CHAPTER 17

Scoring Your Customers

Introduction

coring Process

After a predictive model has been created by using data mining software, the model can then be used to make predictions for new data. The process of using the output of the data mining (the model) is separate from the process that creates the model. Typically, a model is used multiple times after it is created to score different databases. For example, consider a model that has been created to predict the probability that a customer will purchase something from a catalog if it is sent to them. The model would be built by using historical data from customers and prospects that were sent catalogs, as well as information about what they bought (if anything) from the catalogs. During the model-building process, the data mining application would use information about the existing customers to build and validate the model. In the end, the result is a model that would take details about the customer/prospects as inputs and generate a number between 0 and 1 as the output. This process is illustrated in Figure 17-1.

After a model has been created based on historical data, it can then be applied to new data in order to make predictions about unseen behavior. This is what data mining (and more generally, predictive modeling) is all about. The process of using a model to make predictions about behavior that has yet to happen is called "scoring." The output of the model, the prediction, is called a score. Scores can take just about any form, from numbers to strings to entire data

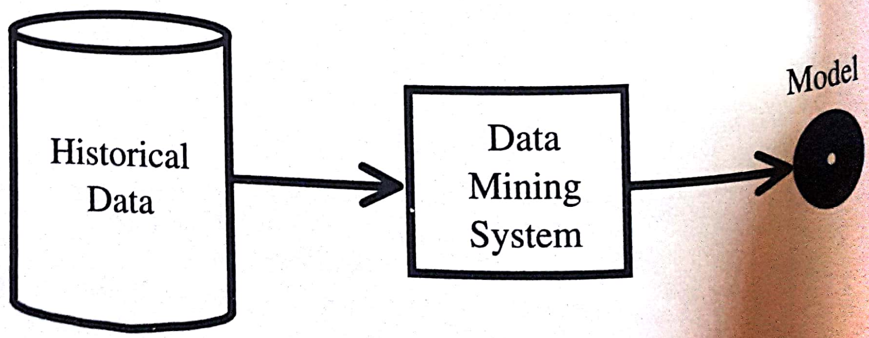


Figure 17-1
The data mining process

structures, but the most common scores are numbers (for example, the probability of responding to a particular promotional offer).

Scoring is the unglamorous workhorse of data mining. It doesn't have the sexiness of a neural network or a genetic algorithm, but without it, data mining is pretty useless. (There are some data mining applications that cannot score the models that they produce—this is akin to building a house and forgetting to put in any doors.) At the end of the day, when your data mining tools have given you a great predictive model, there's still a lot of work to be done. Scoring models against a database can be a time-consuming, error-prone activity, so the key is to make it part of a smoothly flowing process.

The Process of Scoring

Scoring usually fits somewhere inside of a much larger process. In the case of one application of data mining, database in marketing, it usually goes something like this: there're the steps

1. The process begins with a database containing information about customers or prospects. This database might be part of a much larger data warehouse or it might be a smaller marketing data mart. db
2. A marketing user identifies a segment of customers of interest in the customer database. A segment might be defined as "existing customers older than 65, with a balance greater than $1000 and no overdue payments in the last three months." The records representing this customer segment might be siphoned off into a separate database table or the records might be identified by a piece of SQL that represents the desired customers. segment
3. The selected group of customers is then scored by using a predictive model. The model might have been created several months ago (at the request of the marketing department) in order to predict the customer's likelihood of switching to a premium level of service. The score, a number between 0 and

1, represents the probability that the customer will indeed switch if they receive a brochure describing the new service in the mail. The scores are to be placed in a database table, with each record containing the customer ID and that customer's numerical score.

4. After the scoring is complete, the customers then need to be sorted by their score value. The top 25% will be chosen to receive the premium service offer. A separate database table that contains the records for the top 25% of the scoring customers will be created.

5. After the customers with the top 25% of the scores are identified, the information necessary to send them the brochure (name and address) will need to be pulled out of the data warehouse and a tape created containing all of this information.

6. Finally, the tape will be shipped to a company (sometimes referred to as a mail house) where the actual mailing will occur.

The marketing department typically determines when and where the marketing campaigns take place. In past years, this process might be scheduled to happen once every six months, with large numbers of customers being targeted every time the marketing campaign is executed. Current thinking is to move this process into a more continuous schedule, whereby small groups of customers are targeted on a weekly or even daily basis.

When marketing campaigns are infrequent, manual selection and scoring of the data is not a significant impediment to the process. There is usually significant lead time to allow for the various parties to do their work before the actual mailing will take place. When someone in marketing needs to have a segment of customers selected for the campaign, they simply call someone in IT. When the scores are needed, the statistician who created the model is asked to apply the model to the customers in the desired segment. Because the processing is performed manually, the possibility of an error being introduced into the system is considerable. To minimize errors,

- [Make sure that you] select the right customers for scoring. An error in this process is usually due to an incorrect translation from the marketing user's vocabulary to the syntax of an SQL statement executed by someone in IT.
- Make sure that the correct customers are scored. The correct database table needs to be scored. There is confusion sometimes regarding which table, among hundreds, is supposed to be scored. When the names of the tables are cryptic, as they often are (for example, JF432_IPG), the possibility of using the wrong data for scoring is possible.
- Make sure that the correct model is used to do the scoring. Assuming that the targeted selection of customers is a success, the number of models available could be quite large. In addition, multiple models might be similar (for example, one model predicts responses to a particular catalog for women aged 50–55, whereas another model predicts responses for men aged 50–55).
- Make sure that the scores are put in the right place. Just as confusion sometime exists with the data that is going to be scored, there can also be some confusion about the tables that contain the scores.
- Make sure that you understand how the scores are ordered. Are high values good or bad? If you want to select the best customers, you will need to know what score values represent those customers.

