

Using reporting and data mining techniques to improve knowledge of subscribers; applications to customer profiling and fraud management

Jean-Louis Amat

Abstract — One of the main issues of operators today is to be able not only to store and manage the huge amount of data generated by the applications and customer contact points, but also to give value to these data. But this implies using tools for storing the data, to manage it, look at it, understand it, exploit it, generate actions such as marketing campaigns. It is therefore obvious that using one tool for each of these functions will lead to a too big and unusable solution. We will discuss here the technical issues involved and show how we turned them into an easy-to-use solution for business users.

Keywords — *data mining, customer profiling, OLAP reporting, fraud management, CRM, real-time marketing.*

1. Introduction

Knowing the subscribers is one of the major issues of today business. It could seem easy to do because we have to deal with a great amount of data sources recording every transaction, preference or behavior pattern. But these data are precisely too numerous and too complex. So if we want business-oriented people to be able to deal with them, it is necessary to provide not only powerful and complex processing methods, but also easy-to-use, and this is where the actual issue is [1].

The data understanding process follows a complex chain in which we have to master every step. It goes from data acquisition to learning and generating knowledge. Here are the main identified steps [2]:

- Data acquisition depends on the contact points with subscribers, where the actual interaction is, at any level: switches (CDR – call detail records), billing, customer care, points of sales, etc. The usable knowledge should be disseminated in all these data sources and the solutions we are talking about here will have to be able to get the valuable data from any of them.
- Data storage: this is a quite technical issue, but very crucial. As we are dealing sometimes with millions of subscribers, over time periods ranging from a few months to a few years, important data storage means are needed.
- Data management: this is where the system intelligence should start, allowing users to access easily to what they want, the way they want. Market-

ing databases today are very often structured around a data mart, using a very convenient representation for automatic requesting.

- Data processing: dealing with mathematical functions able to correlate data, to discover patterns, to compute trends or to predict them. We are in the field of data analysis and statistics, but also data visualization (reporting) which is very important in the process of knowledge discovery.
- Data understanding: getting knowledge from data is the part for what we call data mining. It is a complex process using any of the results obtained in the preceding steps, sometimes requiring new processing, in order to understand deeply the data, to get usable and hopefully unknown knowledge.
- Learning: the last part of the process is the most important one and has to remember what was useful in the data/information/knowledge about the subscriber in order to feed future analysis and understanding process. It is obvious today that this part is not present in most of the existing tools but they will have to deal with this in the future.

We are presenting here two solutions we built in order to apply these techniques to fraud management, and to marketing automation (customer analysis and marketing campaigns management).

2. Getting decisions from data

Usable information that will give knowledge about one subscribers is contained in the data we have about his transactions and history. But it is hopeless to try to use every data available; some choices have to be made depending on the final purpose of the application. If we are interested in churn management, we will have to select relevant data to understand the reasons why people are leaving; this is one of the most complicated phenomena to analyze in this field and the solution should be completely specific. But basically, the behavioral data is very important to look at, as well as demographic and financial information about the subscribers. The profiling analysis process is very complex because whatever the data we can put in the analysis,

we can almost always discover patterns. But are these patterns relevant? This is the main concern one has to deal with. For the fraud analysis, it is well known that the calling patterns are often significant in this respect, so a fraud management tool will have to cope with these data, which means analysis of the CDR.

Customer profiling and behavior analysis often rely on the same data sources, present in the telco world. The mostly used data sources are for example the billing system and the customer care.

- **Billing:** this is where the usage information is stored, as well as some personal information about the subscribers. Depending on the billing implementation, this information can be very accurate, or more general. A good recommendation would be to think of the further profiling application when designing the system.
- **Customer care:** data is stored also about customer information, interactions with the operator, and mostly short term data history.

