Translating advances in data mining to business operations: The art of data mining in retailing

Henry G. L. Dillon
Beverley G. Hope
Victoria University of Wellington
Box 600, WELLINGTON, New Zealand
hgld@technologist.com Beverley.Hope@vuw.ac.nz

Abstract

Plummeting data storage costs mean that businesses can now hold more data than ever before. However, increased volume and detail of data necessitates effective and efficient analytical tools. Knowledge Discovery in Databases (KDD) is a field of research that studies the development and use of various data analysis tools and techniques. Data mining is one such tool. 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 IT managers say that data mining products are too difficult to use in a business context. This paper 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 – the money that bought 40 megabytes of storage in 1989 buys 8 gigabytes today [2]. The arrival of holographic storage will enable further reductions in the cost of storage capacity [3] and fuel the already popular data warehousing phenomenon. As the volume and detail of stored data increases, the demand for effective and efficient analysis tools also increases [4].

KDD has been rigorously researched, particularly in the area of data mining [5]. This has resulted in an array of models, theories, functions and methodologies for producing knowledge from data. However, despite these advances, nearly two thirds of IT managers say that data mining products are too difficult to use in a business context [1].

Aside from the scientific aspects of KDD, lies the artistic application of KDD to business. Limited academic attention has been paid to the business implementation of specific data mining techniques [6]. This paper focuses on the broad issue of how advances in data mining translate into business use. We first present a review of the relevant academic literature and then apply this knowledge to a business case study in a major petrol (gas) service station corporation

LITERATURE REVIEW

The purpose of this literature review is to provide an understanding of the concepts of data mining to be used in discussing the case study.

Data Warehousing

In 1990, William Inmon coined the term 'data warehousing' [7]. Inmon's concept differed from previous data storage concepts. Instead of focusing on the storage of raw production data from individual sources, data warehouses focused on data extracted from a variety of production databases [8]. The intent was to construct an architecture that improved data analysis and decision support tasks. Inmon identified four properties of a data warehouse:

> SUBJECT ORIENTED: *There is a shift from application-oriented data to decision-support data. If designed well, subject-oriented data 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. Data can be compared along a time axis unlike operational data which capture a moment in time.*
>
> NON-VOLATILE: *The database absorbs new data, integrating it with previous data, that is, new data are appended rather than substituted.* Inmon cited in [9]

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.

Knowledge Discovery in Databases

The concept of data warehousing originated with practitioners. Academics focused on KDD [7]. KDD is a way of better exploiting the potential of Data Warehouses through improved analysis. It has been defined as:

> *the non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data* [5].

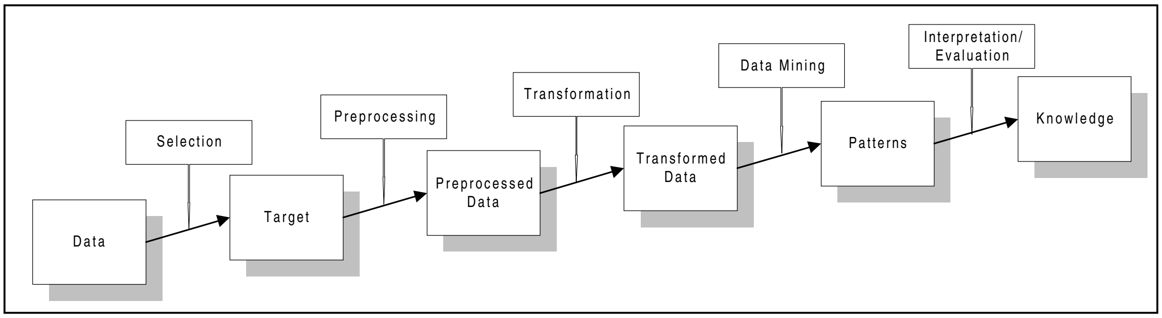

Fig. 1 KDD process overview [9, p.10]

A key point to draw from this definition is the concept of process. KDD is not an technique or method, rather it is an elaborate methodology, of which data mining is one part. A classic five step process for drawing knowledge from data is shown in Fig.1. Each step of the process is imperative but data mining has received the most attention.

Data Mining

Two key objectives of any data mining application are prediction and description. To achieve these analysts commonly employ six data mining methods: classification, regression, clustering, summarization, dependency modeling, and deviation detection [5]. These are discussed next.

