Analytics-driven solutions for customer targeting and sales-force allocation

R. Lawrence
C. Perlich
S. Rosset
J. Arroyo
M. Callahan
J. M. Collins
A. Ershov
S. Feinzig
I. Khabibrakhmanov
S. Mahatma
M. Niemaszyk
S. M. Weiss

Sales professionals need to identify new sales prospects, and sales executives need to deploy the sales force against the sales accounts with the best potential for future revenue. We describe two analytics-based solutions developed within IBM to address these related issues. The Web-based tool OnTARGET provides a set of analytical models to identify new sales opportunities at existing client accounts and noncustomer companies. The models estimate the probability of purchase at the product-brand level. They use training examples drawn from historical transactions and extract explanatory features from transactional data joined with company firmographic data (e.g., revenue and number of employees). The second initiative, the Market Alignment Program, supports sales-force allocation based on field-validated analytical estimates of future revenue opportunity in each operational market segment. Revenue opportunity estimates are generated by defining the opportunity as a high percentile of a conditional distribution of the customer's spending, that is, what we could realistically hope to sell to this customer. We describe the development of both sets of analytical models, the underlying data models, and the Web sites used to deliver the overall solution. We conclude with a discussion of the business impact of both initiatives.

INTRODUCTION

Improving sales productivity is an essential component of growth for major companies. Although hiring the best sales representatives is an obvious first step, there is an increasing realization that for a sales force to achieve its potential, sales representatives and executives must be equipped with relevant information technology (IT) tools and solutions.[1] The past decade has seen the development of a number of customer relationship management (CRM) systems[2,3] that provide integration and management of data relevant to the marketing and sales process. Sales force automation (SFA) systems[4] enable sales executives to better balance

sales resources against identified sales opportunities. Although it is generally (but not uniformly[5]) accepted that such tools improve the overall efficiency of the sales process, major advances in sales force productivity require not only access to relevant data, but also informative, predictive analytics derived from this data.

In this paper we develop analytical approaches to address two issues relevant to sales force productivity and describe the deployment of the resulting solutions within IBM. The first solution addresses a common problem faced by sales representatives: identifying new sales opportunities at existing client accounts and at noncustomer, or *whitespace*, companies. The analytical challenge is to develop models that predict—based on analysis of previous transactions and other available third-party data—the likelihood or propensity that a company will purchase an IBM product. These modeling results, with the underlying data, have been integrated in a Web-based tool called OnTARGET.

A second but related business challenge is to provide quantitative insight into the process of allocating sales representatives to the best potential revenue-generation opportunities. In particular, we are interested in the allocation of resources to existing IBM client accounts. Here the analytics challenge is to develop models that estimate the true revenue potential (or opportunity) at each account within IBM product groups. These models were developed as part of an internal initiative, the Market Alignment Program (MAP), in which the model-estimated revenue opportunities were validated by means of extensive interviews with frontline sales teams. This process and the Web-based MAP tool are described later in this paper.

Although they address different business problems, the OnTARGET and MAP tools share a common architecture. Both employ a data model that effectively joins historical IBM transaction data with external third-party data (such as revenue and number of employees, which we refer to as *firmographic* data), thereby presenting a holistic view of each client. Both systems exploit this linked data to build the models described here. Given the different business objectives, the tools employ different Web-based user interfaces; however, both interfaces are designed to facilitate easy navigation and location of the relevant analytical insights and underlying data.

ONTARGET: A CUSTOMER TARGETING SOLUTION

We begin with a discussion of the OnTARGET business objectives, describe the overall system design and the data model developed to meet these objectives, and conclude with a description of the Web-based user interface to the OnTARGET tool.

OnTARGET business objectives

Corporate revenue is likely to grow at normal rates for the next few years. Because the broad market is likely to grow in aggregate at rates only slightly higher than the gross national product, companies will need to generate revenue growth (excluding acquisitions and divestitures) at rates greater than the overall market to remain competitive. Pursuit of growth opportunities in emerging markets is one approach, but companies will need to generate significant growth in their core businesses and markets as well. This requires a renewed focus on identifying and closing new sales opportunities with existing clients and finding new companies that will be receptive to the core offerings. Improving sales-force productivity is essential to both objectives.

Early in the OnTARGET project, we spoke to a number of leading sales professionals and sales leaders about potential IT-enabled tools that they believed could enhance sales productivity. One common sentiment was that salespeople are often forced to use multiple tools and processes that not only fail to provide the relevant information needed to do their jobs better, but also take valuable time away from actual sales activities. In contrast, the same people were open to using a new tool, provided that the tool offered the following capabilities:

1. References a large universe of existing clients and potential new clients
2. Incorporates relevant data that may require multiple existing tools to access
3. Includes analytical models to help identify the best sales opportunities
4. Integrates all such data for each company under a single user interface designed by end users to facilitate easy navigation

These were the primary objectives for the design of the OnTARGET tool. In the remainder of this section, we discuss specific design decisions and implementations in light of these requirements.

Architecture and data

In this section, we provide an overview of key architectural aspects of the OnTARGET Web-based application.

Design objectives and system overview

The broad business objective for OnTARGET is to provide sales professionals with a single user interface through which they can obtain relevant data and analytics upon which they can act. From a design perspective, this requirement led to decisions on the specific data and linkages to be incorporated and the criteria for specifying the universe of companies to be made available within the tool. After discussions with sales professionals, the following sources of data were selected for inclusion:

1. All transactions executed by IBM with its clients over the past five years
2. Dun & Bradstreet (D&B**) firmographic data,[1] such as company revenue, number of employees, and corporate organizational hierarchy
3. Information on installed hardware and software at IBM client sites
4. Contact information for both customers and noncustomers
5. Competitive information from external vendors
6. Assignments of companies to sales territories.

It is required that OnTARGET include all significant IBM clients and potential new customers drawn from the universe of companies available in the D&B database. Using the historical IBM transactional data, we select client companies for inclusion based on a minimum threshold of their spending with IBM over the past five years. Noncustomer prospects are selected by filtering for minimum thresholds for company sales and the number of employees, based on the D&B data. Using these criteria, OnTARGET currently contains well over one million D&B company sites for the United States; over two million sites are included worldwide.

OnTARGET was developed with a Web-based front end that was flexible enough to allow end users to execute complex queries directly from the user interface. OnTARGET is implemented as a Java** application running on an IBM WebSphere* application server, with IBM DB2* as the relational database. A response time of less than 7 seconds is required for all transactions executed.

The OnTARGET architecture (***Figure 1***) can be viewed as three key elements: the data store (comprised of a production database and a staging database), the analytical models, and the user interface. The architecture somewhat isolates these elements to provide flexibility during development and deployment. It also allows the transformation and updating of data to occur in the staging database area, with subsequent deployment to the production database. These operations are resource-intensive; thus, executing them outside of the production environment eliminates any impact to the production application. The analytical models are developed outside the OnTARGET system and are imported to the staging server and integrated with the other data sources. Separate cross-sell rules are specified by salespeople and are integrated in much the same way as the analytical models.

Data model

The principal design objective of the OnTARGET data model was to facilitate the support and maintenance of key data drawn from multiple data sources across all major geographic regions. Because some of the data from each geographic region came from disparate data sources, commonality of data elements had to be designed into the model. For instance, contact data from one source could have different fields and lengths than data from another source, or an element might have a common field name with different domains. An analysis of the domain and length of each data element was performed to ensure that a common data model could be created to allow the user interface to work more efficiently and to standardize queries and provide a standard code base worldwide. All relevant pieces of data from each of the required entities were gleaned and assembled in a computer-aided software engineering (CASE) tool with which a logical and physical data model was designed. OnTARGET used IBM Rational* Data Architect[6] for its CASE tool.

OnTARGET was initially deployed to several countries in the Americas, followed by 15 countries in Europe and 3 in the Asia Pacific region. Hence, another key requirement of the common data model was that it readily support integration of new countries as data became available. The standardization of data structures allowed the user interface to remain untouched in many instances, even as additional countries were being added.

