2013 IEEE 13th International Conference on Data Mining Workshops

Prescriptive Analytics for Allocating Sales Teams to Opportunities

Ban Kawas, Mark S. Squillante, Dharmashankar Subramanian, and Kush R. Varshney
Business Analytics and Mathematical Sciences, IBM Thomas J. Watson Research Center
1101 Kitchawan Road, Yorktown Heights, NY, USA
Email address: {bkawas, mss, dharmash, krvarshn}@us.ibm.com

***Abstract*—For companies with large salesforces whose sellers approach business clients in teams, the problem of allocating sales teams to sales opportunities is a critical management task for maximizing the revenue and profit of the company. We approach this problem via predictive and prescriptive analytics, where the former involves data mining to learn the relationship between sales team composition and the revenue earned for different types of clients and opportunities, and the latter involves optimization to find the allocation of sales resources to opportunities that maximizes expected revenue subject to business constraints. In looking at the overall salesforce problem, we focus on the interplay between the data mining and optimization components, making sure to formulate the two aspects in a jointly tractable and effective manner. We perform a sensitivity analysis of the optimization component to provide further insight into the interaction between prediction and prescription. Finally, we provide an empirical study using real-world data from a large technology company's salesforce. Our results demonstrate that by using these analytics, we can increase revenue by 15%.**

***Keywords*-customer relationship management; optimization; resource allocation; salesforce analytics.**

I. Introduction

Selling has evolved considerably over the last fifty years from the big, creative, polished in-person pitches depicted in the television series Mad Men, to email and user profile-based offerings, to the use of customer relationship management (CRM) systems, to the emerging trend of turning to intelligent sales automation technologies that allow the optimization of sales processes [1]. In order to enable such a trend in business-to-business (B2B) selling, novel descriptive, predictive, and especially prescriptive salesforce analytics are needed [2].

It is critically important for a business to manage its salesforce in an effective manner because this is the conduit by which revenue is earned and this is also a large sink for selling, general, and administrative expense (SG&A), thereby having a direct influence on the bottom-line profit of the business. It was estimated in 2006 that $800 billion dollars is spent per year on salesforces in the United States, approximately three times as much as on advertising [3]. Many decisions go into running the salesforce of a large enterprise with a diverse product and service portfolio. There are several analytics-based approaches to meeting selling objectives such as growing revenue, selling more to existing clients, and increasing sales productivity, including sales territory optimization, sales team optimization, talent planning, and compensation plan improvement [2], [4].

In this paper, we focus specifically on sales team optimization. Specifically, the problem that we investigate is determining the allocation of sales resources to opportunities in order to maximize revenue in B2B selling. There are two components to this analytics problem, a predictive one and a prescriptive one. The predictive component consists of mining historical selling data to learn *sales response functions* that capture the behavioral relationship between the size and composition of the sales team and the revenue earned for different types of clients and opportunities. The prescriptive component consists of a mathematical programming problem that, using the sales response functions in the objective, determines the allocation of salespeople's effort to client opportunities that maximizes the overall revenue earned by the salesforce subject to business constraints. In formulating these two components, it is important to take their interplay into account: the form of the learned response function must be such that the optimization problem can be formulated in a tractable and effective manner, and the learning of the response function must itself be tractable and effective. Also, it is important to understand the sensitivity of the optimization results to small changes in the estimation.

The optimization aspect of sales resource allocation has been studied in the management science literature for quite a number of years, but the focus has only been on the optimization, not on the data mining [5], [6], [7], [8]. When even discussed, the estimation of the sales response function is treated as an aside in this literature; often simplified procedures are followed, including typically considering discretized sales response functions rather than continuous ones, or only considering very simple forms.

A piece of work that focuses more heavily on learning, looks at the relationship between product offerings and client characteristics, and at the estimation of revenue potential of clients rather than the relationship between the sales team composition and revenue as mediated by client characteristics [9]. The work of [10] considers a specialized form of a B2B sales response function in which there are only two options and in [11], although continuous sales response functions are learned, optimization is not considered. There is an extensive literature [12] in the field of marketing science in which various selling strategies are characterized

978-0-7695-5109-8/13 $31.00 © 2013 IEEE
DOI 10.1109/ICDMW.2013.156

and optimized, but the focus is on the business-to-consumer (B2C) domain rather than the B2B domain.

In this paper, we gather B2B selling data from a technology company listed in the Fortune 500 and develop a tractable joint predictive and prescriptive analytics solution for intelligent sales automation that is appropriate for this particular company by taking much subject matter expertise into account. We ask and answer the question of how this company should invest incremental dollars of SG&A to maximize revenue. Moreover, our work reveals some general insights into how data mining combined with optimization should be pursued in resource allocation problems.

