



Chapter III

Data Mining and Knowledge Discovery

Andi Baritchi, Corporate Data Systems, USA

ABSTRACT

*In today's business world, the use of computers for everyday business processes and data recording has become virtually ubiquitous. With the advent of this electronic age comes one priceless by-product — **data**. As more and more executives are discovering each day, companies can harness data to gain valuable insights into their customer base. Data mining is the process used to take these immense streams of data and reduce them to useful knowledge.*

Data mining has limitless applications, including sales and marketing, customer support, knowledge-base development, not to mention fraud detection for virtually any field, etc. "Data mining," a bit of a misnomer, refers to mining the data to find the gems hidden inside the data, and as such it is the most often-used reference to this process. It is important to note, however, that data mining is only one part of the Knowledge Discovery in Databases process, albeit it is the workhorse. In this chapter, we provide a concise description of the Knowledge Discovery process,

from domain analysis and data selection, to data preprocessing and transformation, to the data mining itself, and finally the interpretation and evaluation of the results as applied to the domain.

We describe the different flavors of data mining, including association rules, classification and prediction, clustering and outlier analysis, customer profiling, and how each of these can be used in practice to improve a business' understanding of its customers. We introduce the reader to some of today's hot data mining resources, and then for those that are interested, at the end of the chapter we provide a concise technical overview of how each data-mining technology works.

INTRODUCTION

In today's business world, the use of computers for everyday business processes and data recording has become virtually ubiquitous. With the advent of this electronic age comes one priceless by-product — *data*. Virtually every large corporation now records all transactions that take place, no matter how small or insignificant, and stores them in a large and complex data warehouse.

These data warehouses are growing at an ever-increasing rate, and there is no end in sight. Due to their sheer bulk, data warehouses are generally impossible to directly analyze by humans looking for interesting patterns or trends. In most cases, the data contained in the data warehouses is too valuable to simply purge or expire. Paradoxically, the data itself is worthless unless there is a method of analyzing it from the “big picture” perspective. Only from this perspective can the researcher or business person gain the very valuable secrets locked deep within the data.

Hence, there is an urgent need for tools that can analyze the data and search for interesting patterns or information that may be embedded below the surface. It is important to note that, in most large databases, there are virtually an infinite number of patterns available to extract, and extracting all of them would be no more helpful than browsing the large collection of raw data itself. The key is to mine the data for interesting patterns. These are the less obvious patterns in the data, those that provide some insight into underlying market trends, customer preferences, fault and fraud detection, and so forth.

To fulfill this need, researchers from the fields of artificial intelligence and database systems collaborated to design various approaches in the field of knowledge discovery to extract hidden patterns from data. In this chapter, we describe the knowledge-discovery process and highlight the procedures

11 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/data-mining-knowledge-discovery/6064

Related Content

Augmenting Analytical CRM Strategies with Social BI

Daniel Beverungen, Mathias Eggert, Matthias Voigt and Michael Rosemann (2013). *International Journal of Business Intelligence Research* (pp. 32-49).
www.irma-international.org/article/augmenting-analytical-crm-strategies-with-social-bi/83477

Identifying Technical Requirements for a Mobile Business Analytics Application

Tuncay Bayrak (2021). *International Journal of Business Analytics* (pp. 91-103).
www.irma-international.org/article/identifying-technical-requirements-for-a-mobile-business-analytics-application/288060

Influence Estimation and Opinion-Tracking Over Online Social Networks

Luis E. Castro and Nazrul I. Shaikh (2018). *International Journal of Business Analytics* (pp. 24-42).
www.irma-international.org/article/influence-estimation-and-opinion-tracking-over-online-social-networks/212633

Sustainability in Transformational and Transactional Leadership: Turkish Textile Industry Example

Korhan Arun (2024). *Applying Business Intelligence and Innovation to Entrepreneurship* (pp. 241-257).
www.irma-international.org/chapter/sustainability-in-transformational-and-transactional-leadership/342323

Bayesian Variable Selection

Oleg Okun (2014). *Encyclopedia of Business Analytics and Optimization* (pp. 241-250).
www.irma-international.org/chapter/bayesian-variable-selection/107231