Data mining algorithms and techniques research in CRM systems

Conference Paper · July 2011

3 authors, including:

Some of the authors of this publication are also working on these related projects:

SIPAMER - Intelligent System for prediction, analysis and monitoring of performance indicators of technological and business processes in the field of renewable energies View project

Intelligent system for trading on wholesale electricity market (SMARTRADE) View project

Data Mining Algorithms and Techniques Research in CRM Systems

ADELA TUDOR, ADELA BARA, IULIANA BOTHA
The Bucharest Academy of Economic Studies
Bucharest
ROMANIA
{Adela_Lungu}@yahoo.com
{Bara.Adela, Iuliana.Botha}@ie.ase.ro

Abstract: - Data mining has become an increasingly powerful technology, being applied in a variety of areas, from investment management to astronomy. The importance and potential of this technology has been recognized in marketing, banking, health, telecommunications, etc. for applications such as cost analysis, efficiency promoting, customer vulnerability analysis, customer relationship management. Data mining can be considered an interdisciplinary research which employs applied mathematics and computational statistics. The main goal of the paper is to illustrate the importance of the optimization methods used in data mining process, as well as the specific predictive models and how they work in this field. Also, the customer relationship management systems have been developed lately, offering new opportunities for a strong profitable relation between a business and clients.

Key words: **-** data mining, data warehouses, attribute selection, optimization algorithms, preprocessing, clustering, classification, association, customer relationship management, OLAP

1. Introduction

All over the world, we are surrounded by data. There are large volumes of data, but not enough information. This is a problem that many companies and industries are facing. A solution to this issue could be data mining, also known as "knowledge discovery in databases". Many institutions hold large amounts of data, due to their work activity and also due to the rapidly evolution of technology. Useful information is hiding behind these data, but traditional methods of data analysis are outweighed by large volumes of data, too. Data mining offers the possibility of extracting previously unknown and potentially useful knowledge and patterns from large databases. The process consists of numerous steps such as integration of data from several databases or data warehouses, preprocessing of the data and induction of a model with a learning algorithm. The model is then used to identify and implement some decisions within the company.

Strong competition on the market is forcing organizations to identify innovative ways to increase their market share while reducing cost. Therefore, data mining can provide significant competitive advantage to a company by exploiting the potential of large data warehouses. It has an important role in helping the companies to understand their clients'

behavior, anticipating the stocks, to optimize the sales policy and other benefits.

The aim of this technology is usually to find hidden but significant relationships that can lead to a bigger profit. The essential difference between the data mining techniques and traditional methods with databases is that in the second case, the database become passive and it is used only for storing large data amounts. In the first case, the database is no longer passive. Through an automated process of data analysis, it could offer useful information for the business plans.

The process of data mining involves multiple steps (see fig. 1). It starts with the selection of data incorporated in a training set that consists of observed values of certain attributes, generally historical data. The selected data are then cleaned and preprocessed. Cleaning is made in order to remove the discrepancies and preprocessing is responsible for consolidation of relevant information to the mining algorithm, trying to reduce the problem complexity. Among the steps in preprocessing, attribute selection has a special role. The data set is then analyzed to identify patterns, so that different inductive learning algorithms are applied. The model is finally validated with new data sets to ensure its generalizability. The steps in

the mining process are performed iteratively until meaningful business knowledge is extracted.