CLASSIFICATION involves the mapping of data values into classes using a derived function. This provides a class representation of past trends enabling predictions based on classification dependencies. For example, banks could use classification to influence decisions involved in loan approval. By constructing a function that illustrates debtor integrity based on applicant characteristics, banks could use historic classification models to determine risk categories.

REGRESSION is similar to classification in that it uses a function to provide grounded predictions. However, regression enables measures of correlation between two variables. This can be translated to the business context through uses such as predicting future sales over a time series. For example, a bank might use regression to determine the strength of relationship between income and debt.

CLUSTERING is a descriptive tool used to group common data instances. The concept has classic applications in the marketing discipline, where customers with common attributes are grouped in the process of ‘segmentation’ [10]. For example, banks might use clustering to identify customers who are heavy users of particular services.

SUMMARIZATION uses an array of techniques to describe data. Simple techniques include tabulation of aggregates or averages. Advanced techniques stretch to summary rules, functional relationships, or associations derived from a data source [11]. For example, a bank may use summarization to determine the mean and standard deviation of overdraft account balances.

DEPENDENCY MODELING incorporates probability and uncertainty into a model to determine the significance of dependency between variables. It offers stronger grounds for predicting uncertain outcomes such as human behavior. For example, banks with loyalty programs could develop a dependency model to determine the probability of a customer visiting three retail outlets in the coming month.

DEVIATION DETECTION examines the comparative dynamics of patterns in data, usually over a time series. This task delivers both a predictive and descriptive benefit [12]. The description of past data trends can highlight seasonality or periodic market characteristics. It may then be possible to predict and prepare for changes. For example, a bank could identify trends in home loan applications. By identifying new application seasons, banks could schedule advertising to maximize their application opportunities.

Data mining has advanced in the last five years. The added complexity has allowed scientists to predict and describe reality in a richer form. How can these complex advances into be applied to the business context? We seek to answer this question through a case study in the petrol (gas) service station industry.

A NEW ZEALAND CASE EXAMPLE

To examine the application of data mining concepts to business, we examined a data warehouse implementation in the New Zealand division of a large international corporation.

The Organization Background

The company is a global energy company with retail operations in 125 countries. It has revenues of $65.9 billion and average investment spending of $4.9 billion a year over the last five years. Over 15,000 company-branded retail outlets operate internationally. In New Zealand the company owns 70 service stations and supplies hundreds of independent outlets. During 1997, a new entrant to the local industry triggered a price war on wet-stock product (petrol, diesel, and alternative fuels). Throughout this price war, unprofitable wet-stock margins were sustained, resulting in heavy dependence on alternative profit centers such as dry-stock merchandising (automotive supplies, groceries, and fast foods) to maintain profitability.

The Retail Stores Marketing department is responsible for decision making regarding dry-stock merchandising. The department can be divided into three managerial levels based on the nature of decision making: operational, tactical, and strategic. The relationships between levels is shown in Fig. 2. The lowest level of merchandising decision making, operational, is the responsibility of site managers. Site staff have the closest relationship with customers, consequently, search

they are given some autonomy in decision making to ensure the needs of localized customers are met.

Supporting the site mangers are staff in the Resale Stores Marketing group. Decisions made at this tactical level are less site specific. Here, nationwide tactics are designed with the intention of increasing profitability and volume for the organization as a whole.

Guiding the direction of the group's efforts, is the senior management. Decisions at this strategic level range from corporate visions to specific quantifiable goals.

Case Methodology

This research uses a single organization as a case study to demonstrate the application of data mining techniques to a business context. Data were collected through interviews, observations and document examination. Among these, interviews provided the main sources of data.

1. INTERVIEWS: Semi-structured interviews were conducted at three levels: operational, tactical and strategic. At each level, a selection of employees were interviewed with interviews tailored to each level.
2. OBSERVATION: During the project, the principal researcher had a specialist technical role with the firm which gave him an opportunity to observe and reflect on events in context. A journal of these observations was kept. This captured the unusual and unexpected aspects outside of formal interactions.
2. DOCUMENTATION: Material, such as newspaper articles, public documents, and industry literature, were used to supplement and validate data gained through interviews. Data were interpreted to generate a story. Through reflection, chunks of meaning were identified across data sources. These branded observations were then compared against academic literature. The data analysis methodology is consistent with current recommendations [13][14][15].

