DATA MINING AS A NEW PARADIGM FOR BUSINESS INTELLIGENCE IN DATABASE MARKETING PROJECTS

Filipe Pinto, Pedro Gago

School of Technology and Management, Polytechnic Institute of Leiria, Portugal

Manuel Filipe Santos

Department of Information Systems, University of Minho, Guimarães, Portugal

Keywords: Database Marketing, Knowledge Discovery from Databases, Data Mining, Business Intelligence.

Abstract: Information technologies provide not only the ability to collect and register in databases many kinds of signals external to the organization, but also the capacity to use them in different ways at different organizational levels. Database Marketing (DBM) refers to the use of database technology to support marketing activities in order to establish and maintain a profitable interaction with clients. Currently DBM is usually approached using classical statistical inference, which may fail when complex, multi-dimensional, and incomplete data is available. An alternative is to apply Data Mining (DM) techniques in a process called Knowledge Discovery from Databases, which aims at automatic pattern extraction. This will help marketers to address customer needs based on what they know about them, rather than a mass generalization of their characteristics. This paper exploits a systematic approach for the use of DM techniques as a new paradigm in Business Intelligence in DBM projects, considering analytical and marketing aspects. A cross-table is proposed to associate DBM activities to the appropriate DM techniques. This framework guides the development of DBM projects, contributing to improve their efficacy and efficiency.

1 INTRODUCTION

Due to the advances in information and communication technologies corporations can obtain and store transactional and demographic data on individual customers at reasonable costs (Naik, 2003). The challenge is how to extract important knowledge from these vast databases in order to gain competitive advantage (Cohen, 2004). However, database (DB) usage in many organizations remains complex and sometimes unavailable, not only because database management systems requests relevant background knowledge, but also because data are not ready to be used outside DB management systems purposes.

Nowadays, organizations are increasingly realizing the importance of understanding and leveraging customer level data, and critical Business Intelligence (BI) decision models are being built upon analyzing such data. Emphasis on customer relationship management makes the marketing function an ideal application area to greatly benefit from the use of Data Mining (DM) tools for decision support in a BI context. Through Database Marketing (DBM) organizations can identify valuable customers, predict future behaviours, and make proactive, knowledge-driven decisions by means of a statistical calculus or development of sample queries to marketing DBs. However, that approach is not structured and there is a need for a unified view guiding marketing practitioners in their quest for relevant knowledge. This includes understanding the customers' preferences and customers' behaviour through analyzing their data. There has been much research done in this direction, and DM techniques have been used with success in several areas, such as fraud detection (Wheeler, 2004), bankruptcy prediction (Cielen, 2004), intensive care medicine (Silva, 2004) and engineering (Santos, 2003), just to name a few.

Indeed, the old model of "design-build-sell" (a product-oriented view), is being replaced by "sell-build-redesign" (a customer-oriented view)

(Drozdenko, 2002). The traditional process of mass marketing is being challenged by the new approach of one-to-one marketing. As a marketing strategy definition support, DBM activity has changed significantly over the last several years. The current approach relies on predictive response models to target customers for offers. These models accurately estimate the probability that a customer will respond to a specific offer and can significantly increase the response rate to a product offering. Their use for marketing decision support highlights unique and interesting issues such as customer relationship management, real-time interactive marketing, customer profiling and cross-organizational management of knowledge (Shaw, 2001).

Normally DBM projects face several types of constraints:

- *Organizational culture*: Scattered around the organization it is frequent to find DBs with redundant information or noise in their registers;
- *Data quality*: Data quality depends of the operational usage of the data. They are considered ready to be used if they are free of defects or cleaned from errors;
- *Data access*: data usage requests a practical data access facilitating the algorithms use.
- *Data quantity*: having lots of data may hamper the data analysis work. Huge DBs to not guarantee that the available data has the information needed for any particular objective;
- *Technical limitation*: Know-how to handle with DBs in order to extract unknown information that is hidden in data.

Some contributions to overcome these constraints have been published, addressing data pre-processing (Pinto, 2004), data quality aspects (Oliveira, 2004) or others. Nevertheless there are some important aspects that still remain without answers, like those that refer to data integration, pre-processing, usage and exploration in marketing activities.