When the frequency of the marketing campaigns is increased so that they occur on a daily or weekly basis, there are two significant impacts on the campaign. First, the decreased time between mailings means that there is much less room for error when carrying out the individual steps in the process. If a mistake is found, there is less time to correct it compared to the less frequent campaigns. Second, the sheer number of scoring "events" will increase dramatically, due to both the increased frequency of the [...]

marketing and IT, as well as between marketing and the modelers. The best approach to solving this problem is to use the campaign management software that is integrated with the scoring engine. If integrated software is not available, care will need to be taken so that difficulties are minimized.

Scoring Architectures and Configurations

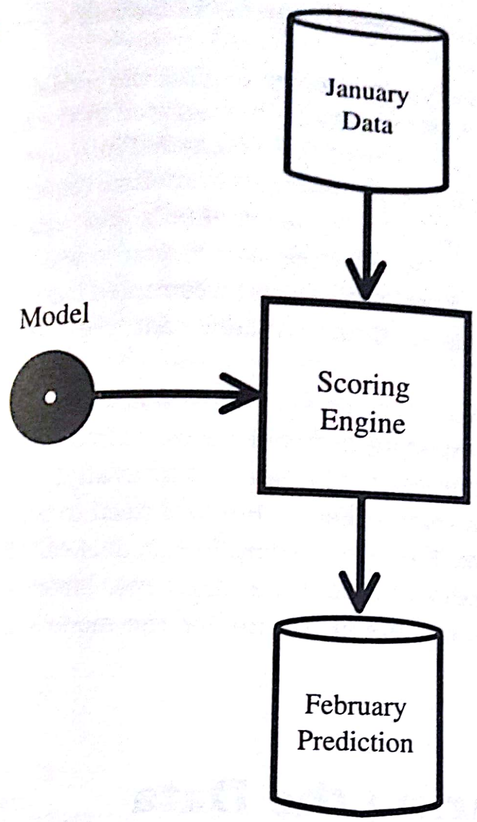
The software systems that are used to carry out the scoring process are usually simpler than the applications used to build the models. This is because the statistical functions and optimization procedures that were used to create the model are no longer needed; all that is required is a piece of software that can evaluate mathematical functions on a set of data inputs.

Scoring involves invoking a software application (often called the scoring engine, Figure 17-2), which then takes a model and a dataset and produces a set of scores for the records in the dataset. There are three common approaches to scoring engines:

- A scoring engine software application that is separate from the model-building application.
- A scoring engine that is part of the model-building application.
- A scoring engine that is produced by compiling the model code (for example, C++ or Java) that is output by the data mining application. In this case, a model is itself the scoring application because it is an executable piece of software (once it is compiled).

The type of model generated will depend upon the data mining system that is used. Some data mining systems can produce multiple types of models, whereas others will generate only a single type.

In the first two cases, the scoring engine is a software application that needs to be run by the user. It might have a graphical user interface or it might be a command line program, in which the user specifies the input parameters by typing them onto a console interface



when the program is run. There are usually three inputs to the scoring engine: the model that is to be run, the data that is to be scored, and the location where the output scores should be put.

In some cases, a data mining system might generate a model that can be executed by another software vendor's scoring engine. Although there are currently no standards for the specification of a predictive model, some data mining vendors have decided to use the modeling formats created by established statistical software vendors. As of the writing of this book, at least two data mining software vendors have optional model output formats that are compatible with the modeling language supported by the SAS Institute's software. Models that are written out in the SAS modeling format can then be executed by the SAS Institute's scoring engine (known as SAS/Base).

In the last type of scoring engine, the model acts as its own scoring engine. After the model is generated by the data mining software application, it will need to be compiled into an executable form. This step is usually done manually and often requires knowledge of system and programming level details (for example, linking ODBC database drivers). The primary reason to use a compiled model is to increase performance because a compiled model will usually run significantly faster than a model that requires a separate scoring engine.

There are obvious downsides to this approach, though. First is the fact that preparing a model for execution (compiling, linking, etc.) requires expertise that might not be available. Second, if the models change on a regular basis, they will need to be recompiled whenever they change. The use of compiled models can significantly increase the complexity of model management, especially if there are large numbers of models in use and/or the models change on a frequent basis.

Preparing the Data

Before you can score a model, you need to prepare the data on which the model is going to operate. Key to this process is the concept of consistency. The customers that are to be scored by the model should be consistent with the customer data that was used to build the model. For example, if a model was built using response data from low balance customers aged 40 to 50, it should not be used on customers aged 50 to 60.

A second type of consistency involves the type of interaction that will take place with the customer or prospect. The interaction needs to be consistent with the original data, or else the results might not be correct. The historical data that was used to build the model had a context that needs to be considered. The color of the envelope, the wording used in the offer, the type of offer, and other variables will affect the results of the interaction. If your model was built from historical response data for a mailing that used a blue envelope, the results that you will see if you send out a new offer in a green envelope could be different from what the model predicts. Care must be taken so that any assumptions, both from the marketing and modeling sides of the fence, are not lost when the implementation of a model takes place. A process (possibly part of a corporate knowledge base) should be maintained to describe customer segments, as well as the types of offers that are made to those customers/prospects.

