Chapter 54

DATA MINING FOR SELECTION OF MANUFACTURING PROCESSES

Bruno Agard
École Polytechnique de Montréal

Andrew Kusiak
The University of Iowa

Abstract Data Mining tools extract knowledge from large databases. The data generated in manufacturing has not been entirely exploited. This chapter discusses applications of Data Mining in a manufacturing environment. A methodology for selection of manufacturing processes is proposed and illustrated with an industrial scenario.

Keywords: manufacturing, process selection

1. Introduction

Enterprise Resources Planning (ERP) systems generate large volumes of data. The data collection efforts are often driven by productivity improvements. This large quantity of data makes it almost impossible for a person to develop a complete understanding of the entire process without using any tools (see Figure 54.1).

Examples of data sources in manufacturing include:

- Schedules.
- Production capacity, efficiency, failures, etc.
- Manufacturing parameters.
- Process quality.
- Process plans.
The knowledge discovered from the ERP data may benefit a company. The focus of this chapter is on extracting knowledge from industrial databases in support of selection of manufacturing processes.

Some of the data-mining applications of interest to the methodology presented in this chapter are reviewed in Section 2. Section 3 presents a methodology for the selection of manufacturing processes with Data Mining, an example scenario is included. Section 4 concludes the chapter.

2. Data Mining in Engineering

Anand and Büchner (1998) defined Data Mining as the discovery of non-trivial, implicit, previously unknown, and potentially useful and understandable patterns from large data sets.

Westphal and Blaxton (1998) categorized Data Mining functions as classification, segmentation, description and estimation.

- Classification involves assigning labels to new data based on the knowledge extracted from historical data.
- Segmentation (called also clustering) divides a population into smaller sub-populations with similar behaviour according to a predefined metric. It maximizes homogeneity within a group and maximizes heterogeneity between the groups.
- Description and visualization are used to explain the relationships among the data. Frequent patterns may be extracted in the form of \( A \implies B \) rules with two measures of quality: the support which represents the number of times \( A \) occurs as a fraction of the total number of examples and confidence which expresses the number of times \( B \) exists in the data when \( A \) is present.
- Estimation focuses on filing in a missing value in a particular field of an incoming record as a function of other fields in the record (usual statistical regression techniques and neural network are most often employed).