Nevertheless, the majority of contributions to the DBM field refer both:

- To a simple methods usage in specific cases e.g., market basket analysis (Chen, 2005), cross-selling and up-selling activities (Cohen, 2004), or customer relationship management (Shaw, 2001); or
- To a particular set of techniques to improve specific results e.g., segmentation, one-to-one marketing activities or clustering analysis (Drozdenko, 2002), (Russell, 1999), (Shepard, 1998).

In order to help marketers make use of the knowledge obtained through the KDD approach in their marketing activities and improve their results, we propose a framework for the efficient systematization and integration of the involved processes.

This paper is organized as follows. First, we present a brief description about DBM and relevant issues regarding the KDD process (section 2). We then continue with marketing activities and data mining objectives, closing this section with a cross-table that integrates marketing activities and DM techniques (section 3). On section 4 a proposal for DBM systematization with a KDD approach is presented. A case study is presented in order to illustrate the framework use. Finally we discuss the framework proposed identify some of the emerging issues to be addressed in the process of managing the discovered marketing knowledge.

2 DATABASE MARKETING AND KDD

2.1 Database Marketing

In this article, DBM is referred as the use of database technology for supporting marketing activities, while marketing DB it is referred to the database system it self. Coopers & Lybrand (1996), proposed three different levels of DBM in order to better organize these concepts:

- *Direct Marketing* – Organizations manage lists and conduct basic promotion performance analyses;
- *Customer Relationship Marketing* – Companies apply a more sophisticated, tailored approach and technological tools to manage their relationship with customers;
- *Customer-centric Relationship Management* – Customer information drives business decisions for the entire enterprise, thus allowing the retailer to dialogue directly with individual customers and ensure by this way, loyal relationship.

DBM has been defined has the establishment of a customers and prospects DB with which it is possible to the organization to communicate with them in a personalized way (Wolf, 1999). Others consider DBM as a medium to use consumer information with the objective of incrementing marketing activities efficacy and efficiency (Roberts, 1997). Finally it is possible to define DBM

as the usage of customer information with benefits both them and to the organization (Berson, 2001).

These definitions emphasize DB technologies as a support to the marketing activities, and establish as DBM definition, a set of processes based in marketing DBs exploring and analysing them looking for new insights (Pinto, 2004).

2.2 Data Mining Objectives

Data Mining, also popularly known as Knowledge Discovery in Databases (KDD), refers to the nontrivial extraction of implicit, previously unknown and potentially useful information from data in databases (Fayyad, 1996). While DM and KDD are frequently treated as synonyms, DM is actually part of the knowledge discovery process (Zaïane, 1999).

In short, DM aims at building models from data. There are many available algorithms; each with specific characteristics. The major DM activities are (Povel, 2001):

- Predictive modelling: mapping a set of "input" values (independent variables) to an "output" value (dependent variables). This kind of models takes two forms depending on the type of the output, as follows:
  - Classification: learning a function that associates with each data object one of a finite number of pre-defined classes (e.g., customer profile)
  - Regression: learning a function that maps each data object to a continuous value (e.g., amount spent)
- Descriptive modelling: discovering groups or categories of data objects that share similarities and help in describing the data space (e.g., customer segments);
- Dependency Modelling: learning a model that describes significant associations or dependencies among features (e.g., contents of subscription orders, market baskets);
- Change and deviation detection/modelling – Detecting the most significant deviations from previous measurements/behaviour or norms (e.g., fraud detection);

DM activities selection is directly dependent on the marketing objectives initially defined

2.3 Supporting Marketing Activities with DM Models

Marketing activities refer to the exchange of products and services and are oriented by the major marketing objectives. There are five important questions to which marketing activities must be able to respond (Suther, 1999):

- *Who* should I target?
- *What* should I target them with?
- *When* should I do it?
- *Which* promotion channel should I use?
- *How* should the promotion be done?

DBM is a process oriented to the marketing objectives (Pinto, 2004), which will determine all information collection process. From here, adopting the above model it is possible to suggest at least a DM task for each one of the objectives.

Effectively finding **"***Who***"** means using DM techniques to segment likely responders, repeat users, non-promiscuous acquisition targets, customers with profit upside, likely defectors and those customers to will refer business your way.