After you are sure that the data is consistent with the historical customer data and interaction details, you need to map the individual columns (the variables) in your data set to the inputs of the model, Figure 17-3. The data that is to be scored using an existing predictive model needs to match the data that was used to build the model. Matching means that all of the data fields that were used as inputs to the model need to be made available for the model during the scoring process. It should be noted that not all fields that were

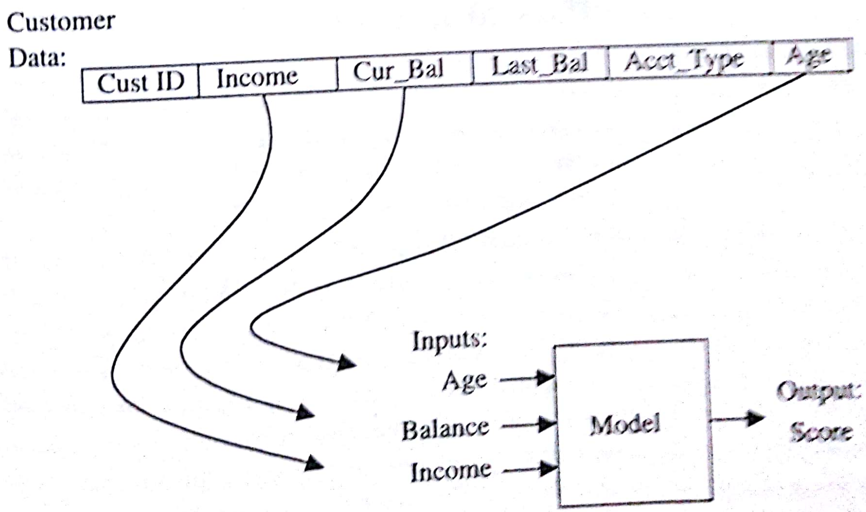


Figure 17-3 Mapping the data to the model

used to build the model are necessary when scoring the model. It is likely that many of the available fields were not used as inputs to the model because the data mining process determined that they did not provide any predictive information. Only the fields that were actually used in the in the model need to be included. This can usually improve performance because not all data needs to be passed to the scoring engine.

When mapping the data in the database to the inputs of the model, there are two types of mapping that can take place: direct and offset.

Direct Mapping

In a direct mapping approach, a variable that was used to build the predictive model and is included as an input is mapped to the same variable. For example, if the variable "Account Type" were an input to the model, it would simply map to the same variable. This approach is best used for input variables that are not part of a time series.

Offset Mapping

In offset mapping, the variables that were as inputs to the model are mapped to variables that are different from those used to build the model. This is often the case when input variables are part of a time series. For example, if a model was built using data from January, there might be inputs that are specific to that month (for example, "Outstanding_Balance_Jan"). When this model is applied to data after January, the inputs will need to be offset to match the time period for which the predictions are being made. When applied to February data, the input should be mapped to "Outstanding_Balance_Feb." The easiest approach, if the data is in a database, is to use a database view to re-direct the inputs to the appropriate table and column. The view would be updated to whenever new monthly data was made available so that it pointed to the latest outstanding balance.

In the real world, the scoring process would probably use a combination of both direct and offset mappings.

The last step in preparing the data, if necessary, is to transform the input to conform to any requirements specific to the model. For example, an account type in the database that is represented as a string (i.e., "checking," "savings") might need to be transformed into numbers before it can be fed to the model. The form of the transformation is usually specific to the model type and should be specified by the person who created the model. Although this functionality should be incorporated into the model itself by the data mining system, some applications require the user to do any transformations manually.

Integrating Scoring with Other Applications

Scoring isn't something that takes place in a vacuum. After a model has been produced, other applications need to know that they exist and make use of the scores that they generate. Tight integration of data mining applications with other software systems is relatively new, but it is a trend that will continue for some time. Some of the software categories that are likely to embrace integration with data mining applications include *enterprise resource planning* (ERP), *Campaign Management*, and tools such as *Online Analytical Processing* (OLAP) and data visualization.

As an example, consider how a data mining system might be integrated with a marketing campaign management system. Marketing managers are interested in using the output of a data mining model in order to further refine the customer segments that they have specified. The simplest example might involve segregating a group of customers into separate yes/no categories. The customers that fall into the yes category will end up receiving a marketing offer, whereas the other group will not receive the offer. The marketing department will use a campaign management software system to manage the selection of the customers and the segments they fall into.

The closer that the data mining and campaign management software work together, the better are the business results. In the past, the use of a model within campaign management was often a manual, time-intensive process. When someone in marketing wanted to run a campaign that used model scores, he or she usually called someone in the modeling group to get a file containing the database scores. With the file in hand, the marketer would then solicit the help of someone in the information technology group to merge the scores with the marketing database.

Integration is crucial in two areas:

- First, the campaign management software must share the definition of the defined campaign segment with the data mining application to avoid modeling the entire database. For example, a marketer may define a campaign segment of high-income males, between the ages of 25 and 35, living in the Northeast. Through the integration of the two applications, the data mining application can automatically restrict its analysis to database records containing just those characteristics. This is important for the sake of data consistency between the data that was used to build the model and the data that will be scored by the model. By using the same definition, it will be more difficult to make a mistake and score records that are inconsistent with the records used to build the model.
- Second, selected scores from the resulting predictive model must flow seamlessly into the campaign segment in order to form targets with the highest profit potential. Any manual process involved with the movement of scores from the output of the model to a separate software complicates the overall process. Besides being a source of possible errors (for example, using the wrong score table), the delay associated with the manual processing could limit the frequency of marketing efforts.

Creating the Model

In the case of data mining for a marketing campaign, an analyst or user with a background in modeling creates a predictive model using the data mining application. This modeling is usually completely separate from the process of creating the marketing campaign. The complexity of the model creation typically depends on many factors, including database size, the number of variables known about each customer, the kind of data mining algorithms used, and the modeler's experience.

Interaction with the campaign management software begins when a model of sufficient quality has been found. At this point, the data mining user exports his or her model to a campaign management application, which can be as simple as dragging and dropping the data from one application to the other. This process of exporting a model tells the campaign management software that the model exists and is available for later use.

