

Measuring the impact of data mining on churn management

Miguel A.P.M. Lejeune

The author

Miguel A.P.M. Lejeune is a PhD Student in the School of Management at Rutgers University, Newark, New Jersey, USA.

Keywords

Data collection, Electronic commerce, Customer care, Sensitivity analysis, Customer loyalty

Abstract

Churn management is a fundamental concern for businesses and the emergence of the digital economy has made the problem even more acute. Companies' initiatives to handle churn and customers' profitability issues have been directed to more customer-oriented strategies. In this paper, we present a customer relationship management framework based on the integration of the electronic channel. This framework is constituted of four tools that should provide an appropriate collection, treatment and analysis of data. From this perspective, we pay special attention to some of the latest data mining developments which, we believe, are destined to play a central role in churn management. Relying on sensitivity analysis, we propose an analysis framework able to prefigure the possible impact induced by the ongoing data mining enhancements on churn management and on the decision-making process.

Electronic access

The research register for this journal is available at http://www.mcbsp.com/research_registers

The current issue and full text archive of this journal is available at <http://www.emerald-library.com/ft>

Internet Research: Electronic Networking Applications and Policy
Volume 11 · Number 5 · 2001 · pp. 375–387
© MCB University Press · ISSN 1066-2243

1. Introduction

Retention and acquisition of new clients are matters of great concern for businesses. While incoming companies focus on acquiring new customers, established, mature ones strive towards keeping existing or loyal customers. Over the long haul, these customers are more likely to be more profitable for the company. Companies have greater opportunities for cross-selling with loyal clients.

The rapid extension of the digital economy has implied different modifications in the acquisition-retention customer process. Electronic commerce has induced an exponential increase in the amount of available information. The rise in available information enlarges the customers' awareness of the different marketed solutions or products. Customers tend to be more demanding and more price- or characteristics-sensitive and are able to take more relevant decisions. The information availability impacts companies' business models. Companies have new means to customize the goods and services they offer. They can opt for strategies based on one-to-one marketing or mass customization. We also observe that companies deciding to compete through the electronic channel can vary the amount of information disclosed on the different channels, thereby creating an additional differentiation tool (Zettelmeyer, 2000).

In this paper, we propose a customer relationship framework (CRM) to help deal with churn issues. This model integrates the electronic channel and involves four tools for enhancing data collection, data treatment, data analysis and data integration in the decision-making process.

We explain the adequacy of data mining for alleviating churn issues. More especially, we underline the possibility of detecting the set of the most relevant features for the different clusters of customers. We examine the ongoing developments in data mining methods (machine-learning algorithms, logical analysis of data, ...) and address the consequences resulting from the increase in efficiency due to new data mining methods. We purport to come with a causal assessment scheme that helps understand the impact of upgrades in the data mining on customer relationship management.

We use sensitivity analysis which provides principles for assessing how the optimum solution can be affected by changes in data. The goal is to provide an analysis framework for ascertaining how customer relationship management may evolve depending upon improvements in data mining methods. We focus upon providing such a global analysis framework that is applicable to very diverse situations and motives (data mining for segmentation, for association, for prediction, ...).

2. Structure and related work

In this section, we describe the structure of the paper and the goals we pursued. We also proceed to the review of the literature we refer to.

In section 3, we define the churn concept and some other notions related to customer-oriented management by referring to Brookman (1998), Komenar (1997), Strouse (1999), Rowley and Dawes (2000) and Peppard (2000). We state the impact of electronic commerce on the customer relationship management. We also cite different figures for giving prominence to the churn amplitude, be it in the USA or in Europe. These estimations are provided by Prahbaker (2000), Strouse (1999), Groth (1999), Hoffman and Novak (2000) and the SAS Institute (2000).

In section 4, we introduce the concept of customer relationship management (CRM). We argue the need and the importance of integrating the electronic dimension in the CRM. We describe the purposes and the characteristics associated to the CRM model proposed. For this section, we refer to the SAS Institute (1999), Strouse (1999), Ansari *et al.* (2000), Kolesar and Galbraith (2000) and Wang *et al.* (2000).

In section 5, we address the notion of data mining (Berry and Linoff, 2000) and we contrast data mining and statistical data analysis (Sato, 2000). We describe the two types of data mining (SAS Institute, 1999) and the extended use of data mining for marketing purposes (Gomory *et al.*, 1999). We provide a description of the data to be harvested and describe the contribution of the electronic

channel for data collection (Mena, 1999; Novak *et al.*, 2000; Gomory *et al.*, 1999). We refer to Dick and Basu (1994), Jones and Sasser (1995) and Rowley and Dawes (2000) to provide a customer classification grid based on the customer loyalty “level” or “type”. We briefly discuss the importance of creating a trust environment and the issues related to private information disclosure. Mena (1999) and Wang *et al.* (2000) give valuable insights for ensuring security and integrity of the data.

In section 6, we show the outcomes, in terms of resource allocations, resulting from an efficiency increase in data mining. We address this point by using the simplex method developed by Dantzig (1963) and we have recourse to sensitivity analysis (Kleinberg *et al.*, 1998; Vanderbei, 1998). Section 7 provides some concluding remarks and finally the references used in this paper are listed.

3. Churn management

3.1 Definition and characteristics

Relationship building and customer-oriented management are key factors to which companies’ success or failure are closely linked. Customer management requires the collection of a significant amount of information and set up of procedures for treating this information. The harvested information must enable companies to assess and monitor the evolutions regarding the following concepts (Komenar, 1997):

- customer acquisition is relative to the measures taken by firms for gaining clients and, this, as efficiently as possible;
- customer retention is relative to the measures taken to keep the current customers;
- customer extension describes methods to increase the customer return (cross-selling, ...);
- customer selection concerns the identification of the most profitable customers.

