Advances in Computer Vision and Pattern Recognition

Chris Aldrich
Lidia Auret

Unsupervised Process Monitoring and Fault Diagnosis with Machine Learning Methods

Springer

Advances in Computer Vision and Pattern Recognition

For further volumes:
http://www.springer.com/series/4205

Chris Aldrich • Lidia Auret

Unsupervised Process Monitoring and Fault Diagnosis with Machine Learning Methods

Springer

Chris Aldrich
Western Australian School of Mines
Curtin University
Perth, WA, Australia

Department of Process Engineering
University of Stellenbosch
Stellenbosch, South Africa

Lidia Auret
Department of Process Engineering
University of Stellenbosch
Stellenbosch, South Africa

Series Editors
Sameer Singh
Research School of Informatics
Loughborough University
Loughborough
UK

Sing Bing Kang
Microsoft Research
Microsoft Corporation
Redmond, WA
USA

ISSN 2191-6586 ISSN 2191-6594 (electronic)
ISBN 978-1-4471-5184-5 ISBN 978-1-4471-5185-2 (eBook)
DOI 10.1007/978-1-4471-5185-2
Springer London Heidelberg New York Dordrecht

Library of Congress Control Number: 2013942259

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Although this book is focused on the process industries, the methodologies discussed in the following chapters are generic and can in many instances be applied with little modification in other monitoring systems, including some of those concerned with structural health monitoring, biomedicine, environmental monitoring, the monitoring systems found in vehicles and aircraft and monitoring of computer security systems. Of course, the emphasis would differ in these other areas of interest, e.g. dynamic process monitoring and nonlinear signal processing would be more relevant to structural health analysis and brain–machine interfaces than techniques designed for steady-state systems, but the basic ideas remain intact. As a consequence, the book should also be of interest to readers outside the process engineering community, and indeed, advances in one area are often driven by application or modification of related ideas in a similar field.

In a sense, the area of process monitoring and the detection and analysis of change in technical systems are an integral part of the information revolution, as the use of data-driven methods to construct the requisite process or systems models becomes dominant over first-principle or higher knowledge approaches. This revolution has changed the world as we know it and will continue to do so in as yet unforeseen ways.

Rightly or wrongly, there is a perception that the mining engineering environment is conservative as far as research spending is concerned, reluctant to embrace future technologies that do not have an immediate proven impact on the bottom line, also as far as process automation is concerned. However, this is rapidly changing, with large mining companies investing considerable sums of money in the development of advanced process automation systems with no immediate benefit. These new automation systems will have to sense changes in their environment and be able to react to these changes, consistently, safely and economically. Apart from the development of advanced sensors, process monitoring technologies would play a central role in the success of these automated mining systems. For example, in underground mining, these systems would have to be able to differentiate between mineral and the surrounding gangue material in real time or be able to differentiate between solid rock and rock that might be on the verge of collapse in a mining tunnel. Humans have mixed success in these tasks, and current automation systems are too rudimentary to improve on this.

These new diagnostic systems would have to cope with the so-called *Big Data* phenomenon, which will inevitably also have an impact on the development and implementation of the analytical techniques underpinning them. In many ways, *Big Data* can simply be seen as more of the same, but it would be unwise to see it simply as a matter that can be resolved by using better hardware. With large complex data sets, the issues of automatically dealing with unstructured data, which may contain comparatively little useful information, become paramount. In addition, these data streams are likely to bring with them new information not presently available, in ways that are as yet unforeseen. Just like video data can simply be seen as a series of images, if taken at a sufficiently high frequency, these data can reveal information on the dynamic behaviour of the system that a discontinuous series of snapshots cannot. It is easy to see that in some cases this could make a profound difference on our understanding of the behaviour of the system.

In the same way that *Big Data* can be seen as data, just more of it, machine learning can arguably be seen as statistics, simply in a different guise, as in many ways it is without a doubt. However, looking into the future, as systems rapidly grow in complexity, the ability of machines to truly learn could also be influenced in unforeseen ways. By analogy, one could consider a novice chess player, who has learnt the rules of chess and knows how to detect direct threats to his individual pieces on the board. However, it is only by experience that he learns to recognize the unfolding of more complex patterns or emergent behaviour that would require timely action to avoid or exploit.

