Integration of Analytical CRM in Business Processes: *An Application*

Tom Breur, ING Bank (Netherlands)

Abstract

The business value of data mining technology is widely accepted these days. Data mining has proven to hold a lot of promise in Customer Relationship Management (CRM). Deploying data mining models in multi-channel, large-scale organizations is quite complex. Often, deployment turns out to be even more challenging than predicting customer behavior itself. In an effort to manage the complexities of organizational processes, we implemented an application at ING Bank labeled **"Process Automation Center" (PAC).** We feel that CRM strategy has to do more with getting processes right than the technical issue of implementing a tool or technique. The main function of our PAC is to deal with deployment issues in analytical CRM. To these ends, we determined critical inputs and outputs in business processes and carefully filtered the essential and meaningful triggers in the process flow. As a metaphor, PAC performs the function of a CRM "nerve center". It interprets inputs and translates them into meaningful outputs. The PAC distributes all essential, yet only the minimal subset of information to parties involved and it can also execute a wide range of actions needed to put together a mature CRM strategy.

Introduction

The use of data mining models is becoming more and more widespread. In the field of CRM, data mining is turning from specialized technology into mainstream practice. Analytical CRM has proven to be the killer application for data mining. Probably the most important challenge for unleashing the power of data mining technology is seamless integration with business processes.

When data mining models are used on a regular basis, issues of maintenance and meta data become increasingly important. Sometimes a distinction is made between analytical versus operational CRM. **Analytical CRM** has to do with modeling, campaign management, and long-term decisions on customer development strategies. **Operational CRM** deals with execution of customer contact strategy.

It might be interesting to predict customer behavior, but only when one manages to act upon these insights does this knowledge become commercially relevant. Prediction and analysis is important, but only deployment will make this useful. Exactly this is where the most difficult challenge seems to lie at the moment. When deployment results are consistently fed back, the organization will learn from its past actions and truly adapt to customer needs as displayed by response. We make the assumption here that one may infer relevance of the offer to the customer from response.

In large organizations, specialties of analysis versus execution are likely to become dispersed -- even more so as the level of sophistication in modeling advances. The interface between operational and analytical CRM relies heavily on good meta data for using and reusing data mining model score code. Also, the ability to evaluate and monitor models in a time efficient and error-free manner will lead to improved targeting. As an invaluable side effect of this, more knowledge on customer behavior will be gained. This customer knowledge is considered by many to be the most important asset in competitive markets these days.

2. CRM Processes

The use of data mining models to achieve improved target marketing is becoming more and more commonplace. Some examples where data mining is effectively used in analytical CRM are:

- Targeting direct mail offers for acquisition, cross-sell, deep-sell, up-sell, retention
- Analytically deriving cross-sell suggestions for call-center agents, e.g., proposing which offer to make for which customer
- Credit scoring: who to grant credit, and how much. Also early warning systems for potential defaulters
- And many more.

2.1 Interfaces in the Direct Mail Process

PAC streamlines and automates business processes. As an illustration of the interfaces in the CRM process, the schema in Figure 1 shows the steps on the left hand side, with primary responsible parties on the right. To manage all these interfaces, PAC operates as a linking pin between operational and analytical CRM. In this paper, we will illustrate the steps in this process by running through a direct mail campaign.