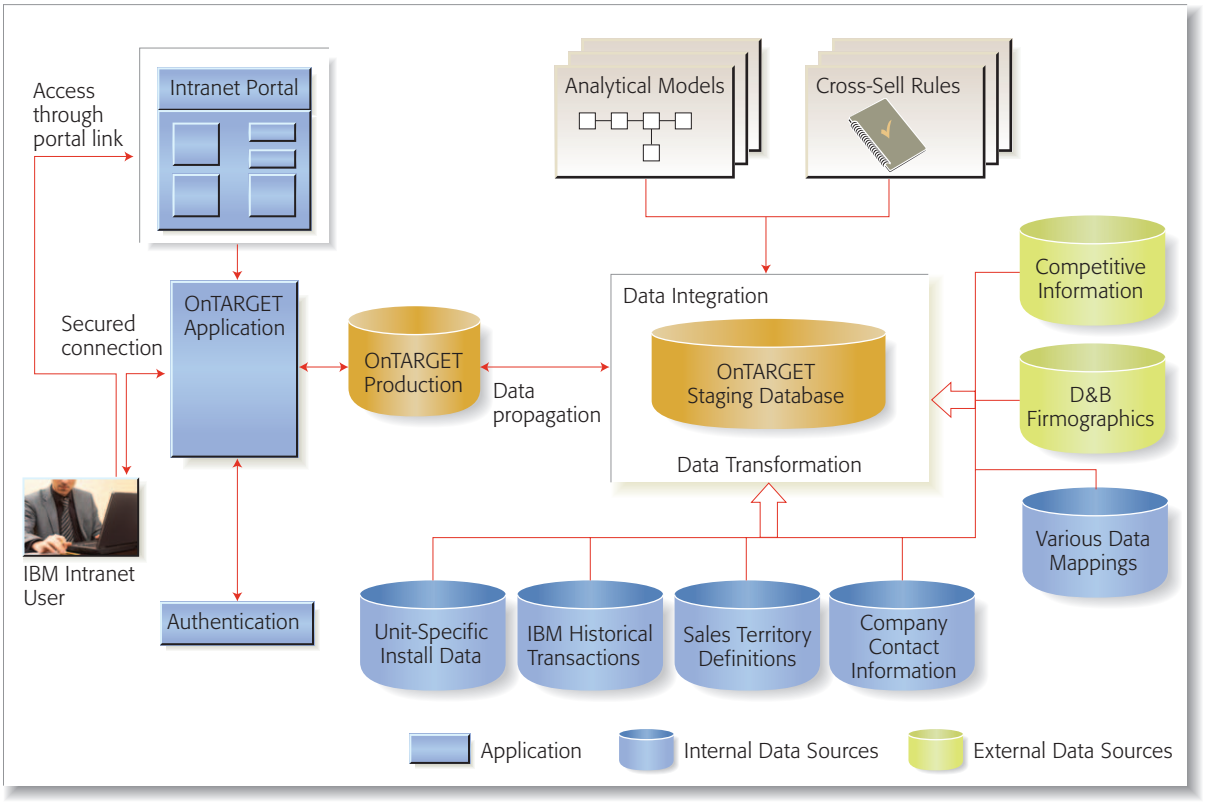

Figure 1
OnTARGET architecture and data

OnTARGET uses both internal IBM data (e.g., transactional data) and external reference data (e.g., D&B firmographic information along with competitive data from multiple vendors). Because IBM uses an internal reference number to identify customers, it was necessary to introduce a unique database key in order to join internal data with the external reference data for each company. An external reference number, the D-U-N-S** number (Data Universal Numbering System),[4] was chosen as the main key primarily because OnTARGET also includes companies that are not currently IBM customers. We developed a flexible process to transform all data to this common key.

Transformation algorithms were developed by using a transformation tool, WebSphere DataStage*,[7] to allow for consistent data presentation within the application. This helped to give the OnTARGET user interface a more consistent look and feel, regardless of the geographic location in which it was being used. This tool also helped document the data flows within the application and was useful for ongoing maintenance and training.

Major updates to the OnTARGET data are made each quarter. During each update cycle, the historical IBM transactional data is refreshed, and updated populations of noncustomers are extracted from the D&B tables. All models are rebuilt using this data, and hence the predictive-model scores are always consistent with the latest financial and firmographic data. Updates to the other information, including company contact information and product installation records, are made more frequently.

OnTARGET user interface

The purpose of the OnTARGET user interface is to help sales personnel quickly identify the best potential revenue opportunities in their sales territory. *Figure 2* shows a simplified, conceptual view of the OnTARGET user interface. The basic objective is to allow the user to build a focused customer-targeting list composed of companies that meet

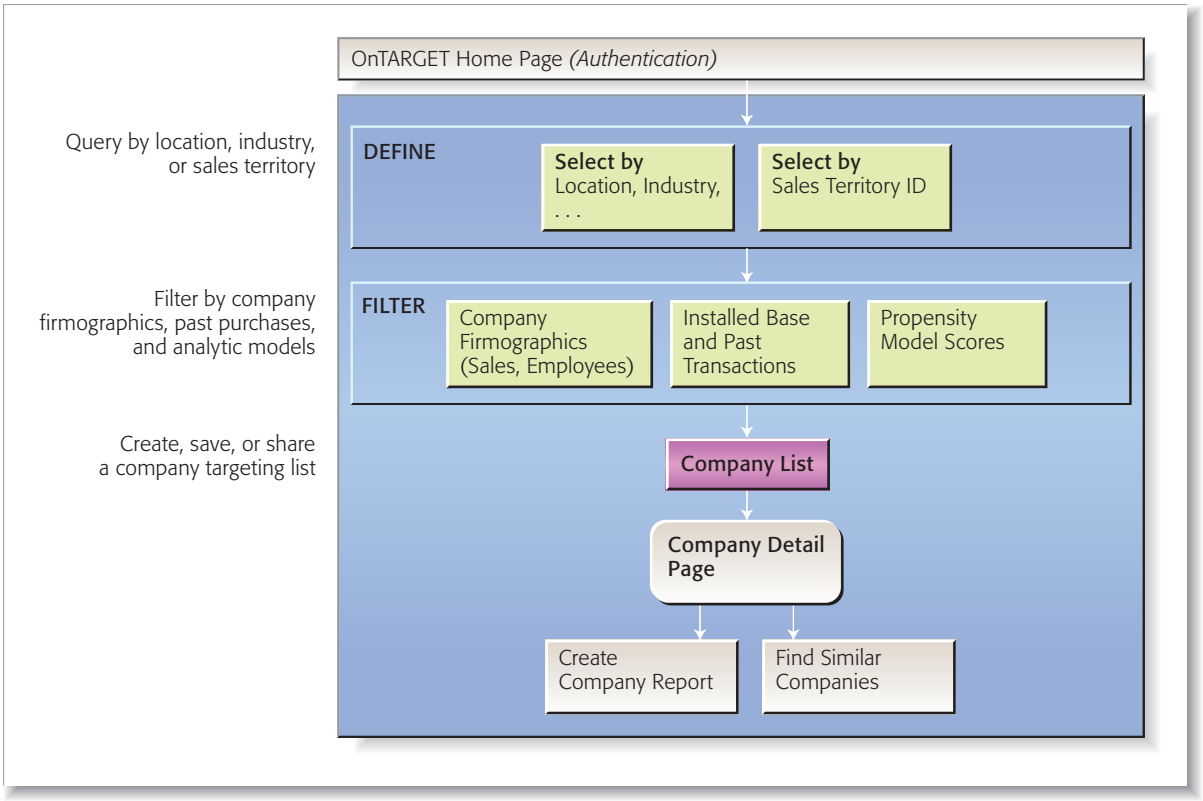

Figure 2
OnTARGET user interface

criteria specified by the user. In the first step, the user creates a broad set of companies by filtering for location, industry, or sales territory, or a combination of criteria.

In the second step, additional criteria can be used to further filter the initial set of companies. For example, it is possible to filter the list based on the upper and lower limits of company size as given by the D&B values for company sales and number of employees. It is possible to select only companies that have made purchases within IBM product groups (e.g., Lotus* software). Furthermore, using the OnTARGET propensity models, a user can select companies that have a high propensity to purchase in one of 10 different IBM product groups (e.g., Tivoli* software or System x* servers). The interface allows Boolean operators like AND, OR, and NOT to be applied in specifying the query. The various selection criteria are easily entered in the interface by means of standard pull-down and selection menus.

The result of the above define-and-filter process is the creation of a query that is executed against the OnTARGET database. All companies that meet the specified criteria are displayed as the resulting targeting list. This list can be further modified by adding and removing companies directly or by modifying the filter criteria in an iterative process that yields a list of key potential opportunities as a focus for the sales process. Selecting any company in the list takes the user to a company detail page, which includes a comprehensive view of all the information in the OnTARGET database about the company; for example, D&B firmographics, contact information, installed base, competitive information, and propensity scores. This holistic view facilitates the sales process by providing all relevant information in one place, and the user can easily generate a report of this information.

An interesting feature incorporated in OnTARGET is the capability to identify companies that are similar to a target company. *Similarity* is defined by a

distance metric constructed using only firmographic information, for example, companies in the same industry with comparable sales and numbers of employees. This feature is useful in two ways: to identify additional sales prospects and to analyze the IBM product mix that was purchased by other companies of comparable size in the same industry.

The targeting list can be saved for future reference or used as a basis for applying other criteria. OnTARGET also provides a capability for sellers to collaborate by sharing the targeting lists with others. Users can receive targeting lists and then refine the filters to meet their specific requirements. In many cases, this function enables persons in sales operations to define criteria and pass them on to representatives in their region.

An essential feature of the user interface is the enforcement of appropriate security and privacy rules to ensure that all information is protected according to IBM and country-specific policies. This capability is managed from a separate administrative interface that allows the specification of rules to restrict display of sensitive data only to users with the appropriate authorization.

OnTARGET includes the capability to collect usage statistics, such as the number of logins by each user. It also records time-stamped user accesses to each company detail page. This data is essential to quantify both the acceptance of the tool and the extent to which subsequent revenue for a specific client can be linked to the use of the tool. We discuss these metrics in the section "Solution deployment and business impact."

ONTARGET PROPENSITY MODELS

OnTARGET and MAP employ different predictive models based on their business objectives. For OnTARGET, propensity models predict the probability of purchase within a specific product group; the MAP models estimate the potential revenue opportunity at each client account. The MAP models are described in a subsequent section.

The goal of the propensity models is to differentiate customers (or potential customers) by their likelihood of purchasing various IBM products. These models are built to predict purchases within broad product groups or brands rather than at the level of individual products. Examples of these brands are Lotus or Tivoli (IBM software brands) and System p* or IBM System Storage* (IBM server brands). Currently, we develop separate propensity models for 10 product groups.