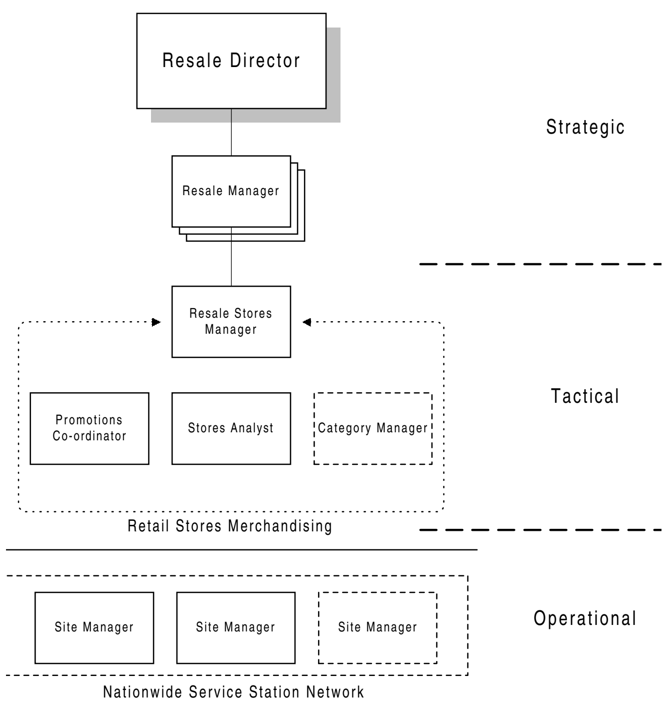

Fig. 2 Organizational structure

The Opportunity for Change

Prior to 1997, the company had few strategies for effective dry-stock merchandising. The primary business focus had been on wet-stock volumes. However, the 1997 wet stock price war forced management to focus on dry-stock sales as a compensatory profit source. To maximize dry-stock opportunities, they had to adopt better merchandising practices.

The then existing nationwide merchandising strategy, had several inherent weaknesses (Table 1). A key weaknesses was a bloated nationwide product range. Over time, sites had accumulated a large range of products and this weakened profitability. This damage to profitability was explained by:

1. Inappropriate or unprofitable products were stocked.
2. Market leader products or flavors were not stocked.
3. Guaranteed nationwide product ranging deals could not be negotiated.

Another weakness was the lack of sales information about individual sites caused by localized dry-stock decision-making. In addition, Head Office had difficulty in executing decisions affecting sites. Reasons for this difficulty varied, but common factors included: poor communication, decisions undermined by rival political opinions, and site negligence.

Management sought to overcome these problems by implementing centralized management of merchandising operations. The objective was to increase uniformity of sites creating opportunities to leverage deals with suppliers and to employ advanced merchandising practices.

The decision to centralize retail decision making required two key developments in merchandising strategy: improved Head Office knowledge about sites and customers and improved execution of Head Office decisions. These improvements were to be achieved through the initiation of two projects, the data warehouse project and the Category Management System (CMS) Project.

The Data Warehouse Project

Centralized decision making required centralized data. To satisfy the new need, a data warehouse project was initiated. An initial objective was to provide information to allow

TABLE 1
MERCHANDISING WEAKNESSES AND OPPORTUNITIES

WEAKNESSES	OPPORTUNITIES
▪ Site control of product ranges ▪ Limited company sales information ▪ Vast range of individual products stocked across the country	▪ Centralized Management of Merchandising operations ▪ Ability to leverage deals from suppliers ▪ Advanced merchandising practices, such as dynamic advertising, knowledge based-ranged

range rationalization; the longer term purpose was to facilitate tactical and strategic decision making.

The data warehouse would help meet the first goal by improving Head Office knowledge about sites and customers. However the second objective, improved execution of decisions by Head Office, was still unaddressed. To solve this, management decided to implement an new digital communication infrastructure to automatically implement Head Office decisions. This was called the CMS project.

The Category Management System Project

One of the roles of the Merchandising group at Head Office, is to leverage special deals and negotiate promotions for sites. However, the execution of, and compliance with, these plans at site was never absolute. The promotions coordinator in Head Office, noted that although the data warehouse would provide crucial, timely information to support decisions, it would be worthless if the decisions weren't executed at site. To overcome this problem, a Category Management System (CMS) was proposed.