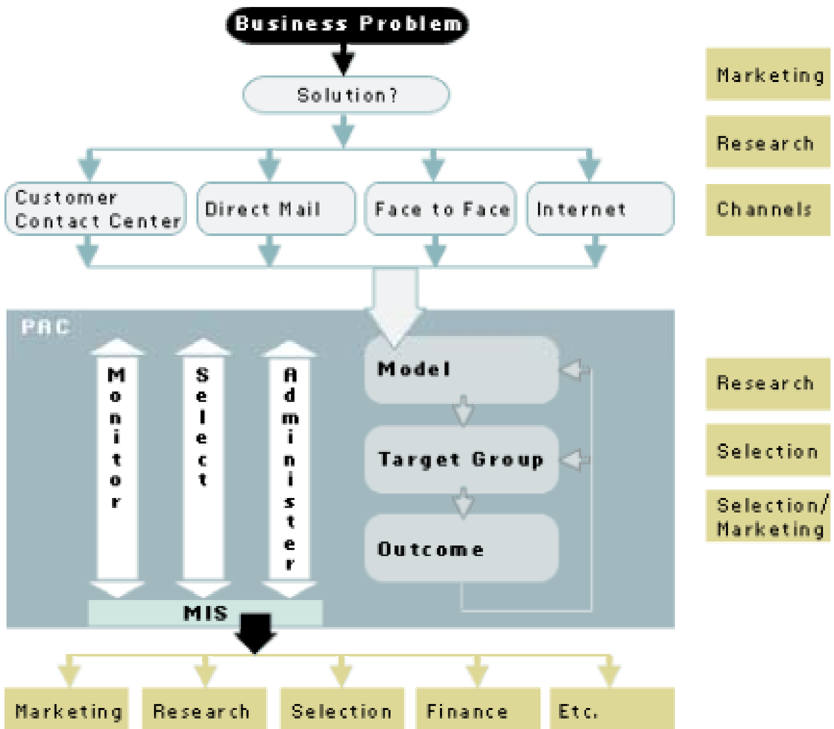

Figure 1. Marketing processes with the PAC

Producing selection code for direct mail can be quite hard.

First of all, a base population at which an offer will be aimed needs to be specified. In our company (a retail bank), targeting may be aimed at the account, person, or household level, each requiring different deduplication processes.

Then, some segments may need to be excluded on the basis of (potentially very complex) business rules.

Thirdly, considerations from the account manager or branch office about excluding certain customers from an offer are gathered and maintained in a central database.

The PAC environment is a crossroad of connected parties, involved in either operational or analytical CRM. As such it performs the following functions:

- Administration of data mining models that are in use
- Selection for direct mail campaigns, possibly using data mining models
- Evaluation of direct mail campaigns
- Evaluation of the effectiveness of data mining models
- Automatic reporting of steps in the process flow and results along the way.

PAC has a number of interfaces:

1. There's an interface with the department that develops data mining models
2. An interface with the selection department
3. There's an interface between PAC and marketing.

The PAC application requires input of only the essential parameters to be controlled by all parties. When we take up the example of running a direct mail campaign, there's only a limited set of parameters that need to be controlled by the marketer. These consist of:

- Determining to whom the offer might in principle be applicable (choice of base population)
- Deciding on the mail depth, preferably on the basis of ROI calculations (choice of target group) **(1)**
- Deciding when the campaign is going to be run.

So marketing decides when to run the campaign, and who to make the offer. They really don't want to be bothered by the impact this has on the choice of database tables, or other strictly technical process issues. The inputs are based on how the campaign should be run, and how data mining models are developed. PAC generates the following outputs.

1. At each step in the process, reports are generated with all (and only!) the essential decisive parameters for each party involved
2. Score code is being generated for the production database, as well as the accompanying meta data.

All essential files in this process are automatically written to and organized in a dedicated directory structure.

As a result of this, PAC "knows" how to automatically generate all the code needed for subsequent evaluation and monitoring of model performance because it calls upon files that PAC itself has written to designated directories. Therefore, the tasks of selecting and evaluating campaigns now become error-free and highly automated (time efficient).

All this needs to be taken into account at selection time. Other businesses will have their own idiosyncrasies. Together, these factors can make the needed selection code very hard (and time consuming) to write, difficult to check and therefore error prone. It also relies heavily on complex database structures.

One of the triggers for this project was the desire to be able to automatically translate data mining models in database score code. Except for the score code, our PAC also provides the accompanying meta data (broadly defined). As such it does what PMML (Predictive Modeling Mark-up Language) might do, but the primary function really is to integrate data mining models in business processes. The purpose of automatic database scoring is to ensure the validity of models being applied, and to support evaluating results and subsequent monitoring when models are being reused.

(1) Size/depth of the target group should ideally be based on cost/benefit considerations. But if expected revenue from the product is not available, one can calculate the implied expected revenue of the lowest propensity customer included in the target group just from the marginal cost.

