

2011

The BTS Group

Philip M. Porreca

DATA MINING CONCEPTS

A White Paper

The BTS Group, LLC
5616 231st Ave SE
Issaquah, WA 98029
info@thebtsgroup.com

Introduction3
Benefits of Data Mining3
Data Mining Overview3
 Tool Comparisons3
 Examples5
Data Mining Technical Concepts7
 Unsupervised Learning Algorithms7
 Supervised Learning or Predictive Algorithms8
Summary9

DATA MINING CONCEPTS WHITE PAPER

INTRODUCTION

This paper is intended for anyone who is interested in applying Data Mining techniques in practical business and commercial situations. Specifically, it describes the value that data mining brings to the area of business analysis and how it complements traditional reporting systems and techniques. The foundation of Data Mining techniques will be described and followed with concrete examples of use.

BENEFITS OF DATA MINING

Today's competitive marketplace demands that companies aggressively manage their customers and the invaluable information they possess related to their customers. It is in this area that data mining can be of tremendous value by analyzing the massive amounts of collected data and finding hidden information, which can allow you to better understand your customers and anticipate their behavior. BTS builds data mining software applications that help you to uncover the hidden information about your customers that resides in the multitude of transactional systems. Equipped with this information, organizations will be capable of enhancing your customer relationships and increasing your understanding of their habits, which enables them to:

- Profile customers and understand their behaviors
- Target profitable customers with the correct set of offerings
- Increase retention of existing customers and avoid churn
- Reduce costs related to customer acquisition
- Improve profit margins

DATA MINING OVERVIEW

In a nutshell, Data Mining (DM) enables organizations to find hidden patterns in, and develop predictions based upon, the large volumes of data stored in their operational systems and a variety of independent data stores. A company can discover associations of events, patterns, and attributes that would normally be unseen by the naked eye via the implementation of a variety of DM algorithms and their associated methodologies. DM can also help organization to perform a variety of complex analysis including demand forecasting, what if analysis, and marketing program effectiveness.

TOOL COMPARISONS

There are a number of categories in which to classify BI tools found in the marketplace (See Figure1). It is important to first understand the various categories and how DM compares. The first grouping is that of

current state reporting which provides a surface level view of information and is useful for tactical operational decisions. This can be sub-divided into two areas traditionally known as Ad-hoc and Query & Reporting. The Query & Reporting tools generally provide information derived from SQL queries that are preconfigured by a vendor or IT department, and can be characterized as static and parameterized. Ad-hoc reports, while also based on SQL queries, differ in that they are configured by a business analyst on an as-needed basis. These reports can be characterized as inflexible and user-centric. The next category is that of OLAP Reporting, which is capable of providing an additional level of sophisticated analysis. OLAP Reporting applications use the multi-dimensional characteristics of Data Marts to provide more diverse capabilities to assist with business functions such as enterprise budgeting and financial reporting, demand planning, and market analysis. The characteristics of Data Marts allow the flexible traversing of data by enabling the easy retrieval of summarized data, "slicing & dicing" data into separate marts based on business unit or operational function, and simplifying the reporting process through user interfaces that are more flexible than writing SQL-based queries.

Trend and Prediction tools take the sophistication of OLAP Reporting to yet another level. While OLAP Reporting could potentially detect some patterns in the company data, it is, by definition, based on a slice of information. This limiting factor can be thought of as a filter. The fact that an analyst uses a-priori business knowledge, or assumptions, to create the originating dataset can also be considered a filter. By contrast, DM algorithms directly analyze all data, removing the inhibiting filters, and finding patterns that would otherwise be unseen. These "hidden" patterns are not only useful for understanding trends and behaviors, but can be used for predicting the outcome of business decisions.

Query & Reporting	Data Marts	Data Mining
Detail, Ad-hoc Reporting	Sliced/Diced Summarized/Drill Down	Cluster, Associations, Pattern, Prediction
Information	Analysis	Insight
<i>Who made purchases through the online store last year?</i>	<i>Summarized online sales by Location and Month</i>	<i>Who will make online purchases net year?</i>

EXAMPLES

DM is particularly useful in the area of customer management. This includes understanding customers past purchasing behaviors, creating a custom offering and predicting the outcome of specialized marketing campaigns. A couple concrete examples of CRM applications that can be offered using DM are as follows:

- Customer Profiling Application – allowing customers to be segmented based on their subscribed services, activities and behavior. Figure 2 below shows a Decision Tree analysis for call-forwarding subscription. As deduced from the tree, a targeted campaign directed for similar services can be launched to target those households with 3 adults or more since that particular segment has the most consistent success with past call-forwarding subscription.