The remainder of the paper is organized as follows. In Section II, we provide an in-depth account of the application domain of B2B selling. Section III discusses in detail the specific optimization problem we are considering in this work. We mathematically formulate appropriate and tractable learning and optimization problems in Sections IV and V, respectively. In Section VI, we provide an analysis of the sensitivity of the optimization results to deviations in the estimation. Section VII presents results based on real-world data from the technology company's salesforce. Finally, we conclude in Section VIII.

II. Background on B2B Selling

In this section, as domain background, we provide a high-level description of B2B selling, staying true to the procedures followed by the technology company whose data we mine. The major component of a business's SG&A is usually the expense associated with selling. Selling can be delineated into two categories: *inside sales* and *field sales*. Inside sales is the selling activity conducted by salespeople in a remote manner using only telecommunications technologies whereas field sales involves sellers meeting with clients face-to-face. Within field sales, there are different types of sellers and sales job roles as well. The particular technology company in our investigation has two different salesforces within field sales: an organization devoted only to sales and distribution, and a sales organization within the product development group. There are also further categorizations of sellers, including technical salespeople and client-facing salespeople.

Salespeople have one full-time equivalent (FTE) of effort per year to allocate among their activities, which can be described using the sales pipeline model. The selling process starts with lead generation, which consists of identifying entities that have the interest and authority to purchase a product. After this step, leads are evaluated and some are qualified to become prospects or opportunities. A sales opportunity consists of a set of one or more products or services that the seller or team of sellers is attempting to convert into an actual sale with the client. They do so by sending information, meeting with the client, demonstrating the product/service, and conducting other such activities. In the end, an opportunity is either won or lost. A won opportunity results in revenue, which combined with the FTEs expended, goes into calculating profit. Lost opportunities contribute only SG&A, not revenue. The FTEs expended by the sellers on the various stages of opportunities are recorded in CRM systems.

Guiding the stages of the sales pipeline are several pieces of information. The management team of the salesforce estimates the revenue potential of the various clients at the beginning of the year to yield aspirational revenue amounts associated with each client. They set the territories for their salespeople. They also categorize clients along various dimensions, constructing regions, segments, sectors, and so on. Such information is also recorded in CRM systems and can be used in developing salesforce analytics solutions.

III. Problem Description

As previously noted, our focus here is on the optimization of the B2B selling process in a large technology company. Assuming there is a direct connection between FTEs expended and revenue enabled, our objective is to optimally allocate sellers to opportunities in order to maximize total revenue. (A profit-based objective can be optimized in a similar manner.) In this section, we will give a detailed description of the specific business optimization problem at hand and introduce some notation.

Let $\mathcal{S}$ denote the set of client segments as characterized in the company's CRM systems and $\mathcal{I}^s$ the set of sales opportunities associated with segment s. Each opportunity $i \in \mathcal{I}^s$, for all segments $s \in \mathcal{S}$, has an estimated revenue potential A_i^s. Sales teams are formed as a combination of sellers from different categories $k \in \mathcal{K}$ and assigned to opportunities based on expert opinions and business constraints. Let $\mathcal{T}^s$ be the set of teams that can be assigned to opportunities within segment s and $\mathcal{K}^t$ the set of seller categories of which team t is composed. The company aims at having an ideal portfolio of sellers (an ideal number of sellers per each category $k \in \mathcal{K}$) such that it is able to assign the most effective teams—in terms of revenue generation—to sales opportunities.

In order to evaluate the effectiveness of a team t, the first part of our analysis will focus on learning the sales response signal of each team $t \in \mathcal{T}^s$ and segment $s \in \mathcal{S}$ by mining relevant data available in CRM systems—i.e., identify the predictive relationship between FTEs expended by sellers from categories $k \in \mathcal{K}^t$ in team t and the corresponding earned revenue. In the second part, we will utilize these sales response functions to identify the ideal headcount in each category of sellers $k \in \mathcal{K}$ such that the company's future revenue is maximized, and business and strategic constraints are satisfied.

The company wants to optimize its salesforce portfolio without incurring *additional* SG&A expenses—particularly, without increasing the total expense dedicated to sellers' salaries and compensation, denoted as C—but by identifying an improvement on the current distribution of C amongst the different categories of sellers. This may require hiring new sellers in some categories and/or transitioning sellers in other categories to different parts of the business. Total compensation c^k of a seller in category k may differ from one category to another. The company is not able to drastically change the headcount in each category, denoted as H^k, and only allows a change $\epsilon^k \in [0,1]$, $\forall k \in \mathcal{K}$, due to business and strategic constraints, and due to the fact that hiring/transitioning sellers is a complex and costly process.