The emergence of electronic commerce has multiplied the amount of available information and thus offers new ways for companies to efficiently respond to clients’ expectations.

Simultaneously, customers can more easily inquire about the market opportunities. They become more demanding and tend to switch from their previous supplier to another retailer. This gave birth to the notion of churn. Churn or customer attrition is defined as “the annual turnover of the market base” (Strouse, 1999). This phenomenon has been magnified by electronic commerce. The Internet channel returns control and power to customers who are no longer confined to the decisions of a single company. The outcomes are increase in customer power (Peppard, 2000) and competition exacerbation. Competitors are only one “click away”. Customer empowerment is likely to persist and amplify customer attrition issues. On the other hand, Zettelmeyer (2000) asserts that companies competing on multiple channels get information from multiple sources and can decide to communicate different amounts of data to different clusters of customers, thereby creating new differentiation opportunities. As a result, companies augment their market power, impede the emergence of a competitive strategy essentially based on the cost dominance and thus can design strategies that aim at softening churn problems. This supposition is, however, subject to some restrictive assumptions.

Churn management consists of developing techniques that enable firms to keep their profitable customers and it aims at increasing customer loyalty. However, a long-term customer is not necessary a loyal customer. Customer loyalty is used for individuals who remain clients of their original supplier even if a competitor proposes more advantageous conditions. Loyalty is also not unfailing. Companies have to cultivate it. Loyal customers are the most profitable. They are a free marketing channel in terms of the benefits retrieved by companies from word-of-mouth. These customers are the most likely to purchase other items of a company’s product line.

The impact of churn is often correlated with the industry life-cycle. When the industry is in the growth phase of its life cycle, sales increase exponentially; the number of new customers largely exceeds the number of churners. Companies aim at getting more and more new customers. Nevertheless, the ratio (new customers/churners) tends towards one over

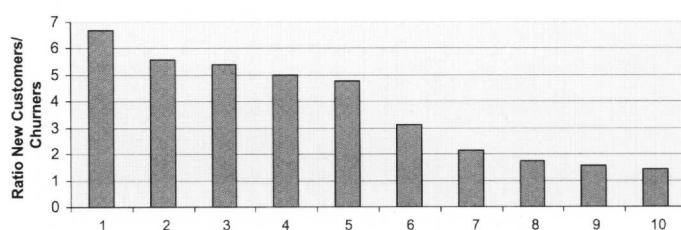
time (Figure 1). The impact of churn becomes then markedly more sensitive. For products in the maturity phase of their life cycle, companies put the focus on the churn rate reduction.

3.2 Churn magnitude: some figures

Let us illustrate the scope of the problem with some recent figures:

- SAS (2000) reported that the telecommunications sector endures an annual rate of churn, ranging from 25 per cent to 30 per cent. This churn rate could still continue to increase in correlation with the growth of the market. Another factor that could push churn rates to higher summits is the deregulation trend.
- The churn costs for European and US telecommunications companies are estimated to amount to US\$4 billion annually (SAS, 2000).
- The ratio (customer acquisition costs/customer retention or satisfaction costs) would be equal to eight for the wireless companies (SAS, 2000). It is generally admitted that companies need three years to amortize the cost (US\$400 in USA and US\$700 in Europe) induced by the replacement of churners and the acquisition of new customers (SAS, 2000). Not less than 10 per cent keep the same supplier while opting for another package. While this kind of churn is not as alarming as the loss of a client, the effect nonetheless remains negative provided that the amount of money spent to recapture the customer in that case is more or less equal to the costs associated with normal churn.
- The Strategis Group asserts that the Internet service providers endure a five-times higher churn rate, culminating to 10 per cent monthly. The main reasons

Figure 1 Churn rate evolution



Source: Berry and Linoff (2000)

invoked were “busy signals, connection speed and poor customer service” (Strouse, 1999). Churn rate is more marked for private customers than for business ones.

- Groth (1999) underlines that the cellular telephone market experienced a 30 per cent annual rate in the USA. In this industry, the cost of acquiring new customers amounts to \$400 per new subscriber.
- Hoffman and Novak (2000) confirm that costs incurred for obtaining new customers are much higher than costs linked with customer retention. Average customer acquisition costs supported by retailers on the Internet range between US\$100 and US\$500 per customer.

These different examples underline why companies pay such attention to the churn rate. Companies that solely rely on onerous marketing campaigns to replace the lost customers are prone to serious survival problems before long. Churn management increasingly becomes crucial and generates a variety of customer-oriented activities.

4. Customer relationship management (CRM)

4.1 Potential of the e-customer relationship management

The effectiveness of customer-oriented marketing techniques involves customization and one-to-one marketing. The Web represents a unique opportunity to develop such a customer relationship framework. Thanks to instantaneous connections, companies have the capacity to receive immediate feedback, to collect data concerning customer tastes, needs and interests, to adapt their product or service line to real-time changes. It is crucial to integrate the electronic dimension in the conception of customer relationship management framework. Indeed, the electronic channel conveys an abundance of information that could not be harvested by only having recourse to traditional information channels. Information stemming from the Internet can be transposed to other, non-Internet based marketing efforts. This relationship is non-reciprocal. Companies cannot, for instance, ask

in a phone questionnaire which ads or banners have led customers to visit the company's Web site. Starting from Net information, one can extrapolate strategies for non-Internet marketing events but the opposite is not likely.

We now describe and summarize the pursued objectives and the desired characteristics and procedures to incorporate within the e-customer relationship management. We refer to suggestions voiced by Ansari *et al.* (2000) and by SAS (1999).