Perth, WA, Australia Chris Aldrich

Acknowledgements

In many ways, this book can be regarded as a product of the Anglo American Platinum Centre for Process Monitoring and the research work of a large number of postgraduate students that have passed through the Process Systems Engineering group at Stellenbosch University over the last decade or more. The collaboration between academia and industry has been especially productive in this respect.

Our special thanks therefore to Dr. J.P. Barnard and Ms. Corné Yzelle for making available the Centre's *Process Diagnostics Toolset* software without which the methods outlined in Chap. 6 in the book could not have been implemented.

In addition, we would also like to express our sincere gratitude to Dr. Gorden Jemwa, not only for his contributions to the Process Systems Engineering group over many years but also specifically for his major contribution as main author of Chap. 8 in the book.

Finally, it may be a cliché, but it does not make it less true that a book like this does not write itself, and the authors would like to make use of this opportunity to thank their families and friends for their understanding and active support in this respect.

Chris Aldrich and Lidia Auret

Contents

Acronyms

Acronym	Description
ACF	Autocorrelation function
ADALINE	Adaptive linear element
AHPCA	Adaptive hierarchical principal component analysis
AID	Automatic interaction detection
AKM	Average kernel matrix
AMI	Average mutual information
AR	Autoregressive
ARL	Alarm run length
ARMA	Autoregressive moving average
ARMAX	Autoregressive moving average with exogenous variables
AUC	Area under curve
BDKPCA	Batch dynamic kernel principal component analysis
BDPCA	Batch dynamic principal component analysis
BZ	Belousov–Zhabotinsky
CART	Classification and regression trees
CHAID	Chi-square automatic interaction detection
COW	Correlation optimized time warping
CSTR	Continuous stirred tank reactor
CUSUM	Cumulative sum
CVA	Canonical variate analysis
DD	Detection delay
DICA	Dynamic independent component analysis
DISSIM	Dissimilarity
DKPCA	Dynamic kernel principal component analysis
DPCA	Dynamic principal component analysis
DTW	Dynamic time warping

(continued)

(continued)

Acronym	Description
EEMD	Ensemble empirical mode decomposition
ELM	Extreme learning machine
EMD	Empirical mode decomposition
EWMA	Exponentially weighted moving average
FAR	False alarm rate
FS	Feature samples
ICA	Independent component analysis
INLPCA	Inverse nonlinear principal component analysis
IOHMM	Input–output hidden Markov model
JITL	Just-in-time learning
k-DISSIM	Kernel dissimilarity
KICA	Kernel independent component analysis
KKT	Karush–Kuhn–Tucker
KPCA	Kernel principal component analysis
KPLS	Kernel partial least squares
LCL	Lower control limit
MA	Moving average
MADALINE	Multiple adaptive linear element
MAID	Multiple or modified automatic interaction detection
MAR	Missing alarm rate
MCEWMA	Moving centre exponentially weighted moving average
MEB	Minimum enclosing ball
MHMT	Multi-hidden Markov tree
MICA	Multiway independent component analysis
MKICA	Multiscale kernel independent component analysis
MPCA	Multiway principal component analysis
MPLS	Multiway partial least squares
MSDPCA	Multiscale dynamic principal component analysis
MSE	Mean square error
MSKPCA	Multiscale kernel principal component analysis
MSPC	Multivariate statistical process control
MSSA	Multichannel singular spectrum analysis
MSSPCA	Multiscale statistical process control
MSSR	Mean sum of squared residuals
MVU	Maximum variance unfolding
MVUP	Maximum variance unfolding projection
NIPS	Neural information processing systems

(continued)

(continued)

