Data Mining at a Major Bank: Lessons from a Large Marketing Application

Petra Hunziker, Andreas Maier, Alex Nippe, Markus Tresch, Douglas Weers, and Peter Zemp

Credit Suisse
P.O. Box, CH–8070 Zurich
Switzerland

e-mail contact: peter.zemp@credit–suisse.ch

Abstract. This paper summarizes experiences and results of productively using knowledge discovery and data mining technology in a large retail bank. We present data mining as part of a greater effort to develop and deploy an integrated IT-infrastructure for loyalty based customer management, combining data warehousing, and campaign management together with data mining technology. We have completed a first campaign where potential customers were selected using the new built data warehouse together with data mining. Because of the better insight we have used a decision tree as selection method.

1 Introduction

Recent developments of technology, like for example storage management or the Internet, have made it very easy and cheap to collect tera bytes of data and make them on-line accessible in very large databases. However, these valuable assets are still not comprehensively and systematically exploited as part of daily business processes.

The growing competition, the increased speed of business changes and developments of new businesses has dramatically shown the need for knowledge about data and domain-spanning quantitative data analysis. Today, understanding a small detail in data or understanding data faster can make a difference and improve productivity.

This paper summarizes experiences and results at Credit Suisse, a large Swiss retail bank, in developing and deploying knowledge discovery and data mining technology, applications, and solutions. We intentionally don't consider the technology stand-alone (i.e., for the sake of its beauty), but present it as part of a greater effort to develop and deploy in a very short time an integrated IT-infrastructure based on data mining. Notice that we furthermore don't present theoretical experiences with data mining technology, but experiences in actually using it as part of a productive system.

From a business point of view, the general goal of the project is to establish loyalty based customer management (LBM), that is,

- to strengthen customer acquisition by direct marketing and establish multichannel contacts,
- to improve customer development by cross selling and up selling of products, and
- to increase customer retention by behaviour management.

The paper is organized as follows: In Section 2, we present the project architecture and show how data mining fits as one piece of technology into a whole system for marketing campaign management. In Section 3, we discuss experiences in integrating data mining in business processes. Chapter 4 draws conclusions for the future.

1.1 LBM Project and Architecture

From a technical point of view, the goal of the LBM project is to set up an IT infrastructure bringing together data warehousing, data mining, campaign management and online analytical processing (OLAP) technologies. In a first release, the infrastructure will be able to run direct marketing campaigns. The difference to traditional marketing campaigns is that data mining is used for customer selection to find likely targets [2].

The LBM technical architecture is shown in Figure 1. Its major components are operational and external data sources (feeder systems), a comprehensive data staging system (extraction, transformation, cleansing, and integration), the central data repository (warehouse), and the data mining, campaign management, and OLAP systems with their own data stores. It is important to see that the project works only if all pieces - warehouse, mining and campaign - work together.

Logically, the data flow is driven by the target campaign to be launched. Hence, the first step is to define the campaign (1) and identify the data required (2). Data from several operational systems is loaded, including customer, product, transaction, and business structure data, as well as customer profitability data. External data complements the repository with information about credit worthiness.

Getting consistent and high quality data is important, therefore data used by all LBM parts is extracted from operational and external sources (3) and feed into the repository (4). Usually, this extraction will be a repetitive, incremental process. A major task of the cleansing is to create a subject-oriented customer view, which requires de-duplication and merge of data records from the same customer. Cleansing creates added value, for example, information about who shares the same household, which is not contained in the operational systems.

Now the data mining model is built (5) and customer data records are scored. The campaign management system selects data records based on these scores (6) to run the campaign (7). The result and evaluation of the campaign flows back