The electronic customer relationship management framework helps companies increase their profitability. Companies attempt to create a personal, specific communication and a long-term relationship with each customer, thus attempting to lessen churn and to raise retention and especially loyalty. The main objective is to maximize the “lifetime value of a customer to the organization” (Peppard, 2000). This can be split into four needs (Strouse, 1999):

- (1) discovery of motives prompting existing customers to leave and of the features characterizing these customers;
- (2) ability to increase loyalty and to provide services and products dovetailed with expectations of the targeted clients;
- (3) ability to design actions for recapturing profitable customers;
- (4) customization of sales and marketing campaigns directed to customers identified as probably profitable and loyal.

4.2 Objectives and integrated tools

The principal functionalities sought would be to implement tools that are able to gather customer data information, analyze and draw conclusions from the available data, visualize the scores and statistics and recommend or propose strategies designed for facing customer attrition issues (Figure 2).

The Data Warehouse tool makes it possible to collect, manage and sort customer-oriented information. Data stem from customers themselves or from instances (legislative body, ...) whose decisions can impact the customer or the company. The data sources are represented in Figure 3. As underlined by Ansari *et al.* (2000), it is important that the data warehouse possesses transforming, extracting and loading functionalities. Such properties can

Figure 2 Electronic customer relationship framework

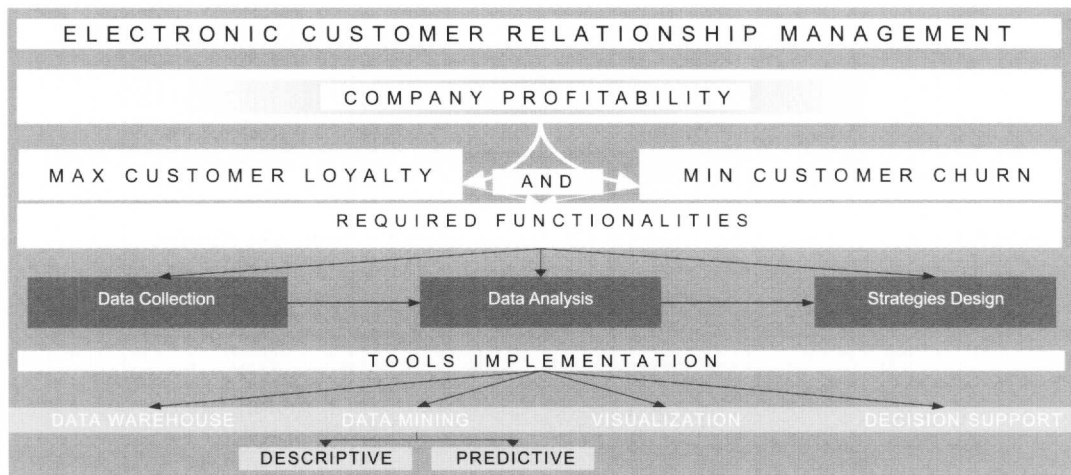
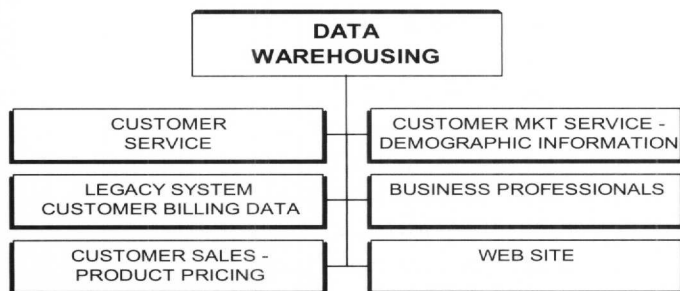


Figure 3 Types of collected data



Source: Groth (1999)

significantly reduce the data preparation time. Processed data are then transmitted to the Data Mining tool. Data mining techniques (decision tree, clustering methods, ...) are used to discover hidden information features in the original data and to construct predictive models. Statistics provided and data patterns discovered are then graphically represented by visualization techniques such as, for instance, scatterplot graphs or parallel coordinates (Gomory *et al.*, 1999). The Decision Support tool reports the discovered data specificities and the computed models to the executive layer. This constitutes the basis for designing advertising campaigns, launching research and development projects, designing the company's Web site, modifying product or improving the after-sales service. In the following section, we focus on the characteristics and potential benefits of data mining methods.

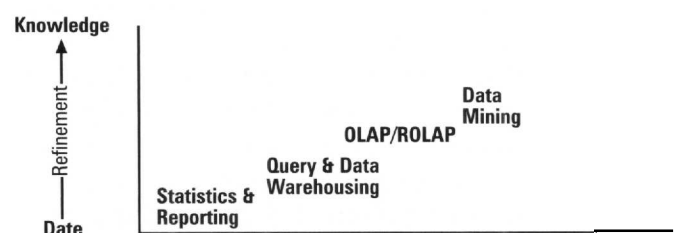
5. Data mining

5.1 Data mining and statistical analysis

Due to the constant increase in the amount of data efficiently operable to managers and policy makers through high speed computers and rapid data communication, there has grown, and will continue to grow, a greater dependency on statistical methods as a means of extracting useful information from the abundant data sources. Statistical methods provide an organized and structured way of looking at and thinking about disorganized, unstructured appearing phenomena. Figure 4 illustrates the different stages involved in the never-failing quest for more refined information.

From this perspective, data mining methods appear to be prominent. Data mining techniques allow the transformation of raw data into business knowledge. The SAS Institute defines data mining as "the process of selecting, exploring and modeling large amounts of data to uncover previously unknown data patterns for business advantage" (SAS Institute, 2000). Berry and Linoff (2000) propose another

Figure 4 Evolution in the quest for information



definition: “data mining is the process of exploration and analysis, by automatic or semi-automatic means, of large quantities of data in order to discover meaningful patterns and rules”. Whatever the definition, the data mining process differs, as underlined by Sato (2000), from the statistical analysis of data.