Acronym	Description
NLPCA	Nonlinear principal component analysis
NN	Neural network
NOC	Normal operating conditions
OOB	Out of bag
PAC	Probably approximately correct
PCA	Principal component analysis
PDPCA	Partial dynamic principal component analysis
PLS	Partial least squares
RBM	Restricted Boltzmann machine
RF	Random forest
ROC	Receiver operating curve
RQA	Recurrence quantification analysis
SBKM	Single batch kernel matrix
SI	Subspace identification
SOM	Self-organizing map
SPC	Statistical process control
SPE	Squared prediction error
SPM	Statistical process monitoring
SSA	Singular spectrum analysis
SSICA	State space independent component analysis
SVD	Singular value decomposition
SVDD	Support vector domain description
SVM	Support vector machine (1-SVM one class SVM)
SVR	Support vector regression
TAR	True alarm rate
THAID	Theta automatic interaction detection
TLPP	Tensor locality preserving projection
UCL	Upper control limit
VARMA	Vector autoregressive moving average
VC	Vapnik–Chervonenkis

Chapter 1
Introduction

1.1 Background

Technological advances in the process industries in recent years have resulted in increasingly complicated processes, systems and products that pose considerable challenges in their design, analysis, manufacturing and management for successful operation and use over their life cycles (Maurya et al. 2007). As a consequence, not only do the maintenance and management of complex process equipment and processes, and their integrated operation, play a crucial role in ensuring the safety of plant personnel and the environment, but they are also crucial to the timely delivery of quality products in an environmentally responsible way. Since the management of process plants remains a largely manual activity, the timely detection of abnormal events and diagnosis of its probable causes to enable appropriate supervisory control decisions and actions to bring the process back to a normal, safe operating state become all the more important. Without a doubt, there is still major scope for process improvement in all these aspects of plant operation, including safety, profitability and environmental responsibility, as discussed in more detail below.

1.1.1 Safe Process Operation

Industrial statistics show that about 70 % of industrial accidents are caused by human errors (Venkatasubramanian et al. 2003). Recent events have shown that large-scale plant accidents are not just a thing of the past. Two of the worst ever chemical plant accidents, namely, Union Carbide's, Bhopal, India, accident and Occidental Petroleum's Piper Alpha accident, happened relatively recently (1980s). Such catastrophes have a significant impact on safety, the environment and the economy. The explosion at Kuwait Petrochemical's Mina Al-Ahmedhi refinery in

C. Aldrich and L. Auret, *Unsupervised Process Monitoring and Fault Diagnosis with Machine Learning Methods*, Advances in Computer Vision and Pattern Recognition,
DOI 10.1007/978-1-4471-5185-2_1,

June 2000 resulted in damages estimated at $400 million. Likewise, the explosion of the offshore oil platform of Petrobras, Brazil, in March 2001 resulted in losses estimated at $5 billion (Venkatasubramanian et al. 2003).

Although the occurrence of major industrial accidents such as mentioned above is not common, minor accidents are very frequent and occur almost daily, resulting in many occupational injuries and sickness and costing billions of dollars every year (Venkatasubramanian et al. 2003). This suggests that there is still a long way to go to enhance the performance of human operators to improve their diagnostic capability and good judgment.

1.1.2 Profitable Operation

Industrial processes are under increased pressure to meet the changing demands of society. For example, in the mining sector, processes have to be adapted to deal with more complex or refractory ores, as more accessible resources dwindle. The same applies in the oil industry, where the search for large repositories is increasingly focusing on deep-sea beds, as many of the world's largest fields, from Ghawar in Saudi Arabia to Prudhoe Bay in Alaska, are becoming depleted. At present, deep-sea rigs are capable of reaching down more than 12 km – twice as deep as a decade ago.

With globalization and increased competition, profit margins of companies are under pressure, and companies have to be more responsive to varying customer demands, without sacrificing product and process quality. This has led to the development of quality control management methodologies, like Six Sigma and ISO 9000, and other management programs to assist organizations in addressing some of these challenges.

In addition, modern process operations have become more complex owing to plant-wide integration and high-level automation of a large variety of process tasks (Jämsä-Jounela 2007). For example, recycling of process streams is widely established to ensure efficient material and energy usage. Process plants have become intricate virtual information networks, with significant interactions among various subsystems and components. Such interconnectivity facilitates the integration of operational tasks to achieve broader business strategic goals but invariably complicates other tasks, like planning and scheduling, supervisory control and diagnosis of process operations.