We have at our disposal several major data sources to use in this task. The two major sources, available for the largest number of companies, are the historical IBM transactions for all IBM customers and publicly available firmographic data from D&B and other sources. Our goal is to make use of this data to build propensity models that consider all potential customers, are widely applicable, and accurately differentiate between high-propensity and low-propensity customers on a product-by-product basis. For this purpose, for every product brand Y, we first divide the universe of OnTARGET companies into three distinct groups:

1. Companies that have already purchased Y in the past. These companies are eliminated from the propensity modeling.
2. Companies that have a relationship with IBM but have never purchased Y. For these companies, we can use both data sources—internal and external—to build our existing customer model.
3. Companies that have never purchased from IBM. For these companies we have only the firmographic data, and the model for them is called the *whitespace model*.

With multiple geographic areas (the Americas, Europe, Asia Pacific), multiple countries within each geographic area, and multiple product brands, a large number of propensity models (currently about 160) are built in each quarter. In what follows, we summarize our modeling approach and the considerations leading to it, demonstrate its evaluation process during modeling, and show results of actual field testing. Finally, we discuss the modeling automation put in place to handle the overwhelming number of models built each quarter.

Propensity-modeling methodology

We begin by specifying a geographic area, a brand Y, and a modeling problem (existing customer or whitespace). Our first step is to identify positive examples and negative examples to be used for modeling. In each modeling problem, we try to understand what drives the first purchase decision for brand Y and to delineate companies by the likelihood of their purchase. Assume the time period

t (typically last year or the last two years), and our modeling problem is formulated as:

Differentiate companies that had never bought brand Y until period t, then bought it during period t, from companies that have never bought brand Y.

Of the companies that had never bought brand Y before period *t*, some will have bought other products before *t*. These companies form the basis of the existing customer model for Y. The companies that had never bought any brand before *t* are the basis for the whitespace model. Thus, for the whitespace problem, our positive and negative examples are:

Positive: *Companies that had never bought from IBM before t, then bought Y during t.*
Negative: *Companies that had never bought from IBM before or during t.*

The definitions for the existing customer problem are similar, except that a previous purchase from IBM is required for inclusion. For some combinations of geographic area and brand and modeling problem, the number of positives may be too small for effective modeling (we typically require at least 50 positive examples to obtain good models). In that case, we often choose to combine several similar modeling tasks (where similarity can be in terms of geographic area, brand, or both) into one meta-model with more positives. In Reference 8, we discuss in detail the trade-offs involved in this approach and demonstrate its effectiveness.

Next, we define the variables to be used in modeling. For existing customers, we derive multiple variables from historical IBM transactions, describing the history of the IBM relationship before period *t*. Examples of these features are the following:

- Total amount spent on software purchases in the two years before *t*
- Total amount spent on software purchases in the two years before *t* compared to other IBM customers (rank within IBM customer population)
- Total amount spent on storage-product purchases in the four years before *t*

For both existing and whitespace customers, we derive variables from the D&B firmographic data such as the following:

- Company size indicators (revenue, employees), both in absolute and relative terms (rank within industry)
- Industry variables, both raw industry classification from D&B and derived sector variables
- Company location in corporate hierarchy (e.g., headquarters or subsidiary)

We then build a classification model (more accurately, a *probability estimation model*), which uses these variables to differentiate the positive examples from the negative examples. For each example, the model estimates the probability of belonging to the positive class. For presentation in the OnTARGET tool, these continuous scores are binned from 1 to 5, with bin distributions specified such that only 15 percent of existing customer examples receive the highest rating of 5. For the whitespace model, only five percent get a rating of 5, reflecting the observation that it is generally more difficult to sell into a noncustomer account.

Example of modeling results

The resulting models can be examined and, to some extent, interpreted as scorecards that describe the effect of different variables on the likelihood of converting a company into a customer for brand Y. We describe here a detailed example from a recent round of existing customer models built for North America and discuss possible interpretations.

The example we give is the existing customer model for the Rational software brand. ***Figure 3*** shows the predictive relationships found for this model. Green arrows signal a positive effect (an increase in propensity from an increase in the value of the variable), and red arrows signal an adverse effect on propensity. The width of the arrows indicates the strength of the effect, as measured by the magnitude of the regression coefficient. We show in the figure only statistically significant (as measured by p-value) effects. We see several interesting effects, and most seem to be explainable:

- Industrial sector (IT), geographic area (California) and company's corporate status (Headquarters) seem to have a strong predictive effect. This seems consistent with Rational being an advanced software development platform, which medium-sized IT companies in California (and thus, likely at the leading edge of technology) might be interested in purchasing.

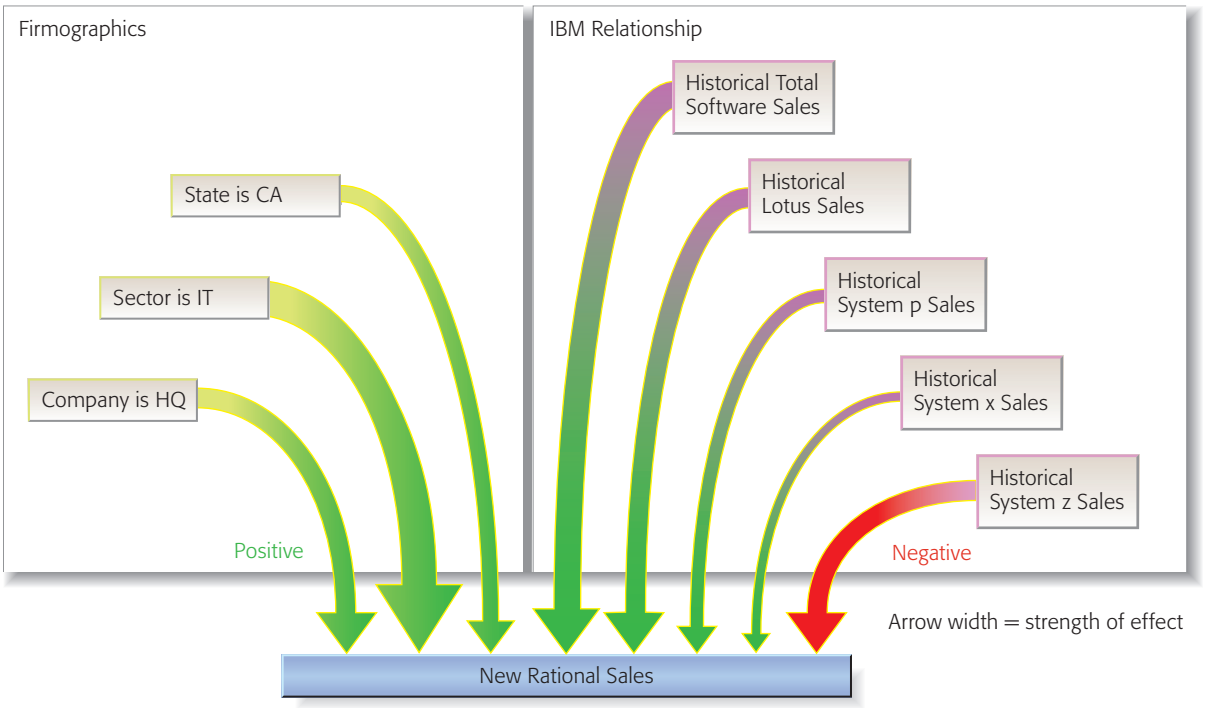

Figure 3
Predictive relationships between some derived variables and new Rational sales to existing customers

- The size of the total prior software relationship with IBM seems to be a strong indicator of propensity to buy. Additionally, having a strong relationship in Lotus seems to afford additional power.
- Although the total size of prior nonsoftware relationship does not have a strong effect, some specific nonsoftware brands seem to be important. System p (and System x, somewhat) seem to promote Rational sales, while IBM System z* seems to dampen them. Although this last fact may seem puzzling, it may be explainable by the fact that System z customers often manage their software relationship with IBM in conjunction with the System z relationship. More analysis would be required to clarify this point.

Evaluating model performance

Statistical model validation is performed using a 10-fold cross-validation approach, whereby we divide our data (positive plus negative examples) into 10 equal-sized bins and build 10 models, each time using nine-tenths of the data (nine of 10 bins) for modeling, and then applying the resulting model to the *leave-out* tenth bin. After repeating this 10 times, we have all data scored as leave-out by the different models, and we can use it to evaluate the modeling success. (For a detailed description of cross-validation, see Reference 9). We then evaluate our model by the lift performance on the holdout data. Lift of the model at percentile x is defined as

$$\text{Lift}\,(x) = \frac{\text{fraction of total positive examples in top } x\%}{\text{fraction of positives plus negatives in top } x\%}.$$

As the denominator is simply x, we can write

$$\text{Lift}\,(x) = \frac{\text{fraction of total positive examples in top } x\%}{x}.$$

The lift is a natural measure in the marketing context because it measures how much more successful our model is than a model that simply assigns random scores. The lift is also quite robust with respect to the ratio of positive-to-negative examples used, which is important because our learning samples are typically biased in favor of positive examples, compared to the full population. (For discussion of these biases and their effect on evaluation, see Reference 10).