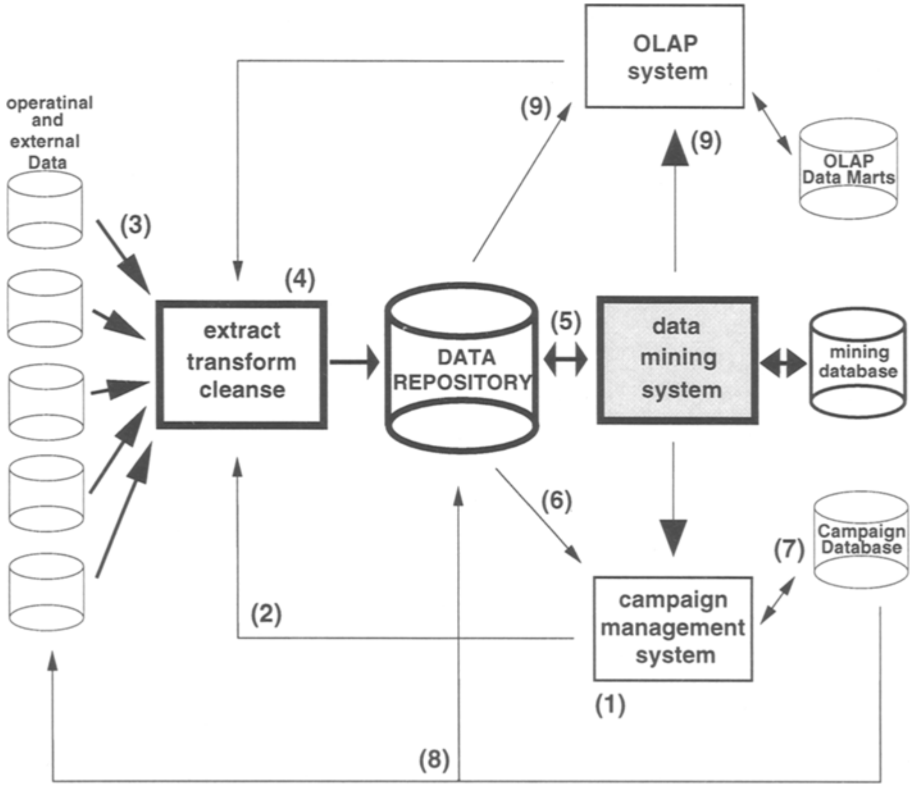

Fig. 1. The overall LBM system architecture

into the data repository and operational systems (8). Alternatively, an OLAP system supports ad-hoc and prepared data analysis and reporting (9) based on scored data.

Notice that this architecture is highly dynamic. The launch of a new campaign or the request of a new analysis may require inclusion of new feeder systems and hence the extension of the data repository.

The LBM infrastructure was set up in a record time of only two months. We defined a pilot campaign as driving force for the system development. In addition to the modeling process the whole environment had to be set up. The infrastructure is based on a Sun Microsystems Enterprise Server 10000 symmetric parallel processor running an 8 CPU license of the data mining suite Darwin from Thinking Machines Corporation. An Oracle 8 relational database is used as the data repository. The campaign management software is Vantage from Prime Response.

2 Data Mining Experiences and Results

The following steps can be distinguished in the mining process:

- data extraction
- construction of target variables
- data set building
- simple statistics, distributions
- initial modeling
- refined modeling.

The first four steps are a major part of the work and can take up to 80% of the whole time used for mining. It is important that one has a lot of different cross checks to be sure about the quality and correctness of the data. The different items are now explained in more detail.

The data is extracted from several tables of the warehouse using SQL statements and joining them into one big flat file to be used by the mining tool. Depending on the business requirements, aggregations of variables are computed. Data preparation turned out to be particularly cumbersome and time consuming, due to the lack of tool support to deal with the 100 GB warehouse data.

The target variable is constructed from the extracted fields. The target typically indicates whether a customer owns a certain product or not (yet). The purpose of data mining is to find a model, predicting potential customers of a product, depending on the information on customers actually having the product.

Four data sets have to be constructed to build the model: three balanced data sets for training, testing and evaluating the model, and one unbalanced set for the validation process. Balancing data sets means having the same number of records for each target value. This is necessary since there is a large discrepancy between targets (customers owning the product) and non–targets (customers not owning the product). The amount of targets can be as low as only 1% of the records. Nevertheless, all balanced data sets must still have reasonable size (more than 10'000 records) to guarantee statistical significance, which obviously requires a huge amount of data.

The analysis starts with simple statistics of single variables (e.g., mean, std. deviation). Analysis is then extended to pairs of variables, in order to detect correlation. One of the variables is typically the target field. There is a tradeoff in the use of variables appearing with a high correlation to the target variable. On the one hand, they are suited for building the mining model. On the other hand, an information leaker might be found that need to be excluded, because it is likely to produce bad models. The only way to solve this problem is a first interaction with business people, discussing the meaning of the high correlation.

Then the modeling process starts. A straightforward way is to build a decision tree with all non information leaking variables. The data mining suite DARWIN uses a parallelized implementation of the CART algorithm [1].

Decision trees have the advantage (in contract for example to artificial neural networks) of being able to show understandable rules as well as getting a quick overview of the quality of the model. Evaluation is required on how well the model selects target fields and how many of them are selected. Good rules show almost no error in classifying the target and are at the same time applicable to a large number of records. An optimum of lift, coverage and error rate has to be chosen. Notice however, that maximum lift decides which model to use and not minimal global error. We used the false positive assessment method in addition.

Decision tree building and evaluation can well be parallelized, which is a big advantage. There are quick turnover times such that multiple alternatives mining models can be built and assessed.

