Chapter IX

Performance Implications of Knowledge Discovery Techniques in Databases

Balaji Rajagopalan
Oakland University, USA

Ravindra Krovi
University of Akron, USA

ABSTRACT

This chapter introduces knowledge discovery techniques as a means of identifying critical trends and patterns for business decision support. It suggests that effective implementation of these techniques requires a careful assessment of the various data mining tools and algorithms available. Both statistical and machine-learning based algorithms have been widely applied to discover knowledge from data. In this chapter we describe some of these algorithms and investigate their relative performance for classification problems. Simulation based results support the proposition that machine-learning algorithms outperform their statistical counterparts, albeit only under certain conditions. Further, the authors hope that the discussion on performance related issues will foster a better understanding of the application and appropriateness of knowledge discovery techniques.
INTRODUCTION

The volume of data collected by businesses today is phenomenal and is increasing exponentially. The challenge is to integrate and correlate data related to both online and offline sales, customer satisfaction surveys, and server log files. To this end, data mining (DM) - the process of sifting through the mass of organizational (internal and external) data to identify patterns, is critical for decision support. Effective data mining has several applications, like fraud detection and bankruptcy prediction (Tam & Kiang, 1992; Lee, Han, & Kwon, 1996; Kumar, Krovi, & Rajagopalan, 1997), strategic decision-making (Nazem & Shin, 1999), and database marketing (Brachman, R.J. Khabaza, T. Kloesgen, W. Piatetsky-Shapiro, G. & Simoudis, E, 1996).

Today, businesses have the unique opportunity for using such techniques for target marketing and customer relationship management. Analysis of massive data collected by businesses can support intelligence-gathering efforts about their competition, product, or market. Intelligent tools based on rules derived from web mining can also play an important role in personalization related to site content and presentation. Recently, there has been considerable interest on how to integrate and mine such data (Mulvenna, Anand, & Buchner, 2000; Brachman et al., 1996).

Business databases in general pose a unique problem for pattern extraction because of their complex nature. This complexity arises from anomalies such as discontinuity, noise, ambiguity, and incompleteness (Fayyad, Piatetsky-Shapiro & Smyth, 1996). Historically, decision makers had to manually deduce patterns using information generated by query reporting systems. One level of analytical sophistication above this was the ability to look at the data and perform analyses such as What-If and goal seeking. More recently, online analytical processing

Figure 1: A framework for analytical processing

Copyright © 2003, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.
A Unified Fuzzy Data Model: Representation and Processing
www.igi-global.com/article/unified-fuzzy-data-model/62033?camid=4v1a