Dynamically Scoring the Data

Dynamic scoring is a type of software integration that allows the scoring process to be invoked by another software application that will use the scores for some other purpose. In our database marketing example, the campaign management system will interface with the scoring engine so that the scores are generated when the campaign manager needs the scores. Further, only the required records will be scored because the campaign management system determines when and what to score. Dynamic scoring avoids mundane, repetitive manual chores and eliminates the need to score an entire database. Instead, dynamic scoring marks only relevant record subsets, and only when needed. Scoring only the relevant customer subset and eliminating the manual process shrinks the overall processing time significantly. Moreover, scoring records segments only when needed assures fresh, up-to-date results.

After a model is in the campaign management system, a user (usually someone other than the person who created the model) can start to build marketing campaigns using the predictive models. Models are invoked by the campaign management system.

When a marketing campaign invokes a specific predictive model to perform dynamic scoring, the output is usually stored as a temporary score table. When the score table is available in the data warehouse, the data mining engine notifies the campaign management system, and the marketing campaign execution continues.

18

Optimizing the CRM Process

Introduction

All customers are not created equal. Some are profitable; some are unprofitable; some will become highly profitable; some will become unprofitable; and some will never be profitable throughout their entire customer lifecycle. Optimization is the science of "optimally" determining what can be done to make a customer as profitable as possible for as long as possible.

Despite the business opportunity, optimization techniques are not generally applied to customer relationship management, or sales and marketing. Optimization is thought to be too esoteric a science to be applied in situations in which intuition can be as important in making the right decision as cold hard facts about customer behavior. In fairness to the skeptics, it is harder to apply optimization to customer relationship management than it is to other areas of business. But, with advances in database technology, data mining, and the CRM systems themselves, optimization can be applied more often than not. When it is applied, it can be the single most important technique in your CRM system for increasing customer profitability.

Figure 18-1 shows an example of how a particular promotion as part of the CRM process could be optimized.

Improved Customer Profitability through Optimization

What Is Optimization?

Many people are familiar with the general idea of optimization, but may not know that there is a science of optimization that is available to businesses today. Typically, optimization is applied to problems where there is a well-defined measure of success. For instance, optimization is applied to airline scheduling (although it may be hard to believe that there is anything optimal about flight times). Constraints such as having the right plane ready at the right time

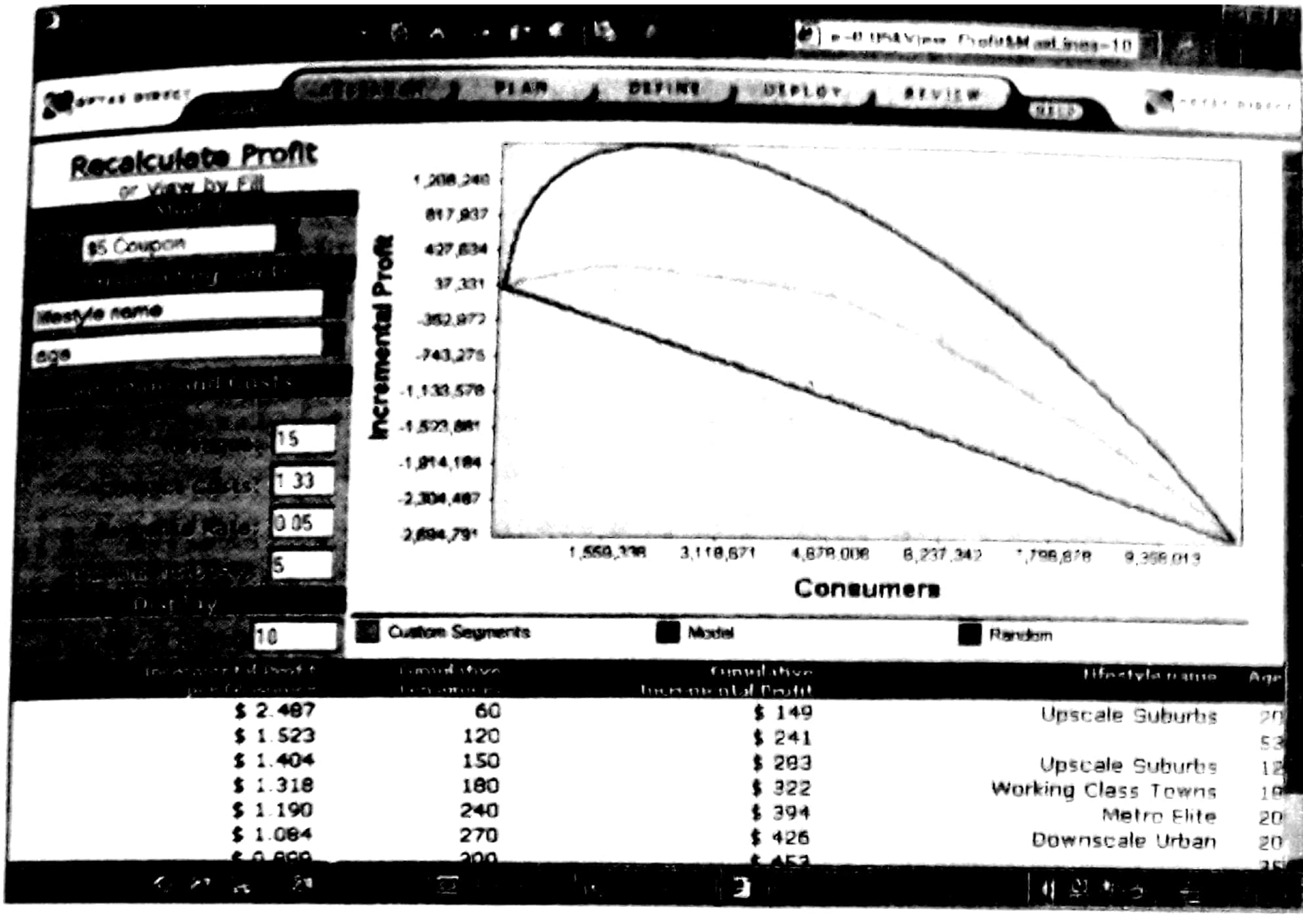


Figure 18-1 Profit of a particular promotion as part of an overall CRM process can be optimized by targeting the right consumers with the right promotion. In this case, $1.3 million can be achieved by targeting the top two million consumers from a data mining predictive model.

in the right airport to accommodate a given scheduled flight can difficult because the cost of leaving an airplane idle for seve hours when it could be in flight could mean critical dollars savec lost. Sometimes, the constraints can be more complex when the ficulties of scheduling a well-rested crew is included. Sometim the constraints for optimization are difficult to accommodate, matter what system you are using. For instance, in the early day the airline industry, all stewardesses from India needed to be ho to sleep in their home country each and every day. This constra needed to be added in to all of the other constraints to provide optimal schedule for the Indian airline.