The data warehouse and CMS were intended to be used together. Fig. 3 illustrates how the two projects fit in context. Sales information from the sites is electronically retrieved and loaded into the data warehouse at Head Office. Head office merchandising personnel then make decisions based on the information stored in the warehouse. CMS stores the changes made to each service station and sends commands electronically back to sites. This information is held on site for the site manager to review. Subsequent changes are automatically made to the site computer and point-of-sale units. Metaphorically speaking, the data warehouse acts as company's eyes and ears for observing reality, while CMS acts as the arms and legs for reacting to the observations.

During the writing of this case, both the data warehouse and CMS projects were stalled. A temporary data mart solution, called Thomas, was developed to meet the information requirements of the Head Office Merchandising staff. Thomas did not provide the data mining techniques planned for use in the data warehouse but it did provide substantial insight into sales information.

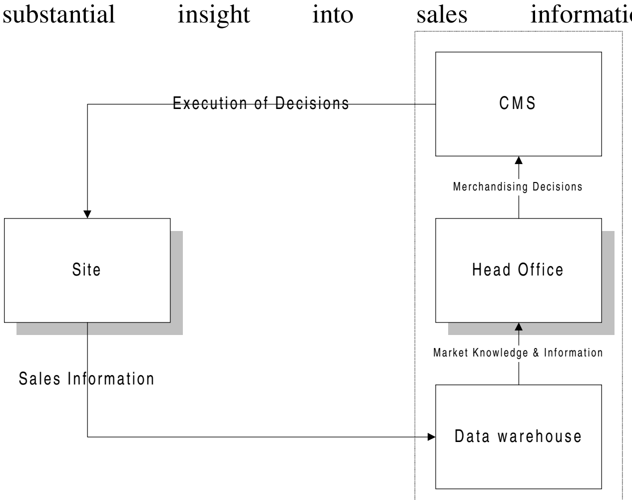

Fig. 3 The new merchandising opportunity

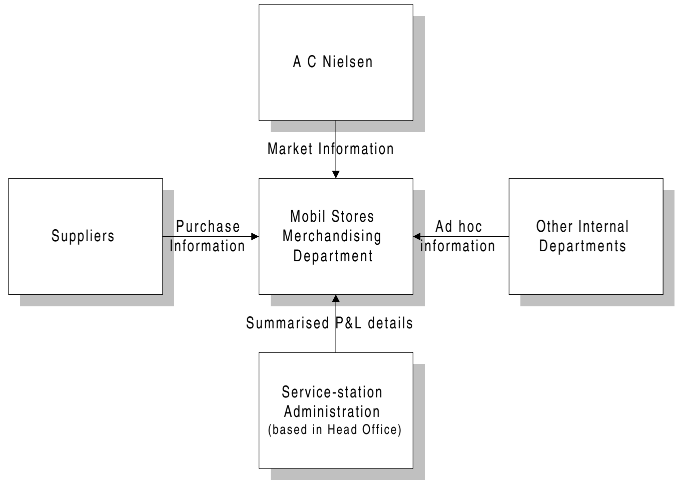

Fig. 4 Past Data Analysis Information Sources

RESULTS

Data from respondents, were analyzed with respect to the grand tour question, namely *'How do today's advances in data mining translate to business operations?'*. To answer this question we examined the past data analysis function within the Stores marketing department, and compared these to the new developments such as the proposed data warehouse project, and Thomas, the substituted data mart.

Past Data Analysis Infrastructure

Prior to the new initiatives, data analysis processes at Head Office were limited. Information came from a variety of sources including AC Nielsen, Suppliers, and internal departments (Fig. 4). However, data were not suitable for mining. AC Nielsen supply a detailed source of market performance information at product level. This information is aggregated into monthly figures for the entire retail network rather than on a site level basis. Supplier information is variable. It can be very rich, for example, identifying product trends and projections, or lean. Information from internal departments is also variable. Under the old regime, four factors hindered data analysis potential: access to data, level of detail, timeliness, and accuracy.

Head office had very limited access to data. Although the technology at sites was capable of capturing rich data, the communications infrastructure and storage capabilities at Head Office could not support increased data accessibility. The Stores Marketing Manager at Head Office explains.

> *'Retail business is all about having access to information and using it. Because we were decentralized, the stores themselves had access to a sophisticated system that provided quite a bit of information to make their own decisions. However centrally, … I had no access to detailed data, I only had access to department, or aggregated level data, through A C Nielsen. Consequently, until we really started to get access to the summary information*

through suppliers we weren't able to make rational decisions. And as a consequence we didn't make any – autonomy was seen to best reside with the site manager.'

