Chapter 3.5
Application of Decision Tree as a Data Mining Tool in a Manufacturing System

S. A. Oke
University of Lagos, Nigeria

ABSTRACT

This work demonstrates the application of decision tree, a data mining tool, in the manufacturing system. Data mining has the capability for classification, prediction, estimation, and pattern recognition by using manufacturing databases. Databases of manufacturing systems contain significant information for decision making, which could be properly revealed with the application of appropriate data mining techniques. Decision trees are employed for identifying valuable information in manufacturing databases. Practically, industrial managers would be able to make better use of manufacturing data at little or no extra investment in data manipulation cost. The work shows that it is valuable for managers to mine data for better and more effective decision making. This work is therefore new in that it is the first time that proper documentation would be made in the direction of the current research activity.

INTRODUCTION

General Overview

In today's digital economy, knowledge is regarded as an asset, and the implementation of knowledge management supports a company to developing innovative products and making critical management strategic decisions (Su, Chen, & Sha, 2005). This digital economy has caused a tremendous explosion in the amount of data that *manufacturing organizations* generate, collect, and store, in order to maintain a competitive edge in the global business (Sugumaran & Bose, 1999). With global competition, it is crucial for organizations to be able to integrate and employ intelligence knowledge in order to survive under the new *business environment*. This phenomenon has been demonstrated in a number of studies, which include the employment of artificial neural network and decision tree to derive knowledge about the job

attitudes of "Generation Xers" (Tung, Huang, Chen, & Shih, 2005). The paper by Tung et al. (2005) exploits the ART2 neural model using the collected data as inputs. Performance classes are formed according to the similarities of a sample frame consisting of 1000 index of Taiwan manufacturing industries and service firms. While there is a plethora of *data mining techniques and tools* available, they present inherent problems for end-users such as complexity, required technical expertise, lack of flexibility, and interoperability, and so on. (Sugumaran & Bose, 1999). Although in the past, most data mining has been performed using symbolic artificial intelligence *data mining algorithms* such as C4.5, C5 (a fast variant of C4.5 with higher predictive accuracy) and CART (Browne, Hudson, Whitley, Ford, & Picton, 2004), the motivation to use *decision tree* in this work comes from the findings of Zhang, Valentine, & Kemp, (2005). The authors claim that *decision tree* has been widely used as a modelling approach and has shown better predictive ability than traditional approaches (e.g., regression). This is consistent with the literature by considering the earlier study by Sorensen and Janssens (2003). The authors conduct an exploratory study that focuses on the automatic interaction detection (AID) — techniques, which belongs to the class of decision tree data mining techniques.

Decision tree is a promising new technology that helps bring *business intelligence* into manufacturing system (Yang et al., 2003; Quinlan, 1987; Li & Shue, 2004). It is a non-parametric modelling approach, which recursively splits the multidimensional space defined by the independent variables into zones that are as homogeneous as possible in terms of response of the dependent variable (Vayssieeres, Plant, Allen-Diaz, 2000). Naturally, *decision tree* has its limitations: it requires a relatively large amount of training data; it cannot express linear relationships in a simple and concise way like regression does; it cannot produce a continuous output due to its binary nature; and it has no unique solution, that is,

there is no best solution (Iverson & Prasad, 1998; Scheffer, 2002). Decision trees are *tree-shaped structures* that represent sets of decisions. Specific *decision tree* methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID) (Lee & Siau, 2001).

Figure 1 is a good illustrative example of potential *sources of data for mining* in manufacturing. The diagram shows the various areas of manufacturing where massive data are generated, managed, and used for decision making. Basically, nine aspects of the *manufacturing organization* are discussed: production system, customer relations, employee database, contractor/supplier unit, product distribution, maintenance, transportation, research and development, and raw materials.

The production system is concerned with transformation of raw materials into finished goods. *Daily production* and target figures are used for mining purposes. Trends are interpreted and the future demand of products is simulated based on estimation from historical data. Data on quality that are also mined relate to the number of accepted products, the number of scraps, and reworks, and so forth. The maintenance controller monitors trends and predicts the future *downtime* and *machinery capacity* data. Customer relations department promotes the image of the company through programs. This department also monitors the growth of the company's profit through the number of additional customers that patronize the company, and also monitors libel suits against the company in the law courts.

Data are also mined from the *employee database*. Patterns observed in this database are used to predict possible employee behaviour, which include possibility of absence from duty. Practical *data mining information* could be obtained from an example of a production supervisor who was last promoted several years ago. If a new employee is engaged and placed higher than him, he may reveal the frustration by handling some of the company's resources and equipment carelessly

Related Content

Functional Dependencies For Value Based Identification In Object-Oriented Databases
Jochens Raschand Hans-Joachim Klein (2002). *Database Integrity: Challenges and Solutions (pp. 250-292).*
www.irma-international.org/chapter/functional-dependencies-value-based-identification/7884

Object-Oriented Database Design
Esperenza Marcosand Paloma Caceres (2001). *Developing Quality Complex Database Systems: Practices, Techniques and Technologies (pp. 210-239).*
www.irma-international.org/chapter/object-oriented-database-design/8277

On the Representation of Temporal Dynamics
Salvatore T. Marchand Gove N. Allen (2003). *Advanced Topics in Database Research, Volume 2 (pp. 37-53).*
www.irma-international.org/chapter/representation-temporal-dynamics/4340

Clustering Schema Elements for Semantic Integration of Heterogeneous Data Sources
Huimin Zhaoand Sudha Ram (2004). *Journal of Database Management (pp. 89-106).*
www.irma-international.org/article/clustering-schema-elements-semantic-integration/3322

A Service Architecture Using Machine Learning to Contextualize Anomaly Detection
Brandon Laughlin, Karthik Sankaranarayananand Khalil El-Khatib (2020). *Journal of Database Management (pp. 64-84).*
www.irma-international.org/article/a-service-architecture-using-machine-learning-to-contextualize-anomaly-detection/245300