# Chapter 6.1
# Data Mining and Decision Support for Business and Science

**Auroop R. Ganguly**
*Oak Ridge National Laboratory, USA*

**Amar Gupta**
*University of Arizona, USA*

**Shiraj Khan**
*University of South Florida, USA*

## INTRODUCTION

### Analytical Information Technologies

Information by itself is no longer perceived as an asset. Billions of business transactions are recorded in enterprise-scale data warehouses every day. Acquisition, storage, and management of business information are commonplace and often automated. Recent advances in remote or other sensor technologies have led to the development of scientific data repositories. Database technologies, ranging from relational systems to extensions like spatial, temporal, time series, text, or media, as well as specialized tools like geographical information systems (GIS) or online analytical processing (OLAP), have transformed the design of enterprise-scale business or large scientific applications. The question increasingly faced by the scientific or business decision-maker is not how one can get more information or design better information systems but what to make of the information and systems already in place. The challenge is to be able to utilize the available information, to gain a better understanding of the past, and to predict or influence the future through better decision making. Researchers in data mining technologies (DMT) and decision support systems (DSS) are responding to this challenge. Broadly defined, data mining (DM) relies on scalable statistics, artificial intelligence, machine learning, or knowledge discovery in databases (KDD). DSS utilize available information and DMT to provide a decision-making tool usually relying on human-computer interaction. Together, DMT and DSS represent the spectrum of analytical information technologies (AIT) and provide a unifying platform for an optimal

*Table 1. Analytical information technologies*

*Data Mining Technologies*
- Association, correlation, clustering, classification, regression, database knowledge discovery
- Signal and image processing, Nonlinear systems analysis, time series and spatial statistics, time and frequency domain analysis
- Expert systems, Case-based reasoning, System dynamics
- Econometrics, Management Science

*Decision Support Systems*
- Automated Analysis and Modeling
  - Operations Research
  - Data Assimilation, Estimation and Tracking
- Human Computer Interaction
  - Multidimensional OLAP and spreadsheets
  - Allocation and consolidation engine, alerts
  - Business workflows and data sharing

*Table 2. Application examples*

- Bio-Informatics
- Genomics
- Hydrology, Hydrometeorology
- Weather Prediction
- Climate Change Science
- Remote Sensing
- Smart Infrastructures
- Sensor Technologies
- Land-use, Urban Planning
- Materials Science

*Business and Economics*
- Financial Planning
- Risk Analysis
- Supply Chain Planning
- Marketing Plans
- Text and Video Mining
- Handwriting/Speech Recognition
- Image and Pattern Recognition
- Long-range Economic Planning
- Homeland Security

combination of data dictated and human-driven analytics.

## BACKGROUND

Tables 1 and 2 describe the state of the art in DMT and DSS for science and business, and provide examples of their applications.

Researchers and practitioners have reviewed the state of the art in analytic technologies for business (Apte et al., 2002; Kohavi et al., 2002; Linden & Fenn, 2003) or science (Han et al., 2002), as well as data mining methods, software, and standards (Fayyad & Uthurusamy, 2002; Ganguly, 2002a; Grossman et al., 2002; Hand et al., 2001; Smyth et al., 2002) and decision support systems (Carlsson & Turban, 2002; Shim et al., 2002).

## MAIN THRUST

### Scientific and Business Applications

Rapid advances in information and sensor technologies (IT and ST) along with the availability of large-scale scientific and business data repositories or database management technologies, combined with breakthroughs in computing technologies, computational methods, and processing speeds, have opened the floodgates to data-dictated models and pattern matching (Fayyad & Uthurusamy, 2002; Hand et al., 2001). The use of sophisticated and computationally-intensive analytical methods is expected to become even more commonplace with recent research breakthroughs in computational methods and their commercialization by leading vendors (Bradley et al., 2002; Grossman et al., 2002; Smyth et al., 2002).

Scientists and engineers have developed innovative methodologies for extracting correlations