1.1.3 Environmentally Responsible Operation

More recently, regulatory frameworks have become more stringent to force better control of the environmental risks posed by industrial activities. Likewise, safety and health policies and practices are now priority issues in the modern process plants. As a result, systematic frameworks have been initiated, including process hazard analysis and abnormal event management and product life cycle management. Process hazard analysis and abnormal event management are aimed at ensuring process safety, while product life cycle management places obligatory stewardship responsibilities on an organization throughout the life cycles of its entire product range, that is, from conception to design and manufacture, service and disposal (Venkatasubramanian 2005).

1.2 Trends in Process Monitoring and Fault Diagnosis

1.2.1 Instrumentation

As a result, companies are making substantial investments in plant automation as a means to achieve their operational and business goals. This includes heavy investment in instrumentation to enable real time monitoring of process units and streams. New sensor technologies such as acoustic or vibrational signal monitoring and computer vision systems have been introduced in, among other, milling plants, multiphase processes, food processing and combustion processes (Zeng and Forssberg 1992; Das et al. 2011; Chen et al. 2012; Germain and Aguilera 2012). In large process plants, these instruments have enabled the observation of many hundreds or even thousands of process variables at high frequency (Venkatasubramanian et al. 2003). As a consequence, huge volumes of data are increasingly being generated in modern process plants. These data sets do not only contain massive numbers of samples but can also contain very large numbers of variables.

For example, in spectroscopy, data are obtained by exposing a chemical sample to an energy source and recording the resulting absorbance as a continuous trace over a range of wavelengths. Digitization of the trace at appropriate intervals (wavelengths) forms sets of variables that in pyrolytic mass spectroscopy, near-infrared spectroscopy and infrared spectroscopy yield approximately 200, 700 and 1,700 such variables for each chemical sample, respectively (Krzanowski and Marriott 1994). In these cases, the number of variables usually exceeds the number of samples by far. Similar features arise with the measurement of acoustic signals, such as may be the case in online monitoring of process equipment (Zeng and Forssberg 1992) and potentiometric measurements to monitor corrosion. Likewise, where image analysis is used to monitor particulate feeds or products in comminution systems, power plants or metallurgical furnaces, each pixel in the image could represent a variable, which could easily lead to millions of variables where high-resolution two-dimensional images are concerned.

1.2.2 Information Technology Hardware

The well-documented sustained exponential growth in computational power and communication has led to profound change in virtually all areas of technology in recent decades and will apparently continue to do so in the foreseeable future. In 1965, Gordon Moore, a co-founder of Intel, first observed that the density of components in computer chips had doubled each year since 1958, and this trend was likely to continue for at least a decade. In 1975, Dr Moore modified his prediction, observing that component density was doubling every 2 years. As a consequence, the performance of personal computers has also roughly doubled every 18 months since then, conforming to what has becoming known as Moore's law.

More recently, in what might be referred to as Koomey's law, Koomey et al. (2011) have shown that since the era of the vacuum tube, computers have also approximately doubled their electrical efficiency every 1.6 years since the mid-1940s. This trend reinforces the continued explosive growth in mobile computing, sensors and controls (Koomey et al. 2011).

The cost of computer memory is showing as pronounced a decrease as that of the other computer components, with roughly cost halving annually. For example, whereas at the beginning of the decade, 40 GB was the highest hard disk drive capacity generally available in personal computers, this has increased to 3 TB at present.

These developments have had a considerable impact on the development and maintenance of advanced process monitoring and control technologies. For example, unlike a mere decade ago, it is now possible to maintain complex instrumentation and process monitoring systems remotely via the Internet. This has led to a breakthrough in the application of instruments, such as the implementation of Blue Cube's inline diffuse reflectance spectrophotometer in remote areas, where calibration of the instrument is maintained from the company's headquarters in Stellenbosch, South Africa. The same applies to Stone Three's maintenance of their computer vision monitoring systems for particulate feeds on belts.