An important issue here is the attribute selection which is preferable to be done before applying the learning algorithm. This involves a process for determining which attributes are relevant in that they predict or explain the data, and conversely which attributes are redundant or provide little information. A subset of M attributes out of N is chosen, complying with the constraint M   N, in such a way that characteristic space is reduced according to some criterion. Attribute selection guarantees that data getting to the mining phase are of good quality [1]. Identifying and keeping the attributes which are relevant to the decision making often provides valuable structural information.
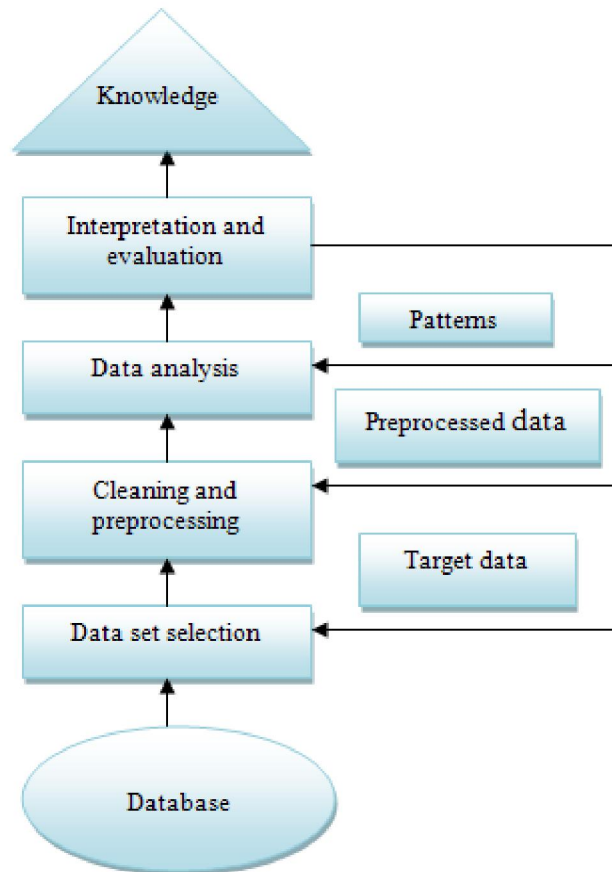
Figure 1 - An overview of the data mining process
Source: adapted from U. Fayyad, G. Piatetsky, P. Smyth, 1996

Many attribute selection methods are based on optimization approach that includes genetic algorithms (Yang and Honavar, 1998), logical analysis of data (Boros, 2000), mathematical programming [2] (Bradley, 1998). One of the most efficient optimization methods for data mining is support vector machines or kernel methods and the most common concepts learned in data mining are classification, clustering and association. The following section includes a data mining application, namely customer relationship management systems (CRM).

2. Customer relationship management systems optimization by using data mining techniques

A customer relationship management system (CRM) is a bucket of IT applications and procedures whose target is to identify the main expectations and preferences of the clients and to use efficiently the gathered information in order to improve the relationships between the business and the customers. The implementation of such system implies two components:

- The managerial component, consisted of the total methods and techniques used for the

integration and usage of data related to the customers behaviour;

- The IT component, which includes the hardware and software equipment used for data collection, storing and management.

The main components of CRM systems are:

- A stop shop is the input point in the system for the data, meaning the requests and claims of the clients, which then are processed within a documents management system process;
- Contact Center/ Help Desk offers special assistance to the clients who ask for information regarding the specific products and services. Developing such a component provides many advantages: reducing the number of missed calls by intelligent distribution of calls, increasing the productivity of the marketing and sales departments, enhancing customer satisfaction by increasing the value that he perceives, monitoring the satisfaction of customers.
- eCRM meaning the internet technology using specific instruments, such as: personalized e-mail addresses, chat or interactive dialogs, forums.

According to [3], CRM consists of four dimensions:

(1) Customer Identification;
(2) Customer Attraction;
(3) Customer Retention;
(4) Customer Development.

They share the common goal of creating a deeper understanding of customers to maximize customer value to the organization in the long term. Data mining techniques, therefore, can help to accomplish such a goal by extracting or detecting hidden customer characteristics and behaviours from large databases.

The main advantages of CRM implementation are: more efficient activities of the orders received from consumers, improving the quality of services provided to the clients, a qualitatively higher level communication with the client by using multiple communication channels (telephone, stop shop, web, e-mail), reducing the communication costs with clients, reduce time consuming for claims, achieving a better image of the organization in front of clients.

In practice, especially in the large companies, applying CRM techniques implies the following steps:

1. Identify the organization's clients and including them in different categories depending on their preferences and behaviors. We can split the clients in four categories:
   a) Clients with general requirements and an uniform character;
   b) Clients with specific requirements and an uniform character;
   c) Clients with general requirements and no uniform character;
   d) Clients with specific requirements and no uniform character

2. Establishing the necessary information and design the system architecture. In this phase, there is planning the clients management database which includes, in general, information related to: identification of person, professional training, social status, embership in a particular category of clients, attitudes and perceptions, behaviors in different situations, requests, complaints submitted by customer.

3. Identifying ways of information gathering which involves developing a toolbox of methods and techniques whereby information describing customer behaviors to be collected and entered into the database.

4. Gathering information and updating the database that consists of applying the techniques defined in the second stage, with the scope of the consolidation of customer database.