The *"What"* question suggests finding the key characteristics of the highest value company customers. This goal may be achieved by analyzing data about products and consumer behaviour

Associated with the "how" question there are a set of prediction activities, e.g., predicting the product sales for a specific period, or how many customers may leave the company.

The time variable in marketing activities is represented here by the *"When"* question. This includes all marketing activities that concerns temporal tasks, like when the company should send promotional e-mail to their clients.

The "which" objective it is one of the most used keyword in marketing activities definition, due the selection characteristics associated, e.g., in market basket analysis, the marketer wants to known which products are associated;

Due their nature all marketing questions include some prediction in their results, hence here the fact that it is possible to assign to each one of them a DM prediction activity. Descriptive DM models are likely best to respond to the *"who"* and *"which"* questions not only by their classification characteristics but also by the kind of desired results. Dependency analysis models have a vast application in marketing activities as it is possible to include them in *"when"*, *"who"* and *"which"* objectives. Finally deviation analysis modelling may be used to answer *"how"*, *"when"* and *"who"* marketing questions

Table 1 presents the combination of marketing activities, here represented by their questions, with DM activities.

Table 1: Data Mining Activities applied to marketing questions.

Marketing	Data Mining Activities			
Questions	Predictive	Descriptive	Dependency	Deviation
Who	☑	☑		
What	☑			
How	☑			☑
When	☑		☑	☑
Which	☑	☑	☑	

3 FRAMEWORK FOR KDD BASED DBM

To address these important questions, *how, when, who, what* and *which*, DM can be useful. On the other hand, DM is not enough by itself requiring a set of related activities to ensure the results quality. Therefore an approach to the development of DBM systems should adhere to an ordered set of steps and requirements. A framework is proposed to explore the concepts and characteristics of the KDD process and cross them with the marketing activities and the questions inherent to the integration of the Data Mining models.

Figure 1 represents the framework for DBM based on KDD. The system has three main phases: Information collection, Knowledge discovery and Evaluation and deployment.

First, data is gathered from different sources. After their registration and analysis a marketing DB is created, in order to support all knowledge discovery process.

As DBM is characterized by marketing strategies based on the great volume of information available in large customer DBs, it is possible to point out the following areas as major candidates for the application of Knowledge Discovery in Databases for knowledge based marketing (Povel, 2001):

- *Customer Acquisition*;
- *Cross- and Up-selling;*
- *Product Development;*
- *Churn Prediction*;
- *Fraud Detection*;
- *Market-basket Analysis*;
- *Risk Assessment*;
- *Prediction/Forecasting*;

The KDD process integrates the second phase proposed and comprises a few steps leading from marketing DB data to some form of new knowledge.
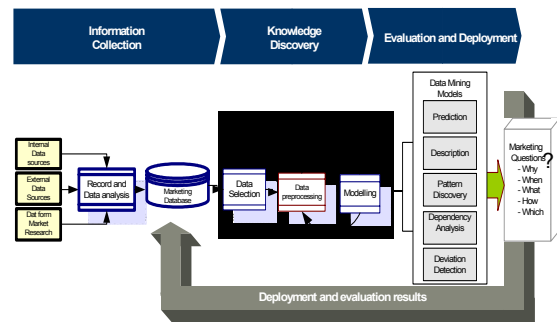

Figure 1: Framework for DBM.

- *Data Selection*. Consists on the selection of the subset of data to mine. This is not the same as sampling the DB or choosing predictor variables, rather, it is a gross elimination of irrelevant or unneeded data.;
- *Data pre-processing* and transformation, is a phase in which the selected data is transformed into forms appropriate for the mining procedure;
- *Modelling:* it is the crucial step in which some techniques are applied to uncover potentially useful patterns. Here, different techniques are used in order to achieve the objectives initially established e.g., classification or segmentation.

The final phase, Evaluation and Deployment refers to the integration of the knowledge obtained from the KDD process in the marketing models. Hence, the answer to marketing questions is sustained by these models.

Table 2 synthesizes the proposed framework by associating to a marketing activity objective a set of DM tools.

Since *"who", "what", "when", "how"* and *"which"* change with customer life events and competitors market activity, the analyses suggested must be continuously refreshed strengthening the case for a systemized view to DBM.