The limited detail meant that only high level summarization and limited trend analysis could be used to examine field information. This information was used for periodic profit and loss measurement and understanding tactical positioning and performance. Timeliness of data was another problem identified by Head Office. Information was collected and presented on a monthly basis which was in many cases too late. The promotions coordinator at Head Office, pointed out that since a standard promotion duration is four weeks, no feedback could be gained regarding the success or failure of a promotion before the promotion was already completed. Ad hoc information collection from sites was possible, however this required considerable time and resources and consequently was requested in only special cases. Tactical staff were forced to make uninformed or gut-feeling decisions.

Finally, information accuracy and completeness under the old system was questionable. It is common KDD practice to expect data to be 'unclean' and consequently to undergo a scrubbing process. However, such processes for field data were absent from the previous information infrastructure.

The reader will recall that one of the fundamental objectives of this paper was to understand how the science of data mining analysis could be applied to the art of retailing. This was not an easy task. We first had to educate the users on the nature of each data mining technique. Once this was done we were able to summarize the interviews and identify common retail principles and examples that were compatible with particular data mining techniques. Findings are summarized in TABLE 2.

TABLE 2
SUMMARY OF DATA MINING APPLICATIONS

Data Mining Technique	**Application Example**	**Value**
Dependency Modeling	Provides grounds for the cross merchandising of products	Good
Deviation Detection	Discovering seasonal or unique sales trends.	Good
Clustering	Provides a model for rationalizing product ranging and pricing.	Good
Summarization	Supplies a more detailed understanding of the market reality	Average
Regression and Classification	Determining Price Elasticity and evaluating product cost/benefit	Low

Dependency Modeling

A number of respondents saw the retail concept of cross merchandising as a sound application of dependency modeling. Cross merchandising is the bundling or grouping of different individual products together at a special or exclusive price. The rationale for doing this is to maximize the total gross profit dollars per customer. The Stores Marketing Manager at Head Office explains.

The opportunity, or the times that [a customer] visits you with a particular product in mind won't produce volumes large enough to make merchandising profitable. Instead, [merchandisers] rely on the fact that humans are relatively impulsive. As a result we must be able to gauge how this impulsion can be controlled or manipulated. The fundamental objective is to increase the basket amount or gross profit dollars per customer.

Dependency modeling can give merchandising analysts valuable information regarding the probability of given product combinations. This provides a grounded understanding of potential impulse buying triggers, and allows merchandising analysts to prepare marketing tactics for firing these triggers.

Deviation Detection

Respondents identified a number of business applications which could use deviation detection techniques. These included; discovery of new trends, monitoring of current and past performance, and prediction of the future. Cyclical or seasonal surges in demand are largely unresearched at a micro-level, however these can be valuable for managers serving impulsive consumers in a convenience market. A Stores Analyst at Head Office recalls one such gem

In the week before Christmas, we will always see enormous lift in the number of battery sales. And do you know why? Its because people buy their kids presents but forget the batteries. So they rush out to one of our sites and buy some. Because we know and understand this trend, we can prepare for this sales lift and increase our returns.

The promotions coordinator also notes the significance of deviation detection in determining the effectiveness of promotions. Throughout a promotion period, all sites should experience some lift in sales. If this isn't the case then some investigation and explanation is required. For example, have they got the signage up? Is the promotion at the right price? Is the promotion receiving good placement relative to foot traffic? The promotions coordinator explains that many sites lose opportunities for higher profits due to failure to comply with promotion guidelines.

Change or deviation detection can supply micro-level analysis of future trends and past history. Thus that analysts can obtain more detailed knowledge about convenience market trends and as a result make grounded decisions on time based site activities. A site advertising example of time based activities is discussed later in the section titled 'Strategies Enabled by Data Mining'.

Clustering

Respondents identified a variety of hypothetical merchandising applications that were appropriate for the clustering technique. Previously, many of these were infeasible due to limited information on environmental dimensions such as customer demographics, psychographics,