5. Making operational the changes in the organizational plan for enhancing the customer satisfaction by improving and diversifying provided services, acting simultaneously both in terms of coverage general requirements and individual ones. Studies reveal that the amplification of satisfaction degree generates an improved image of the organization on the market, but only up to a maximum point, beyond which the image begins to deteriorate.

Data mining plays an important role in CRM by identifying customer behaviour patterns from customer usage data and predicting which customers are likely to respond to cross-sell and up-sell campaigns, which is very important to the business [4]. Regarding former customers, data mining can be used to analyze the reasons for churns and to predict churn [5].

Optimization also plays an important role in CRM and in particular in determining how to develop proactive customer interaction strategy to maximize customer lifetime value. A customer is profitable if the revenue from this customer exceeds company's

cost to attract, sell and service this customer. This excess is called the customer lifetime value [6].

E.W.T. Ngai in [7] proposes a graphical classification framework on data mining techniques in CRM as shown in figure 2:
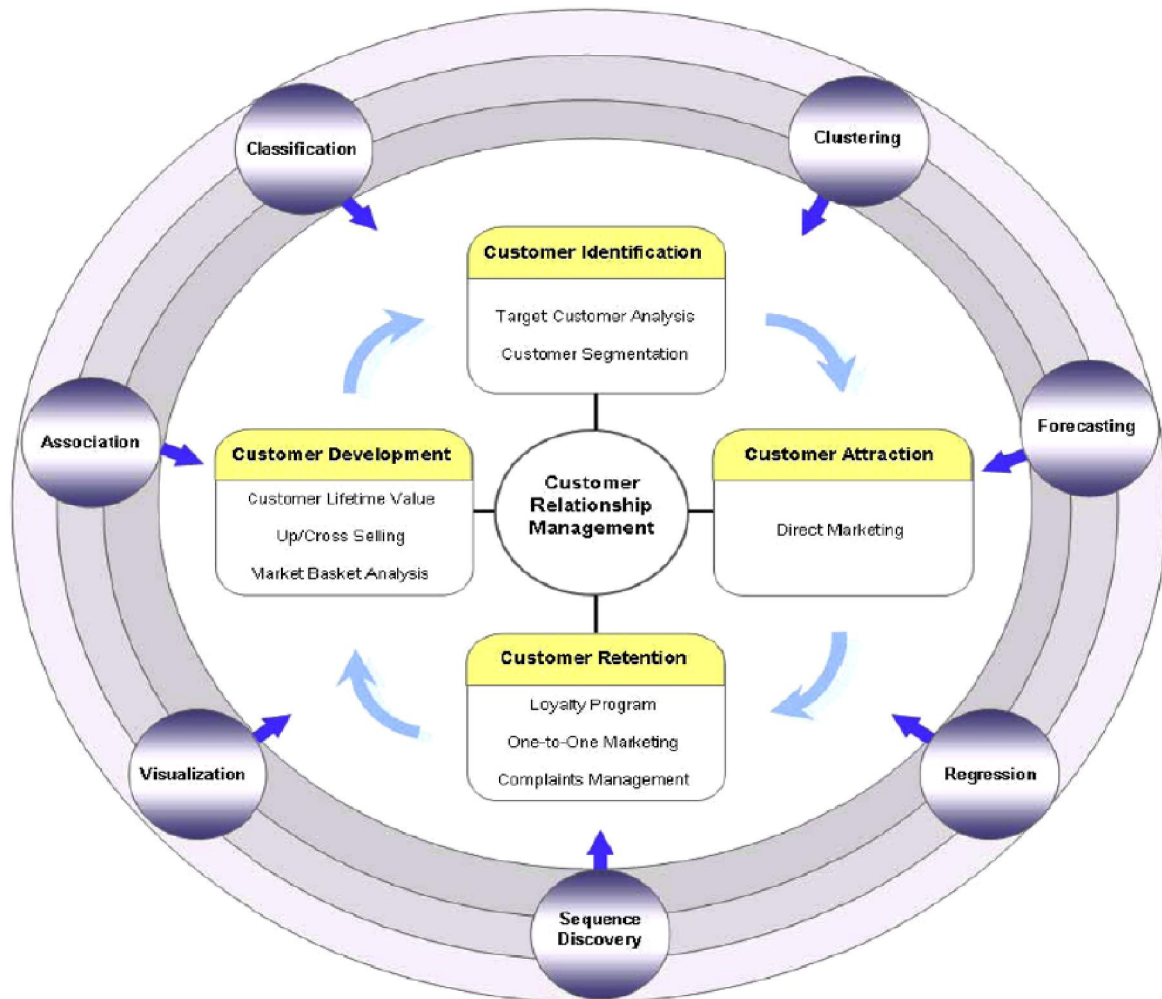

Figure 2 - Classification framework for data mining techniques in CRM
Source: [7]

Data mining techniques can be used successfully, especially because CRM implies a multidimensional approach which can, by instance, include three dimensions:

- Hierarchy of products (brand, class, category, product);
- Hierarchy of periods (years, quarters, months, dates);
- Customer hierarchy (regions, areas class customers).