IV. Sales Response Functions – Predictive Component

In order to estimate the functional relationship between FTEs expended e_{ik}^{st} for each category of sellers $k \in \mathcal{K}^t$ in a given team $t \in \mathcal{T}^s$ towards a sales opportunity $i \in \mathcal{I}^s$ within segment s, and revenues earned r_i^{st}, we use multiple linear regression. This choice was made with tractability in mind —there are multiple explanatory variables, e_{ik}^{st}, $\forall k \in \mathcal{K}^t$, and one response variable, r_i^{st}, and there is a need to use the estimated response functions in an optimization model—and based on the business intuition that there is a monotonic relationship between efforts expended towards a sales opportunity and revenues enabled.

When real data—collected from the CRM systems of the company—was analyzed, both revenues and FTEs spanned several orders of magnitude and there was evidence of substantial skew in the data; therefore, to improve linearity and induce both normality and symmetry, we apply a logarithmic transformation on both FTEs and revenues. Accordingly, the functional form of the sales response signal for each team $t \in \mathcal{T}^s$ and segment $s \in \mathcal{S}$, estimated using training samples of historically won sales opportunities $i \in \mathcal{I}_{won}^s$, is given by

$$\ln(r_i^{st}) = \sum_{k \in \mathcal{K}^t} \beta_k^{st} \ln(e_{ik}^{st}) + \beta_0^{st}, \qquad (1)$$

where β_k^{st}, $\forall k \in \mathcal{K}^t$, and β_0^{st} are the coefficients to be estimated using the training data.

These response functions will be used in both the objective function and constraints of the optimization model, which will devise an effective allocation of teams to sales opportunities in terms of revenue growth. As will be seen in the empirical results, using cross-validation, these response functions perform reasonably well. However, variations in the coefficients of these functions, along-side other parameters in the optimization program, may affect optimal solutions (either the objective function value or teams allocation), see Section VI.

V. Optimization Model – Prescriptive Component

After obtaining the sales response functions that correspond to different teams $t \in \mathcal{T}^s$ within each segment $s \in \mathcal{S}$, we are able to utilize them in the prescriptive component of our analysis. The goal of the technology company is to identify the ideal headcount for each seller category $k \in \mathcal{K}$ such that future sales opportunities $i \in \mathcal{I}^s$ within each segment $s \in \mathcal{S}$ are assigned to the team $t \in \mathcal{T}^s$ that can generate the maximum revenue. Given that there are business constraints (mainly cost and headcount variation constraints as described above), the company cannot simply choose teams solely based on the sales response functions in an unconstrained fashion; therefore, the company requires a more advanced prescriptive analytics solution.

In this section, we devise an optimization model that identifies the ideal team per segment, based on the sales response functions and the historical win-rate performance, such that the total potential revenue from projected sales opportunities is maximized while business and strategic constraints are satisfied. We first introduce some additional notation.

Let w^{st} denote the win rate of team $t \in \mathcal{T}^s$ when assigned to a sales opportunity within segment $s \in \mathcal{S}$. To evaluate the performance of a team, we use win rates in conjuction with sales response functions. Doing so avoids bias in favor of teams that have a high potential for generating revenues—as evaluated by their response functions—but a low win rate. (Given that we derive response functions by analyzing historically won opportunities only without considering lost opportunities, we may evaluate a team that has a low win rate but generally generates high revenues with moderate FTEs expended as a better team than one that has a much higher win rate but generates less revenues with the same effort.)

We introduce the decision variable x_{ik}^{st} as the logarithm of total FTE expended by sellers from category $k \in \mathcal{K}^t$ in team t towards an opportunity i within segment s. Using the response function estimated via (1), we have a linear expression in x_{ik}^{st} for the logarithm of revenue, which will be included in the objective function and constraints of the optimization model. Utilizing multiple linear regression in the predictive component of our analysis enabled us to retain linearity in expressing revenues, which is crucial for the tractability of the presecriptive component.