2.2 Parties Involved and Responsibilities

At ING Bank, a central marketing department is responsible for the profitability of the customers. In this respect, they "own" the problem of making customers more profitable, of devising and executing customer contact strategies that will develop the customer relationships in the best possible way. On the lower end of the customer base, this will have to do with evoking more cost efficient ways of transacting; on the higher end, this typically involves cross-sell and deep-sell.

The research department is responsible for identifying business facts that point to opportunities for customer development. A strategy can then be put together for migrating customers to more profitable segments. Data mining models can be used to optimize targeting for this. Our bank has a number of channels through which we may interact with customers:

- The branch office
- Mobile sales teams
- Our Customer Contact Center (phone/fax/e-mail)
- Internet
- Direct mail.

Each channel tries to optimize its workings.

Within the database marketing department, we have a number of sub-departments. Besides the research group there's a team to perform database selections, and a group of direct distribution managers who closely interact between marketing, communication, selection, research, and at the same time manage the fulfillment process.

A number of roles can be identified to make best use of research capabilities. [4] These roles may sometimes be combined within one person:

- Communicator
- Data archaeologist
- Data programmer
- Statistician
- Technologist.

2.3 Direct Mail Selections

Currently, the following process is in place when data mining models are used for targeting direct mail. The base target population for a certain offer is determined. Basically, this comes down to defining all possible prospects within the database for whom the offer could possibly be relevant. At the outset, the business will have an estimate of the size of this population.

As an example of how the signal function of PAC works, one of the first signals in the selection process could be a surprisingly different size of the base population from what was expected. A trigger goes off, and this discrepancy will be signaled back to marketing. They can now decide whether to proceed anyway, or whether to make adjustments first. After the base population is determined, a random mailing will be sent out to this target audience. When response to this pilot campaign is gathered, a predictive model can be built. **(2)**

Next, all customers in the target population are scored using the model. To begin with, a random selection from the entire target population is made. This group will receive the intended offer ("random group"). Then the "target group" is selected. By "target group" we mean the group that, according to the model, has the highest propensity to respond (segments 1 & 2 in Figure 2). From this target group then, another random subsample is drawn (which we label the "reference group") that will be *excluded* from the offer. This way the reference group has the exact same properties as the target group.

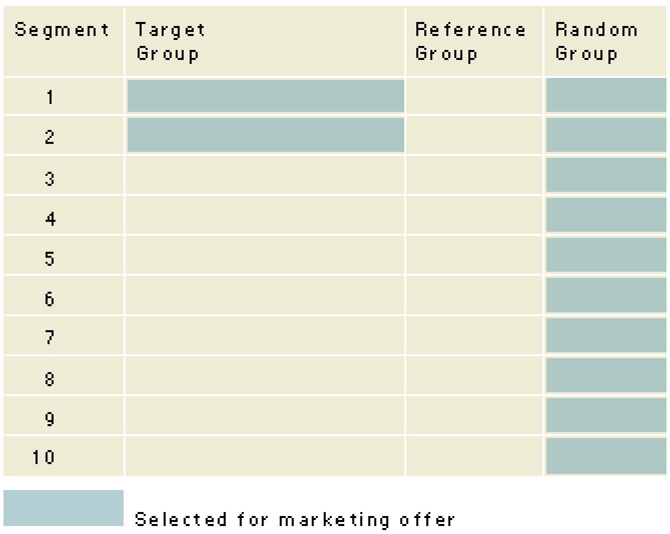

Figure 2. Direct mail groups

There are two reasons why selections are made in this way.

1. Comparing the overall response rate in the random group with the target group will give an empirical estimate of the lift that is achieved using the model. Also, post hoc overlaying the model over the random group will give an evaluation of which regions of the problem space the model worked as expected, and where possibly it didn't.

2. Comparing the target group with the reference group shows the extent to which the marketing offer was successful in eliciting response. This way one can empirically test the effects of the model versus the marketing offer in terms of raising response.

