, Koller D and Biswas G 2000 Bayesian fault detection and diagnosis in dynamic systems. *17th National Conference on Artificial Intelligence (AAAI-00)* , July 30–August 3, 2000, Austin, Texas.

Li Q, Whiteley JR and Rhinehart RR 2003 A relative performance monitor for process controllers. *International Journal of Adaptive Control and Signal Processing* **17**, 685–708.

Mah RSH and Tamhane AC 1982 Detection of gross errors in process data. *AIChE Journal* **28**(5), 828–830.

McNabb CA and Qin SJ 2005 Projection based MIMO control performance monitoring: II – measured disturbances and setpoint changes. *Journal of Process Control* **15**, 89–102.

Nallasivama U, Babjib S and Rengaswamy R 2010 Stiction identification in nonlinear process control loops. *Computers and Chemical Engineering*. **34**(11), 1890–1898.

Nimmo I 2003 Adequately address abnormal situation operations. *Chemical Engineering Progress* **91**(9), 36–45.

Ozyurt D and Pike R 2004 Theory and practice of simultaneous data reconciliation and gross error detection for chemical processes. *Computers and Chemical Engineering* **28**, 381–402.

Pernestal A 2007 *A Bayesian Approach to Fault Isolation with Application to Diesel Engines*. PhD thesis. KTH School of Electrical Engineering, Sweden.

Qi F 2011 *Bayesian Approach for Control Loop Diagnosis*. PhD thesis. University of Alberta.

Qi F and Huang B 2010a A Bayesian approach for control loop diagnosis with missing data. *AIChE Journal* **56**(1), 179–195.

Qi F and Huang B 2010b Dynamic Bayesian approach for control loop diagnosis with underlying mode dependency. *Industrial & Engineering Chemistry Research* **49**, 8613–8623.

Qi F and Huang B 2011a Bayesian methods for control loop diagnosis in the presence of temporal dependent evidences. *Automatica* **47**, 1349–1356.

Qi F and Huang B 2011b Estimation of distribution function for control valve stiction estimation. *Journal of Process Control* **28**(8), 1208–1216.

Qin S 1998 Control performance monitoring – a review and assessment. *Computers and Chemical Engineering* **23**(2), 173–186.

Qin S and Li. W 2001 Detection and identification of faulty sensors in dynamic process. *AIChE Journal* **47**(7), 1581–1593.

Rault A, Jaume D and Verge M 1984 Industrial fault detection and localization *IFAC 9th World Congress, Budapest*, vol. 4, pp. 1789–1792.

Romessis C and Mathioudakis K 2006 Bayesian network approach for gas path fault diagnosis. *Journal of Engineering for Gas Turbines and Power* **128**, 64–72.

Scali C and Ghelardoni C 2008 An improved qualitative shape analysis technique for automatic detection of valve stiction in flow control loops. *Control Engineering Practice* **16**, 1501–1508.

Selvanathan S and Tangirala AK 2010 Diagnosis of poor control loop performance due to model-plant mismatch. *Industrial Engineering Chemistry Research* **49**, 4210–4229.

Seppala CT, Harris TJ and Bacon DW 2002 Time series methods for dynamic analysis of multiple controlled variables. *Journal of Process Control* **12**, 257–276.

Shafer G 1976 *A Mathematical Theory of Evidence*. Princeton University Press.

Shah SL, Patwardhan R and Huang B 2001 Multivariate controller performance analysis: Methods, applications and challenges *Sixth International Conference on Chemical Process Control*, 7–12 January 2001, Tucson, Arizona.

Singhal A and Salsbury T 2005 A simple method for detecting valve stiction in oscillating control loops. *Journal of Process Control* **15**, 371–382.

Srinivasan R, Rengaswamy R, Narasimhan S and Miller R 2005 A curve fitting method for detecting valve stiction in oscillation control loops. *Industrial Engineering Chemistry Research* **44**, 6719–6728.

Thornhill NF, Huang B and Shah SL 2003 Controller performance assessment in set point tracking and regulatory control. *International Journal of Adaptive Control and Signal Processing* **17**, 709–727.

Tipping M and Bishop C 1998 Probabilistic principal component analysis. *Journal of the Royal Statistical Society* **61**(3), 611–622.

Tipping M and Bishop C 1999 Mixtures of probabilistic principal component analysers. *Neural Computation* **11**(2), 443–482.

Venkataramanan G, Shukla V, Saini R and Rhinehart RR 1997 An automated on-line monitor of control system performance. *Proceedings of the American Control Conference*, 4–6 June 1997, Albuquerque, New Mexico, USA.

Venkatasubramanian V, Rengaswamy R and Kavuri SN 2003a A review of process fault detection and diagnosis part ii: Qualitative models and search strategies. *Computers and Chemical Engineering* **27**, 313–326.

Venkatasubramanian V, Rengaswamy R, Yin K and Kavuri SN 2003b A review of process fault detection and diagnosis part i: Quantitative model-based methods. *Computers and Chemical Engineering* **27**, 293–311.

Venkatasubramanian V, Rengaswamy R, Yin K and Kavuri SN 2003c A review of process fault detection and diagnosis part iii: Process history based methods. *Computers and Chemical Engineering* **27**, 327–346.

Willsky A 1976 Detection instrument malfunctions in control systems. *Automatica* **12**, 601–611.

Yamashita Y 2006 An automatic method for detection of valve stiction in process control loops. *Control Engineering Practice* **14**, 503–510.

Yu J and Qin SJ 2008 Statistical MIMO controller performance monitoring. part ii: Performance diagnosis. *Journal of Process Control* **18**, 297–319.