First, predictive data mining is governed by the need to uncover, in a timely manner, emerging trends, whereas statistical analysis is related to historical facts and based on observed data. Companies cannot rely on remote data to establish their priorities. Besides, the length of time between the emergence of a trend and the detection of and reaction to this trend has to be shortened given the competition increase and the time reduction in the product life cycle. Second, statistical analysis focuses on finding and explaining the major sources of variation in the data. For example, the main contribution of factor analysis stems from the first factor(s) that account(s) for the major part of the variation in data. In contrast to this, data mining endeavors to discover, not the obvious sources of variation, but rather the meaningful, although currently overlooked, information. The retrieval of information in this way is unusual but valuable provided that it is a source of new, previously undetected business knowledge. Therefore, statistical analysis and data mining are complementary. Statistical analysis explains and removes the major part of data variation before data mining is used. This explains why the data warehousing tool not only stores data but also contains and executes some statistical analysis programs.

Data mining has a broad application scope. The tasks assigned to data mining can be divided into two kinds of objective:

- (1) descriptive: data mining aims at increasing the understanding of the data and their content;
- (2) predictive or prescriptive: data aims at forecasting and devising, at orienting the decision process.

5.2 Usefulness of electronic data for applying data mining techniques to marketing goals

Data mining techniques have a widespread use in marketing and have been considerably developed following the advent of electronic

commerce. Data mining methods and the Web are associated with pursuing a variety of objectives (Gomory *et al.*, 1999) ranging from evaluation of overall store performance, promotions’ contribution to sales and banner ads effectiveness, identification of the most visited Web pages, Web search engines leading to the company’s Web site, determination of cross-selling strategies, to segmentation of the customer base. This led to the introduction of the concept of “e-mining” which is defined by Mena (2000) as “the iterative process of analyzing the patterns of your online transactions and extracting knowledge about who is buying what, when, and, most importantly, why from your Web site. It is about extracting previously unknown, actionable intelligence from your Web site interactions.” In addition to the traditional questions to be answered, information stemming from the Web helps evaluate the design of the Web site, gives insights to the partnerships or co-advertising to be developed, identifies the cross-selling opportunities ... (Mena, 1999). In the rest of the paper, we continue to use the generic term data mining.

The development of electronic commerce is confronted with data privacy issues. Customers are reluctant to learn that companies collect every snatch of information. Furthermore, they do not know how this information is handled and who can access it. Prahbaker (2000) insists on the need to explain that the collected information is used by businesses to better serve their clients. It is very important not to breach the reliance customers have on the Internet. The most negative impact possibly induced would be the belief that the Internet and electronic commerce are perceived as “structurally flawed” (Geng and Whinston, 2000). Businesses have to do their best to create a trustworthy environment. They have to develop privacy policies that clearly state which information is collected and how this policy is implemented. Wang *et al.* (2000) reveal that 96 percent of customers find it essential that Web sites explain in detail how personal information disclosed during the purchasing process will be used. Although this paper does not intend to focus on the privacy concern, it seems appropriate to report some general and important suggestions provided by

Mena (1999) and Wang *et al.* (2000) for developing trust:

- to build a Web site that is well organized and comprehensive;
- to ensure that private information is protected and kept secret;
- to have contingency plans in case of security or data integrity problems;
- to focus upon the fact that customer information will not be disclosed;
- to emphasize that the primary goal that underlies information gathering is to improve the service provided;
- to highlight the benefits customers receive in exchange for the information they transmit.

Companies can harvest information at the server, client or proxy levels. This information could be obtained through (Berry and Linoff, 2000):

- *Clickstreams*: when a net user visits a Web site, clicks on a Web page, the user browser gives information to the server managing the visited address. This information (browser type, operating system ...) enables the server to fit the Web page to the needs of the customers. More important, it enables the server to identify the address of the referring page and to list the different web pages visited within this Web site.
- *Cookies*: these are text files sent by a server to a browser that stores this file. The “sender” server has access to this information and can know which Web pages the user had visited in his previous visits, can update customer information. Cookies turn out to be very useful for the implementation of mass customization principles. However, cookies are not adapted for computers used by a variety of users, provided that each user is identified by the same cookie.
- *Customer registrations* involve subscription of the customer, assignment of personal login name and password. Customer information is stored on the server computer. Therefore, problems of multi-user computer can be avoided. Some businesses use both registration and cookie processes.

Instead of using ID name or cookies and to have to bear the problems involved (privacy,

customer reluctance, multi-user computers), Wu *et al.* (1998) propose the adoption of a system that uses information provided by log files, that allows the session identification and that reconstructs the users’ “most common traversal paths”.

5.3 Data mining and churn management

Data warehouse tool provides us with the customer database and is assumed to have handled the data by taking into consideration their static, dynamic or event-like features. Data have been sorted out in diverse categories. The database would consist of data provided by:

- (1) Traditional sources (customer marketing service, business professionals ...). We adopt the classification proposed by the SAS Institute (2000):
 - customer demographic data refer to variables such as gender, age ...;
 - raw usage data relate to the way customers use their product or have recourse to services delivered by the company;
 - Contract, account and customer relationship data comprise attributes such as payment method, customer’s creditworthiness, type of contact the customer initiates with the company, customer behavior in response to promotions ...;
 - Technical quality data includes the product options chosen by the customer, the quality of service demanded.
- (2) The Web: we adopt the classification given by Srivastava *et al.* (2000):
 - data relative to the content of the Web page;
 - data relative to the structure of the content (HTML tags ...);
 - data relative to the usage of the Web page (referrer, date of access ...);
 - data relative to the user profile (customer profile ...).