There are several ways to use the data and to provide operational results for marketing purpose or fraud analysis, depending on the users. Back office people need solutions to get quickly and easily reports about their favorite subjects, or deeper analysis of the data/information, while front office management is more concerned with managing customer interaction. This leads to separate functions:

- **Reporting and analysis.** This kind of analysis relies on quite simple tools and solutions, and is mainly based on OLAP modeling (on-line analytical processing). The users are decision makers such as general management, marketing manager, etc. We are actually in a decision support process, but where the intelligent part is completely managed by the user. The only intelligence within the system is the way data has been modeled with multidimensional representation. Nevertheless, this kind of solution is very useful to broadcast automatic reports to decision makers. Alcatel CMI (customer management intelligence) solution includes these features as a first analysis layer.
- **Profiling and segmentation.** The output here is not only information (high values, trends) but also models for prediction or classification. The goal of this layer is to provide more than information, i.e. knowledge of the subscribers. The techniques involved are therefore mainly based on artificial intelligence, and we are using decision trees to generate explicit knowledge (rules), neural networks or Bayesian techniques to model the subscribers buying patterns. Usually, the main issue with these techniques is that users need to understand what they are doing in order to get valuable knowledge; but some algorithms can be predefined with standard parameters in order to provide a global result. Alcatel solutions are in this re-

spect among the most easy to use; CMI solution offers “click and play” data mining features fully integrated, while FMS (fraud management system) uses neural networks in a totally transparent way to generate alarms and help decision.

- **Real-time decision making.** This is the original part of Alcatel offer, relying on E.piphany software. The real-time platform is able, during an interaction with a subscriber, to compute a decision profile in seconds and to push an offer through the selected channel (call center, web, chat, etc), without human intervention. In this case the decision (selecting and proposing an offer) can be taken by the system itself (on a web site for example) or by an agent who can select among several propositions (call center). This intelligent layer includes an automatic real-time learning engine able to make profit of any transaction and to update itself the predictive model. The solution is a combination of rule-based system, self-learning analytics (relying on Bayesian techniques) and collaborative filtering.

The Alcatel CMI solution [3] includes the three presented layers with the first and second ones working for back office people and the last one deployed in front office, facing the subscriber (through an agent on a call center, or directly on the web). The fraud management tool FMS contains also reporting and analysis facilities devoted to fraud alarms.

3. Segmentation, classification and profiling

The functions we described rely on a set of techniques that are able to assist the users in their tasks [4]:

- Segmentation is the process to find classes in the data (Fig. 1).

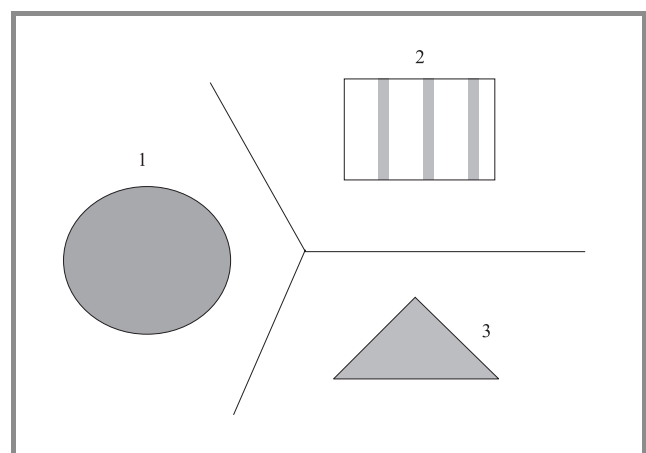


Fig. 1. Classes in the data.

- Classification assumes that a segmentation already exists and tries to attach an element to an existing class or category (Fig. 2).

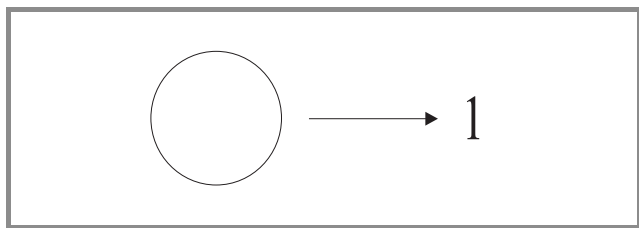


Fig. 2. Classification.

- Profiling consists in describing the elements of a class (Fig. 3).

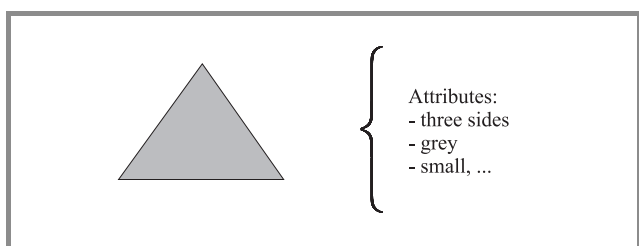


Fig. 3. Profiling.