Then, choosing different input parameters and fields refines the model. For example, only a subrange of a field may be used as input to the tree building, due to noisy data or outliers. Other data transformations, for example from continuos into some few discrete values or introducing derived variables, have a high influence on the quality of the data mining.

Subsequent pruning of the decision tree is quite standard. However, the traditional cost complexity pruning method didn't show convincing improvements, because it simply considers the global error rate of a subtree. Alternatively, we are using lift based pruning. This method determines the lift of each node and step–by–step cuts the subtrees of the node with the highest lift. With this method, those rules are optimized that really predict the target. As usual, models are verified with validation data sets.

Due to lack of time, the good understanding of the data needed and the necessity of tractability and explanation of proposed decisions to business experts, no neural network model was deployed to production so far. As soon as one gets more experience with our data and with business requirements we will start using in addition other methods (like neural networks or nearest neighbour) in production models.

During the model building process we have developed around 50 to 60 models. The time consuming factor was not the computing time but the time we needed to understand the results and to discuss the next steps with business experts. Such discussions are crucial because metadata knowledge is often not written down or even available in the warehouse.

For our first campaign we used a data sample of about 1.1×10^6 records (corresponding to about the same number of customers) and around 150 fields. The density of the target was about 13%. After splitting the data into the different sets we obtained a training set of about 80'000 records (40'000 targets). From this roughly 40 field where excluded during data cleaning or we found out during the discussions with business experts that they where information leakers.

In the end we had around 8'000 false positives, these means customers who do not have the product but for which the model predicts that they have the profile of a potential customer.

In the future, mining agents will take over real-time relationship discovery and scoring of data and will produce (email) messages for the customer advisor. For this purpose, mining models will be turned into C++ code, compiled, and deployed to the computers of the customer advisor's desk. Hence, data mining tools need to be open. It must be possible to export scripts and program code, and deploy it to other machines.

For our first campaigns we loaded back the scores into the data warehouse. The scores are based on the accuracy of the model. Finally the campaign was started using the scored records of the customers.

3 Conclusion and Outlook

Time seems to be right for data mining. The technology is available, namely affordable parallel computers, cheap storage technology, and fast pattern finding algorithms. The data is available in huge enterprise-wide data warehouses. The business need is there, as a consequence of increasing competition.

Data mining is not a stand-alone technology, but can be an important piece in many business processes. Hence, a real challenge is to make it work together with other components, like the warehouse, OLAP, and campaign management system. Managing all these dependencies and interfaces, in a short time of only two months, was a real challenge. A large effort was required to extract the data from the warehouse and turn it in a format to be used by the data mining.

Permanent interaction with business case leaders is important. It is crucial that the business supports the idea. They must understand what data mining can do for them and what it can't (the right expectations). The results of the data mining have to be explained to business. In depth discussions and presentation using visualization of the results is a good approach.

One relies on close contacts to the people that acquired the data in the past. Input and feedback from data owners is crucial. They must explain the data and comment for meaningful relationships and information leakers. The data sets we analyzed had hundreds of poorly documented variables from different OLTP systems. Data mining can not be done without a business question in mind. However, the combination of exploratory and confirmative knowledge discovery turned out to be very promising. Whereas business users provide a hypothesis to be confirmed by data mining, data miners can complement by newly discovered relationships. The support of iterative knowledge discovery (model building, model assessment, model refinement, ...) by the data mining tool is crucial.

Data mining tools of today are still too much of a bare-bone technology. More efforts should be made to integrate the technology in a comprehensive knowledge discovery method [3], providing better guidelines to miners.

In this paper, we focused on application of data mining for marketing campaigns, which is one of the most useful and promising applications [4]. However, due to the growing competition in retail banking, at Credit Suisse we see many more potential application areas, like for example risk management, credit fraud detection, or cross-selling of all-finance products [5] (mortgages together with life insurance).

Acknowledgements

We are grateful to M.Beddows and P. Ossadnik for elucidating discussions.

References

1. Breimann L., Friedman J.H., Olshen R.A. and Stone C.J.: Classification and Regression Trees. Monterey, CA, Wadsworth (1984)
2. Berry M. and Linoff G.: Data Mining Techniques for Marketing, Sales and Customer Support. John Wiley & Sons (1997)
3. Fayyad U., Piatetsky–Shapiro G, Smyth P., Uthurusamy R., editors: Advances in Knowledge Discovery and Data Mining. AAAI/MIT Press (1996)
4. Foley J. and Russell J.D.: Mining Your Own Business. Information Week, March 16 (1998)
5. Kietz J.-U., Reimer U. and Staudt M.: Mining Insurance Data at Swiss Life. Proc. 23^{rd} VLDB Conference, Athens, Greece (1997)