Two different conceptions for the evaluation of customer churn and profitability have been developed. Ansari *et al.* (2000) advocate the importance of data related to recency, frequency and monetary (RFM) attributes for evaluating customer churn and profitability.

From this point of view, companies have to determine whether a client is still “active”. Groth (1999) points out that customers seldom let companies know that they have stopped shopping there. As a result, companies consider the recency parameter for evaluating churn as a rule of thumb. Companies have a tendency to rely on very simplistic rules such as: “should a customer have purchased within the last eight months, he is considered an active customer”. Intuitively, we see the bias involved in such classification rules. Infrequent shoppers are likely to be misrepresented.

Such rules neglect the purchasing behavior that may significantly differ across segments and individuals. Groth grounds his criticisms on the non-integration of the customer behavior in the RFM model. He prefers opting for a methodology called “Value, activity and loyalty method (VAL)”. The VAL method addresses the aforesaid deficiency by “pooling the entire customer base to robustly estimate individual customer behavior based on limited individual purchase history” (Groth, 1999). This method is based on “hazard rate” type of models and uses transactional data. It has been validated by different studies. Groth proposes that the VAL method is superior to the RFM method because the former is forward-looking as compared to the latter which is described as being backward-looking.

5.4 Customer classification

Descriptive data mining methods appear useful to understand differences and particularities of the various categories of clients. It allows customer segmentation, formation of homogeneous clusters or categories of customers characterized by a small variance within groups, and a high variance between groups. Based on these clusters, data mining methods have the ability to detect common features of the individuals belonging to the same cluster.

Different taxonomies have been proposed to describe the different “levels” of “customer loyalty”. Dick and Basu (1994) suggest the conceptualisation of customer loyalty as an interaction between attitude and behavior. This conception of customer loyalty leads to the definition of four types of loyalty: loyalty, latent loyalty, spurious loyalty and no loyalty. We do

not opt for this classification since it seems too restrictive to only take into account attitude and behavior for analyzing customer loyalty.

Moreover, Dick and Basu (1994) content themselves with a single category for non-loyal customers. Rowley and Dawes (2000) provide a classification grid for disloyalty. They split the “no loyalty” category of Dick and Basu (1994) into four new categories for non-loyal customers: disturbed, disenchanted, disengaged and disruptive customers (Figure 5).

Finally, Jones and Sasser (1995) proposed splitting the customer base into four categories. These four categories depend on the behavioral customer attitudes, the degree of (dis)satisfaction and the will or ability to act on their (dis)satisfaction. This classification scheme has the advantage of giving insights regarding what marketing strategy to implement for dealing with the specificities of the different categories.

The four categories defined by Jones and Sasser (1995) are listed below:

- (1) For the loyal, retained customers:
 - *Loyalists and apostles* are particularly satisfied with services and products delivered by the company. Besides, the company benefits from very positive word-of-mouth behavior from these customers.
 - *Hostages* are not satisfied. But they are tied to the company, either because there are no substitute products, or switching costs are too high, or hostages are not conscious of competitors’ offers.
- (2) For the churners:
 - *Defectors* are customers who are merely satisfied, neutral or quite dissatisfied. Should their frustration increase, they could tell others about their dissatisfaction. They are then called terrorists.
 - *Mercenaries* are customers whose degree of satisfaction has no impact on their fidelity. They are versatile, difficult to monitor and buy on impulse.

After completion of the descriptive data mining analysis, we obtain a set of essential attributes for each customer category, allowing us to

Figure 5 Types of disloyal customers

Disloyalty Category	Purchase Pattern and Behavior
Disengaged	No intention to purchase Do not know the product
Disturbed	May purchase in the future, previous customer Bad recent experience
Disenchanted	Not likely to purchase in the future, previous customer Number of negative experiences
Disruptive	No intention to purchase, previous customer Discourage other people from buying the considered product

Source: Rowley and Dawes (2000)

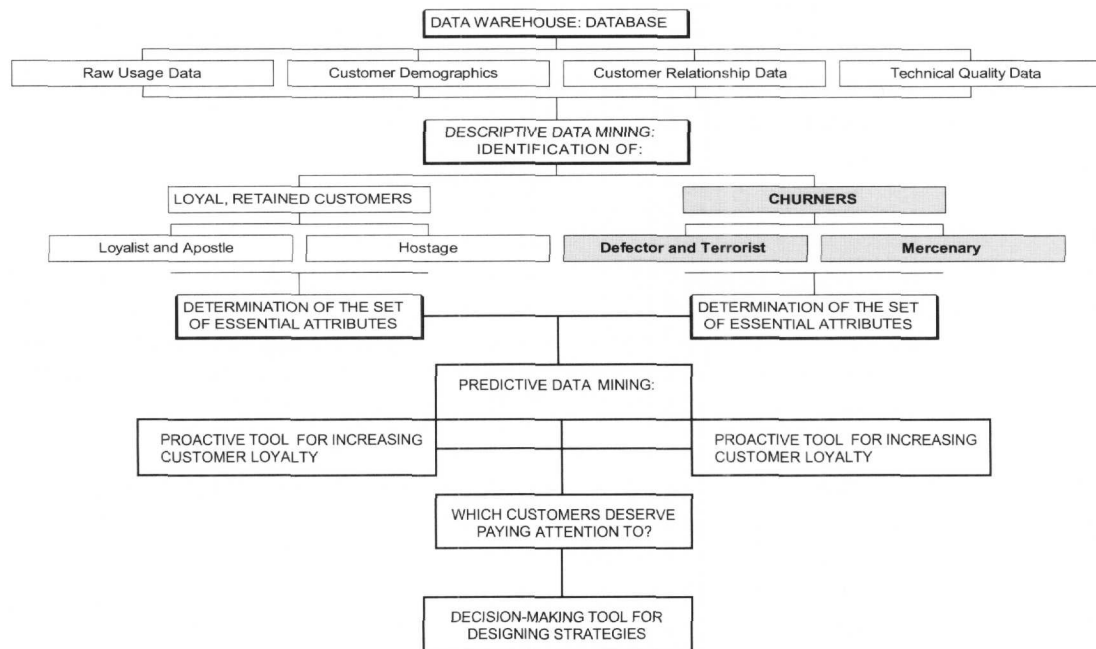
determine, on the basis of the current data, which category each current customer belongs to. We are therefore able to analyze the current clientele of the company and classify each customer regarding the different categories defined by Jones and Sasser (1995). We now use data mining in a prescriptive perspective. Starting from this classification and the particularities associated with the different categories, we are able to predict which customers are likely to leave and which customers are and will remain loyal. Companies can then decide which customers deserve paying attention to and are able to design proactive marketing and/or merchandising strategies that alleviate or remove the motives pushing churners to leave or that reinforce the loyalty of the retained clientele. Jones and Sasser (1995) give some general insights for each customer category.