The ordering in this selection process (first determining random, then target group) could be reversed in theory. However, that would be impractical because eventually the same customers will be mailed, and it would only make the model monitoring tests computationally more complex, and therefore unnecessarily cumbersome. Both the marketing offer and the data mining model can cause an increase in response. This distinction is not trivial. In some cases, one may be very well able to predict which customers are likely to respond.

(2) Alternative procedures for model development are in place when it's not feasible to obtain a random sample of responders first. More on this topic in [2]

However, only when response rates between target and reference group differ substantially will the marketing effort be worthwhile. So when the response rate in the target group is much higher than the random group, one might be able to predict autonomous response quite well. Even though the lift of the model may be very good, the net added value of the campaign is poor. Without this reference group test, the campaign would falsely be deemed a success, but of course only triggered (as opposed to autonomous) response is worth spending marketing dollars on.

When the groups in Figure 2 are selected, all necessary tests can be performed. Group sizes should be determined by confronting business interests with statistical power.[3] The trade off here is that smaller groups will provide a more detailed description of how response percentages decay as a function of customer characteristics. But the price for this more detailed description comes in the form of a larger required random sample in order to be able to find statistically significant results. Since the overall response rate in the random group will be lower than the target group, this knowledge comes at a price. When reusing models one can track over time whether the model keeps performing as expected. An excellent treatment on the effectiveness measures of models can be found in [7]. All groups selected are essential to this end.

Imagine the following situation: from one campaign to the next, the response in the target group drops. Without a random group, there is no way to determine whether to attribute this drop in response to deterioration of the model, or whether the product has just gotten less attractive for the target population. Comparing random response rates in subsequent campaigns will provide the answer here.

3. PAC FUNCTIONS

3.1 PAC as an Expert System

As people and data marts become dispersed across the organization, it's increasingly important that communication between departments is ensured. For example, there are two vital flows of communication between the selection (operational CRM) and research department (analytical CRM):

- One direction of the communication contains the meta data going along with each model that is provided by researchers to the selection department
- In the opposite direction, the selection department provides feedback from the monitoring of models.

Empirical results on the stability of data mining models over time is extremely valuable for improvement on the models (learning how to develop them) as well as the data (e.g. derived variables). The selection department might signal irregularities, and this will call for a second look at the performance of the model to see whether it is (still) valid. Such irregularities might have to do with sample fluctuations or population drift. Sometimes they may just be the result of erroneous selections or faulty data. These issues can sometimes be delicate and require considerable statistical expertise for making judgement calls.

An important issue in the deployment process of data mining models is that different parties have wildly different information requirements. The full breadth of information that is available in the entire process can hardly be comprehended by any single person involved. The model builders can get "lost" with regards to database administration issues in the selection. People on the selection department might experience an information glut when confronted with all the statistical measures available in the model building process. Marketers only need and want to worry about issues relevant to them. They should not be bothered with database administration issues, nor with statistical information about data mining models. Instead, they should really only need to care about what the used data mining models will bring them, and what the consequences of particular choices are. The data warehouse people need to be involved, but should only be provided with information relevant to them like usage of tables and fields, I/O, response times, etc. So the common denominator is that:

- One needs to determine exactly which minimal subset of information is the critical decisive information, relevant for whom
- Information should be presented in a way that is easily comprehendible, as an "image" that is meaningful to the relevant user to decide and or act upon.

In this respect, PAC plays the role of an expert system. A collection of inputs are combined and subsequently lead to a diagnosis. Which diagnoses should be performed, and how they should be reached was derived from interviews with specialists in each field.

As an example, let's consider segment stability at scoring time, using a data mining technique that appoints discrete groups. When the model is applied to the production database, the relative size of the groups can vary, for a number of reasons. The person performing selections need not be a statistical expert to evaluate this. What is needed is a diagnostic tool. In this case, our diagnostic tool is based on statistical considerations. Is the data mining set (still) representative for the population that the model is being applied on [6]? As an answer to this question PAC comes up with one of three possibilities:

1. It's OK proceed
2. It's not OK: stop the process
3. Undecided: contact a statistician to check how serious the problem is.