It can be difficult to conserve fuel and optimally serve the c tomers. For example, hubs are convenient for airlines as pla where they can congregate passengers in order to more efficien have full flights. But the multiple hops they produce for travel who mostly prefer direct flights can also have an effect on custon satisfaction and hence on customer profitability. Also, each schedule that is produced needs to also allow for some error. For instance, if the schedule is perfectly optimized, but a thunderstorm occurs in Boston that closes the airport for 45 minutes, the scheduling should be able to accommodate it without breaking down and causing massive delays throughout the system.

Although optimization is used in marketing today, it is used only within well-defined problems such as optimally conserving mailing costs for a direct-mail package while optimally mailing to only those people who are most likely to be interested in the offer. Determining what is optimal is performed by using data mining to build models to forecast what a customer will do in the future. In some ways, this is similar to other business-optimization problems such as airline scheduling. The behavior of the system cannot be so easily controlled, however—although an airline can mostly control when its flights are ready to take off, a direct marketer has limited control over whether a consumer will react positively to an intervention, and not all data is available for all consumers.

Optimization is used on small islands of applications within marketing, but the function of marketing is still viewed as more art than science. Customer relationship management is an attempt to bring some science to the marketing function. Overall marketing is still viewed as something that cannot be optimized, however, because there is so much that needs to be done by gut feel, by hunch, or by understanding larger global issues than are captured in the customer database.

The question about whether marketing is ready for optimization is really a question of whether marketing is "clean" enough so that the way that it is performed can be modified (hopefully improved) and the results then measured. The fundamental difficulty in marketing as it differs from scheduling airplanes is that with airplanes, success or failure can be measured in a much more real sense. The time airplanes spend idle on the runway, the fuel cost, and fuel consumption are all relatively easy to measure. With much of marketing, the game is harder to play because success or failure is harder to measure. The old adage, "I know I'm wasting half of my advertising budget, but I just don't now which half," is a good example of the promise and problems of applying the technique of optimization to customer relationship management.

If a thirty-second commercial during the Super Bowl costs $750,000, how do you know if that money is well spent? Even if you could collect all possible data, you would need to make some hard decisions about how to measure. For instance, if a person saw the commercial, did not buy within 30 days of having seen the commercial, but did buy a year later, how would you measure the impact of that $750,000 marketing investment? Is it based on the revenue that occurred one year later, or is that an independent event and the consumer would have made the purchase with or without investment in the Super Bowl commercial?

If optimization is going on within marketing, it is not necessarily performed where it is needed most. Instead, it is performed where the results can be most easily measured. It is similar to the old man on the darkened city street who was looking beneath a street lamp for the keys that he had just dropped. When a passerby offered to help him look and asked if he had dropped his keys near the street lamp the man replied: "No, I dropped them over there, but there is no light over there. So, I am looking under the street lamp here where I can see."

Why Not Optimize Customer Relationships?

To many people, speaking of optimizing customer relationships or optimizing marketing processes doesn't make sense. How could you optimize something like a customer relationship or overall customer profitability? There are so many things that could go wrong. So many intangibles that can't be measured. So much of the win or lose is dependent on the creative nature of the marketing people, a good slogan, or a new color scheme. It would be like trying to predict and optimize next year's fashions based on a mathematical mode. Many believe that it is not possible. But why not? What makes it so hard?

As we have shown, optimization is already being used on certain small islands of functionality within the customer relationship process—mostly around the "lamp posts" (where the results can be seen and measured). These islands of optimization include the following:

when optimisation is used in CRM?

- Targeted marketing
- Fraud detection
- Attrition prediction
- Cross-selling
- Acquisition of new customers

To Optimize Something, You Must Have Control over It

One of the most important points in optimizing something is that you must be able to not only measure the effect of your changes, but also to affect the changes in the first place. This is true, for instance, in the optimization performed in nature. The genetic material (DNA) makes a change in the way an organism develops, and then the process of survival of the fittest measures the value of that change and determines whether the change is for better or for worse.

In classical marketing, this has been difficult because it has been difficult to measure. (How do you really measure all of the impact of a television commercial campaign?) It has also been difficult to test for different effects. For instance, in the past it would have been difficult to test out different looks and feels for a catalog or a magazine ad. The cost of design and printing for different variations of the same ad would be cost-prohibitive. Today, however, it is not uncommon for companies to try a variety of different possible messages and artwork for the mailing of a catalog. It is possible today to control and differentiate the type of the marketing intervention that is being provided across many different delivery channels.

Why Now?

