A General Paradigm for Applying Machine Learning in Automated Manufacturing Processes

Wei Zhang[1] and Rod Tjoelker[2]

The Boeing Company

P.O. Box 3707, MS {7L-66[1]|7L-22[2]}

Seattle, WA 98124-2207

Abstract. This paper introduces a general data-analysis paradigm for automated manufacturing processes. A number of tasks within the paradigm are discussed to encourage development of better machine learning and analysis solutions.

Introduction

Increasingly factories are applying *advanced manufacturing technologies* in production processes, leading to the use of automated processes called *flexible manufacturing systems* (or *workcells*). A flexible manufacturing system generally consists of a set of devices such as NC (Numerical Control) machines, sensors, and robots (Lauzon *et al.* 1997). Each device may have different levels of flexibility which are configurable or programmable by process designers . During production, operator intervention is required only when problems occur (often signaled by an alarm) or in some designated manual operations such as inspection.

Often advanced automated manufacturing processes have many process monitoring sensors collecting data at frequent intervals, leading to massive amounts of data as the manufacturing process continues. This allows various data-analysis techniques to be applied to provide insight into process improvement.

This paper presents a general data analysis paradigm for an automated manufacturing process. It defines the form of the data and discusses various tasks that are required to process and analyze the data. Following this, it presents a brief summary of common machine learning and statistical methods that can be applied to these tasks and typical technical challenges involved in applying them and making them more efficient. As a short position paper, details and extended discussions are left out.

The Paradigm

There are generally three types of data collected in an automated manufacturing process.

Time-series sensor data: A process may require dozens or as many as several hundred sensors. The sensor data are often collected in high frequencies, such as every second.

Alarm event data: A complicated process is often encoded with hundreds of different types of alarms. Alarms may be categorized as informational, warning, or critical according to the significance of the event outcomes.

Process log data: An automatic process is run by a designated program. A process is often defined as a sequence of tasks where each task can be defined as a sequence of sub-tasks, each with several levels of sub-tasks. Data in the process log often record process-running information such as control parameter settings, start time and end time of each task, outcome of each task (such as whether a task is completed), and operator interventions.

The primary tasks of the data analysis are as follows:

Process monitoring: A central, non-trivial task in process monitoring is deciding what data should be collected. Sensor types and locations are often decided based on engineering knowledge of the process. These decisions should be revisited as data is analyzed and processes are changed. We would like to see this work become more technically guided by using the results of the following analysis.

Assess process efficiency: The goal is to assess information such as machine run-time, down-time, and quality for individual process tasks to understand process efficiency. The results should be presented in a clear, often short form for daily summary use, without missing key efficiency factors.

Find sensor patterns: In order to make good use of sensor data, patterns in multiple sensors as well as individual sensors must be identified. As the number of sensors grows large, automatic learning methods are needed to find patterns that are too numerous or too complex for human analysts to find.

Find causes for alarms: This task is the natural follow-up of the previous task. The goal is to find the conditions that lead to alarms by examining patterns in alarm and sensor data. Finding these conditions and good alarm prediction models may help understand key factors affecting process efficiency and suggest process changes to reduce alarms and improve a process.

Evaluate and improve process operations: This task evaluates the operations on both control parame-

ter settings and operator interventions with the goal of improving them to achieve better product quality and process efficiency, and reduced cost. Finding an "optimal" way to run a process is desirable. This task requires analysis on existing data to understand current operations as well as active explorations to fine tune an existing process.

Adapt a process to changing requirements: A process run by a flexible manufacturing system is used to produce different products or a product under different specifications. This requires that evaluation and comparison of process operations be conducted across different process configurations, which will allow a process to adapt to changing requirements. An active exploration is often taken when a new process is initiated.

Solutions and Issues

The tasks described above are complicated and involve massive amounts of data. It is not uncommon to collect several hundred megabytes of data each day. Efficient automatic or semi-automatic machine learning methods are essential to complete these tasks.

This section attempts to provide a brief summary of common machine learning and statistical methods that can be applied to these tasks and various issues that make the application challenging. The discussion is grouped into three categories:

Understanding: The fundamental goal of the data analysis is process understanding. Various statistical and machine-learning tools can be applied to understand data. A common approach that can be applied to various tasks above is to develop a set of meaningful features and then apply automatic methods such as rule induction or decision trees to come up with a set of rules to explain the relationships (Quinlan 1993, Breiman et al. 1984). Along this process, tools with powerful visualization functions are very useful.

However, even with the help of these tools, this understanding process is still very long. The nature of this work is largely exploratory.

It is a big challenge to speed up the understanding process! For a given application domain, we believe there are domain-specific knowledge and methods that can be acquired or used to help speed up or automate the understanding process.

Prediction: Machine learning has achieved the most progress in developing automatic methods to learn an accurate model for prediction. These models can be decision-trees, neural nets, memory-based models, or stochastic models. Various hybrid models or ensemble methods are also being developed. These provided a rich set of choices for the complicated tasks involving sensor pattern recognition and alarm prediction. One big challenge is how to determine which is the "right" method (note, even an experiment using one method can be very expensive due to the amount of data). Another challenge is to develop methods specific for dependency analysis of time-series data with both discrete and continuous variables so it can be done easily and

efficiently. Research in this area is still primitive while various new algorithms are being developed. One interesting algorithm toward this direction is multi-stream dependency detection (Oates 1996).

Control: The last two tasks involve learning to better control a process. While *statistical process control* and *design of experiment* are often being applied in factories for process engineers to learn how to run a process "optimally", the learning curve is often very slow. *Reinforcement learning* is a novel, advanced approach to process optimization (Kaelbling et al. 1996, Zhang and Dietterich 1995). Based on dynamic programming, this approach learns to evaluate operations to come up with near-optimal process-running policies automatically during process execution. Reinforcement signals are acquired whenever good or bad "responses" are received and then used to adjust the evaluation of operations. Learning can be conducted "actively" by suggesting when an exploratory action can be taken to find better ways to run a process and when an persistent action has to be used in order to accomplish the current work—making a good part (this is very important during production!). With reinforcement learning, "off-line" experiments can be largely limited and a potential improvement of process operations can be found promptly. Furthermore, learning can be conducted for evaluation on different process configurations. This helps make a process much easier to adapt to changing requirements. While this approach seems very promising, much more practice in real industrial settings will be required.

In conclusion, advanced automatic learning methods are required to accomplish the goals of understanding and improvement of automated manufacturing processes. Complex processes generating massive amounts of data are becoming more common and more efficient analysis methods are needed.

References

L. Breiman, J. H. Friedman, R. Olshen, and C. J. Stone. *Classification and Regression Trees.* Wadsworth Internatinal Group, Belmont, CA, 1984.

L. P. Kaelbling, M. L. Littman, and A. W. Moore. Reinforcement learning: a survey. *Journal of AI Research*, 4, 1996.

S.C. Lauzon, J.K. Mills, and b. Benhabib. An implementation methodology for the supervisory control of flexible manufacturing workcells. *Journal of Manufacturing Systems*, 16(2):91–101, 1997.

T. Oates. Search for structure in multiple streams of data. In *Proceedings of the Thirteenth International Conference on Machine Learning*, pages 346–354. Morgan Kaufmann, 1996.

R. Quinlan. *C4.5: Programs for Machine Learning.* Morgan Kaufmann, San Mateo, 1993.

W. Zhang and T. Dietterich. A reinforcement learning approach to job-shop scheduling. In *IJCAI-95*, pages 1114–1120, 1995.