To seriously evaluate the success of a model, we should always judge it against a reasonable baseline model that a knowledgeable salesperson might employ, rather than against the random model. For this we adopt a baseline model that ranks prospects by a measure of company size. We refer to this as the *Willy Sutton* model. (Willy Sutton was the infamous bank robber who reportedly said that he robbed banks because "that's where the money is.") We rank existing customers by the size of their relationship with IBM (largest to smallest, based on total revenue with IBM), implicitly assuming the largest customers are most likely to buy a brand in which they have no current relationship. For whitespace companies, we rank them by their company size (revenue and employees) as reported by D&B.

Our cross-validated evaluation indicates that our models almost invariably do significantly better than the Willy Sutton model. For the most recent existing customer problems, our models do significantly better than the Willy Sutton model on nine out of 10 problems, with the sole exception being System z, for which the performance is only slightly better. Indeed, this may not be surprising, as this is the brand where size of IBM relationship is indeed most likely to be critical for new purchases. A common graphical display, representing the lift of a model at all values of x, is the lift curve. ***Figure 4*** illustrates an example lift curve for the whitespace model built for the Rational software brand. The lift at 5 percent and 10 percent is calculated explicitly, and the model is compared to the Willy Sutton and random models. Note that the OnTARGET model significantly outperforms both baseline methods.

A more interesting evaluation, however, is to judge the models by their actual success in predicting new sales. We have been able to do this by considering new sales recorded in 4Q 2006 and investigating the scores that our previously built models in 3Q 2006 assigned to these sales. At the time these models were built, these sales were not visible in the data; however, they were most likely initiated before the results of the models were available, and thus not affected by these results. Hence, we are getting a clean evaluation of the success of the models in identifying actual sales as high propensity opportunities.

Table 1 shows the evaluation results for the 10 existing customer models. We compare the performance of our model to that of the Willy Sutton model in terms of the area under the curve (AUC) and the lift at the 5-percent point (Figure 4). For each modeling problem, the higher number (better performance) is shown in boldface.

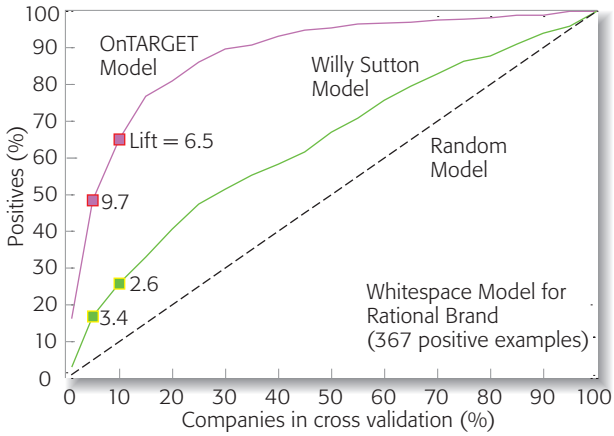

Figure 4
Example lift curve comparing OnTARGET model performance to some baseline models

Table 1 Model evaluation results based on predicting new sales

Product Brand	Number of Positives	AUC		Lift at 5%	
		Model	Willie Sutton	Model	Willie Sutton
Information Management	26	**0.82**	0.69	**9.23**	6.15
Lotus	25	**0.81**	0.72	7.20	**8.80**
Rational	42	**0.91**	0.79	**12.38**	10.00
Tivoli	56	**0.82**	0.77	8.21	**8.57**
WebSphere	37	**0.90**	0.80	12.43	12.43
System i*	69	**0.77**	0.73	**5.22**	2.90
System p	74	**0.83**	0.78	**8.92**	8.38
System x	27	**0.80**	0.74	7.41	**9.63**
System z	6	0.78	0.78	6.67	6.67
Storage	194	**0.88**	0.80	**10.31**	6.91

We observe that for nine of the 10 modeling problems, our models do better than the baseline model in terms of AUC, which is between 0 and 1 and roughly measures the success of a model in ranking all the data successfully. The tenth,

System z, has very few positives, and the models end up in a tie. In terms of lift at 5 percent, we observe that both modeling approaches do very well (often finding as many as 50 percent of sales in this top portion), but our models generally outperform the baseline models, sometimes significantly. Overall we can conclude that our models clearly do a better job of identifying sales opportunities, but that this advantage is less pronounced at the top of the ranked lists (top 5 percent), where the Willy Sutton policy of just taking the biggest customers seems to be a reasonable approximation of the best model we can build.

Model automation

Because we generate at least 160 new models each quarter, a reliable, repeatable, and automated modeling methodology is needed. Initially, we performed several modeling iterations manually to understand the problems and variables involved. Then we created an automated system whose main characteristic was that it had a large collection of possible predictive variables from which it selected some for each prediction model. The main considerations in choosing variables for each model were:

- the number of positive examples (if there were too few examples; we could not use too many variables),
- our experience with the specific modeling problem (some variables are more important than others in specific geographic areas or for specific brands, or both), and
- data availability (some variables were not available in specific geographic areas).

This almost fully automated system generated the models in the two quarters preceding this writing. The main need for manual intervention in this process is to examine the output and evaluate the models, thereby making sure that changes in data, bugs, or other unexpected phenomena have not adversely affected the predictive performance.

MARKET ALIGNMENT PROGRAM

The second initiative, MAP, is used to allocate the sales force. An integral part of the MAP process is the validation of analytical estimates by means of an extensive set of workshops conducted with sales leaders. These interviews rely heavily on a Web-based tool to convey the relevant information and to capture expert feedback on the analytical models.

MAP business objective

Sales organizations face the challenge of effectively aligning their sales force with market opportunity. The objective of the MAP initiative is to address this challenge by focusing on the three main problem areas in sales-force deployment methodology:

1. Lack of a uniform, disciplined approach to estimating revenue opportunity at a customer level. This problem leads to alignment of sales resources with past revenue rather than the opportunity for future revenue. MAP links the sales-force allocation process to field-validated analytical estimates of future revenue opportunity in each operational market segment.
2. Lack of frontline input into the planning process. Purely top-down planning processes, although easier to manage, do not typically result in an optimal allocation of sales resources and revenue targets. They also often result in disenfranchised sales teams. The MAP business process explicitly requires detailed input from the frontline sales teams.
3. Lack of an easily accessible common base of facts and analytical methodology for making resource shift decisions. This problem limited the scope and impact of the previous deployment optimization efforts because sales leaders in different parts of an organization found it difficult to arrive at rational fact-based trade-off decisions in the absence of a common fact base. MAP solves this problem by delivering the properly aggregated information to decision makers through a Web-based tool.

MAP business process

The MAP business process can be broken into four main steps:

1. Prepare input for frontline sales workshops. This step includes populating the Web-based tool with data on past revenue, model-estimated revenue opportunity, and deployment of sales resources.
2. Conduct frontline workshops. This is the most time-consuming phase of the planning process. In this phase we validate the model-estimated future revenue opportunity per customer by product division, validate current coverage, and capture future resource requirements. All workshops are conducted using the MAP tool.
3. Conduct workshops with regional sales leaders within each product division and each industry

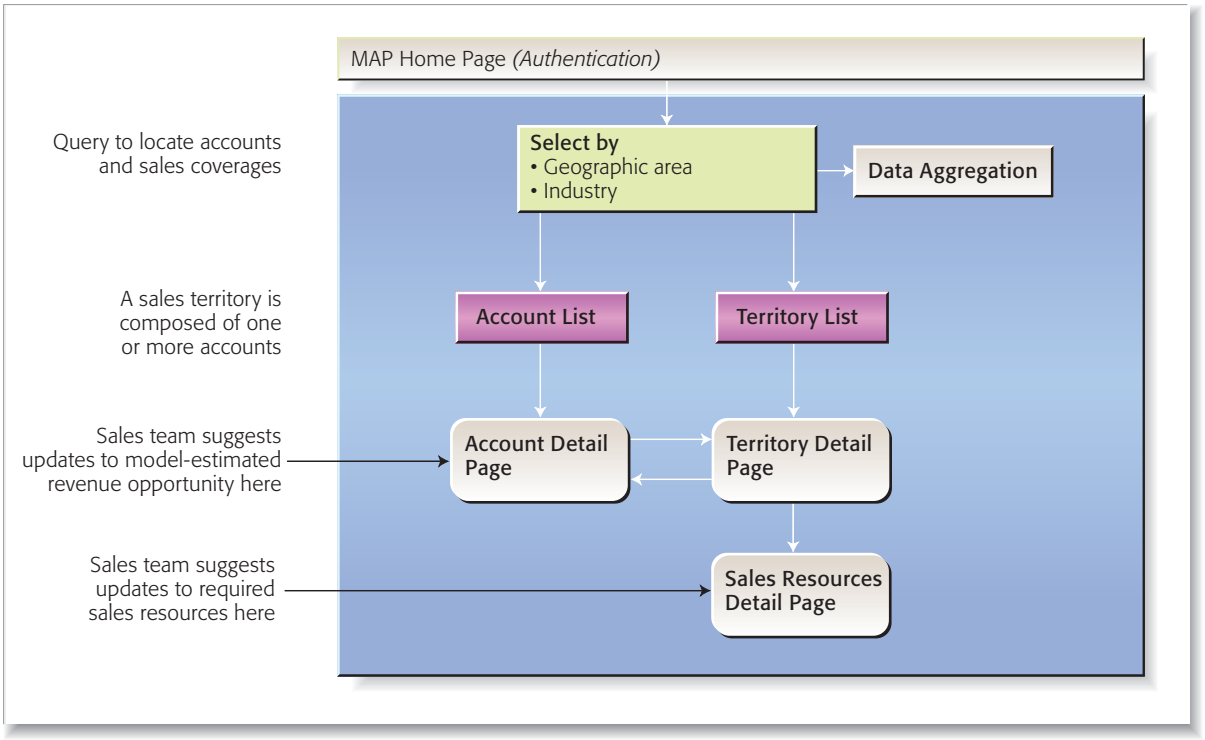

Figure 5
MAP user interface

sector. In this phase we prioritize customers within each product division and industry sector and validate coverage requirements.