However, the fact that we also use a logarithmic transformation introduces nonlinearity to the constraints of the optimization model, as we need to use the exponential of x_{ik}^{st} in order to express the costs associated with the corresponding FTE. This leads to a nonlinear expression for the total FTE-related costs in terms of a positive linear combination of various exponential functions, $\exp(x_{ik}^{st})$. Nonetheless, since the exponential is a convex function of its argument, such a positive linear combination is also a convex function of the various decision variables x_{ik}^{st} [13]. The exponential function also admits a piece-wise linear approximation that is tractable with respect to the budget limitation constraint imposed as an upper bound on the total costs related to FTEs expended. Specifically, we use a piece-wise linear approximation to estimate each exponential expression $\exp(x_{ik}^{st})$ as

$$\exp(x_{ik}^{st}) = \max_{j \in \mathcal{J}} \left\{ a_j \, x_{ik}^{st} + b_j \right\}, \tag{2}$$

where $|\mathcal{J}|$ is the number of pieces used in the piece-wise linear approximation over the range that x_{ik}^{st} spans (note that this range is the same for all decision variables, as it represents lower and upper bounds on the level of effort that can be expended towards a given sales opportunity).

The choice of teams is expressed using binary decision variables $z^{st} \in \{0, 1\}, \forall t \in \mathcal{T}^s, s \in \mathcal{S}$, which take on a value of 1 when a team is selected and 0 otherwise, and is limited to a single team per segment s using the constraint $\sum_{t \in \mathcal{T}^s} z^{st} = 1$. Initially, we have modeled the choice of teams at the level of sales opportunities by using binary decision variables $z_i^{st} \in \{0, 1\}, \forall i \in \mathcal{I}^s, t \in \mathcal{T}^s, s \in \mathcal{S}$, and limiting the choice to a single team per sales opportunity i, $\sum_{t \in \mathcal{T}^s} z_i^{st} = 1$. However, the problem was hard to solve due to its large-scale aspect. Therefore, to obtain an approximation, we analyze the LP relaxation of the corresponding optimization model—using duality theory—and observe that the choice of teams will be the same for all opportunities $i \in \mathcal{I}^s$ within a given segment s, which is due to having the sales response functions at the level of segments rather than opportunities [14]. Motivated by this observation, we model the choice of sales teams using binary decision variables at the level of segments rather than opportunities. This improves the tractability of the model and allows us to devise approximate solutions as described below.

Based on all of the above, the prescriptive component of our analysis is given by the optimization model in Fig. 1. The objective function, to be maximized, is given in (3), in which we utilize teams' win rates and the sales response functions in (1) to express the total potential revenues (in logarithmic form) that are governed by the choice of teams assigned to sales opportunities within the different segments. Note that when a team t is *not* chosen, i.e. $z^{st} = 0$, associated revenues across all opportunities within a given segment s are set to zero via both the variable z^{st} in the objective function (3) and the right-hand side of the constraints in (3f). The latter forces all decision variables x_{ik}^{st}, $\forall i \in \mathcal{I}^s, k \in \mathcal{K}^t$, associated with team t and segment s to their lower bound $lb < 0$ (a very large negative number representing an effort (FTE) that is very close to 0).

The constraints in (3a) bound potential revenues—in logarithmic form, utilizing the response functions in (1)—per each sales opportunity by their corresponding estimate of aspirational revenue (also in logarithmic form $\ln(A_i^s)$). Constraints (3b) and (3g) are introduced to model the piece-wise linear approximation in (2), where y_{ik}^s will be equal to $\exp(x_{ik}^{st^*})$ (i.e., FTE of seller category k in the optimal team t^* to be assigned to opportunities in segment s); therefore, they can be used to express cost and headcount variation constraints (3d) and (3e), respectively. If team $t \in \mathcal{T}^s$ is chosen to be assigned to opportunities in segment s, the constraints in (3c) guarantee a minimum relative FTE (in logarithmic form), $\rho^t \in [0, 1]$, that is required from each category $k \in \mathcal{K}^t$.

Note that having both binary and continuous decision variables makes our optimization model a mixed-integer program, which is a harder class of optimization problems to solve than linear programs (LPs). However, as mentioned above, since the binary decision variables are at the level of sales segments instead of opportunities—where the number of segments is orders of magnitude smaller than the number of opportunities, $|\mathcal{S}| \ll \sum_{s \in \mathcal{S}} |\mathcal{I}^s|$—we are able to solve the proposed optimization problem exactly by enumerating all different combinations of team choices for the different segments and solving the same number of LPs, then choosing the solution with the maximum objective function value. Note that we are using a brute-force search of the binary space only. For our particular application both the number of segments $|\mathcal{S}|$ and the number of associated teams $|\mathcal{T}^s|$, $\forall s \in \mathcal{S}$, are small; therefore, enumerating all different combinations of team choices is viable and still tractable. Moreover, some of the teams may not be considered in the optimization model, further reducing the complexity of the solution method. There are two reasons why we would exclude a team; either the response function associated with it (a) has negative coefficients—as it does not map to the business perspective that removing efforts off of an opportunity may cause a revenue increase—or (b) has a poor prediction ability as measured by ten-fold cross validation. Further details will be discussed in the empirical results section, Section VII.