Optimization of customer relationships is now a possibility, whereas before it was not possible. This has occurred for three primary reasons:

- New technology and its maturation
- Data, and its collection and storage
- Changes in business processes

There are, however, some trends that are now taking place that will allow the use of optimization within customer relationship management to expand. Because optimization requires measurement, any tools that allow for better and more consistent measurement can now be used to improve marketing. There was a good reason why marketing couldn't be optimized in the past—it wasn't that it wasn't a good idea back then, and now it is a good idea. It was always a good idea, but now with the changes in the way that marketing is conducted, it is possible to perform optimization. The time is right also because of the maturation of several new technologies:

- Data warehousing
- Data mining
- OLAP
- World Wide Web

We are also seeing today a dramatic increase in the quantity and types of data that are being collected. Some of this increase is due to the technologies that allow for its collection and its storage, but the data itself comes from a wide variety of sources. The time has come for optimization in the marketing function because of changes in business processes:

- Data-based marketing
- Enterprise resources planning (ERP)
- Electronic commerce
- Customer relationship management

These business processes are critical for the possibility of marketing optimization. Data-based marketing was historically limited to direct mail campaigns; the only "database" part about it was the fact that the name and address of the recipient was stored in the database. Partly because of the ease with which direct mail can now be generated, data-based marketing has become more important. It is no longer broadcast marketing where the same marketing message is sent to the entire consumer base.

The move toward enterprise resource planning has also had an impact on the possibility of performing the optimization of customer relationships. Previously, ERP systems laid the foundation for supply chain optimization in much the same way that database marketing has laid the foundation necessary to support electronic order capture and front office support for the sales force and for telemarketing. New data sources are also improving marketing performance because customer relationships can be managed better only when there is more information. Some of this data comes directly out of ERP systems, which are excellent sources of cleaned data pertinent to customer behavior. Likewise, data captured from electronic commerce systems is also valuable in that the data captured is fairly well structured because it was always in electronic form. There are not issues involved in the manipulation of the data or the original data entry.

Finally, the business process of customer relationship management is also a driving factor in the timing of the use of optimization. CRM is important because of the great disparity between the cost of acquiring a new customer and that of holding onto a current customer. For some industries such as wireless communications (cell phones), the ratio between the two can be ten to one. Also, the value of a current customer may be much greater than that of a new customer because in many industries it can take several years before a customer begins to purchase at their full potential.

These business processes both lay the foundation for optimization as well as build a business mind set (through CRM) for the importance of an optimal marketing function because of the importance of the customer relationship and keeping it profitable.

Optimized CRM

Today, there are a lot of ways that marketing is performed. For instance, we notice that the data that is coming in about our customers comes in from a variety of sources. But, also note that there are a large number of marketing organizations that are not measurement-based at all.

These organizations generate television and radio advertising, mailings, magazine advertising, and a significant Web presence, and yet have no way of collecting the information—either about their customers as they are or even about how they had promoted to them. Therefore, a catalyst is necessary. In scientific terms, a catalyst in a chemical reaction speeds up a process that might well have occurred so slowly without the catalyst or that would, for all intents and purposes, not have happened at all.

We believe, then, that this scientific term is valuable for capturing any and all interventions that a marketing organization might take with their customer prospects. For instance, one catalyst might be the sending out of coupons to likely users of the product. Another catalyst might be a television commercial (not a particularly measurable catalyst). Another catalyst might be a sales call that is made on the prospect by a sales person.

Fundamentally, marketing and sales are relatively simple. Consider, in the broadest view, that every person or corporate entity in existence is a possible prospect if only the right set of catalysts could be applied. The first question then is the following: What is the right catalyst or sequence of catalysts to apply? This then begs the real question, which is the following: What is the most profitable catalyst that can be applied?

When viewed in this way, the seemingly unconstrained and complicated world of marketing and sales starts to look more like a system that could be optimized. There is a collection of data, there is something that you can do to change the system, and there is a way to measure the impact of that change. If you can do this, then you are in a situation in which you can begin to optimize by recognizing what has worked in the past and trying subtle variations on those successes. Each time, try to become more and more effective in creating a profitable customer relationship.

If we look at things this way, then there are really two different aspects of the optimization problem. The first is to come up with a selection of possible useful catalysts from which to choose, and the second is to optimally assign a catalyst based on what you learned from the past use of the catalysts. The first step is creative and exploratory; the second is analytical.

The Complete Loop

As we have stated before, in order to optimize, one must first be able to measure. Then, one must have the ability to try different actions in order to compare and contrast which method is better. If you cannot measure, you cannot differentiate between good and bad outcomes; if that is the case, then you cannot improve. Likewise, if your system cannot incrementally improve based on past experience to better handle future cases, then that also does not allow for optimization. If the system can be measured and options can be proposed, but there is no accountable action taken to test out that improved strategy, then the cycle of improvement is broken.

Like a three-legged stool, there are three important steps to the optimization process as follows:

1. *Measure.* See what happened.
2. *Predict.* Figure out something else to try based on what happened.
3. *Act.* Try it out.

From a marketing perspective, this translates into being able to measure the outcome of whatever marketing intervention is performed (sounds simple, but it is the toughest step). Then based, on those measurements, you can make a prediction about what will happen next and what changes in the marketing intervention might give useful results that could be measured (note that both good and bad results are useful if they shed light on the efficacy of the marketing intervention).

As an example of a company going out of its way to measure the value and the marketing of its product, consider the video distributor who provided users with two bins in which to return the movie that they had just rented. One bin would be used if they liked the movie. The other bin would be used if they didn't like it. A little bit of extra effort in order to measure the customer satisfaction is well worth the effort in order to get the feedback from the users.

If, however, the video retailer does not make use of this information to improve its product, then marketing optimization, even at its most basic level, cannot be performed. It is very much like the three-legged stool, with one of the legs missing.