4. Conduct regional summit meetings with the sales leaders from all product divisions and industry sector teams. In this phase we develop the overall IBM sales coverage and strategy for priority customers.

System overview

The MAP Web-based tool is used to conduct extensive interviews with IBM sales teams, and this process has motivated several design features that differ from OnTARGET. In addition to many of the data sources used in OnTARGET (Figure 1), MAP also includes data on assignments of current sales resources. The MAP revenue-opportunity models are built for each major IBM product brand by using combined transactional and D&B data. Revenue opportunities are estimated for each account, and these results are stored in the database for display during the interview process. Unlike OnTARGET, the MAP tool must capture specific feedback on the revenue-opportunity models presented to the sales teams during the interview process. The tool allows the sales team to input their estimates of revenue opportunity and explain their reasons for recommending a change to the model results. Opportunities validated this way are stored in the MAP database and then post-processed with separate tools after the interviews have been completed.

MAP data model and user interface

Figure 5 shows the overall flow of the MAP user interface. It is essential that participants in the MAP interview process be able to locate their accounts and sales territories within the tool. For this reason, the query interface shown in Figure 5 returns either a list of accounts or a list of sales territories that satisfies query conditions such as geographic area and industry sector. In contrast, the analogous queries in OnTARGET (Figure 2) return lists of companies that are indexed by the D&B D-U-N-S number. Hence, the MAP data model is organized by IBM client accounts, whereas the OnTARGET data model uses the D&B representation of companies. The difference in design is due to different business objectives: OnTARGET must provide easy identification of noncustomer companies, whereas MAP must support an IBM-centric view in order to be

consistent with the account- and territory-based views of the participants in the MAP interview process.

Returning to Figure 5, the generated lists of accounts and territories contain links to pages that summarize all relevant information for each account or sales territory. For example, the account detail page shows a five-year summary of all IBM revenue for each major IBM product brand and the estimated revenue opportunity for each brand. Feedback on the estimated revenue opportunities is collected by means of user inputs on this page. The sales team can also enter changes to existing numbers of sales resources that might be required to achieve future revenue levels; note that the data is entered at the sales-territory level (on the territory detail page) because sales resources are allocated at this level. To facilitate regional and industry-based workshops, the MAP tool also provides a capability to aggregate data such as revenue opportunity within market segments, for example, within a specific geographic area or industry.

MAP REVENUE-OPPORTUNITY MODELS

The total amount of money a customer can spend on a certain product category is a vital piece of information for planning and managing sales and marketing efforts. This amount is usually referred to as the customer's wallet or the revenue opportunity for this product category. For the MAP workshops, we needed an unbiased, realistic estimate of the true revenue opportunity at each account for the purpose of leading an informed discussion with each sales team. Although in this section we develop ideas related to revenue opportunity, modeling approaches, and evaluation of the models in the context of IBM revenue opportunity with its clients, the methodology is applicable to any company with a large volume of historical transaction data.

Definition of opportunity

What do we mean by a customer's wallet or opportunity? We discuss this in the context of IBM as a seller of IT products to a large number of customers for whom we wish to estimate the wallet. We have considered three nested definitions:

1. The total spending by this customer in a particular group of IT products or services. This is simply the total IT spending (by product group) by the customer. We denote this as the *total opportunity*.
2. The total attainable (or served) opportunity for the customer. In the IBM case, this would correspond to the total spent by the customer in IT areas covered by IBM products and services. While IBM serves all areas of IT spending (software, hardware, and services), its products do not necessarily cover all needs of companies in each of these areas. Thus, the *served opportunity* is smaller than the total opportunity.
3. The realistically attainable opportunity is defined by what the best customers (as defined below) spend. This is different from served opportunity because it is not realistic to expect individual customers to spend their entire budget with IBM. We refer to this as *realistic opportunity*. This is also the definition that we use to define revenue opportunity for MAP.

Total company revenue is readily available for all companies from sources like D&B. We also know the total amount of historical sales (IBM sales) made by IBM to its customers. In principle, the relation

$$\text{IBM sales} < \text{realistic} < \text{served} < \text{total} < \text{company revenue}$$

should hold for every company. Note that we expect IBM sales to approach the realistic opportunity for those companies where IBM is the dominant IT provider.

As noted above, MAP uses the realistic definition because it is most consistent with the concept of opportunity held by IBM sales executives. Defining the best customers is essential to estimating realistic opportunity. In what follows, we define *best customers* in a relative sense, as those who are spending with IBM as much as we could hope for given a stochastic model of spending. Thus, a good customer is one whose spending with IBM is at a high percentile of its spending distribution. We describe below some approaches that allow us to build models that predict such a high percentile and, moreover, allow us to evaluate models with regard to the goal of predicting percentiles of individual spending distributions.

Modeling realistic opportunity as quantiles

Under the realistic definition, we are looking for a high percentile (e.g., 80 percent) of the conditional spending distribution of the customer, given all the

information we have about this customer. As an example, consider a customer C_x and imagine that we have not just one customer, but a million customers identical to C_x, except that each customer independently makes its decision about how much it spends with IBM. We could then take just the 80th percentile of this spending distribution (i.e., the quantity O such that 80 percent of these one million identical customers spend \$$O$ or less with IBM) as our realistic wallet estimate for C_x and its identical counterparts. In practice we do not, of course, observe multiple copies of each company, and so our challenge is to get a good estimate of this conditional spending percentile for each company from the data available to us. In general, the approaches for doing so can be divided into two families:

1. Local approaches, which try to take the idea described above (of having a million copies of C_x) and approximate it by finding companies that are similar to C_x, and then estimate the realistic wallet of C_x as the 80th percentile of IBM sales in this neighborhood of companies.
2. Global models, which attempt to describe the 80th percentile as a function of all the information we have about our customers. The simpler and most commonly used approach is quantile regression,[11] which directly models the quantile (or percentile) of a response variable y (in our case, IBM spending) as a function of predictors x (in our case, the firmographics from D&B and the IBM historical transaction data).

   The standard regression approach estimates the conditional expected value $E(y|x)$ by minimizing the sum of squared error $(y - \hat{y})^2$. In the case of quantile regression, we have to find a model that minimizes a piecewise linear, asymmetric loss function known as quantile loss:

$$L_p(y, \hat{y}) = \begin{cases} p \times (y - \hat{y}) & \text{if } y > \hat{y} \\ (1 - p) \times (\hat{y} - y) & \text{if } \hat{y} > y, \end{cases}$$

   where p is the particular percentile. This loss function is appropriate for quantile modeling because the loss is minimized in expectation when the desired quantile is being perfectly modeled (for further discussion, see Reference 11).

Quantile estimation techniques

Within the scope of the MAP project, we explored a number of approaches for quantile estimation and developed some novel modeling techniques. We evaluated the different models both in a traditional predictive modeling framework on holdout data and against the expert feedback collected in the initial round of MAP sales-team workshops. We discuss in more detail below the two approaches most relevant to the MAP project, but we also note that our research has led to three additional techniques for wallet estimation:

1. *Quanting*,[12] which uses an ensemble of classification models to estimate the conditional quantiles.
2. *Graphical decomposition models*,[13] which assume that the IBM revenue is determined by two independent drivers: the relationship of the company with IBM and its IT budget (the opportunity), which itself is determined independent of the IBM relationship based on the IT needs of the company.
3. *Quantile tree induction*,[14] which follows closely the traditional tree induction algorithm[15] using a fast divide–and-conquer greedy algorithm that recursively partitions the training data into subsets. However, the objective of quantile estimation requires a number of alterations, including adjustments of the splitting and stopping criteria and the prediction of a quantile rather than the mean of the values in a leaf node.