VI. Sensitivity Analysis

In this section, we will discuss how the accuracy of the predictive component of our analysis, the multiple linear regression model, affects the prescriptive component of our analysis, the optimization model. As will be shown in the empirical results, the regression model performs reasonably well and is sufficient for predictive analytics. However, when it is used in the optimization model, the quality and reliability of the prescriptive recommendations will be a function of its accuracy. The more accurate the regression model, the more reliable and accurate the solutions obtained from the optimization model, where accuracy is ultimately determined by the prescriptive recommendations.

The purpose here is to test the robustness of the prescriptive recommendations against variation in the regression parameters. We follow a sampling-based approach to capture the effects of changes in the regression parameters, within the statistical confidence of the model, on the solutions of the optimization model in Fig. 1. In particular, we simultaneously and uniformly sample regression parameters

$$\max \quad \sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}^s}\sum_{t\in\mathcal{T}^s} w^{st}\left(\sum_{k\in\mathcal{K}^t}\beta_k^{st}\,x_{ik}^{st}+\beta_0^{st}z^{st}\right) \tag{3}$$

$$\text{s.t. } w^{st}\left(\sum_{k\in\mathcal{K}^t}\beta_k^{st}\,x_{ik}^{st}+\beta_0^{st}z^{st}\right)\le \ln(A_i^s), \qquad \forall i\in\mathcal{I}^s, t\in\mathcal{T}^s, s\in\mathcal{S}, \tag{3a}$$

$$y_{ik}^s \ge a_j\,x_{ik}^{st}+b_j, \qquad \forall i\in\mathcal{I}^s, k\in\mathcal{K}^t, t\in\mathcal{T}^s, s\in\mathcal{S}, j\in\mathcal{J}, \tag{3b}$$

$$x_{ik}^{st}-lb \ge \rho^t\sum_{q\in\mathcal{K}^t}(x_{iq}^{st}-lb), \qquad \forall i\in\mathcal{I}^s, k\in\mathcal{K}^t, t\in\mathcal{T}^s, s\in\mathcal{S}, \tag{3c}$$

$$\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}^s}\sum_{k\in\mathcal{K}} c^k y_{ik}^s \le C, \tag{3d}$$

$$(1-\epsilon^k)H^k \le \sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}^s} y_{ik}^s \le (1+\epsilon^k)H^k, \qquad \forall k\in\mathcal{K}, \tag{3e}$$

$$lb \le x_{ik}^{st} \le lb(1-z^{st}), \qquad \forall i\in\mathcal{I}^s, k\in\mathcal{K}^t, t\in\mathcal{T}^s, s\in\mathcal{S}, \tag{3f}$$

$$y_{ik}^s \ge 0, \qquad \forall i\in\mathcal{I}^s, k\in\mathcal{K}, s\in\mathcal{S}, \tag{3g}$$

$$z^{st}\in\{0,1\}, \qquad \forall t\in\mathcal{T}^s, s\in\mathcal{S}, \tag{3h}$$

$$\sum_{t\in\mathcal{T}^s} z^{st}=1, \qquad \forall s\in\mathcal{S}. \tag{3i}$$

Figure 1. Optimization model – prescriptive component

from their respective 95%-confidence intervals and then use each sample to obtain the corresponding optimal revenue and headcount recommendations.

Relative to the solutions obtained using fitted response functions, we observe from the numerical results that:

- optimal total revenue may change;
- optimal headcount per each category of sellers (prescriptive recommendation) may change;
- or both may change.

Our results show that the reliability of the prescriptive component relies heavily on the quality of the predictive component. In particular, the accuracy of the predictive component needs to be evaluated with respect to the accuracy and quality of the prescriptive component. The interactions between the two is being further analyzed as part of ongoing work [14] in which robust and stochastic optimization models are considered (i.e., optimization models that devise immunized solutions against variability and uncertainty in input parameters).

VII. Empirical Results

Having formulated our mathematical programming problem with predictive and prescriptive components for the business problem of interest in the previous sections, as well as investigating the interplay between the two components through sensitivity analysis, we now apply the proposed methodology to real-world data from the technology company of interest. In particular, we work with opportunity, effort, aspirational revenue, win rate, actual revenue, and seller compensation data from a hardware subdivision of the company in a particular geographic region, for a recent couple of years.