1.2.3 Academic Research into Fault Diagnostic Systems

Figure 1.1 shows recent trends in academic research into fault diagnosis, indicating publications associated with fault diagnosis and neural networks (ANN); expert systems (XS); kernel methods and support vector machines (SVM); multivariate methods, including principal components and latent variables (PCA); artificial immune systems and immunocomputing (AIS/IC); and others, not including the previous categories (OTHER) in the *IEEE Xplore* digital library.

Publications related to fault diagnosis and expert systems have remained more or less constant over the last two decades, since expert systems are mostly associated with qualitative fault diagnosis, while the other approaches are typically associated with data-driven fault diagnosis, which show a sharp rise, especially from 2006 to 2010. Although the publications considered here were selected to belong more or less exclusively to a particular category (e.g. SVM would indicate papers containing "support vector" *or* "kernel", but not "neural network") together with

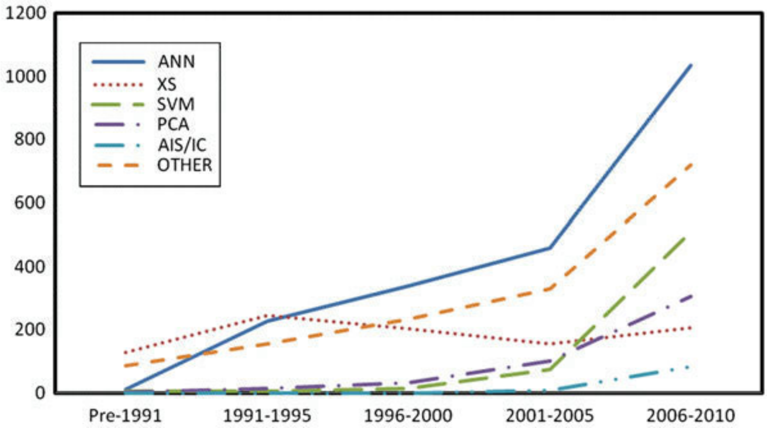

Fig. 1.1 Trends in academic research related to fault diagnosis based on number of publications in the *IEEE Xplore* digital library from 1991 to 2010

"fault diagnosis", the trends should still only be interpreted in an approximate qualitative manner, and some overlap between the categories was unavoidable. Even so, the overall trends indicate the strong growth in data-driven methods in fault diagnosis as well as the strong growth in machine learning in this area.

1.2.4 Process Analytical Technologies and Data-Driven Control Strategies

In turn, the above developments have led to further investment in advanced knowledge-based or data-driven process control strategies, collectively referred to as intelligent control systems, to enhance the information content of the data. Fortunately, advances in the information sciences have yielded data processing and analytical techniques that are very promising with respect to targeted applications in process control.

1.3 Basic Fault Detection and Diagnostic Framework

A fault can be defined as anomalous behaviour causing systems or processes to deviate unacceptably from their normal operating conditions or states. In process plants, faults can be categorized according to their sources, i.e. sensor faults affecting process measurements, actuator faults leading to errors in the operation of the plant, faults arising from erroneous operating policies or procedures as well as system component faults arising from changes in process equipment. These faults

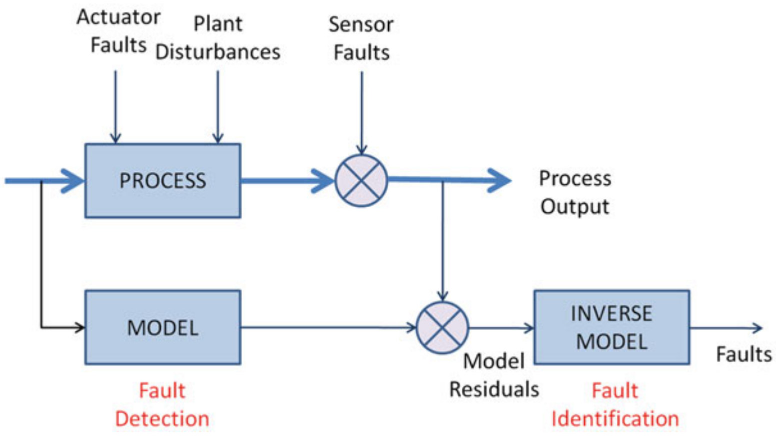

Fig. 1.2 A basic outline of the fault diagnosis problem

can arise abruptly, for example, with the sudden failure of process equipment, or faults can evolve over time, such as associated with gradual wear and tear of equipment or sensor drift.