Similar to the propensity models discussed in the previous section, we estimate revenue opportunity for each company at the major product-brand level by using the firmographics from D&B and the IBM historical transaction data.

k-nearest neighbor

The revenue-opportunity estimates provided in the first release of the MAP tool used a k-nearest-neighbor[16] approach, which follows very closely the definition of realistic opportunity. In particular, for each company we find a set of 20 similar companies, where similarity is based on the industry and a measure of size (either revenue or employees, depending on the availability of the distribution). From this set of 20 firms, we discard all companies with no IBM revenue in the particular product brand and report the median of the IBM product-brand revenues of the remaining companies. The choice of the median (50th percentile) reflects considerations of both the statistical robustness and total market

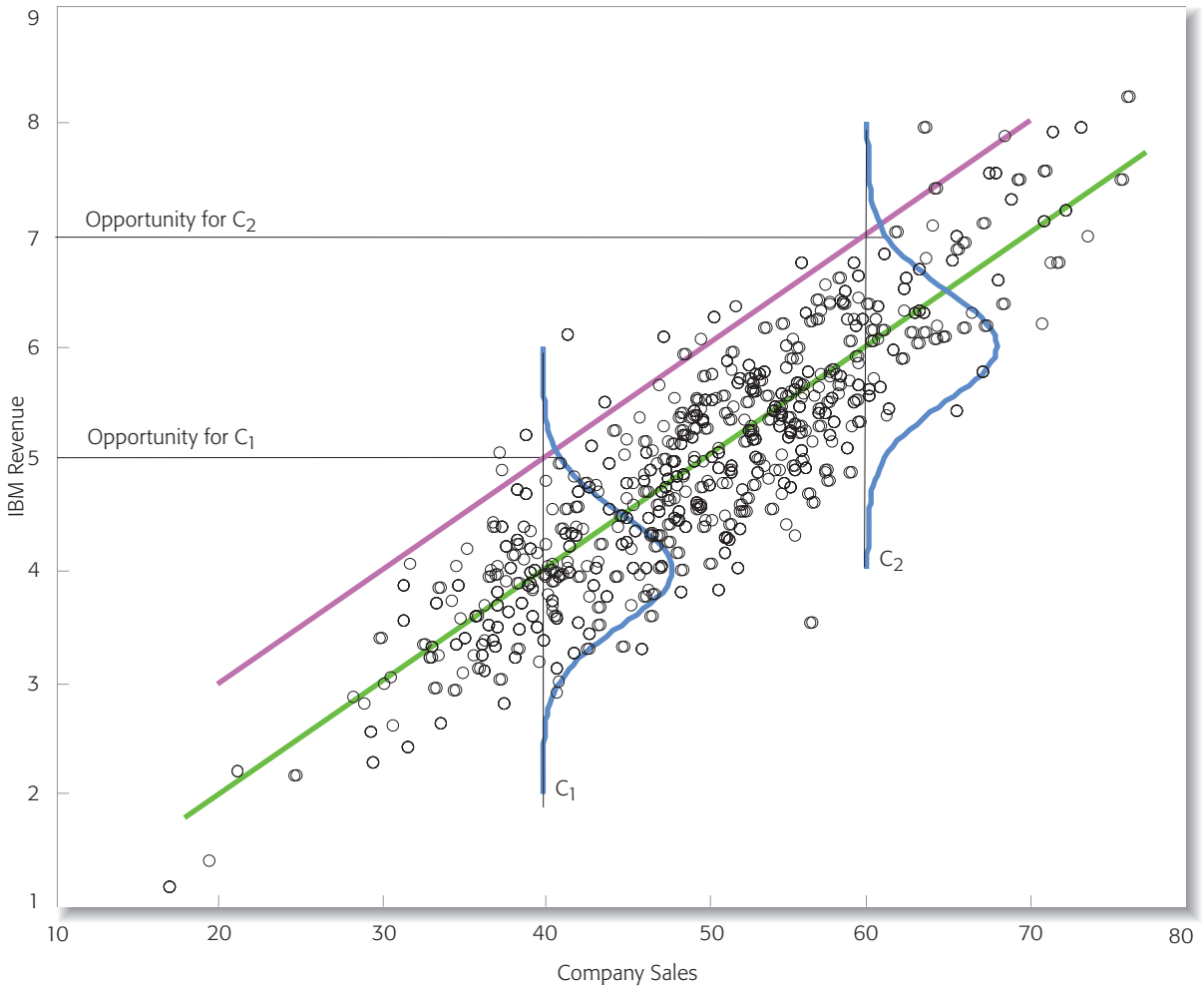

Figure 6
Comparison of quantile regression and linear regression

opportunity (sum over all companies) relative to the total IBM revenue.

Linear quantile regression

A standard technique to estimate the realistic wallet as percentiles of a conditional distribution is linear quantile regression.[11] Similar to standard linear regression models, quantile regression models aim to find a coefficient vector β such that $X\beta$ is close to Y.

The main difference between traditional linear regression and quantile linear regression is the loss function. While linear regression models the conditional expected value by minimizing the sum of squared error, quantile regression minimizes quantile loss as defined earlier. ***Figure 6*** shows conceptually the difference between the linear regression line in green and the quantile regression line for the 90th percentile in red.

Post-processing

Because the realistic opportunity is defined as a high quantile of the conditional distribution, the predicted opportunity will be smaller than the realized IBM revenue for some companies. In particular, for a quantile of 90 percent we would expect about 10 percent of companies to generate IBM revenue that is larger than our opportunity forecast. While we do not know the exact IBM revenue for the next year, we use the revenue of the previous year as a proxy and report in the MAP tool the maximum of the opportunity model and the revenue of the previous year in the brand.

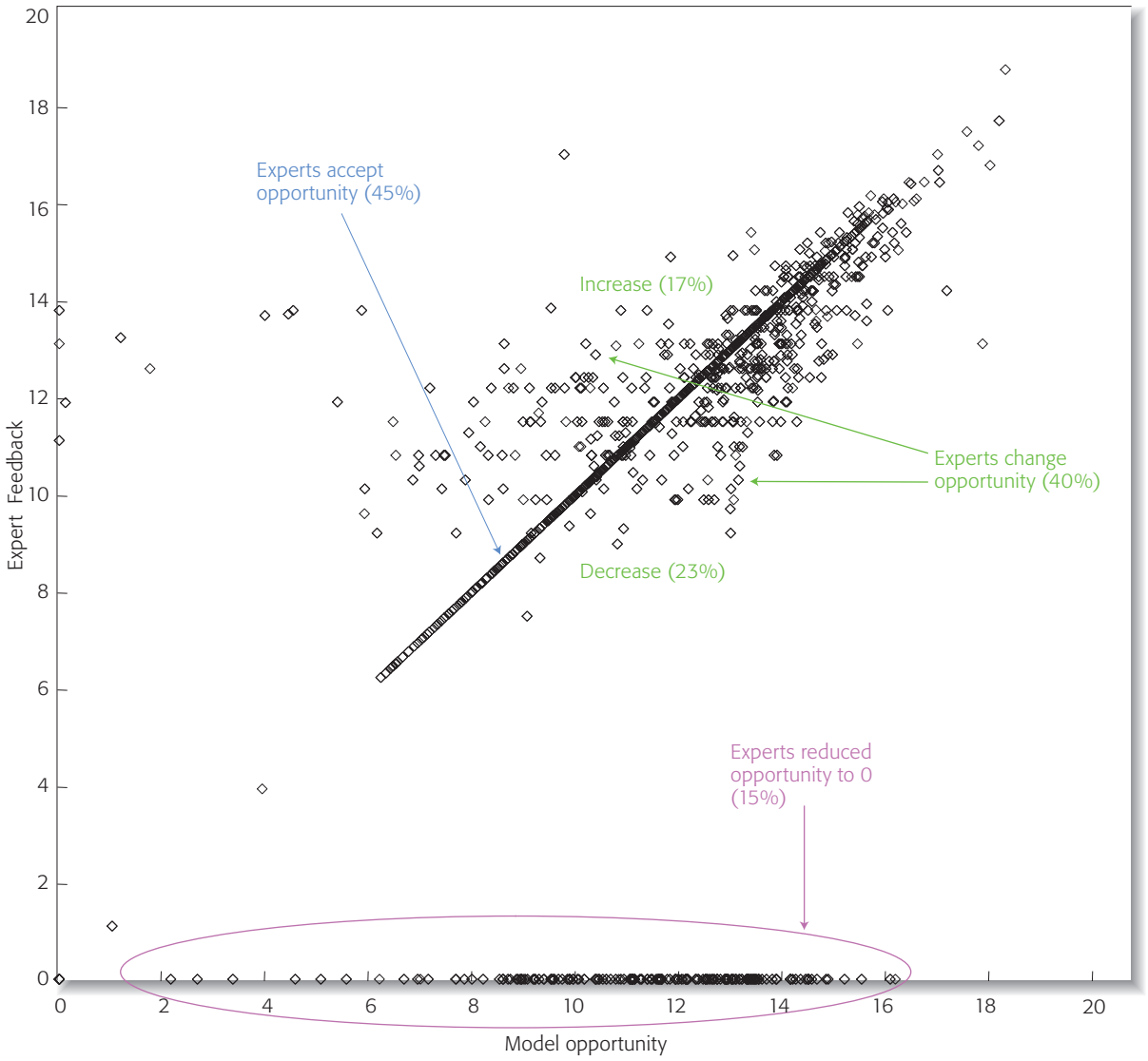

Figure 7
Comparison of predicted and expert-validated opportunity

Evaluation of opportunity models

To decide which method to implement in the MAP tool, we first needed to identify the most appropriate quantile for the opportunity estimation and then choose an appropriate evaluation criterion for model comparison. Because we never directly observe the true realistic IBM opportunity, we need a reference solution. Therefore, rather than use potentially unreliable survey data, we decided to use the expert feedback collected in the initial round of sales workshops conducted in 2005. In particular, we built various opportunity models using D&B firmographics and IBM revenues for the approximately 30,000 companies that compose the 6000 major MAP accounts discussed in the 2005 interviews.