The company's goal is to identify an ideal headcount per each category of sellers that will enable it to assign the most productive (in terms of revenue generation) sales team to opportunities within each clients segment. The experiments will demonstrate (a) the efficacy of the regression model in predicting revenues induced by teams efforts (FTEs)—using 10-fold cross-validation, (b) the utility of the proposed prescriptive optimization model in devising optimal headcount per seller category and associated potential increase in revenue—assuming that the company is re-experiencing the same market environment as the one in which the data was collected, (c) the sensitivity of optimal solutions to deviations in the parameters of the prediction model from their mean values. The latter demonstration is to stress the importance of developing prediction models that are appropriately accurate with respect to their utility purpose (e.g. when used in an optimization model).

The framework described in this paper was implemented

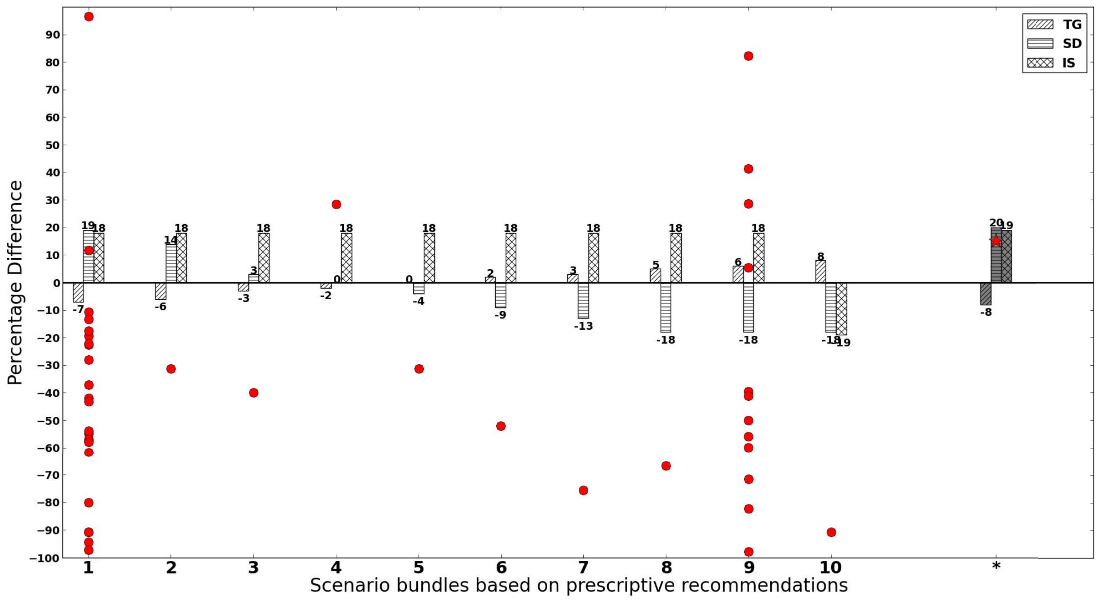

Figure 2. Sensitivity analysis results: scenarios with the same prescriptive recommendations were grouped together. The x-axis is the different scenario bundles, and the y-axis is the percentage difference of: a)[circles] optimized revenues relative to actual revenues b)[vertical bars] optimal head count recommendations relative to actual headcount

and utilized by domain experts in the company as a guiding tool for strategic and operational planning. Solutions obtained from the prescriptive component are based on the mean sales response functions of the predictive component. In future work [14] we investigate (a) alternative regression models that provide more accurate predictions, and (b) robust and stochastic formulations of the optimization model to address data and model uncertainty.

As discussed in Section II and mathematically indicated in subsequent sections, there are different seller categories. This company has three seller categories of interest: inside sales (IS), sales and distribution (SD), and technology group (TG), where the latter two are field-sales categories. From these categories, we can construct up to seven different types of teams through the power set of categories and excluding the empty set.

The company partitions its clients and opportunities into three segments. The first segment contains large accounts in which there has been much historical buying and selling. The second segment represents customers with whom the company hopes to make further inroads and sell more than they have been in the past. The third segment is business that is taken opportunistically.

Our data set contains 3041 different opportunities among 1562 unique clients; 1320 in the first segment, 600 in the second, and 1121 in the third. This data set has been cleansed by removing outliers.