The main functions to perform when trying to find knowledge about the subscribers often fall into one of these categories. Generally, the process consists in applying a segmentation process to the data and then to classify new data, or to understand the content of the classes. For the example of churn management, we can identify segments in the data representing the different cases of churn. Then we can use profiling techniques to understand each category.

The CMI solution provided by Alcatel contains functions for performing segmentation or profiling, but also predictive models. We give here a short list of the main functions provided [5]:

- **Profiles** is a charting, visualization and description application that enables users to inspect and understand their data visually. Profiles can predict for example how customer segments are varying over time, how product sales vary by region, what the most common profession is for each division, etc.
- **Basic trends** provides time series trending analysis capabilities including straight line growth, constant percentage growth, or moving average trends, including extrapolations of one, two or three periods.
- **High and low clusters** is an anomaly detection method that allows to find the highest- and lowest-performing groups within an attribute or set of attributes according to some measurement. It is very helpful to answer questions such as finding the best

and worst customers (the customers who buy much more than expected are probably the best), finding product lines where current year revenue is surprisingly higher than last.

- **Clustering** is a segmentation tool that can be used for customer segmentation, or finding outliers: sometimes the most useful clusters are the smallest ones. These “nuggets” might represent a unique niche or highly profitable (or unprofitable) customers, for example.
- **Influence** is a classification and regression component that is used for two main purposes: first, to find which of the input attributes have the most power to predict the target, and second, to build predictive classification and regression models. Once created, the models can be used to score lists of customers, for example based on the likelihood a customer will respond to a particular campaign.
- **Bayes classifier** allows users to use Naïve or optimal dependency tree (ODT) Bayes classifiers to create classification model, for example helping users to identify characteristics of profitable and unprofitable product lines, select new sales prospects based on the buying patterns of current customers, and perform other types of predictive analyses.
- **Scoring** allows users to use models built in **influence, clustering, and Bayes classifier**, or use a pre-defined measure, to rank customers in a list and to target marketing communications where they can be most effective. For example, suppose a user was concerned about high rates of customer attrition. That user could use **influence or Bayes classifier** to build a model to predict which customers were most likely to defect. Then, with the **scoring** application, the user could generate a scored list that ranked customers based on attrition likelihood, and select only the highest risks, in order to address them a specific offer.

4. Fraud management

In a normal day of activity, a telephone company has the potential of creating many hundreds of thousands or even millions of call records. Within this mass of data are calls being made by people who either are targeting the organisation with the aim of defrauding it or people potentially using the service with no intention to pay. And the only way to detect quickly these people is to monitor their usage which is contained in the data calls (CDR). The system Alcatel built for this purpose gets the CDR directly and provide decision support for fraud managers with the help of a rule-based system containing explicit knowledge, call query facilities, usage variation analysis (neural networks application) and also subscriber fingerprinting analysis to

help recognizing fraudsters through their calling habits [6] (Fig. 4).

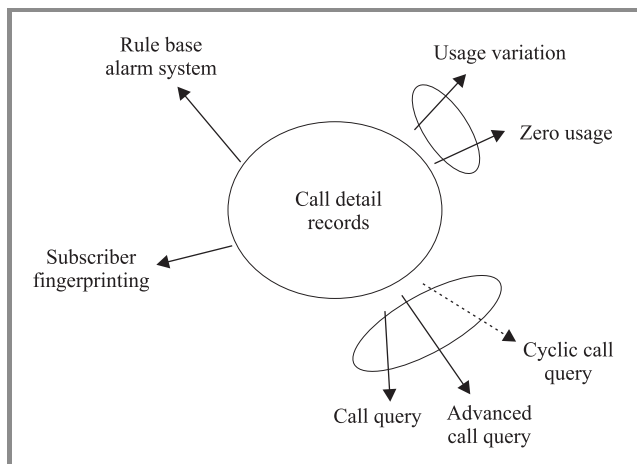


Fig. 4. Various usage of CDR.