Before describing these results, we discuss the expert feedback obtained during the 2005 MAP workshops. Recall that the experts could either accept the model estimate or revise it. ***Figure 7*** shows the validated (expert-specified) opportunity for a major IBM software brand for about 1200 accounts as a function of the original opportunity estimates that were provided in the MAP tool using the initial nearest neighbor model. The plot supports a number of interesting observations:

1. Forty-five percent of the opportunity estimates are accepted without alteration. The majority of the accepted opportunities are for smaller accounts. This shows a strong human bias toward accepting the provided numbers (this is broadly known as anchoring).
2. For 15 percent of the accounts, the experts concluded that there was no opportunity—mostly for competitive reasons that cannot be known to our revenue-opportunity model. For that reason we decided to exclude those accounts from the model evaluation.
3. Of the remaining 40 percent of accounts, opportunity estimates were decreased (23 percent) slightly more often than they were increased (17 percent).
4. The data pattern of horizontal lines reflects the human preference toward round numbers.
5. The opportunities and the feedback appear almost jointly normal in a log plot. This suggests that the opportunities have an exponential distribution with potentially large outliers, and that the sales experts corrected the opportunities in terms of percentage.

Given the high skew of the distribution, residual-based evaluation is not robust.[17] To account for this fact, we evaluated model performance on three scales: original, square root, and log. In addition to the sum of squared errors for each scale, we also considered the absolute error. This provides us with a total of six different performance criteria. For this analysis, we built nearly 100 different models, counting all variations of model parameters. We then ranked all models according to each of the six performance criteria and compared how often a given model appears within the top 10 of all models. Based on this analysis for three major product brands, we concluded that the linear quantile regression model showed the most consistently good performance for a quantile of 80 percent. In other words, this quantile regression model provided the best agreement with the expert feedback collected during the initial 2005 MAP workshops. Hence, this model was selected to provide the revenue opportunity estimates for MAP workshops conducted in 2006.

SOLUTION DEPLOYMENT AND BUSINESS IMPACT

An essential component of initiatives such as OnTARGET and MAP is that we be able to quantify the impact of the delivered solution against the overall business objectives. In general, it is challenging to isolate the impact of a given tool or process when it is injected into a broad, complex, and dynamic business environment. In this section, we describe several measures of business impact for each of these solutions.

OnTARGET business impact

We assess impact in four ways: adoption, productivity gains, impact of the propensity models, and pipeline influence.

Adoption

The user population for OnTARGET has grown steadily over the last two years, from about 1000 users at the end of 2005 to approximately 7000 worldwide users by the close of 2006. These sales professionals are in 21 countries across three major geographic areas. Interest in the application continues to increase, and the growth of the user community is anticipated to continue as new countries and sales personnel are added. As sales representatives are often reluctant to use tools that do not enhance their productivity, the adoption rate of OnTARGET is a significant measure of its impact, especially because use of the tool is not mandated.

Although the initial deployment was for the IBM Software Group division, the fact that the database includes information from most business areas has made the application useful as an enterprise-wide application covering multiple lines of business. Therefore, the application quickly moved to other divisions and has become an enterprise application. It is currently being used by sales professionals from the IBM Software Group division, the Systems and Technology Group division, and the Services organizations.

Productivity gains

Face-to-face and call-center sales personnel are the primary users of OnTARGET, as they look for the best potential opportunities in their space. During a recent survey of our user base, the average productivity gain identified was two hours per week. This productivity gain can be attributed to the fact that the user can quickly create focused targeting lists and does not have to use multiple tools to access additional data and research on prospective clients. OnTARGET users accessed and downloaded over 235,000 company-detail reports in 2006.

Impact of propensity models

An obvious question concerns the degree to which the propensity models provide quantifiable business impact. In the section "OnTARGET propensity models," we examined the scores assigned to closed opportunities (i.e., a sale within a product brand) by models built in the quarter before the sale was recorded. This analysis, summarized in Table 1, provides strong evidence that the models are indeed identifying new sales opportunities that lead to closed revenue.

Pipeline influence

OnTARGET was designed to help identify the best potential opportunities for salespeople; hence we have focused on measuring the impact on the sales opportunity pipeline. We define an opportunity to have been influenced by OnTARGET if the detail page for the prospective company was viewed within a three-month window prior to the opportunity creation date. By this metric, OnTARGET influenced over 17 percent of won opportunities in 2006 across the included lines of business worldwide. These OnTARGET-influenced opportunities represent more than 22 percent of the reported pipeline dollars. Opportunities that were influenced by OnTARGET generated greater average revenue than those in which OnTARGET played no role. For example, in the IBM software business, the average revenue associated with won opportunities influenced by OnTARGET in 2006 was more than 45 percent larger than those not influenced by the tool.

MAP business impact

We view the business impact of MAP from four perspectives: deployment and adoption, revenue growth, sales pipeline growth, and quota attainment.

Deployment and adoption

The MAP initiative has been widely adopted throughout IBM and plays a key role in the deployment of sales resources. Since its initial deployment in the U.S. in 2005, MAP has been deployed to 32 countries around the world, which constitute more than 95 percent of IBM total revenue. During the 2006 deployment, approximately 420 MAP workshops were conducted with sales teams globally, involving nearly 3000 sellers from all IBM sales units. More than 2200 individual accounts, representing approximately 55 percent of the total modeled revenue opportunity, were discussed.

Following the interview process, client accounts for each business sector were classified within a two-dimensional segmentation defined by IBM revenue and validated revenue opportunity. With respect to the business objective of improving resource allocation, the most relevant segment is *invest* accounts, where the validated revenue opportunity is significantly greater than current IBM revenue. Using the MAP segmentation, specific measurable decisions were made to optimize account coverage. As a result of the 2005 deployment, a total of 380 sellers were reassigned to invest accounts, with coverage of approximately 50 lower-opportunity accounts shifted to inside sales teams supporting the ibm.com telecoverage channel.

The MAP prioritization framework has been adopted globally by all IBM business units, resulting in a common, cross-IBM view of clients. The IBM Software Group in particular has embraced the MAP methodology and uses it to identify investment accounts supported by 665 sales representatives dedicated exclusively to invest territories.

As resource deployment investments are expected to drive incremental revenue growth, it is very important to measure the MAP impact from a revenue growth perspective. At the same time, we would not expect the impact of those investments to occur quickly, as any shifted resource will need time to ramp up to full productivity. The impact can, therefore, be assessed by comparing the year-over-year revenue growth. In particular, the growth of the invest accounts relative to the growth for U.S. accounts as a whole is a key measure of performance. As further measures of impact, we can use sales pipeline growth and sales quota attainment for the population of sellers who were shifted to cover these investment accounts.

Revenue growth

If we look at large client accounts in the United States, the year-over-year revenue growth during the first half of 2006 was 5 percent higher in the MAP-identified investment accounts than for the background of all United States accounts. Although we cannot state unequivocally that all of this 5-percent growth is due to the MAP process, internal analysis suggests that there is some causal effect. It is

expected that this contribution will increase as shifted resources are given more time to produce results.

Sales pipeline growth

The sales pipeline, as derived through the opportunity management system, is an important leading indicator of future revenue. Here again, growth of the investment segment and its contribution to the total pipeline is an important indicator of impact. It is also important to recognize that any impact on the sales pipeline resulting from MAP will occur over some period of time beyond the current quarter. We can, therefore, use a rolling four quarters worth of validated pipeline as an appropriate measure. As of week 12 in 3Q 2006, the validated sales pipeline of investment accounts (over a rolling four-quarter period) grew year over year at a rate of 14 percent greater than the total United States sales pipeline. As sales pipeline is a leading indicator of revenue, the fact that pipeline growth is greater than revenue growth in the investment accounts is further evidence of the financial impact of MAP.

Quota attainment

A further measure of impact is the performance of those sales resources that are either shifted or dedicated to investment sales territories as a result of MAP. For the first two quarters of 2006, the year-to-date quota attainment of the shifted resources was 45 percent, compared to 36 percent for resources shifted as a result of other initiatives. This suggests that MAP has identified greater sales opportunities and that movement of resources to these accounts has yielded increased productivity.

CONCLUSIONS

OnTARGET and MAP are examples of analytics-based solutions that were designed from the outset to address specific business challenges in the broad area of sales-force productivity. Although they address different underlying issues, these solutions implement a common approach that is generally applicable to a broad class of operational challenges. Both solutions rely on rigorously defined data models that integrate all relevant data into a common database. Choices of the data to be included in the data model are driven both by end-user requirements and by the need for relevant inputs to analytical models. Both business problems have a natural mapping to applications of predictive modeling: predicting the probability to purchase in the case of OnTARGET, and estimating the realistic revenue opportunity in the case of MAP.

Delivering the underlying data and the analytic insights directly to frontline decision makers (sales representatives for OnTARGET and sales executives for MAP) is crucial to driving business impact, and a significant effort has been invested in developing efficient Web-based tools with the necessary supporting infrastructure. Both solutions have been deployed across multiple geographic regions, with a strong focus on capturing and quantifying the business impact of the initiatives. Indeed, we have field evidence that the analytical models developed for OnTARGET are predictive. MAP is a more recent initiative, but preliminary evidence suggests that sales-force allocations made within the MAP process are leading to measurable improvements in sales efficiency. Finally, although we have implemented these solutions within IBM, we believe that the underlying methodologies, business processes, and potential impact are relevant to enterprise sales organizations in many other global industries.