Table I
10-FOLD CROSS-VALIDATION RESULTS – AVERAGE MAPE AND STANDARD DEVIATION – THE LATTER IN PARENTHESES

	team composition	seg1	seg2	seg3
1	{TG,SD,IS}	8.8 (2.9)	7.2 (5.18)	10.3 (4.0)
2	{TG,SD}	12.6 (3.8)	7.9 (1.5)	12.1 (3.8)
3	{TG,IS}	9.0 (4.5)	24.0 (22.5)	15.1 (16.4)
4	{SD,IS}	34.7 (62)	2.9 (1.5)	25.1 (37.2)
5	{TG}	13.6 (8.0)	8.1 (4.9)	21.6 (19.9)
6	{SD}	8.1 (5.5)	16.6 (19.9)	26.7 (28.5)
7	{IS}	9.8 (5.7)	NA	17.0 (10.3)

The first experiment that we conduct concerns the predictive component. Using the logarithmically-transformed revenues and efforts of different sales teams attending to opportunities in different segments, we use the regression (1) to develop a predictive sales response function. We use ten-fold cross-validation to assess the generalization accuracy of the sales response functions that are obtained. Table I provides the average mean absolute percentage errors (MAPE) across the ten folds for each of the sales teams considered under each of the three market segments. The number in parentheses next to the MAPE is the standard deviation of the MAPE across the ten folds.

The MAPE values that we observe in the table are of good quality for this application domain, as confirmed with domain experts within the technology company. In this domain, percentage error is the most well-received figure of merit as opposed to other potential regression quality metrics. The highest errors are in those combinations of teams and segments that are the most rare from a business perspective. For example, a team of SD and IS salespeople is very rarely called upon to sell segment 1 or 3 opportunities. We note that more sophisticated regression approaches such as kernel regression, support vector regression, or Gaussian process regression could have been used to further improve the predictive quality, but this would have rendered an intractable version of an already difficult optimization problem in the prescriptive component. On the one hand, we will see later in this section how these accuracies translate into monetary predictive accuracies. On the other hand, such cases are rare from a business perspective in the first place because they do not represent a desirable solution from a business perspective, and therefore these cases can be eliminated from the optimization model.

To test the quality of the proposed analytics framework, we next assume that the company is re-experiencing the same market environment as the one in which the data was collected (i.e., the company has the same number of opportunities and clients). The company wants to identify what would have been an optimal allocation of its salesforce to sales opportunities in such a way that revenues are maximized, total SG&A expense is kept constant, and a 20% variation of current headcount per seller category is allowed. We then respectively compare optimal revenues and headcount per seller category against actual revenues and headcount in the data set.

Fig. 2 displays the results of the sensitivity analysis described in Section VI. Given that the optimization problem is difficult to solve, for demonstration purposes, we only use 40 scenarios. We observe that when varying the parameters of the sales response functions, optimal revenues and prescriptive recommendations may change. As can be observed from Fig. 2, multiple scenarios may provide the same optimal prescriptive recommendations, but different optimal revenues. In such cases, we observe a wide range of values for the relative differences between optimized and actual revenues, and similarly observe a wide range of values for the relative differences between optimal and actual headcounts. These results highlight the interactions between the prediction and prescription components. The optimization model is only as good as the input; the more accurate the regression model, the more reliable and accurate the solutions obtained from the optimization model, where —as noted above—accuracy is ultimately determined by the prescriptive recommendations. This motivates exploring alternative regression models and developing stochastic and robust counterparts to the deterministic optimization model described here. As mentioned above, this is ongoing work that will be described in greater detail in [14].

The optimal solution using the fitted mean regression models (sales response functions) is depicted as scenario (*) in Fig. 2. We note that if the decision maker is highly confident in the quality of the mean regression model, the optimal potential revenue induced by optimal headcount recommendations is 15% higher than actual revenues, which represents a significant increase in revenue and profit for any large technology company.

VIII. Conclusion

We devise a general analytics framework based on predictive and prescriptive analytics for salesforce optimization of a large enterprise that has a diverse portfolio of products and sevices. Our focus is on sales team optimization, i.e., identifying optimal allocation of sales teams to sales opportunities in order to maximize overall revenues and maintain the current selling, general, and administrative expenses. The predictive component of our analysis consists of mining historical selling data to learn *sales response functions* that capture the behavioral relationship between the size and composition of the sales team and the revenue earned for different types of clients and opportunities. Using the sales response functions as input, the prescriptive component consists of a linear mathematical programming problem that determines the optimal allocation of salespeople's effort to client opportunities that maximizes total revenue earned by the salesforce subject to business and strategic constraints.