Optimal CRM Process: Measure, Predict, Act

Optimizing the CRM process requires a business best practice that consists of three main steps and an architecture for implementing and supporting those steps. The steps are formally named: "measure," "predict," and "act." They represent the steps in a cycle of customer relationship management that is continuously improving. The methodology is general enough that it is applicable to the vast majority of customer management functions such as the following:

- *Cross-selling.* Selling a new product to an existing customer
- *Acquisition.* Acquiring new valuable customers
- *Retention.* Retaining existing valuable customers

To achieve these goals of better customer management, there are a variety of processes that are used over and over again by the marketing organization:

- *Targeted marketing*
- *Lifetime value prediction*
- *Channel management.* Matching the channel to the customer in the most profitable way

In marketing, there typically exists a certain set of interventions that can be performed against the customer or prospect. Often, this is a fixed list that has been handed down from marketing personnel. Usually, these are new creative offers that have been created based on feedback from customers, either directly through phone interviews or focus groups, or just from the experience of particular marketing managers. These marketing products, campaigns or anything else represent the available arsenal with which to motivate the customer to do more business. If this arsenal of interventions already exists, then the remaining questions are relatively simple: "What do I do to whom, and when?" Assuming that there is some regularly scheduled launching of marketing programs (let's say, monthly), the question can become: "What should I do to whom?" Not doing anything should also be a viable intervention/catalyst, as well as anything else that might be used.

If you want to do a good job of answering the question: What should I do to whom? It pays to create a model of what the likely outcome would be if the variety of different interventions were to be applied. This is the predict step (based on past experience). This is what the technology of data mining is used for.

For a complete marketing optimization system, whatever is predicted must also be measured. For instance, it doesn't help to correctly predict customer attrition if you don't know what the value of a saved customer is. Also, it is important that the step of action is within the infrastructure of running the complete marketing optimization system. For instance, many of the classical decision support or business intelligence systems will leave off on the action step. These systems will often provide useful information about how to improve the marketing process to the customer, but the results are delivered in the forms of graphs or reports that a senior manager can look at and better understand their business. In some way, the CEO can act based on the report or the graph, but typically it happens outside of the systems that should be recording the purpose of the action and the deployment of the action. The CEO may be influenced by what they've seen in the results of the decision support system, but their action is "non-recordable" and non-measurable. For marketing optimization to realize the full potential, this action step must be recorded within the same system that provides the measurement and produces the predicted future behavior.

What Marketing Optimization Is Not

Marketing optimization is not data warehousing. Marketing optimization is an infrastructure in the same sense that data warehousing is an infrastructure, but along with that infrastructure comes a set of best practices that guide the use of that infrastructure specifically along the lines of business issues. Marketing optimization is top down in the sense that it starts with the overall marketing needs (which, in summary, is improved customer profitability), and defines the infrastructure and then the data, which is necessary to support those business problems. Data warehousing is most often created in a bottom-up fashion, starting with the infrastructure (mostly building an infrastructure to collect all available data) and then determining the business problems that could benefit from this infrastructure. For this reason, data warehousing projects are typically quite vast because they start with the attempt to collect all available data. Marketing-optimization systems can be much smaller and easier to manage, and more quickly provide a measurable return on investment because only the infrastructure necessary to support the best-practices methodology and the solution of the business problems is created.

Marketing optimization is not one-to-one marketing. Marketing optimization can be used to make one-to-one marketing more profitable per customer, but the emphasis of one-to-one marketing on individual consumers is somewhat at odds with the mentality of marketing optimization that operates often times at a higher level. It is typically at a level of segmentation rather than at individual customers. The main reason for this is that in order to measure customer behavior to a significant level, you need more than one customer as data input to measure. In another sense, one-to-one marketing also increases the number of possible interventions that can be placed on the customer. But many of the examples given for one-to-one marketing may be so limited in terms of the scope of customers receiving the intervention that measurement may not be valid. Or, as in many of the cases and as in the classical case of business intelligence and decision support, actions may be taken but they are not recordable or measurable. If what you did isn't recorded, then its effect can't be measured and it can't be changed because you don't know what you did in the first place.

Using Data Mining to Optimize Your CRMS

There are really two forms of optimization that are possible. One, in which there is a well-defined value that is being optimized—such as the total number of days that parts for a John Deere tractor sit at the factory after being purchased and shipped from the supplier before they are used to build a new tractor. Or the number of days that completed tractors sit in inventory before they are shipped for sale.

This measure of value can be calculated and optimized fairly accurately because the processes that control the time to build and to ship are fairly well understood, although complex. The other class of optimization problems is not only complex but also much less "causal" in that what affects and "causes" the increase or decrease in the important measure of value is not as direct or predictable. An example of this type of problem would be the typical class of problems in the CRM space. And this class of problems, in which it is difficult to optimize via techniques that make use of causal models, is often difficult because there are human beings in the loop who have very complex behavior. They can react in completely different ways, depending on small differences in the way that the system is constructed.

For example, as you have seen in the cellular telephone example, trying to "optimally" reduce churn by offering customers a lucrative offer of lower rates actually had the unforeseen consequence of increasing churn. Just the effect of sending the offer caused the customers to evaluate all of their options and, because there were competitive options, they moved to another carrier despite the better rates provided in the offer. Without a viable competitive offer, the outcome of the churn-reduction program could have been completely opposite. In a causal system, either these small changes result in small changes in the total value, or they can be controlled to the degree where they are effectively non-existent.

Often, companies are sending out multiple messages to their customers or prospects at the same time. These messages are used to overcome a variety of barriers that the consumer may be facing or feel that they are facing in making a buying decision for the company's product. Matching the right message to the right consumer can be as important as determining the message and the barriers in the first place.

Optimization Techniques

Neural n/w
link wt
genetic Back propg

If you think simply about the training process in the data mining technique called neural networks, you could view it as a game of trying to find the best possible combination of link weights for a given network architecture (for example, the standard three-layer, fully connected between layers). The network is in fact fully defined by the numbers assigned to each link (for a network with five inputs, three hidden and two output nodes, there would be 21 links ($5 \times 3 + 3 \times 2$) and just 21 numbers that defined the full behavior of the neural network). The trick in training is to find those right numbers and this is what the back propagation algorithm accomplishes by changing the numbers each time to improve the accuracy of the network.