3.1 Sample Application

This framework was used in a DBM project carried out by a Portuguese marketing enterprise. The referred company distributes an own-branded magazine which includes discount vouchers to promote products of a great multinational distribution organization (food and beauty products). The main goal of the project was to determine for each product the customer profile. The discovered association rules can be used as filters on the DBs in order to identify prospects for cross-selling.

147

Table 2: standard examples of marketing concepts.

Question	Example	DM Tools
How	Fraud Detection	Instance-based learning
		Unsupervised learning
		Neural Networks
	Churn Prediction	Decision Trees
What	Product development	Neural Networks
		Rule Induction
Who	Customer segmentation	Instance-based learning
		Neural Networks
		Unsupervised learning
	Cross and Up-selling	Decision Trees
When	Outcomes measurement	Neural Networks
		Rule Induction
	Risk assessment	Association learning
	Deviation analysis	Neural networks
		Instance-based learning
Which	Customer acquisition	Decision Trees
		Neural Networks
		Rule Induction
	MarketBasket Analysis	Association learning
	Customer Profile Analysis	Instance-based learning
		Neural Networks
		Unsupervised learning

The project started with the collection of data from diverse data sources, from company owned data regarding previous promotions to acquired external DBs containing extra information. Then, a marketing DB was created after a careful analysis and documentation of these DBs content. Next, a monthly promotional magazine containing several discount vouchers and a questionnaire was sent to each of the prospects on the DB. New issues of the magazine (containing new discount vouchers) continued to be sent only to those that answered the questionnaire. Approximately two hundred vouchers and eight questionnaires were sent. Finally, by registering the available information a DB was created with data from about 630,000 individuals and a total of over eleven million commercial transactions. The analysis was performed on a DB sample containing roughly 10% of the records and selected according to geographic distribution, sex and age significance.

The main DM objective was to find the associated costumer profile for each product (answering the "which" question). By using self organizing maps (Kohonen, 1995) and the C.5 decision tree algorithm (Quinnlan, 2004) we were able to find decision rules that guided the selection of costumers for new marketing activities. Results evaluation lead not only to new similar studies but also to the use of both the DB and the framework procedures in new campaigns (Pinto, 2004).

4 DISCUSSION

KDD is an evolving field that presents interesting challenges for researchers and practitioners, with implications for the DBM function. Even after having presented an integrated framework for KDD in the context of DBM, we realize that there are still a number of research questions to be answered. Some of them are related to the DM techniques and the KDD process, while others are related to the management of knowledge in marketing activities.

Knowledge discovery through DM is an iterative process starting with the data collection process, which includes all activities concerned with data collection and selection and leads to the creation of a Marketing Database. We consider that the definition of the marketing activities objectives should occur before the data collection process starts. Thus, it is possible to orient the data pre-processing and transformation phases to those objectives. The importance of this ordering of steps becomes evident when larger DBs are used.

Next, in the knowledge extraction phase, the selection of data mining algorithms, hypotheses formation, model evaluation and refinement are key components. One of the research challenges is to make this process more structured, easier to use by marketers and thus improve the productivity of the DM efforts. To this purpose we defined a cross-table which illustrates the relation between marketing activities (organized by the pre-defined main questions); DM tasks and respective DM tools available to support their development.

A second challenge is how to use the knowledge extracted from DBs as it is often represented by means that are not easily understood by marketer.

The main difficulty refers to multiple classifications, in cases when marketing activities can belong to more than one DM activity.

Past experiences in DM projects dictated a need for a clear framework to enable better results. The framework proposed tries to achieve that by supplying marketers with a "roadmap" that will consistently guide them through their projects whenever DM techniques are to be used.

5 CONCLUSIONS

In this work, a KDD approach was used to DBM projects, regarding the systematization of overall process in order to simplify her use in marketing activities support. In the current customer centric business environment, it is our firm belief that there is a need for deeper understanding of use of data mining and knowledge management for marketing decision support.

Towards that end, we have shown how data mining can be integrated into a marketing knowledge management framework. With the availability of large volume of data, made possible by modern information technology, a major problem is to filter, sort, process, analyze and manage this data in order to extract the information relevant to the user. The growth in the size and number of existing DBs far exceeds human abilities to analyze such data using traditional tools and thus creates both a need and an opportunity for data mining tools. With the shift from mass marketing to one-to-one relationship marketing, one area that could greatly benefit from data mining is the marketing function itself.