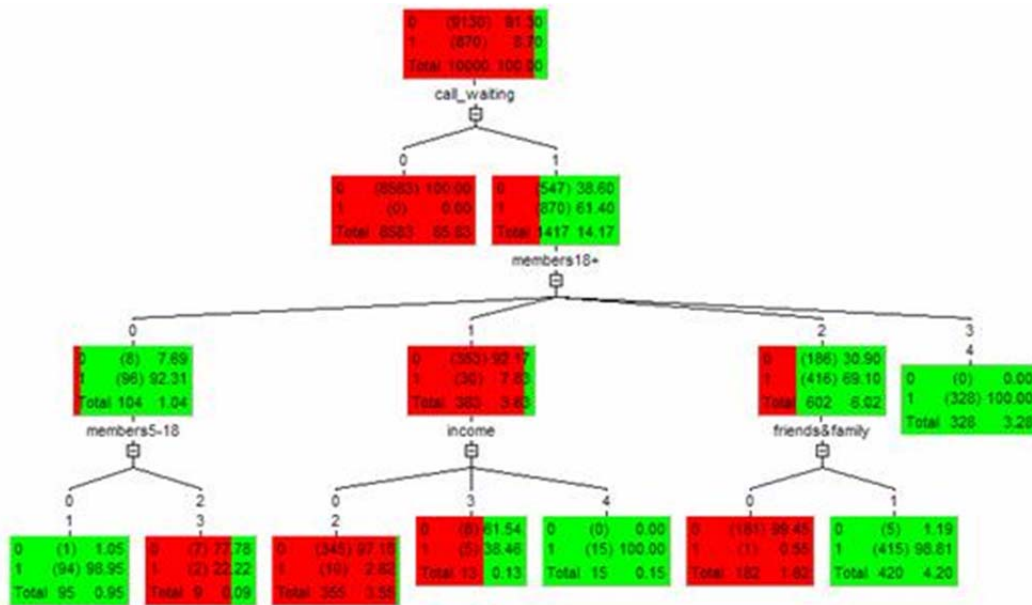


Figure 2 – Customer Profiling Example – Decision Tree

- Another example is a real-time environment used in online sales. As already implemented in many online stores, Real-time Recommendations Applications as depicted in Figure 3 provide cross-sell and up-sell recommendations based on factors such as the customer's profile, past transactions and demographic information. These same techniques can be applied to systems that enable telephonic-based sales and support organizations.

Mastering the SAP Business Information Warehouse
 by Kevin McDonald, Andreas Wilmesmeier, David C. Dixon, W. H. Insko
 List Price: \$60.00
 Price: \$60.00 & This item ships for FREE with Super Saver Shipping. See details.
 Availability: Usually ships within 24 hours

Real-time cross-sell based on the profile and behavior of the customer

Personal Recommendation:

- *SAP BW Certification: A Business Information Warehouse Study Guide* by Catherine M. Roze (Author) (Paperback)
- *Configuring SAP 8/2 FI/CO: The Essential Resource for Configuring the Financial and Controlling Modules* by David Nowak, Quentin Hurst (Hardcover)
- *SAP BW Reporting Made Easy, 2.08/2.10* by SAP Labs Inc. Simplification Group (Paperback)
- *SAP Objects: Introduction to Programming SAP Applications* by Horst Keller, Sascha Kruger (Hardcover)
- *Practical Workflow for SAP - Effective Business Processes using SAP's WebFlow Engine* by Alan Rickayzen, et al (Hardcover)

Figure 3 – Real-time Cross-sell Recommendations

Examples of other usages of Data Mining applications are as follows:

- Predicting the probability that a certain event will occur:
 - *Offering x will be subscribed by group y.*
 - Method: Based on information about past similar services, the new offering can be predicted for its success or failures.
 - *Credit check process*
 - Method: By correlating historical information about customers' profiles and their past performance, a probability of credit problem can be predicted
 - *Likelihood of a positive feedback for a marketing campaign.*
 - Method: Analyze market segment with its attribute and predict the success rate of the new campaign.
 - *Likelihood of customer churn or Employee attrition .*
 - Method: Based on historical information, the probability of a churn or attrition for a specific customer or employee can be calculated
- Profiling or clustering of customers
 - *Customer segmentation for a product or service.* Customers can be segmented and profiled individually for a subscribed service or a product purchase
 - *Association of Product sales X and Y.* Through correlation, sales of products can be individually associated with others. Besides marketing, this helps operational and supply chain planning.

DATA MINING TECHNICAL CONCEPTS

The technical foundation of Data mining is to learn based on past information and decisions. As more samples and cases are processed, the learning development becomes more and more accurate to handle a variety of conditions. The methods in which algorithms can learn are characterized by two categories.

UNSUPERVISED LEARNING ALGORITHMS

Unsupervised learning is a process where goal(s) are specified to direct the learning development of the algorithm to reach its prediction. For unsupervised learning, no specific dependent variable is specified and all independent variables are provided into the system for processing. This is found in Clustering techniques where the application groups similar instances together into clusters of data points. Each cluster is then described in the form of allocation of features within the segment and a fit distribution can be measured. Once a cluster model is created, a new case can be routed into the model to find its best-fit cluster.

As shown Figure 4 below, a visually rich tool can clearly show the feature distributions for each cluster. In the bottom of the diagram distribution and allocation characteristics of the clusters are displayed.

