Chapter 5.12
Data Mining in the Federal Government

Les Pang
National Defense University, USA

INTRODUCTION

Data mining has been a successful approach for improving the level of business intelligence and knowledge management throughout an organization. This article identifies lessons learned from data mining projects within the federal government including military services. These lessons learned were derived from the following project experiences:

- Defense Medical Logistics Support System Data Warehouse Program
- Department of Defense (DoD) Defense Financial and Accounting Service (DFAS) "Operation Mongoose"
- DoD Computerized Executive Information System (CEIS)
- Department of Transportation (DOT) Executive Reporting Framework System
- Federal Aviation Administration (FAA) Aircraft Accident Data Mining Project
- General Accounting Office (GAO) Data Mining of DoD Purchase and Travel Card Programs

- U.S. Coast Guard Executive Information System
- Veteran Administrations (VA) Demographics System

BACKGROUND

Data mining involves analyzing diverse data sources in order to identify relationships, trends, deviations and other relevant information that would be valuable to an organization. This approach typically examines large single databases or linked databases that are dispersed throughout an organization. Pattern recognition technologies and statistical and mathematical techniques are often used to perform data mining. By utilizing this approach, an organization can gain a new level of corporate knowledge that can be used to address its business requirements.

Many agencies in the federal government have applied a data mining strategy with significant success. This chapter aims to identify the lessons gained as a result of these many data mining

implementations within the federal sector. Based on a thorough literature review, these lessons were uncovered and selected by the author as being critical factors which led toward the success of the real-world data mining projects. Also, some of these lessons reflect novel and imaginative practices.

MAIN THRUST

Each lesson learned (indicated in **boldface**) is listed below. Following each practice is a description of illustrative project or projects (indicated in *italics*), which support the lesson learned.

Avoid the Privacy Trap

DoD Computerized Executive Information System - Patients as well as the system developers indicate their concern for protecting the privacy of individuals—their medical records need safeguards. "Any kind of large database like that where you talk about personal info raises red flags," said Alex Fowler, a spokesman for the Electronic Frontier Foundation. "There are all kinds of questions raised about who accesses that info or protects it and how somebody fixes mistakes" (Hamblen, 1998).

Proper security safeguards need to be implemented to protect the privacy of those in the mined databases. Vigilant measures are needed to ensure that only authorized individuals have the capability of accessing, viewing and analyzing the data. Efforts should also be made to protect the data through encryption and identity management controls.

Evidence of the public's high concern for privacy was the demise of the Pentagon's $54 million Terrorist Information Awareness (originally, Total Information Awareness) effort—the program in which government computers were to be used to scan an enormous array of databases for clues and patterns related to criminal or terrorist ac-

tivity. To the dismay of privacy advocates, many government agencies are still mining numerous databases (General Accounting Office, 2004; Gillmor, 2004). "Data mining can be a useful tool for the government, but safeguards should be put in place to ensure that information is not abused," stated the chief privacy officer for the Department of Homeland Security (Sullivan, 2004). Congressional concerns on privacy are so high that the body is looking at introducing legislation that would require agencies to report to Congress on data mining activities to support homeland security purposes (Miller, 2004).

Steer Clear of the "Guns Drawn" Mentality if Data Mining Unearths a Discovery

DoD Defense Finance & Accounting Service's Operation Mongoose was a program aimed to discover billing errors and fraud through data mining. About 2.5 million financial transactions were searched to locate inaccurate charges. This approach detected data patterns that might indicate improper use. Examples include purchases made on weekends and holidays, entertainment expenses, highly frequent purchases, multiple purchases from a single vendor and other transactions that do not match with the agency's past purchasing patterns. It turned up a cluster of 345 cardholders (out of 400,000) who had made suspicious purchases.

However, the process needs some fine-tuning. As an example, buying golf equipment appeared suspicious until it was learned that a manager of a military recreation center had the authority to buy the equipment. Also, casino-related expense revealed to be a commonplace hotel bill. Nevertheless, the data mining results have shown sufficient potential that data mining will become a standard part of the Department's efforts to curb fraud.

Related Content

Algorithms for Data Mining

Tadao Takaoka, Nigel K.L. Popeand Kevin E. Voges (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications (pp. 1301-1319).*

www.irma-international.org/chapter/algorithms-data-mining/7700

Evaluation of Data Mining Methods

Paolo Giudici (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications (pp. 364-370).*

www.irma-international.org/chapter/evaluation-data-mining-methods/7651

DWFIST: The Data Warehouse of Frequent Itemsets Tactics Approach

Rodrigo Salvador Monteiro, Geraldo Zimbrao, Holger Schwarz, Bernhard Mitschangand Jano Moreira de Souza (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications (pp. 3142-3163).*

www.irma-international.org/chapter/dwfist-data-warehouse-frequent-itemsets/7825

Using Data Mining for Forecasting Data Management Needs

Qingyu Zhangand Richard S. Segall (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications (pp. 2088-2104).*

www.irma-international.org/chapter/using-data-mining-forecasting-data/7750

Intelligent Query Answering

Zbigniew W. Rasand Agnieszka Dardzinska (2005). *Encyclopedia of Data Warehousing and Mining (pp. 639-643).*

www.irma-international.org/chapter/intelligent-query-answering/10675