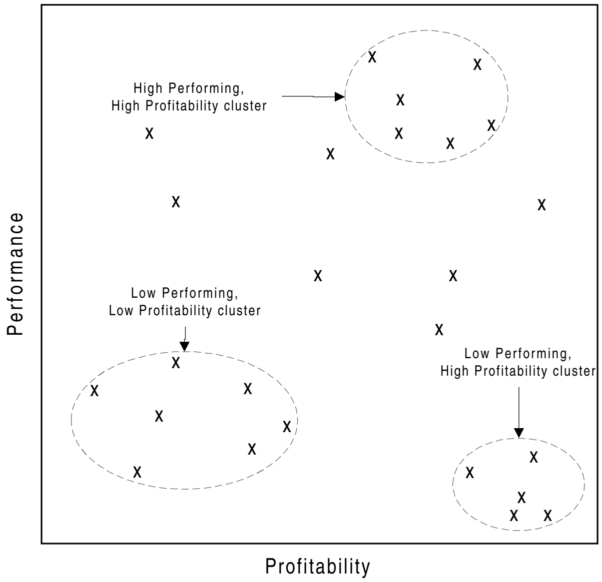

Fig. 5 Two Dimensional Clustering

or geographic location. A potential application of clustering identified by respondents was performance vs. profitability. Thomas already delivers a simple two dimensional cluster analysis (Fig. 5). But data mining enables more advanced analysis by enabling inclusion of additional factors. Such cluster analysis could provide rational grounds for product ranging and pricing. For example, suppose a particular group of sites were identified as high performers in the Asian foods category. Given this knowledge, management might decide to expand the Asian foods range within those sites to include products the basic range.

It was interesting to note that respondents commonly confused clustering with dependency modeling. This is understandable as they do have some interrelated applications. For example, clustering can provide good initial grounds for determining and describing product bundles appropriate for cross merchandising. Dependency modeling provides a more predictive element allowing analysts to determine the probability of a customer buying product X if they also bought Y.

Summarization

Basic summarization analysis was previously carried out as a means of understanding aggregated figures. Thomas applied concepts of aggregation, standard deviation and mean analysis of site figures.

More advanced summarization such as summary rule definition could provide a sound grounding for understanding the dynamics of merchandising reality. Suppose, for example, an analyst identified a number of variables that influenced the sales performance of a site (Fig. 6). Summary rules could provide analysts with information to better understand reality and use this understanding to justify or rationalize future decisions.

Summary rules have limitations, when applied to social behavior phenomena. Because of the enormous number of variables present in the 'real world', social behavior is difficult to quantify and model. This complexity impacts on the ability to formulate stores merchandising summary rules. Many independent variables impact sales performance, so analysts will nearly always be faced with issues of incomplete information. The example shown in Fig. 6 shows only a few of the variables that affect sales performance and fails to consider variables such as weather, customer psychographics, and culture. This complexity issue reduced the value attributed to advanced summarization analysis by respondents.

Regression and Classification

Of all the data mining techniques, regression and classification received the least respondent interest. Many respondents had difficulty understanding the concepts and, consequently, they had difficulty applying them. Price elasticity and classification of product cost and benefit were suggested by one respondent. The ability to calculate price elasticity would mean analysts could model customer responsiveness to price changes. Also the ability to classify product effectiveness would allow analysts to predict the outcome of new product ranging decisions. Despite these applications the availability of such information had limited appeal. Perhaps this limited appeal was due to the lack of understanding of the technique.

The majority of data mining techniques discussed have shown application value to merchandising operation, particularly Dependency Modeling, Deviation Detection, and Clustering. Most examples given to usr elate to tactical activities within the organization, and did not address the strategic relevance of data mining. The following section addresses this issue.

STRATEGIES ENABLED BY DATA MINING

Data mining enables innovative strategies to be justified by providing analysis based on large data stores. Management in our case study is still under-utilizing the data sources at their disposal. To date, no new strategies have been created through the information provided by Thomas. Despite this, many respondents recognized potential strategies enabled by information provided through data mining.

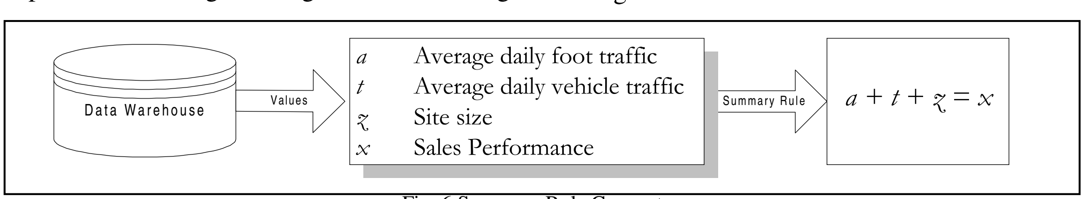

Fig. 6 Summary Rule Concept

One significant strategy is the discovery of cross merchandisable products using dependency analysis. For example, information could be used to justify a strategy to install service stations without fuel. To date, no New Zealand studies have been conducted to determine the dependency of dry-stock sales on wet-stock availability. There are many costs and risks involved in the resale of fuels. Therefore, if data showed that showroom sales were not dependent of fuel sales, management might consider constructing sites that didn't sell fuel.