In practice, this approach is successfully performed through a modern concept that stands today in the majority of process support systems decision, namely the OLAP (On-Line Analytic Processing), which is based on technical multidimensional data analysis [9], [10].

If we refer to the CRM user demands, OLAP systems provide support for real-time satisfaction of specific claims, because they anticipate the timing and content of the interrogation and provides the optimal combination between pre-calculated results and those calculated at the time of information requested. OLAP systems use a specific tool, reason for which most experts believe that they represent the best environment for implementation of functional information models based on systems dynamic principles.

At present, almost all large organizations hold an Intranet platform which, together with some extensions and instruments, provides the basic functionality of Business Intelligence applications, such as organizing information in data warehouses

and processing them using data mining techniques. Numerous specific data mining functions are already implemented as components of the Intranet architecture or like specific solutions such as CRM.

3. Conclusions

The large volume of information that decision makers are facing, requires advanced processing technologies, but also new types of systems to assist decision. Business Intelligence is currently offering solution for the problems in decision making at all managerial levels.

Data Mining, as part of BI systems, has enjoyed great popularity in recent years, with advances in both research and commercialization. Data mining is focused on assessing the predictive power of models and performs analysis that would be too hard-working and time-consuming by using traditional statistical methods. It offers important information which is used to improve customer retention, response rates, attraction, and cross selling. As shown in the paper, through the full implementation of a CRM program, the companies increase the value of their customers, keeping and attracting the right ones.

Although many books and articles have been written on Business Intelligence topic, it still represents a promising research field. Interest in data mining continues to increase and the potential of using optimization methods needs more study. Also, investigating how to combine optimization and data mining techniques, especially in the CRM area, should be encouraged for many reasons. Data mining and optimization can be integrated to build customer profiles, which is absolutely necessary in many CRM applications.

References

[1] Liu, H., Motoda, H., Feature Selection for Knowledge Discovery and Data Mining, Kluwer academic Publishers, 1998.

[2] Bradley, P.S., Mangasarian, O.L., k-Plane clustering. Journal of Global Optimization 16 (1), 23–32, 2000.

[3] Kracklauer, A. H., Mills, D. Q., & Seifert, D. Customer management as the origin of collaborative customer relationship management. Collaborative Customer Relationship Management - taking CRM to the next level, 3–6, 2004.

[4] Chiang, I., Lin, T., Using rough sets to build-up web-based one to one customer services. IEEE Transactions, 2000.

[5] Chiang, D., Lin, C., Lee, S., Customer relationship management for network banking churn analysis. In: Proceedings of the International Conference on Information and Knowledge Engineering, Las Vegas, NV, 135–141, 2003.

[6] Sigurdur Olafsson, Xiaonan Li, Shuning Wu, Operations research and data mining, European Journal of Operational Research 187, 1429–1448, 2008.

[7] E.W.T. Ngai, Li Xiu, D.C.K. Chau, Application of data mining techniques in customer relationship management: A literature review and classification, Expert Systems with Applications 36, 2592–2602, 2009.

[8] Parvatiyar, A., & Sheth, J. N. Customer relationship management: Emerging practice, process, and discipline. Journal of Economic & Social Research, 3, 1–34, 2001.

[9] Bâra A., Lungu I., Oprea S. V. - *Public Institutions' Investments with Data Mining Techniques,* Journal WSEAS Transactions on Computers, Volume 8, 2009, ISSN: 1109-2750, http://www.worldses.org/journals/computers/computers-2009.htm

[10] Bâra A., Lungu I., Velicanu M., Oprea S.V. - *Intelligent Systems for Predicting and Analyzing Data in Power Grid Companies*, TheProceedings of the IEEE International Conf. on Information Society (i-Society 2010) London, july 2010.