Chapter XVIII

Translating Advances in Data Mining to Business Operations: The Art of Data Mining in Retailing

Henry Dillon
Independent Consultant, UK

Beverley Hope
Victoria University of Wellington, New Zealand

ABSTRACT

Knowledge discovery in databases (KDD) is a field of research that studies the development and use of various data analysis tools and techniques. KDD research has produced an array of models, theories, functions and methodologies for producing knowledge from data. However, despite these advances, nearly two thirds of information technology (IT) managers say that data mining products are too difficult to use in a business context. This chapter discusses how advances in data mining translate into the business context. It highlights the art of business implementation rather than the science of KDD.

INTRODUCTION

In the past, high storage and processing costs meant that businesses had to be selective about what data they stored. Today, this restriction has been removed as costs of data storage plummet. In addition, there are now more opportunities for capturing detailed data, particularly with the increase in e-commerce activities, where detailed

business transactions can be traced (Kimball & Merz, 2000). As the volume and detail of stored data increases, the demand for effective and efficient analysis tools also increases (Brachman, Khabaza, Kloesgen, Piatetsky-Shapiro & Simoudis, 1996).

KDD has been rigorously researched, particularly in the area of data mining (Fayyad, Piatetsky-Shapiro, Smyth & Uthurusamy, 1996). This has resulted in an array of models, theories, functions and methodologies for producing knowledge from data. Despite these advances, nearly two thirds of IT managers say that data mining products are too difficult to use in a business context (Foley & Russell, 1998). This is because, along with the scientific aspects of KDD, a business needs the artistic application of KDD to business (Kimball & Merz, 2000). While it is possible to propose strategies and methods for best practice in applying data mining to a business context (Brachman et al., 1996), this would be from an academic perspective. This chapter does not seek to do this. Instead, we focus on the pragmatic issue of how advances in data mining can be used in the business world, using action research in an energy firm to demonstrate the potential. We first present a review of the relevant academic literature and then apply this knowledge to a business case study in a major gas service station corporation.

BACKGROUND

The purpose of this section is to provide an understanding of the concepts of data mining to be used in discussing the case study. Here, we briefly address three concepts: data warehousing, KDD and data mining.

Data Warehousing

In 1990, William Inmon coined the term "data warehousing" (Sakaguchi & Frolick, 1997). The new concept differed from previous data storage concepts by incorporating data extracted from a variety of production databases, rather than focusing on the storage of raw production data from individual sources (Inmon, 1996). The intent was to construct an architecture that improved data analysis and decision support. Inmon identified four properties of a data warehouse:

- **Subject Oriented:** In the data warehouse, there is a shift from application-oriented data to decision-support data. If designed well, subject-oriented data will provide a stable image of business processes, capturing the basic nature of the business environment.
- **Integrated:** The warehouse consolidates application data from different legacy systems and eliminates data inconsistencies.
- **Time-Variant:** Each data point is associated with a point in time. Warehoused data can be compared along a time axis, unlike transactional data, which capture a moment in time.
- **Non-Volatile:** The database absorbs new data, integrating it with previous data, that is, new data is appended rather than substituted.

Data warehouses provide an abundance of data for analysis. However as data warehouses grow in size, users encounter information overload issues and find that their traditional applications are inadequate to access and analyze the data. Advances in KDD seek to remedy this problem.

Related Content

Classification and Visualization of Alarm Data Based on Heterogeneous Distance
Boxu Zhaoand Guiming Luo (2018). *International Journal of Data Warehousing and Mining (pp. 60-80).*
www.irma-international.org/article/classification-and-visualization-of-alarm-data-based-on-heterogeneous-distance/202998

Speckle Noise Filtering Using Back-Propagation Multi-Layer Perceptron Network in Synthetic Aperture Radar Image
Khwairakpam Amitab, Debdatta Kandarand Arnab K. Maji (2016). *Research Advances in the Integration of Big Data and Smart Computing (pp. 280-301).*
www.irma-international.org/chapter/speckle-noise-filtering-using-back-propagation-multi-layer-perceptron-network-in-synthetic-aperture-radar-image/139408

Data Mining System Execution Traces to Validate Distributed System Quality-of-Service Properties
James H. Hill (2013). *Data Mining: Concepts, Methodologies, Tools, and Applications (pp. 751-774).*
www.irma-international.org/chapter/data-mining-system-execution-traces/73469

Algebraic Reconstruction Technique in Image Reconstruction Based on Data Mining
Zhong Qu (2006). *International Journal of Data Warehousing and Mining (pp. 1-15).*
www.irma-international.org/article/algebraic-reconstruction-technique-image-reconstruction/1767

Machine Learning Techniques for Network Intrusion Detection
Tich Phuoc Tran, Pohsiang Tsai, Tony Janand Xiangjian He (2010). *Dynamic and Advanced Data Mining for Progressing Technological Development: Innovations and Systemic Approaches (pp. 273-299).*
www.irma-international.org/chapter/machine-learning-techniques-network-intrusion/39646