Of crucial importance is the interplay between the predictive and prescriptive components of the proposed analytics framework with a focus on joint tractability and effectiveness. Given that the salesforce response functions, obtained from the predictive model, are utilized in the prescriptive optimization model, their structural properties are critical to the tractability of the optimization model, and moreover, their prediction quality directly affects the reliability of the prescriptive recommendations. We provide an empirical study using real-world data from a large technology company's salesforce. We demonstrate that by using a combination of predictive and prescriptive analytics, we are able to potentially increase revenues by 15%. As part of ongoing work, we focus on formulating robust and stochastic optimization models to minimize the effects of prediction quality (i.e., uncertainty present in the parameters of the prediction model) on the prescriptive recommendations and devise solutions that are immunized against data uncertainty.

Acknowledgment

We thank our fellow project team members, most notably: Moritz Baier, Jonathan Debusk, and Boris Peres. We also thank the following executives for their support: Chid Apte, Martin Fleming, Bonnie Ray, Robert Sutor, Christian Toft-Nielsen.

REFERENCES

[1] J. Reese, "How sales has evolved since the Mad Men era," http://blog.leads360.com/infographic-mad-men-era, Apr. 2013.

[2] M. Baier, J. E. Carballo, A. J. Chang, Y. Lu, A. Mojsilović, M. J. Richard, M. Singh, M. S. Squillante, and K. R. Varshney, "Sales-force performance analytics and optimization," *IBM J. Res. Dev.*, vol. 56, no. 6, p. 8, Nov./Dec. 2012.

[3] M. K. Mantrala, S. Albers, F. Caldieraro, O. Jensen, K. Joseph, M. Krafft, C. Narasimhan, S. Gopalakrishna, A. Zoltners, R. Lal, and L. Lodish, "Sales force modeling: State of the field and research agenda," *Market. Lett.*, vol. 21, no. 3, pp. 255–272, Sep. 2010.

[4] M. Singh, D. Bhattacharjya, L. Deleris, D. Katz-Rogozhnikov, M. S. Squillante, B. Ray, A. Mojsilović, C. Barrera, D. Kakrania, J. Richard, A. Saha, and J. Fu, "The growth and performance diagnostics initiative: A multidimensional framework for sales performance analysis and management," *Serv. Sci.*, vol. 3, no. 1, pp. 82–98, Spring 2011.

[5] A. A. Zoltners and P. Sinha, "Integer programming models for sales resource allocation," *Manage. Sci.*, vol. 26, no. 3, pp. 242–260, Mar. 1980.

[6] A. Drexl and K. Haase, "Fast approximation methods for sales force deployment," *Manage. Sci.*, vol. 45, no. 10, pp. 1307–1323, Oct. 1999.

[7] B. Skiera and S. Albers, "Prioritizing sales force decision areas for productivity improvements using a core sales response function," *J. Personal Selling & Sales Management*, vol. 28, no. 2, pp. 145–154, Spring 2008.

[8] M. Golalikhani and M. H. Karwan, "A hierarchical procedure for multi-skilled sales force spatial planning," *Comput. Oper. Res.*, vol. 40, no. 5, pp. 1467–1480, May 2013.

[9] R. Lawrence, C. Perlich, S. Rosset, J. Arroyo, M. Callahan, J. M. Collins, A. Ershov, S. Feinzig, I. Khabibrakhmanov, S. Mahatma, M. Niemaszyk, and S. M. Weiss, "Analytics-driven solutions for customer targeting and sales-force allocation," *IBM Syst. J.*, vol. 46, no. 4, pp. 797–816, Oct. 2007.

[10] J. Wang, M. Singh, and K. R. Varshney, "Does selection bias blind performance diagnostics of business decision models? A case study in salesforce optimization," in *Proc. IEEE Int. Conf. Serv. Oper. Logist. Inform.*, Suzhou, China, Jul. 2012, pp. 416–421.

[11] K. R. Varshney, M. Singh, and J. Wang, "Decision trees for heterogeneous dose-response signal analysis," in *Proc. IEEE Statist. Signal Process. Workshop*, Ann Arbor, MI, USA, Aug. 2012, pp. 904–907.

[12] G. S. Linoff and M. J. A. Berry, *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*. Indianapolis, IN, USA: Wiley Publishing, 2011.

[13] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[14] B. Kawas, M. S. Squillante, D. Subramanian, and K. R. Varshney, "Optimization under uncertainty for prescriptive sales-force analytics," *working paper*, 2013.