Chapter XIX

Impediments to Exploratory Data Mining Success

Jeff Zeanah
Z Solutions, Inc., USA

ABSTRACT

This chapter discusses impediments to exploratory data mining success. These impediments were identified based on anecdotal observations from multiple projects either reviewed or undertaken by the author and are classified into four main areas: data quality; lack of secondary or supporting data; insufficient analysis manpower; lack of openness to new results. Each is explained, and recommendations are made to prevent the impediment from interfering with the organization's data mining efforts. The intent of the chapter is to provide an organization with a structure to anticipate these problems and to prevent the occurrence of these problems.

INTRODUCTION

Organizations of all kinds are experimenting with the application of data mining techniques. They may refer to these projects as data warehousing applications, market research or data mining. Regardless of the terminology, data mining applications are intended to provide an organization with a better understanding of the environment or market in which they operate.

There are two general types of data mining undertakings. Relatively well understood are the traditional scoring applications in which observations are scored to determine if they met certain criteria. In these projects, an organization typically will apply a set of tools to a large database, such as a mailing list. For example, a charity which is considering a mailing to solicit donations will score its list to determine the most likely candidates to be solicited. By using data mining to examine the characteristics of individuals who donated in the past, the charity can reduce a mailing list of 2 million households to a list of the 200,000 households most likely to donate. Soliciting this smaller list will be more profitable than "wasting" a mailing to the 1.8 million households that are very unlikely to respond. In this effort, there is not a great need to understand why the households were selected, only whether or not the refinement of the list leads to a higher response rate and increases the profitability of the mailing.

Conversely, exploratory data mining is designed to provide strategic insights from the data and guidance for future strategic or operational decision-making. Consider the following example. A consumer products manufacturer is interested in characteristics of consumers who buy its products, rather than its competitor's. Are these customers younger or older? Are they married or single? What is their ethnicity? Are their household incomes higher or lower? Simple queries of the company's data warehouse can be used to answer most of these questions. If the data is available, the average household income of the company's customers easily may be compared to the average household income of a competitor's customers. This is exploratory data mining. And for many organizations, the ability to read the databases and perform these queries is the extent of their data mining activities. In fact, for very large databases, this can be nontrivial, requiring substantial effort.

Simple queries may not provide all the answers the company needs. For example, the company discovers that its customers have higher household incomes than the competitor's customers have, are older and are more likely to be married. The organization realizes that many married families have higher incomes than do single households (two incomes versus one), and many older households have higher incomes than do the younger beginning households. The company wonders: "Are the customers who prefer our product older and happen to be married at a higher percentage or are our customers married and happen to be older?" Or are they just higher income people? Whatever the relationship, is this the same as it was five years ago? And most important, what will they be like five years from now? If the overall population ages, will that help the company sales? Will it help the company's sales only if the aging population stays married? Is the company not seeing a hidden trend that may change the company's strategic direction? These questions are not going to be answered through simple methods. The solution will be found only through causal predictive modeling and similar investigations.

Many organizations lack the ability to answer these difficult questions. When they began collecting data in their data warehouse, they expected that they would be able to

18 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/impediments-exploratory-data-mining-success/27922

Related Content

Towards Comparative Mining of Web Document Objects with NFA: WebOMiner System
C. I. Ezeifeand Titas Mutsuddy (2012). *International Journal of Data Warehousing and Mining (pp. 1-21).*
www.irma-international.org/article/towards-comparative-mining-web-document/74753

Implications for Nursing Research and Generation of Evidence
Suzanne Bakken, Robert Lucero, Sunmoo Yoonand Nicholas Hardiker (2013). *Data Mining: Concepts, Methodologies, Tools, and Applications (pp. 1082-1096).*
www.irma-international.org/chapter/implications-nursing-research-generation-evidence/73485

Volunteer Data Warehouse: State of the Art
Amir Sakka, Sandro Bimonte, Francois Pinetand Lucile Sautot (2021). *International Journal of Data Warehousing and Mining (pp. 1-21).*
www.irma-international.org/article/volunteer-data-warehouse/286613

A Workload Assignment Strategy for Efficient ROLAP Data Cube Computation in Distributed Systems
Ilhyun Suhand Yon Dohn Chung (2016). *International Journal of Data Warehousing and Mining (pp. 51-71).*
www.irma-international.org/article/a-workload-assignment-strategy-for-efficient-rolap-data-cube-computation-in-distributed-systems/168486

Document Classification
(2021). *Developing a Keyword Extractor and Document Classifier: Emerging Research and Opportunities (pp. 132-136).*
www.irma-international.org/chapter/document-classification/268466