3.2 Aligning Organizational Goals

Another challenge that integrating data mining in business processes brings is the difference in "culture" that exists between departments. For example, in the research group we take at least days, or possibly weeks to develop models. The eventual selections are prone to be performed in a very short time span: hours, or maybe a day or two at the most, including all tests to see if the selection came out right. The environment in which professionals perform selections should accommodate speedy operations. In particular with regard to capturing meta data the infrastructure should facilitate this in a user friendly manner. But the same goes for support to call center agents or field sales force. CRM applications should support, not hinder service and sales force in direct contact with customers [5]. And of course this same principle applies to data analysts who should feel supported, not hindered by technology.

For operational and analytical CRM alike, rather than rely on discipline, one prefers convenience. In direct mail, professionals performing selections work under constant time pressure. And the same goes for front-office staff in direct contact with customers. Operational CRM applications should support this contact, and no extra effort should be required in order to capture the data that describe the outcome of the interaction with the customer. These contact data need to be fed back into the database. Any system that calls for additional, voluntary effort of sales or service staff is prone to incomplete and faulty input from front-office staff.

All this has to do with aligning organizational goals: an interface that is easy to use for the selection department will decrease time pressure. Ideally, this interface will make selections easier to perform *and* at the same time provide all the wanted meta data. Rather than rely on discipline for inputting meta data, one wants to drive this from within the process by making an interface available that both helps in making selections and at the same time automatically captures and provides all the meta data. PAC sort of "reads this off the hands" of the selector as they punch in keystrokes. Voluntary input is brought back to the absolute minimum. The Process Automation Center helps to:

- Speed up the process of making selections
- Deal with the complexity factor of generating the code
- Make the selection process less dependent on "human factors" for producing meta data

Facilitate timely evaluation of campaigns and the models being used.
We set out to produce a user-friendly environment whose interface should only demand the absolute minimum number of input parameters, and that would return three different sorts of output:

1. Immediate reporting, at each stage, of all results as one advances through subsequent steps in the direct marketing process
2. All desirable meta data for all parties involved
3. Both the selection code as well as the evaluation and monitoring code.

This way we are guaranteed that meta data become centrally available as soon as the model is taken in production. Also, the administration of all the used databases that are called upon within the selection process are controlled by the interface. In the case of the direct mail selection for instance, PAC "remembers" (by administering the necessary tables for this) which customers fell in the random and reference group. Because of this, the code to evaluate campaigns and monitor performance of (potentially reused) models over time is also immediately available and requires no extra effort.

Given proper design and maintenance of the database, changes in the data should have no impact. For instance, when a new attribute value occurs because of the introduction of a new product, this needs to be changed in only one reference table. To this end, we made adjustments to the database architecture for easier maintenance. If changes in database tables take place, like new products or different definitions, this has no impact on PAC. This also safeguards that selection and evaluation code is produced error-free, and at a minimal expense of time and effort.

3.3 PAC and the Nerve Center Metaphor

To illustrate the workings of our PAC we will use the metaphor of a nerve center. The brain receives continuous input from the environment. All these inputs need to be interpreted and processed, and sometimes acted upon. In the brain, a set of inputs (like neurons firing on the retina) are translated and then projected as a meaningful picture. In the same way, PAC will receive and combine a set of inputs (indicators on the state of running processes), and then recombine these inputs to perform a diagnosis. This resembles the human eye looking at a situation and judging it.

While transmitting inputs to points of execution along the process, a number of process checks and execution tests are performed and reported to parties involved. In a similar vein, the brain can process an image of its environment, and while interpreting this image, display reflexes or decide to voluntarily act upon it.

In the brain, the entire range exists from complex translating and communication processes, up to autonomous reflexes requiring little voluntary action. Similarly, PAC may combine and interpret sets of input parameters, and come up with either a diagnosis or a comprehensive report of the situation. This report will then automatically be offered to the person required to make a decision. But sometimes the inputs will only call for a predetermined immediate action. Like the nerve center, it may also send out a signal to invoke an action, as in the case of the knee-jerk reflex when tapped below the kneecap.