Another strategy suggested by respondents was dynamic on-site advertising. Throughout the day, sites serve a range of potential customers with differing characteristics. By changing site advertising to suit the current audience, sites could improve the effectiveness of their advertising. For example, suppose we could establish macro sales volumes trends for two different products over a daily time frame. Such a trend could be represented as in Fig. 7. Notice that Product X has significantly higher volume sales over the morning time frame, whereas Product Y has considerably higher volume sales after midday. Using this information, site staff could change signage to match the expected daily volume fluctuations and hopefully trigger, persuade and motivate more customers to purchase the advertised product.

Data mining will inevitably enable new strategies. The previous paragraphs have outlined some of the possibilities. However, there are many potential applications for any one technique. The only limitation is the analysts' imagination and creativity.

DISCUSSION

A number of implications have emerged from this study. These include findings affecting practitioners, and insight into the implications for further research in KDD.

Implications for Practitioners

The application of data mining in our study seemed to have most relevance at the tactical decision making level. Operational level decision makers had limited use for data mined information because of the decentralized and territorial nature of their responsibilities. Because of their frontline involvement, site managers already had good knowledge and understanding. In addition, individual sites are smaller scale operations with lower data volumes unsuitable for data mining tools. Data mining may have the potential to support strategic decision making. but it was not evidenced in this study. Our findings suggest, that practitioners should initially look at focusing data mining applications at the tactical levels and subsequently at the strategic levels of their business.

A critical issue facing practitioners, is the complexity of the KDD science. Respondents in our study had difficulty understanding the concepts and techniques associated with data mining. In several cases, respondents' understanding was generic or over-simplified, that is, they understood data mining to be a 'black box' solution that returned valuable information without any requirement for direction or instruction. This observation highlights the disparity of understanding between the art and science of data mining. To resolve this gap more intensive training of analysis skills needs to be executed to improve the knowledge or understanding of the qualitative results delivered by data mining. This is particularly important for tactical and strategic decision makers where data mining deliverables seem to be most relevant.

Environmental information such as customer demographics and geographic characteristics need to be fed into the sales data warehouse. A transactional data store does not supply the full richness of data. More environmental information supplies richer knowledge nuggets. Customer loyalty programs are a means by which businesses can gain supplementary information to add richness to customer profiles.

Implications for research

Advances in KDD research have resulted in improved methods of using data warehouses to describe and predict phenomena of interest. The majority of KDD literature focuses on the scientific perspective of KDD, that is, the development of new processes and techniques for discovering knowledge. While the scientific studies have undoubtedly added value to the field of KDD, little academic work has been done on the business application or representation of KDD to commercial reality. Areas for future study include ;

1. Modes of presentation,
2. Analyst education, and
3. Managing organizational culture change .

Better modes of presentation are required to display data mining findings. Many data mining techniques lead to.

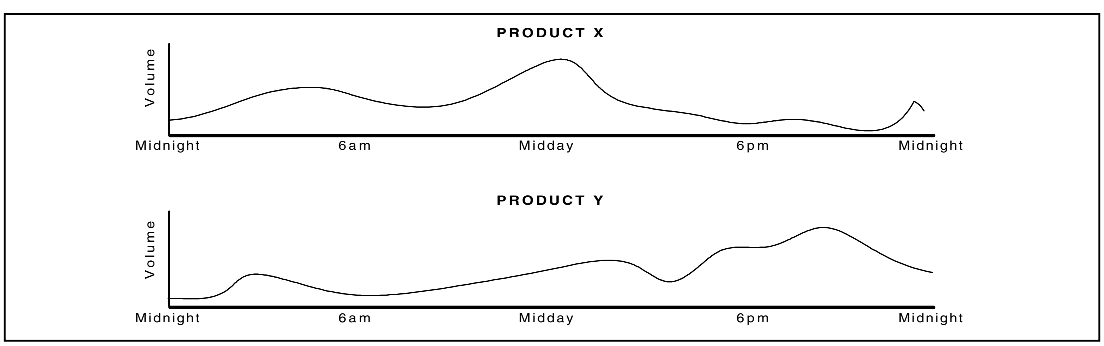

Fig. 7. Product performance by daily time series

search

abstract presentations of results [5]. These are difficult for practitioners to understand, interpret and act upon. Research is needed to develop presentation aids and create new ways of creatively illustrating KDD findings

In addition to the need for better presentation of results, there is a need to improve the analysts' understanding and knowledge of the KDD process. Data mining in particular, requires a special set of skills to allow analysts to exploit the potential of large data stores. Without at least a base understanding of the process and techniques, the benefits of conducting a data mining operation are severely reduced. Research needs to identify the data mining skills relevant to business and offer some suggestion as to how organizations should approach this training issue.