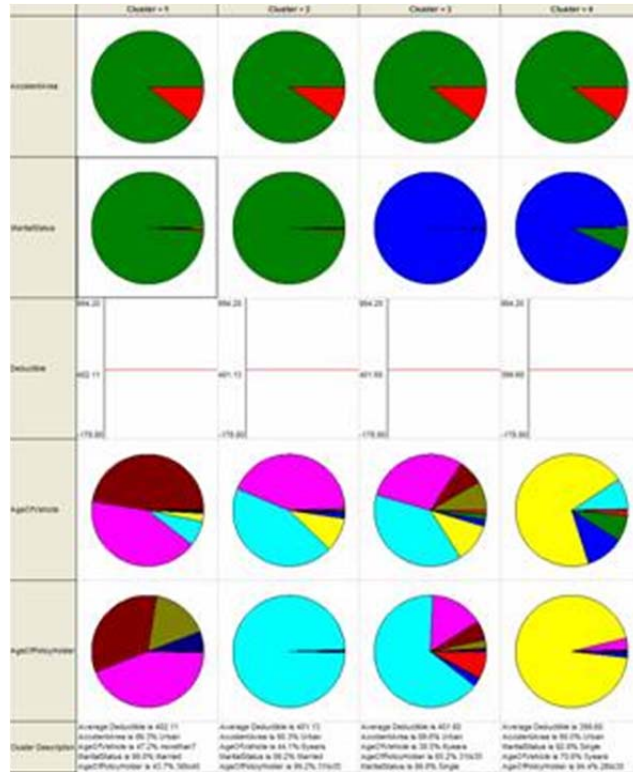


Figure 4 - Clustering Example

Clustering has a wide range of applicability that varies from customer segmentation analysis used for the purpose of targeted marketing campaigns, to profile segmentation of problematic or profitable customers for financial credit or fraud analysis.

An example of a basic clustering algorithm is K-means where, for k clusters, the algorithm iteratively minimizes the distance of a case to the mean by adjusting the position of the mean in each cluster. The iteration ends when no global mean distance can be further improved. There can be a few variations of K-means that are offered by a specific vendor, essentially to improve performance and to support more variability of its inputs.

SUPERVISED LEARNING OR PREDICTIVE ALGORITHMS

With target objectives set by the data analyst, predictive algorithms are capable of learning and predicting new cases based on trends and patterns found in historical information. Classification is a predictive technique that requires its target goals to be presented as true or false. Decision-tree is an example of a Classification technique. More complex and sophisticated mathematical solutions are needed to predict a number, such as prediction of the future revenue or a business performance score. One of the more generally used algorithms in prediction is a family of non-linear solutions called Neural Nets. Neural Nets solve complex regression problems and can be adapted for classification as well. The most widely implemented Neural Nets employ a layer of hidden units or neurons that are fully connected with its inputs and outputs. As the number of hidden unit increases, the complexity of the

solutions becomes more and more non-linear. Figure 5 illustrates an example of a prediction result of BMW owners based on customer profiling and purchasing habits.

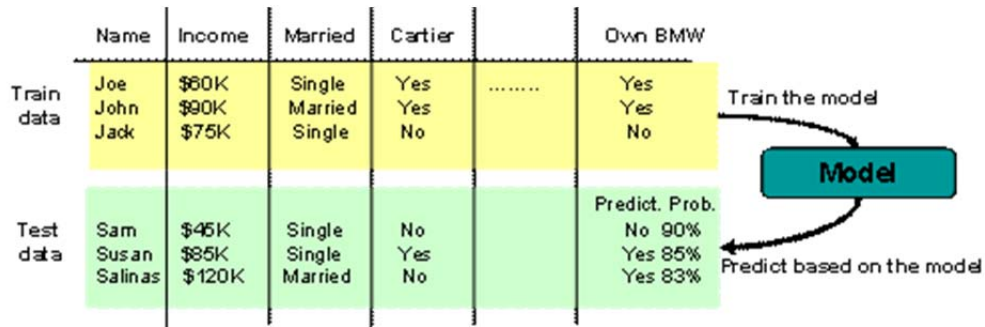


Figure 5 – Prediction Result

SUMMARY

Data Mining can be of tremendous benefit to organizations by enhancing their customer management capabilities and, thereby, improving their capabilities in a competitive marketplace. The phenomenal amount of data stored by today's companies prohibits manual analysis by even the most skilled personnel. The powerful algorithms that are employed by Data Mining solutions enable an organization to:

- Find hidden patterns in the data
- Better understand and predict customer behavior
- Better understand and predict the outcome of business decisions

The BTS Group, LLC is a technology services company that assists business owners and executives use technology and business process analysis to reach their strategic objectives. We implement the appropriate combination of consulting, managed and sourced services that enable successful completion of their initiatives and the quantifiable measurement of that success. For more information on how BTS can help you get more from your technology, contact us at info@thebtsgroup.com.