- *Loyalists and apostles* are the ground layer for the company. They deserve to receive constant attention. It is important to give customers ways to manifest their dissatisfaction. By properly responding, the company can conserve its loyal customers or reinforce their satisfaction.
- *Hostages* can, in some circumstances, leave and even transform themselves into terrorists. They can also decide not to leave and continue complaining. Companies have a tendency to spend vast amount of money to satisfy them without tangible results. Any optimal solutions can be elaborated for them.
- *Defectors*: the company cannot accept their defection. By treating them well, companies can keep them or even convert them into loyalists. It is once again important to let them express their dissatisfaction motives. However, the desire to satisfy each defector can be too onerous and counterproductive. For example, Southwest Airlines “fires” customers they cannot appropriately serve.
- *Mercenaries* are and will remain out of control whatever the company may do. Companies do not have to care about them.

6. Impact of data mining effectiveness

As described in the preceding section (and Figure 6), data mining is an essential component of the customer relationship management and can be viewed and used as a descriptive and/or a predictive tool. Substantial research has recently been conducted on data mining. Decision trees techniques, machine-learning algorithms (Bshouty and Hellerstein, 1998; Servedio, 2000) or logical analysis of data (LAD) (Boros *et al.*, 1996, 1997, 2000) have appeared or have been further polished. For instance, LAD is a pattern-grounded method and has the ability to discover a minimal set of essential features making it possible to appropriately represent and explain the set of observations as well as the detection of anomalies or hidden information in the original data set. LAD generates a discriminant function

Figure 6 Integration of data mining techniques in the decision-making process



that classifies the observations and can be used in a predictive perspective.

These new methods applied to extended databases that incorporate data harvested on the Net are likely to bring efficiency gains in the analysis and understanding of data. This is highlighted by a paper published by Lim *et al.* (2000) who have compared 32 algorithms. It appears that LAD, when applied to the same databases, is always very close to the best solution from the 32 other data mining algorithms and turns out to be the most efficient methods for two of the five analyzed databases. Moreover, LAD has been successfully used for situations as diverse as the diagnosis of psychiatric patients, creditworthiness evaluation and characterisation of breast tumors (Boros *et al.*, 1996, 1997 and 2000).

We analyze the impact of new data mining methods, presumably leading to efficiency gained, on churn management. We aim at developing a framework for appraising the impact of the ongoing data mining enhancements on churn management. This may help the company answer questions such

as: How could it use the information extracted by these methods from the raw data stocked in the data warehousing tool? What is the value of the churn decisions resting on the extracted patterns? What are the consequences on the resources allocation within the e-CRM tools?

The analysis framework we propose is appropriate for a company attempting to minimize the cost (rate) of churn and opting for e-customer relationship management. We make the hypothesis that churn costs may be substantially reduced by developing an adequate CRM framework, that includes data warehousing, data mining and OLAP functionalities. This hypothesis is strongly supported by Strouse (1999), Berry and Linoff (2000), The SAS Institute (2000) and Mittal *et al.* (2000). Data mining, data warehousing and OLAP functionalities are viewed as tools used by a company willing to take advantage information it possesses on its customers and willing to employ it in the decision-making process. We pose the further hypothesis that the relationship between the churn costs savings a company can retrieve are linked to the efficiency of its CRM framework, i.e. with the

contribution of the four CRM tools or, stated differently, with the level of “information interestingness” (Kleinberg *et al.*, 1998) provided.

Let us model this situation by using the simplex method. Any company wants to minimize the churn costs C ; this minimization is subject to m constraints. This can be stated as: $\min_{s.t. Ax=b, x \geq 0} C$, A is the constraint matrix, b is the resource bounds vector, X is the vector of independent variables and c is a vector that stores the objective function coefficients. We assume that the elements of A , c and b are given. We interpret the components of the vector c as being the respective effectiveness or contribution of the CRM tools. The element c_i can be defined as the “churn cost saving” resulting from the investment of one currency unit in the CRM tool i . Churn cost savings are positively correlated with the level of information interestingness provided by the CRM tools, with the propensity to integrate the information provided in the decision-making process. The elements of the vector x denote the amount of money invested in the four CRM tools. The product $(c_i \cdot X_i)$ thus denotes the “total churn cost savings” induced by the investment of an amount of magnitude X in the tool i which contributes to “churn problem alleviation” at a rate c_i .

The objective function consists of maximizing the benefits (cost reduction) retrieved from applying the CRM framework churn management: $\min C \approx \max_{s.t. Ax=b, x \geq 0} X_{CRM}$, with X_{CRM} being the total cost savings induced by the adoption of the CRM model.