New skills need to be applied within a commercial culture. To facilitate the change, from a traditional mindset to data mining mindset requires a supportive culture. Research is needed into factors that foster or inhibit cultural acceptance of KDD within organizations. This transition is particularly relevant with the dawn of the new information age, where the role of analysts dealing with information, is extended to the new role of knowledge workers dealing with knowledge.

Summary

In this study we sought to determine how advances in data mining can be applied to modern business operations. To do this we conducted a case study. Our analysis showed that not all data mining techniques are applicable at all business operation levels. There is particular value in applying KDD at tactical and strategic levels. The application of each technique will also likely differ across businesses and industries. The inherent complexities of KDD require businesses to train users so that an understanding of KDD can be grasped. Once users are familiar with data mining techniques, finding business applications should not be difficult.

The findings of this research clearly indicate the need for further work in this growing field. Issues regarding tactical worker training and evolution of an organizational mindset to support KDD need to be addressed to ensure data mining benefits are maximized. The next logical step from this research would be an examination of the new skills and understanding required by knowledge workers of the future.

REFERENCES

[1] J. Foley and J. Russell, "Mining Your Own Business," *InformationWeek,* March 1998. Accessed from: 16http://www.techweb.com/se/directlink.cgi?IWK19980316S0022

[2] M. Garvey, J. Governor and M. Hayes, "Serious About Storage," *InformationWeek,* August 10, 1998. Accessed from: http://www.informationweek.com/695/95iusto.htm

[3] M. Halper, "Holographic Storage, A terabyte inside a crystal," *Wired*., vol. 5, no. 11, pp. 108, 1997.

[4] R.J. Brachman, T. Khabaza, W. Kloesgen, G. Piatetsky-Shapiro, and E. Simoudis, "EMining business databases," *Communications of the ACM*, vol. 39, no 11, pp.42-48, 1996.

[5] U.M. Fayyad, G. Piatetsky-Sharpiro, P. Smyth, and R. Uthurusamy, *Advances in knowledge discovery and data mining.* Massachusetts: MIT Press, 1996.

[6] P. Fabris, "Advanced Navigation," *CIO,* May 15, 1998. accessed from http:\\www.cio.com/archive/051598_mining.html

[7] T. Sakaguchi, and M.A. Frolick, "Review of the Data Warehousing Literature," *Journal of Data Warehousing*, vol. 2, no. 1, pp. 34-54, 1997.

[8] W.H. Inmon, "The data warehouse and data mining," *Communications of the ACM,* vol. 39, no. 11, pp. 49-50, 1996.

[9] R. Hackathorn, "Data warehouse energizes your enterprise," *Datamation*, vol. 41, no. 2, pp. 38-45, 1995.

[10] J. Guiltinan, G. Paul, and T. Madden, *Marketing Management: Strategies and Programs.* New York: McGraw-Hill, 1997.

[11] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, and A. Verkamo, "Fast Discovery of Association Rules," in *Advances in knowledge discovery and data mining*, U. M. Fayyad, G. Piatetsky-Sharpiro, P. Smyth, R. Uthurusamy, Eds. Massachusetts: MIT Press, 1997, pp. 307-328.

[12] D.J. Berndt, and J. Clifford, "Finding patterns in time series: A dynamic programming approach," in *Advances in knowledge discovery and data mining,* U. M. Fayyad, G. Piatetsky-Sharpiro, P. Smyth, R. Uthurusamy, Eds. Massachusetts: MIT Press, 1997, pp. 229-248.

[13] R.K. Yin, *Case Study Research: Design and methods.* Newbury Park, CA: Sage, 1990.

[14] G. McCracken, *The long interview.* Beverley Hills: Sage, 1988.

[15] J.W. Cresswell, *Research Design: Qualitative and quantitative approaches.* Thousand Oaks, CA: Sage, 1994.