The primary objective of fault diagnosis is the timely detection of aberrant process or system behaviour, identification of the causes of the fault and elimination of these causes with as little disruption to the process as possible. This is typically accomplished by comparing the actual behaviour of the process with a model representing normal or desirable process behaviour. The detection of process faults is based on monitoring of the deviation between the actual process behaviour and that predicted by the model, with a fault condition flagged when these deviations exceed certain predetermined limits. Once a fault is detected, identification of the root cause(s) of the problem is generally based on an inverse model. Correction of the problem depends on engineering expertise and is typically less well automated than the detection and identification problems. Figure 1.2 shows a schematic outline of the fault detection and identification problem.

1.4 Construction of Diagnostic Models

From a philosophical point of view, all fault diagnostic activities depend on models in one form or another. Models are simply compact representations of knowledge, which can either be explicit or tacit. Explicit knowledge exists in the form of documented equations, facts, rules, heuristics, etc. In contrast, tacit knowledge is more difficult to define and consists of all those things that humans know how to do, but not necessarily how to explain (Polanyi 1958). From a process perspective, it is the best practices, experience, wisdom and unrecordable intellectual property that reside within individuals and teams.

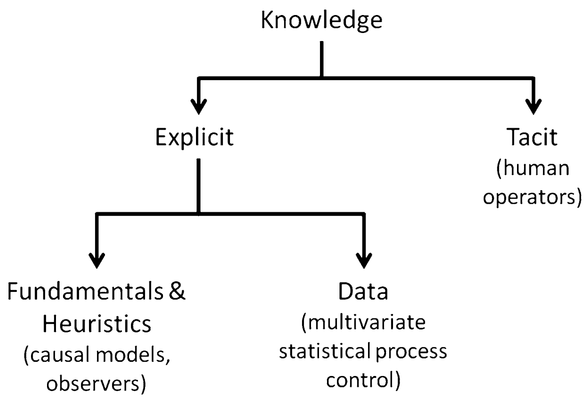

Fig. 1.3 Process fault diagnostic models as representations of process knowledge

Figure 1.3 shows a diagrammatic representation of approaches to fault diagnostic models based on different forms of knowledge. According to this diagram, process fault diagnostic methods can be categorized into models based on formal knowledge (causal models, observers), data (multivariate statistical process control) as well as manual approaches based on the tacit knowledge of human operators.

Classically, models have been derived from first principles or phenomenological models, requiring extensive knowledge of the behaviour of the process and interactions between the components of the process. These include Fickian or non-Fickian diffusion models used in the description of transport processes in leaching or adsorption process, heat conduction in warm plates, etc. Unfortunately, complete knowledge of real processes is often not available or very expensive to acquire. Under these circumstances, explicit knowledge in the form of data or process observations can be used to construct suitable models.

In some instances, tacit process knowledge or operator experience is also used to detect faults in plants. Tacit knowledge is subjective heuristic knowledge that cannot be expressed in words or numbers, often because it is context specific. For example, in froth flotation processes used in the recovery of metals, expert operators are often called upon to diagnose the condition of the process based on the appearance of the flotation froth. Similarly on food processing, the taste of the food is also sometimes used as an early indicator of the quality of the final product.