In the framework developed above (Figure 3), CRM involves the use of four tools: data warehouse tool (DW), data mining tool (DM), visualization tool (V) and decision support tool (DT). We assume that the relationship between CRM and its four components is linear. However, the developed framework can be extended to non-linearities by using Kuhn-Tucker conditions.

As explained in section 5, we suppose that the data mining tool has to be considered from two angles, i.e. descriptive data mining X_{DMd} and predictive data mining X_{DMp} . The objective function can now be rewritten as:

$$\begin{aligned} \min_{s.t. Ax=b, x \geq 0} C &\approx \max_{s.t. Ax=b, x \geq 0} X_{CRM} \\ &\approx \max_{s.t. Ax=b, x \geq 0} C_{DW} \cdot X_{DW} + C_{DMd} \cdot X_{DMd} + \\ &\quad C_{DMp} \cdot X_{DMp} + c_V \cdot X_V + C_{DT} \cdot X_{DT} \end{aligned}$$

Under these assumptions, A is a $[m, 5]$ matrix, c is a r -row vector and b is a m -dimensional vector. Assume that we solve this linear program. From the 5 initial variables, we get a subset of m basic variables; we set the remaining non-basic variables at zero. The matrix A is now partitioned into two sub-matrices B and N : $A = [B \ N]$. Let B be the $[m, m]$ square, non-singular matrix whose columns are those of A associated to the basic variables and N the rectangular matrix associated with the non-basic variables. We get a similar partition for the vector $c = [c_B \ c_N]$. The optimal level for each variable is given by the components of the vector $B^{-1}b$.

We decide to substitute new data mining algorithms for traditional data mining techniques. We pose the hypothesis (as discussed above) that they are able to augment the efficiency of the data mining tool, either for descriptive or predictive goals. For evaluating the impact of the data changes, i.e. the augmented contribution due to recent data mining algorithms, on the benefits of customer relationship management, we use the simplex sensitivity analysis. Sensitivity analysis consists of principles and rules for assessing how changes in input data alter the optimal solution.

For performing the sensitivity analysis, we introduce the variables c'_{DMd} and c'_{DMp} such as: $y = c'_{DMd} - c_{DMd}$, with $y > 0$ and $c'_{DMp} - c_{DMp}$, with $z > 0$. The insertion of c'_{DMd} and c'_{DMp} gives the vector c' . We are confronted with two possibilities, depending on whether DM_d and DM_p were basic or non-basic variables:

- (1) DM_d and DM_p were non-basic variables: we have to calculate the minimal values for y and z allowing respectively DM_d and DM_p to enter into the basis. Let A_{DMd} and A_{DMp} the columns of A associated respectively with DM_d and DM_p . For DM_d becoming a basic variable, the expression $(c'_B B^{-1} A_{DMd} - DM_d')$ has to be equal to 0. This is equivalent to: $(c_B B^{-1} A_{DMd} - c_{DMd} - y) = 0$. The same reasoning is applicable to DM_p .

- (2) DM_d and DM_p were basic variables: we have to calculate the upper bounds for y and z so that DM_d and DM_p remain basic variables. This condition is expressed by: $(c'_B B^{-1} N - c'_N) \geq 0$. The same reasoning is applicable to DM_p .

Data used may have been uncertain or may have been accurate and updated, companies may wish to consider various scenarios. Such situations are very common and lead to the question whether it is possible to have recourse to the previously compiled optimal solution to obtain easily and rapidly the resulting new optimal solution that takes into account the most recent data changes. Sensitivity analysis allows integrating changes and updating of the current solutions as a consequence of data changes. Sensitivity analysis is adapted for being applied to rapidly evolving contextual situations. In this paper, sensitivity analysis permits to build a quantitative framework for assessing to what extent data mining evolutions affect customer relationship management and how this could lead to a modification in the resources affectation within activities involved in customer relationship management. The sensitivity analysis model we propose characterizes itself by its generality and is valuable for being adaptable and refined for specific scenarios and goals.

7. Conclusion

We discuss the churn concept and illustrate its magnitude with different figures. We introduce and characterize a conceptual framework for alleviating attrition problems faced by many companies from different industrial sectors. We advocate the construction of a customer relationship management that would incorporate the electronic dimension and would be enhanced by the development of adequate tools for the collection, the treatment and the analysis of data. The four tools involved are related to data warehousing, data mining, visualization and decision support.

While mentioning the importance of data warehousing and OLAP functionalities, we focus on data mining, be it descriptive- or prescriptive-oriented. Parallel to the broader

databases made available by the electronic channel, data mining has recently been marked by substantial improvements. This is demonstrated by the results provided by Lim *et al.* (2000).

Finally, we use sensitivity analysis for developing a general causal assessment scheme caused by data mining enhancements on customer relationship management. This allows the modeling of how ongoing developments in the data mining field may affect churn rates, decision-making processes and budget allocations. We recall that our analysis is based on the supported assumption (Strouse, 1999; SAS Institute, 1999, 2000) that churn amplitude is negatively correlated with the efficiency of data mining tools and that the relationship between churn and CRM tools is linear. The linear assumption can be removed and non-linear relationships can be analyzed by having recourse to the Kuhn-Tucker conditions. The application of our analysis framework is not contingent on specific situations and has the property to be applicable for various data mining purposes and to be, if necessary, refined to capture specificities of situations under study.