A systematic application of data mining techniques enhances the knowledge management process and arms the marketers with better knowledge of their customers leading to better service to customers. To us, it is also clear that the Web technology will have a major impact on the practice of data mining and knowledge management and that should present interesting challenges for future information systems research.

REFERENCES

Naik, A. P., Tsai, C., 2003. Isotonic single-index model for high-dimensional database marketing, Computational Statistics and Data Analysis.

Cohen, M. D., 2004. Exploiting Response Models – Optimizing cross-sell and up-sell opportunities in banking, Information and Systems, 29, 327-341

Wheeler, R., Aitken, S., 2004. Multiple algorithms for fraud detection, Knowledge-Based Systems, Volume 13, Issues 2-3, 93-99

Cielen, A., Peeters, L., Vanhoof, K., 2004. Bankruptcy prediction using a data envelopment analysis, European Journal of Operational Research, Volume 154, Issue 2, 526-532

Silva, A., Cortez, P., Santos, M., Gomes, L., and Neves, J., 2004. Multiple Organ Failure Diagnosis Using Adverse Events and Neural Networks, In I. Seruca et al. Eds., Proceedings of 6th International Conference on Enterprise Information Systems - ICEIS 2004, Vol. 2, 401-408

Santos, M. F., Quintela, H., Cruz, P., 2003. Forecasting of the ultimate resistance of steel beams subjected to concentrated loads using data mining techniques, Data Mining IV, 533-541

Shaw, M. J., Subramaniam, C., Tan, G., Welge, M., 2001. Knowledge Management and Data Mining for Marketing, Decision Support Systems, 31, 127-137

Quinnlan, J., 2004. C5.0 Data Mining Tool, http://www.rulequest.com.

Kohonen, T., 1995. Self Organizing Maps, Springer-Verlag,

Suther T. 1999 "Customer Relationship Management: Why Data Warehouse Planners Should Care About Speed and Intelligence in Marketing", DM Review

Zaïane Osmar R.; 1999. "Principles of Knowledge Discovery in Databases" University of Alberta; USA;

Coopers & Lybrand Consulting; 1996 "Database Marketing Standards for the Retail Industry"; Retail Target Marketing System Inc

Drozdenko, Ronald; Drak Perry D., 2002. Optimal Database Marketing, SAGE Publications, Thousand Oaks, USA

Berson A; Smith, S.; 2001. "Data Warehousing, Data Mining & OLAP" McGraw Hill International Edition.

Roberts, M Lou.1997."Expanding the Role of the Direct Marketing Database." Journal of Direct Marketing 11

Wolf , M. J; Copulsky, J. 1999."Relationship Marketing: Positioning for the Future," The Journal of Business Strategy, July/August, pp. 16–20

Pinto, F., Santos, M.F., Cortez, P., Quintela, H., 2004. "Data Preprocessing for Database Marketing", Data Gadgets 2004, Málaga Spain, pp 76-84

Oliveira, P.; Rodrigues Fátima; 2004. "Limpeza de dados – uma visão geral", Data Gadgets 2004, Málaga.

Chen, Yen-Liang; Tang, Hu, Ya-Han 2005. "*Market basket analysis in a multiple store environment*" Decision Support Systems v40 p339–354

Russell, S.; Lodwick, W 1999 "*Fuzzy clustering in data mining for telco database marketing campaigns*"; Fuzzy Information Processing Society, 1999. NAFIPS. 18th International Conference of the North American p720 – 726

Shepard, David 1998. "The New Direct Marketing: How to Implement A Profit-Driven Database Marketing Strategy"; ed David Shepard Ass; McGraw-Hill, 3ª ed

Povel, O.; Giraud-Carrier C.;2001."Characterizing Data Mining Software"; Intelligent Data Analysis; IOS Press; v.5 p.1-12

Fayyad, U, G. Piatetsky-Shapiro, P. Smyth, & R. Uthurusamy,1996. "Advances in Knowledge Discovery & Data Mining". Cambridge, MA (The AAAI Press/The MIT Press)