These alternative approaches to fundamental modelling based on explicit models derived from data or externalization of tacit knowledge have grown remarkably in the last half of the twentieth century based on learning from experience, such as operator knowledge or process data. Learning from data represents a paradigm shift from classical scientific inquiry in which phenomena were explained in terms of materials within a well-defined metric system. Instead, problems are cast in terms of data representation, information and knowledge. For example, a dominant theme that has emerged from the twenty-first-century computational biotechnology is the upgrade of information content in biological data, with strong parallels to the process control perspective (Aldrich 2000; Ogunnaike 1996; Venkatasubramanian 2005). Deriving knowledge from data can be achieved by statistical inferencing or planned experimental campaigns. An alternative and suitable approach that uses few or no assumptions and exploits the ever-growing volumes of process data accumulating in plant data bases is machine learning. Machine learning is concerned with developing machines and software that can discover patterns in data by learning from examples. It brings together insights and tools of mathematics, theoretical and applied computational sciences and statistics. In particular, it overlaps with many approaches that were proposed separately within the statistical community, for example, decision trees (Breiman et al. 1984; Quinlan 1986). Process fault diagnosis can also be cast as a machine learning problem, as outlined in more detail below.

1.5 Generalized Framework for Data-Driven Process Fault Diagnosis

The data-driven construction of models for process fault diagnosis can be cast in a general framework consisting of a number of elements as indicated in Fig. 1.4. These include a data matrix representing the process or system being monitored (**X**); a feature matrix extracted from the data matrix (**F**), from which diagnostic variables are derived for process monitoring and fault diagnosis; a reconstructed data matrix ($\widehat{\boldsymbol{X}}$) serving as a basis for fault identification, as well as an indication of the quality of the extracted features; and, finally, a residual matrix (**E**) serving additionally as a monitoring space.

More formally, the problem can be considered given a set of sample vectors $\{\boldsymbol{x}\}_{i=1}^{N} \in \Re^M$, drawn from the random vector $\boldsymbol{X}$, find the mapping $\Im\colon \Re^M \to \Re^q$ and $\aleph\colon \to \Re^M$, such that for all $i = 1, 2, \ldots, N$, $\Im(\boldsymbol{x}_i) = \boldsymbol{f}_i$ and $\aleph(\boldsymbol{y}_i) = \widehat{\boldsymbol{x}}_i \approx \boldsymbol{x}_i$, where $\{\boldsymbol{f}\}_{i=1}^{N} \in \Re^q$ denote the corresponding set of reduced sample vectors or features drawn from the random vector $\boldsymbol{F}$. M and q denote the dimensionalities of the original and the feature vector or reduced latent variable space, respectively. For data visualization, $q = 2$ or 3 would be normal; otherwise, $q \ll M$.

Derivation of the mappings $\Im$ and $\aleph$ can be done by optimizing one of several possible criteria, such as the minimum mean square error or maximum likelihood criterion. For instance, with principal component analysis, the forward mapping $\Im$ is computed by eigendecomposition of the covariance matrix of the samples.

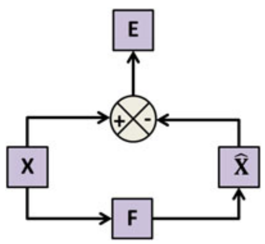

Fig. 1.4 A general framework for data-driven fault diagnosis

The reverse mapping ($\aleph$) is automatically derived from the forward mapping $\Im$. Similarly, in other linear latent variable models, such as independent component analysis, the reverse mapping is first computed from which the forward mapping can then be obtained via pseudo-inverses. Nonlinear methods can be more problematic, since mappings may not be easy to find with nonlinear transformation, and $\aleph$ is usually identified first, after which $\Im$ is defined by some projection operator.

These elements can be generated in various ways. For example, the data matrix can contain measurements of physical process variables in steady-state systems or could arise from the embedding or lagging of coordinates in dynamical systems (trajectory matrix). It could also be a component of a decomposed data matrix associated with multiscale methods.

An overview of data-driven methods to establish process fault diagnostic models is given in Chap. 2. In this book, this generalized diagnostic framework is treated from a machine learning perspective, where feature extraction is viewed as an unsupervised learning problem. Three machine learning paradigms are considered in this context, viz. neural networks, tree-based methods and kernel methods, as discussed in more detail in Chaps. 3, 4 and 5. In the remainder of the book, case studies and applications of the methodologies to different classes of fault conditions are considered.