When the fraud analysis process commences, the raw call records are sent through a data interface programme for conversion into FMS CDR formats. The system calculates the call duration for each completed call record received, rates the call using an internal tariff table (if CDR are delivered unrated) and merges newly received CDR with existing CDR. When all new CDR have been processed, the system compares the activity per subscriber against triggers established in the system and generates an alarm report each time the set threshold is exceeded by any of the prescribed amounts.

4.1. Call query

The data held within the FMS is very high volume; many million records per day can be easily managed by the system. The call query module allows simple interrogation of that data. By entering one or more search criteria, the user can mine for data. The resulting analysis itself could be many thousands of records or it could, if the search criteria were precise enough, bring back a single record.

4.2. Velocity checking

When two calls are made from two different locations and with a time difference greater than the minimum time required to travel from one to the other location, this optional FMS feature is able to identify and display the relevant CDR.

4.3. Usage variation

Fraud is often detected in its earliest forms by the change in usage of a subscriber. Very little fraud of any substance can be undertaken without usage accelerating on a subscriber's account. The usage variation module of the FMS can assist the user to detect major changes in usage very quickly.

The FMS usage variation module is implemented as a feed-forward neural network, fed by input vectors which represent values of parameters assumed by each individual subscriber in the previous 24 hours. The input elements of the vectors are weighted to produce a scalar output value that represents the status of the subscriber for the given day. Over a certain period of time, a reference output value is stored into an alarm reference file, this reference being the maximum output value assumed by the subscriber. The system is thus capable of generating an alarm whenever a subscriber has an output value that exceeds the relevant stored maximum.

4.4. Subscriber fingerprinting

This optional functionality has been introduced in order to address the issue of subscription fraud. In this type of fraud, individuals would subscribe with a telecom operator under false identification. Bills, possibly after a first "quiet" period of normal behaviour, are eventually not paid. After disconnection, they again register using different identification. A possible solution to this problem lies with the assumption that the fraudster should have a specific pattern, like a **signature**, which is defined by some of the numbers called. These numbers could be the majority of his/her calls, or just a few calls, but with a certain degree of **uniqueness**, which might permit to identify the fraudster. **Fingerprints** (identifying patterns), stored in an internal database, made of sequences of called numbers from specified accounts are matched against new account traffic.

5. Customer analysis

Using raw data such as CDR with some explicit knowledge can help to perform powerful fraud management. But for marketing purposes, it is needed to analyze more sophisticated data, running from billing information to clickstream analysis over a web site. In order to get these data and to understand them, the CMI solution is able to store it into a data mart and to provide a fully integrated interface to activate together reporting, data mining, and active marketing campaign management.

For example, suppose a marketing user is concerned about high rates of churn. That user could use **influence** to build a model using historical data to predict which customers were most likely to defect to a competitor. Then, with the scoring application, the user could generate a scored list that ranked customers based on how likely each were to defect. Finally, the user could import that scored list into the campaign manager module to direct a special marketing incentive at the most likely victims of attrition. Within CMI solution, every described operation is performed with the same tool, giving the facilities to switch from OLAP to data mining, then to campaign management, and back to OLAP if needed (Fig. 5).

For performing churn analysis, for example, the basic idea is to detect in a first time the people that has churned in the

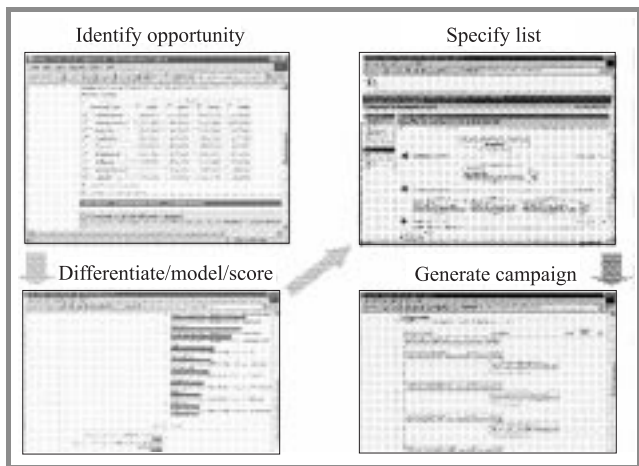


Fig. 5. The process of generating list of clients for marketing campaign.

past, within a specific time range (to be defined). Regarding this, the system helps to extract a list of “churners” from the customer database, which will be used to feed the learning functions. The data mining tools can then be applied on this population, in order to assess two models.