Along these same lines the nerve center "learns" from its environment, and this learning is displayed by adapting to slowly changing environmental conditions. Likewise, PAC can help to adapt to customer needs, displayed by progressively focused targeting.

In PAC, automated processes can serve as building blocks to assemble customer contact strategies. This resembles the way a coping strategy is made up of a set of skills one masters.

The analogy can go on and on. Only when complex routines become automated and more or less straightforward can one begin to master increasingly difficult tasks. After the basics are mastered and processed "subconsciously", one can focus on the next challenges just over the horizon.

4. PROCESS AUTOMATION WITH PAC

4.1 Key Success Factors

Very important for the success of this project was close cooperation between the selection and research department, database administrators and direct distribution managers. All interests need to be aligned. A comprehensive process description was available at the outset of the project. This process description, with all steps written out in detail, used to serve as a guideline to be followed by all parties involved. It proved invaluable as input for process analysis and as guideline for the programmers.

It turned out that in practice some of the steps in this process would sometimes be skipped, mainly due to time pressure. Notably, capturing of meta data and timely evaluation were aspects that suffered when time pressure was highest. However, these problems did drive the PAC development process forward.

Some key success factors for making an application like PAC work are:

- One needs to master the process, and it should be sufficiently standardized
- Connectivity of software, together with an open architecture so that data models can easily "communicate".

When business processes are not standardized yet, it would be too constraining for the organization to be forced within any format. Streamlining of processes implies they are performed in only one or a limited number of ways. Otherwise automation is useless. Connectivity is important, but in practice was rarely a problem. All our common applications could easily interface, at least technically. Logical issues can be a bit more challenging, in particular when data models are not transparent.

4.2 General Process Automation

This article only deals with integrating data mining models in direct mail. Of course one need not be confined to direct mail, nor to direct marketing in general. Our PAC is not even confined to marketing processes. With PAC, we control many more processes, like scheduling monthly database updates and generating data sets for model building (a.k.a. RME [1]).

PAC can also be used to integrate data mining in business processes involving face-to-face contact. Branch office and sales force contact can be included in this methodology. However, multi-channel campaigns are far more complex to deploy.

For instance, sales force automation software (like laptops with software for Relationship Management) can be asynchronically updated with sales opportunity triggers that were derived through data mining.

Or at a higher level, one of the purposes of segmentation is to cluster groups of customers on the basis of presumed needs and development targets. At an aggregate level, the success of attaining segment goals can be tracked within segments. These outcomes serve as the basis for deciding to migrate customers between segments.

The building blocks of marketing treatments typically consist of individual campaigns. By making these building blocks relatively easy to manage, one clears the road for more elaborate testing and complex customer development strategies. To advance one's CRM capabilities, standard campaigns should be running smoothly, requiring as little attention as possible for non-crucial issues. Only then can we begin to deliver on the promises of CRM. In a way this resembles a nervous system adapting to its environment, learning to perform increasingly difficult tasks.

References

[1] Berry, M & Linoff, G. (2000). **Mastering Data Mining: the Art and Science of Customer Relationship Management**. New York, NY: John Wiley & Sons. [2] Breur, T. and Paas, L. (2000). "Building Predictive Models for Direct Mail: a Framework for Choosing Training and Test Data". Journal of Database Marketing, Vol. 8 (1), 9-16. [3] Cohen, J. (1977). **Statistical Power Analysis for the Behavioral Sciences (revised edition)**. New York, NY: Academic Press. [4] Furness, P (2001). "Masterclass Descriptive and Predictive Customer Modelling", March 27th 2001. [5] Lee, D. (2000). **The Customer Relationship Management Survival Guide**. St Paul, MN: HYM Press. [6] Pyle, D. (1999). **Data Preparation for Data Mining**. San Francisco, CA: Morgan Kaufman. [7] Rosset, S. et al (2001). "Evaluation of Prediction for Marketing Campaigns". In: **Proceedings of KDD-01**, pp. 456-461. Menlo Park, CA: AAAI Press.