1.6 Machine Learning

Machine learning is automatic computing based on logical and binary operations to learn tasks from examples. It can also be seen as the study of computational methods designed to improve the performance of machines by automating the acquisition of knowledge from experience or data. Different machine learning paradigms include artificial neural networks (multilayer perceptrons, self-organizing maps, radial basis function neural networks, etc.); instance-based learning (case-based reasoning, nearest neighbour methods, etc.); rule induction; genetic algorithms, where knowledge is typically represented by Boolean features, sometimes as the conditions and actions of rules; statistics; as well as analytical learning. The field of machine learning has originated from diverse technical environments and communities.

1.6.1 Supervised and Unsupervised Learning

A distinction can be made between supervised, unsupervised and reinforcement learning and combinations thereof. In supervised learning, the training data consist of a set of exemplars $\{\boldsymbol{x}, \boldsymbol{y}\}_{i=1}^{N}$, each of which is a pair comprising an input and an output vector, $\boldsymbol{x} \in \Re^{M}$, $\boldsymbol{y} \in \Re^{p}$. If the output is continuous, a regression

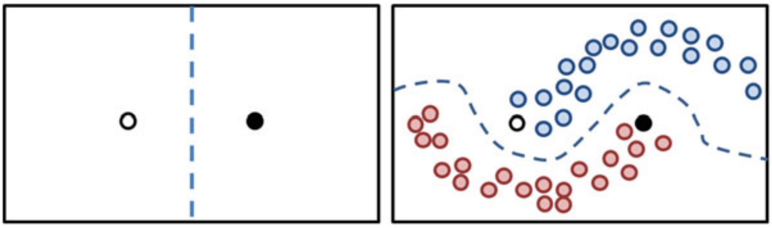

Fig. 1.5 An example of semi-supervised learning where unlabelled data (*red* and *blue markers*) are used in conjunction with labelled data (*black* and *white markers*) for learning the distribution of two classes

function is learnt; otherwise, if the output is discrete, a classification function is learnt. In unsupervised learning problems, unlabelled data are used, i.e. $\{\boldsymbol{x}\}_{i=1}^{N}$. In this case, the outputs represent the structure of the data, which is determined by a cost function to be minimized. In contrast, reinforcement learning uses a scalar reward signal to evaluate input–output pairs by trial and error to discover the optimal outputs for each input. This approach is most suited to problems where optimal input–output mappings are not available a priori, but where any given input–output pair can be evaluated. In this sense, reinforcement learning can be considered to be intermediary to supervised and unsupervised learning, since the use of a reward signal represents some form of supervision.

1.6.2 Semi-supervised Learning

Semi-supervised learning (Zhu et al. 2003; Zhao 2006) is another intermediary to supervised and unsupervised learning and comprises the solution of supervised learning tasks given labelled und unlabelled data are generated by the same distribution or underlying process.

More formally, consider a set of L independently distributed examples, $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_L \in X$, with corresponding labels $y_1, y_2, \ldots, y_L \in Y$. In addition, there are W unlabelled samples available $\boldsymbol{x}_{1+W}$, $\boldsymbol{x}_{2+W}, \ldots,$ $\boldsymbol{x}_{L+W} \in X$. Semi-supervised learning attempts to make use of all the data to improve a *classification* model. This could be accomplished by clustering the unlabelled data and then labelling the clusters with the labelled data, moving the decision boundary away from high-density regions or learning an underlying manifold where the data are located, as indicated in Fig. 1.5. Semi-supervised learning is transductive, when the correct *labels* are inferred for $\boldsymbol{x}_{1+W}, \boldsymbol{x}_{2+W}, \ldots, \boldsymbol{x}_{L+W}$ only and inductive when the correct *mapping* $X \to Y$ is learnt (Vapnik 2006), as shown in Fig. 1.6.

Other approaches to machine learning are linked to the structure of the models. In deep learning, neural network models with a large multiple of layers (deep layering)