References

- Ansari, S., Kohavi, R., Mason, L. and Zheng, Z. (2000), "Integrating e-commerce and data mining: architecture and challenges", *WEBKDD'2000 Workshop on Web Mining for E-Commerce – Challenges and Opportunities*.
- Berry, M.J.A. and Linoff, G. (2000), *Mastering Data Mining: The Art and Science of Customer Relationship Management*, Wiley Computer Publishing, New York, NY.
- Boros, E., Hammer, P.L., Ibaraki, T. and Kogan, A. (1997), "Logical analysis of numerical data", *Rutcor Research Report*, No. 04-97.
- Boros, E., Horiyama, T., Ibaraki, T., Makino, K. and Yagiura, M. (2000), "Finding small sets of essential attributes in binary data", *Rutcor Research Report*, No. 13-2000.
- Boros, E., Hammer, P.L., Ibaraki, T., Kogan, A., Mayoraz, E. and Muchnik, I. (1996), "An implementation of logical analysis of data", *Rutcor Research Report*, No. 22-96.
- Brookman, F. (1998), "Companies invest in customer loyalty", *Marketing News*, Vol. 32 No. 5.
- Bshouty, N. and Hellerstein, L. (1998), "Attribute efficient learning in query and mistake bound models", *Journal of Computer System Science*, Vol. 56, pp. 310-19.

- Dantzig, G.B. (1963), *Linear Programming and Extensions*, Princeton University Press, Princeton, NJ.
- Dick, A.S. and Basu, K. (1994), "Customer loyalty: toward an integrated conceptual framework", *Journal of the Academy of Marketing Science*, Vol. 22 No. 2, pp. 99-113.
- Geng, X. and Whinston, A.B. (2000), "Defeating distributed denial of service attacks", *IT Pro*, Vol. 2 No. 4, pp. 36-41.
- Gomory, S., Hoch, R., Lee, J., Podlaseck, M. and Schonberg, E. (1999), "E-commerce intelligence: measuring, analyzing and reporting on merchandising effectiveness of online stores", *IBM Research Report*, Yorktown Heights, NY.
- Groth, R. (1999), *Data Mining: Building Competitive Advantage*, Prentice Hall, Santa Clara, CA.
- Hoffman, D.L. and Novak, T.P. (2000), "How to acquire customers on the Web", *Harvard Business Review*, pp. 3-8.
- Jones, T.O. and Sasser, W.E. (1995), "Why satisfied customers defect", *Harvard Business Review*, pp. 88-99.
- Kleinberg, J., Papadimitriou, C. and Raghavan, P. (1998), "A microeconomic view of data mining", *Data Mining and Knowledge Discovery*, Vol. 2 No. 4, pp. 311-24.
- Kolesar, M.B. and Galbraith, R.W. (2000), "A service-marketing perspective on e-retailing: implications for e-retailers and directions for further research", *Electronic Networking and Applications*, Vol. 10 No. 4, pp. 424-38.
- Komenar, M. (1997), *Electronic Marketing*, Wiley Computer Publishing, New York, NY.
- Lim, T.-S., Loh, W.-Y. and Shin, Y.-S. (2000), "A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms", *Machine Learning*, Vol. 40, pp. 203-29.
- Mena, J. (1999), *Data Mining your Web Site*, Digital Press, Boston, MA.
- Mittal, V., Kumar, P. and Tsiros, M. (2000), "Attribute-level performance, satisfaction, and behavioral intentions over time: a consumption-system approach", *Journal of Marketing*, Vol. 63, pp. 88-101.
- Novak, T.P., Hoffman, D.L. and Yung, Y.-F. (2000), "Measuring the customer experience in online environments: a structural modeling approach", *Marketing Science*, Vol. 42 No. 4, pp. 80-5.
- Peppard, J. (2000), "Customer relationship management (CRM) in financial services", *European Management Journal*, Vol. 18 No. 3, pp. 312-27.
- Prahbaker, P.R. (2000), "Who owns the online customer", *Journal of Consumer Marketing*, No. 17, pp. 158-71.
- Rowley, J. and Dawes, J. (2000), "Disloyalty: a closer look at non-loyals", *Journal of Consumer Marketing*, Vol. 17 No. 6, pp. 538-47.
- SAS Institute (1999), "The role of e-intelligence in customer relationship management (CRM)", *A SAS Institute White Paper*.
- SAS Institute (2000), "Best practice in churn prediction", *A SAS Institute White Paper*.
- Sato, Y. (2000), "Perspective on data mining from statistical viewpoints", *Knowledge Discovery and Data Mining: Current Issues and New Applications*, 4th Pacific-Asia Conference, PAKDD 2000, Kyoto, Japan.
- Servedio, R.S. (2000), "Computational sample complexity and attribute-efficient learning", *Journal of Computer and System Sciences*, Vol. 60 No. 1, pp. 161-78.
- Srivastava, J., Cooley, R., Deshpande, M. and Tan, P.-N. (2000), "Web usage mining: discovery and applications of usage patterns from Web data", *SIGKDD Explorations*, Vol. 1 No. 2, pp. 12-23.
- Strouse, K.G. (1999), *Marketing Telecommunications Services: New Approaches for a Changing Environment*, Artech House Inc., Norwood, MA.
- Vanderbei, R.J. (1998), *Linear Programming: Foundations and Extensions*, Kluwer Academic Publishers, Boston, MA.
- Wang, F., Head, M. and Archer, N. (2000), "A relationship-building model for the Web retail marketplace", *Electronic Networking and Applications*, Vol. 10 No. 4, pp. 374-84.
- Wu, K.-L., Yu, P.S. and Ballman, A. (1998), "SpeedTracer: a Web usage mining and analysis tool", *IBM Systems Journal*, Vol. 37 No. 1.
- Zettelmeyer, F. (2000), "Expanding to the Internet: pricing and communications strategies when firms compete on multiple channels", *Journal of Marketing Research*, Vol. 37 No. 3, pp. 292-308.