ACKNOWLEDGMENTS

We acknowledge the important contributions of the following people to the OnTARGET and MAP projects: Georges Atallah, Kevin Bailie, Madhavi Bhupathiraju, Mike Burdick, Upendra Chitnis, Steve Garfinkle, Elizabeth Hamada, Katherine Hanemann, Joan Kennedy, Kyle Keogh, Shiva Kumar, John LeBlanc, Keith Little, Agatha Liu, Imad Loutfi, John Pisello, Mike Provo, Ashay Sathe, Colleen Siciliano, Diane Statkus, Shan Sundaram, Nancy Thomas, Ruth Thompson, Lisa Yu, and Brian Zou.

CITED REFERENCES

1. D. Ledingham, M. Kovac, and H. L. Simon, "The New Science of Sales Force Productivity," *Harvard Business Review*, 124–133 (September 2006).
2. W. G. Zikmund, R. McLeod, Jr., and F. W. Gilber, *Customer Relationship Management: Integrating Marketing Strategy and Information Technology*, John Wiley & Sons, Inc., Hoboken, NJ (2002).

3. M. J. A. Berry and G. S. Linoff, *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*, Wiley Computer Publishing, Hoboken, NJ (2004).
4. A. J. Morgan and S. A. Inks, "Technology and the Sales Force–Increasing Acceptance of Sales Force Automation," *Industrial Marketing Management* **30**, No. 5, 463–472 (2001).
5. C. Speier and V. Venkatesh, "The Hidden Minefields in the Adoption of Sales Force Automation Technologies," *Journal of Marketing* **66**, No. 3, 98–111 (2002).
6. Rational Data Architect, IBM Corporation, http://www-306.ibm.com/software/data/integration/rda/.
7. WebSphere DataStage, IBM Corporation, http://www-306.ibm.com/software/data/integration/datastage/.
8. S. Rosset and R. D. Lawrence, "Data-Enhanced Predictive Modeling for Sales Targeting," *Proceedings of the SIAM Conference on Data Mining*, Bethesda, MD (2006), http://www.siam.org/meetings/sdm06/proceedings/063rossets.pdf.
9. T. Hastie, R. Tibshirani, and J. H. Friedman, *Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer-Verlag, New York, NY (2003).
10. S. Rosset, E. Neumann, U. Eick, N. Vatnik, and I. Idan, "Evaluation of Prediction Models for Campaign Planning," *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA (2001), pp. 456–461.
11. R. Koenker, *Quantile Regression*, Cambridge University Press, New York, NY (2005).
12. J. Langford, R. Oliveira, and B. Zadrozny, "Predicting Conditional Quantiles via Reduction to Classification," *Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence*, Cambridge, MA (2006), http://hunch.net/~jl/projects/reductions/median/median_uai.pdf.
13. S. Merugu, S. Rosset, and C. Perlich, "A New Multi-View Regression Approach with an Application to Customer Wallet Estimation," *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Philadelphia, PA (2006), pp. 656–661.
14. C. Perlich, S. Rosset, R. Lawrence, and B. Zadrozny, "High Quantile Modeling for Customer Wallet Estimation with Other Applications," *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Jose, CA (August 2007).
15. L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and Regression Trees*, Chapman and Hall/CRC Press, Boca Raton, FL (1984).
16. T. M. Mitchell, *Machine Learning*, McGraw-Hill Publishing, San Francisco, CA (1997).
17. S. Rosset, C. Perlich, and B. Zadrozny, "Ranking-Based Evaluation of Regression Models," *Proceedings of the 5th IEEE International Conference on Data Mining*, Houston, TX (2005), pp. 370–377.

Accepted for publication April 2, 2007.
Published online July 28, 2007.

Richard Lawrence
IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (ricklawr@us.ibm.com). Dr. Lawrence is a research staff member and manager of Predictive Modeling in the Mathematical Sciences department at IBM Research. He has a B.S. degree from Stanford University and a Ph.D. degree in nuclear engineering from the University of Illinois. He joined IBM Research in 1990, and has managed groups in high-performance computing, e-commerce applications, and predictive modeling. He received IBM Outstanding Technical Achievement awards for his work in parallel computing, scalable data mining, customer targeting, and revenue-opportunity modeling. Dr. Lawrence's recent work has focused on predictive-modeling approaches to solving a broad range of real-life business problems.

Claudia Perlich
IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (perlich@us.ibm.com). Dr. Perlich joined the Data Analytics research group as a research staff member in October 2004. She received an M.S. degree in computer science from Colorado University at Boulder, a Diplom. in computer science from Technische Universitaet, Germany, and a Ph.D. degree in information systems from the Stern School of Business, New York University. Dr. Perlich's research interests are in statistical machine learning for complex real-world domains and business applications.

Saharon Rosset
IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (srosset@us.ibm.com). Dr. Rosset is a research staff member with the Predictive Modeling group in the Mathematical Sciences department. He received a B.S. degree in mathematics and an M.S. degree in statistics, both from Tel Aviv University, and a Ph.D. degree in statistics from Stanford University. He has received IBM Outstanding Technical Achievement awards for his work in revenue and opportunity modeling. Dr. Rosset's recent work focuses on predictive-modeling approaches to solving a broad range of operational business and scientific problems.

Jorge Arroyo
IBM Software Group, 1823 Arrowwood Drive, Carmel, Indiana 46033 (jarroyo@us.ibm.com). Mr. Arroyo is a senior certified executive IT architect in the Global eBusiness Transformation (GeT) organization. He received a B.S. degree from Purdue University. He is the recipient of IBM Outstanding Technical Achievement awards and the IBM IT Architect Profession Excellence Award. Mr. Arroyo is the OnTARGET architect and Enterprise architect for the distributed software value chain.

Matthew Callahan
IBM CHQ, Enterprise On Demand, 1 North Castle Drive, Armonk, New York 10504 (callmatt@us.ibm.com). Mr. Callahan is an executive program manager in the Business Performance Services practice. He received a B.S. degree in mechanical engineering from Virginia Polytechnic Institute and an M.B.A. degree from New York University. Mr. Callahan currently focuses on leading cross-functional teams to drive profitable growth.

J. Matthew Collins
IBM CHQ, Enterprise On Demand, Route 100, Somers, New York 10589 (mattcoll@us.ibm.com). Mr. Collins is a director of Business Transformation Growth initiatives. He received a B.A. degree from Dartmouth College and an M.B.A. degree from Harvard Business School. Mr. Collins joined IBM in 2003, and has led cross-unit initiatives to build and deploy technology solutions to improve productivity and provide decision support.

Alexey Ershov
IBM CHQ, Enterprise On Demand, Route 100, Somers, New York 10589 (ershov@us.ibm.com). Dr. Ershov is an executive program manager in the Business Performance Services practice. He received a B.S. degree in physics from the Moscow Institute of Physics and Technology and a Ph.D. degree in experimental elementary particle physics from Harvard University. He joined the IBM Business Performance Services group in 2004 and has led cross-divisional internal teams in global strategic initiatives focused on revenue growth.

Sheri Feinzig
IBM CHQ, Enterprise On Demand, 1 North Castle Drive, Armonk, New York 10504 (sfeinzig@us.ibm.com). Dr. Feinzig is an executive program manager in the Business Performance Services practice. She received a B.A. degree in psychology and a Ph.D. degree in organizational and industrial psychology from the State University of New York at Albany. Dr. Feinzig has spent the past decade in a management consultant role, advising clients in the areas of sales transformation, productivity enhancements, job design, knowledge management, and teaming and collaboration.

Ildar Khabibrakhmanov
IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (ildar@us.ibm.com). Dr. Khabibrakhmanov is an advisory software engineer with the Predictive Modeling organization in the Mathematical Sciences department. He received B.S. and Ph.D. degrees in theoretical and mathematical physics from the Moscow Institute of Physics and Technology. His focus has been on software development for numerical analysis, data mining, and predictive modeling. Dr. Khabibrakhmanov received IBM Outstanding Technical Achievement awards for his work in customer targeting and revenue-opportunity modeling projects.

Shilpa Mahatma
IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (mahatma@us.ibm.com). Mrs. Mahatma is a staff software engineer in the Statistical Analysis and Forecasting organization in the Mathematical Sciences department. She received a B.E. degree in computer science from Jawaharlal Nehru Engineering College, India. Mrs. Mahatma led the development of the initial OnTARGET prototype and the development of the Market Alignment Program data model and Web interface.

Mark Niemaszyk
IBM Software Group, 5 Bedford Farms Drive, Bedford, New Hampshire 03110 (mn@us.ibm.com). Mr. Niemaszyk is a member of the Software Information Delivery organization within IBM GeT. He is the technical data and database team leader for the OnTARGET application. He received B.S. and M.S. degrees in mathematics from the University of Massachusetts. He received a Business Transformation/ Information Technology Leadership Award for his work on the OnTARGET project. Mr. Niemaszyk's recent technical work has focused on the data, database, and information architecture portions of internal applications within IBM.

Sholom M. Weiss
IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (sholom@us.ibm.com). Dr. Weiss is a research staff member and a professor (emeritus) of computer science at Rutgers University. He is an author and coauthor of numerous papers on artificial intelligence and machine learning, including *Text Mining: Predictive Methods for Analyzing Unstructured Information* (Springer, 2005). His current research interests emphasize innovative methods for data mining and knowledge discovery. Dr. Weiss is a Fellow of the American Association for Artificial Intelligence. ■