Back propagation is not the only way of determining these link weights, however. Genetic algorithms have also been deployed to try to find the best possible link weights. In Figure 18-2, a population of neural networks is being evolved. From one generation to the next, some networks do better and some do worse. Over time, however, the best solution in the population gets better and better.

Genetic algorithms simulate natural evolution on the computer. To do so, they simulate the DNA, which in nature describes how to

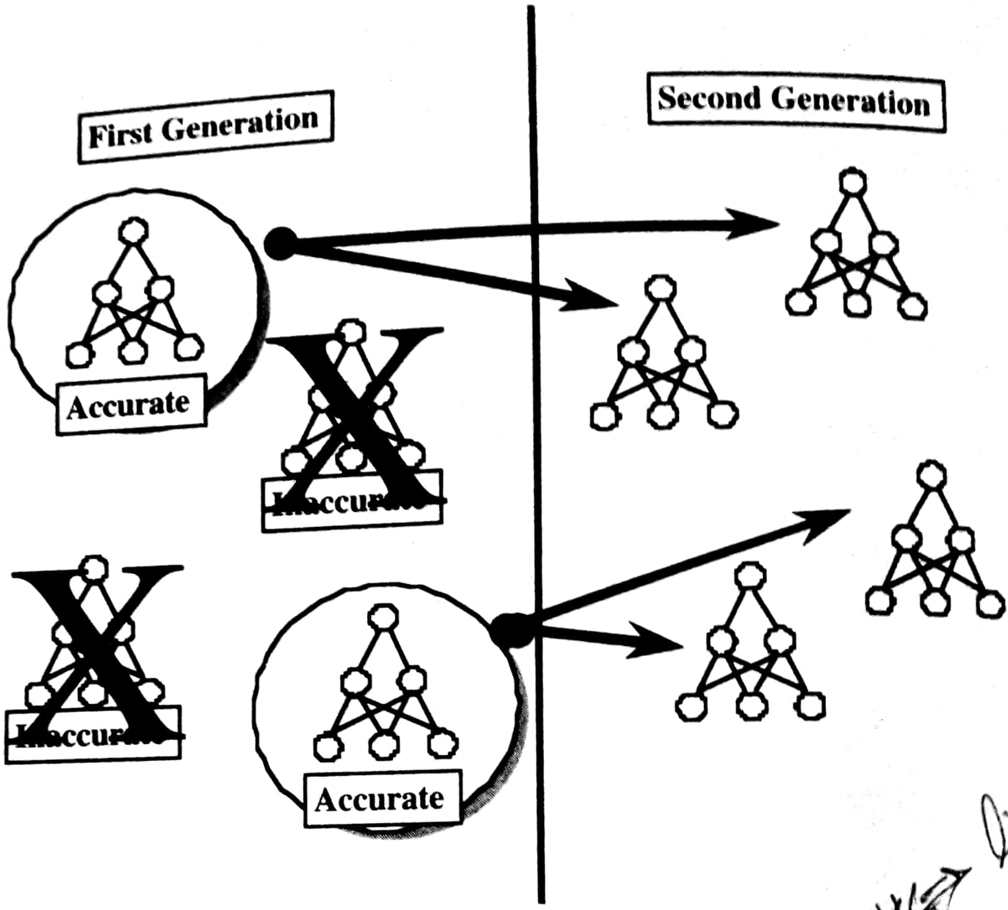


Figure 18-2 Optimization with Genetic Algorithms

DNA → link wt

uniquely grow the animal or plant. The analogue to DNA in genetic algorithms is the list of numbers that represent the link weights in the neural network. In some ways, this list is like a chromosome, and each number is like a gene. Taken together, all of these simulated genes fully describe how to build the given neural network. In genetic algorithms, there are good lists of link weights that represent highly accurate neural networks and there are poor lists of link weights that represent neural networks that are no more predictive than random guessing. Genetic algorithms generate many different genetic guesses at the right link weights, and create a population of different neural networks. Survival-of-the-fittest techniques are then used to weed out the poorly performing networks and reward the more accurate ones.

If a network is particularly accurate, it will be rewarded by allowing it to "reproduce"—by making copies of itself with slight variations. These modifications to the link weights are made at random like genetic mutations, and sometimes result in improved perfortomers to evaluate all of their options and, because there were competitive options, they moved to another carrier despite the better rates provided in the offer. Without a viable competitive offer, the outcome of the churn-reduction program could have been completely opposite. In a causal system, either these small changes result in small changes in the total value, or they can be controlled to the degree where they are effectively non-existent.

Often, companies are sending out multiple messages to their customers or prospects at the same time. These messages are used to overcome a variety of barriers that the consumer may be facing or feel that they are facing in making a buying decision for the company's product. Matching the right message to the right consumer can be as important as determining the message and the barriers in the first place.

Neural n/w
link as
genetic back prop

Optimization Techniques

If you think simply about the training process in the data mining technique called neural networks, you could view it as a game of trying to find the best possible combination of link weights for a given network architecture (for example, the standard three-layer, fully connected between layers). The network is in fact fully defined by the numbers assigned to each link (for a network with five inputs, three hidden and two output nodes, there would be 21 links ($5 \times 3 + 3 \times 2$) and just 21 numbers that defined the full behavior of the neural network). The trick in training is to find those right numbers and this is what the back propagation algorithm accomplishes by changing the numbers each time to improve the accuracy of the network.

Back propagation is not the only way of determining these link weights, however. Genetic algorithms have also been deployed to try to find the best possible link weights. In Figure 18-2, a population of neural networks is being evolved. From one generation to the next, some networks do better and some do worse. Over time, however, the best solution in the population gets better and better.

Genetic algorithms simulate natural evolution on the computer. To do so, they simulate the DNA, which in nature describes how to