The first model computed is a **scoring model**. The model will compute from existing/known churners a way to assign a score to each subscriber, which will depend on selected attributes. Once established, this scoring model will be applied on active subscribers and will therefore assess a propensity to churn. Sorting the list of subscribers with this specific measure will give the most probable churners. It will then be possible to activate a specific retention campaign in order to keep these subscribers.

The second model is a **segmentation model**.

In fact, it might be interesting, before acting on the global list of potential churners, to detect the most profitable ones, and to estimate the global revenue regarding the cost of the campaign. The segmentation model will provide a profiling of the churn population, which will have to be interpreted by the analysts (Fig. 6).

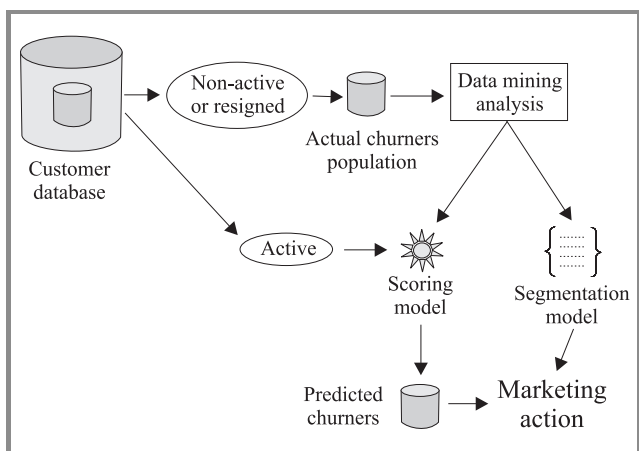


Fig. 6. Segmentation model.

6. Conclusion

Using a combination of the functions described, we are able to propose a fully integrated and operational solution. The decision support part is also integrated with marketing automation and CRM (customer relationship management). This allows to get a unique view of the subscribers, through any interaction channel (Fig. 7).

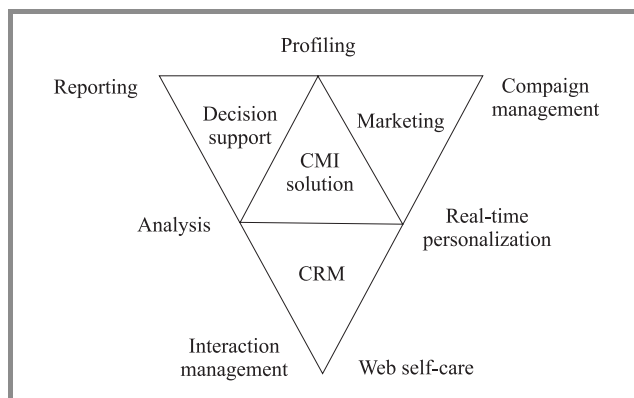


Fig. 7. Integrated solution.

Dealing with complex and heterogeneous data sources is an actual issue in order to perform tasks such as fraud analysis or customer profiling in order to activate marketing campaign.

Alcatel solutions we presented are able to provide already developed interfaces to the standard data sources, and integrated functions for powerful analysis to support decision for fraud detection, customer data management and analysis, and customer relationship management [7].

References

- [1] A. Berson, S. Smith, and K. Threaling, *Building Data Mining Applications for CRM*. McGraw Hill, 2000.
- [2] J. L. Amat and G. Yahiaoui, “Advanced techniques for information processing: neural networks, fuzzy logic, genetic algorithms”, Cepadues Eds., 1995 (in French).
- [3] Alcatel, “Customer management intelligence solution”, 2000.
- [4] J. L. Amat, “Applying customer profiling and segmentation methods”, in *IIR Sem. Customer Profiling & Segmentation for Telcos*, London, July 2000.
- [5] E.piphany, “E.4 Datamining overview”, 1999.
- [6] Alcatel, “Alma FMS: fraud management system”, 1999, 2000.
- [7] J. L. Amat, “Latest CRM and eCRM trends for telcos”, in *IIR Sem. eCRM*, London, Feb. 2001.

Jean-Louis Amat
 e-mail: Jean-Louis.Amat@alcatel.fr
 Network Applications Division
 Alcatel
 1, rue Ampere
 91302 Massy cedex, France