Predicting Resource Usage for Capital Efficient Marketing

D.R. Mani
Massachusetts Institute of Technology, USA, and Harvard University, USA

Andrew L. Betz
Progressive Insurance, USA

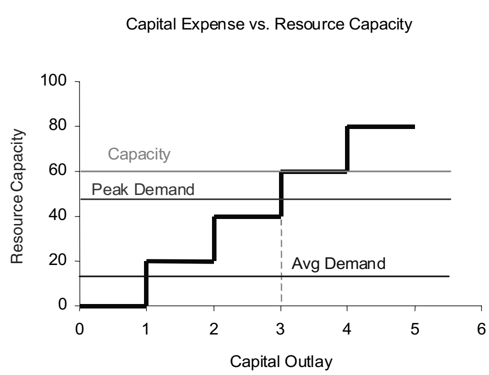
James H. Drew
Verizon Laboratories, USA

INTRODUCTION

A structural conflict exists in businesses that sell services whose production costs are discontinuous and whose consumption is continuous but variable. A classic example is in businesses where capital-intensive infrastructure is necessary for provisioning service, but the capacity resulting from capital outlay is not always fully and efficiently utilized. Marketing departments focus on initiatives that increase infrastructure usage to improve both customer retention and ongoing revenue. Engineering and operations departments focus on the cost of service provision to improve the capital efficiency of revenue dollars received. Consequently, a marketing initiative to increase infrastructure usage may be resisted by engineering, if its introduction would require great capital expense to accommodate that increased usage. This conflict is exacerbated when a usage-enhancing initiative tends to increase usage variability so that capital expenditures are triggered with only small increases in total usage.

A data warehouse whose contents encompass both these organizational functions has the potential to mediate this conflict, and data mining can be the tool for this mediation. Marketing databases typically have customer data on rate plans, usage, and past responses to marketing promotions. Engineering databases generally record infrastructure locations, usages, and capacities. Other information often is available from both general domains to allow for the aggregation, or clustering, of customer types, rate plans, and marketing promotions, so that marketing proposals and their consequences can be evaluated systematically to aid in decision making. These databases generally contain such voluminous or complicated data that classical data analysis tools are inadequate. In this article, we look at a case study where data mining is applied to predicting capital-intensive resource or infrastructure usage, with the goal of guiding marketing decisions to enable capital-efficient marketing. Although the data mining models developed in this article do not provide conclusive positions on specific marketing initiatives, and their engineering consequences, the usage revenues, and infrastructure performance predicted by these models provide systematic, sound, and quantitative input for making balanced and cost-effective business decisions.

Figure 1. Marketing initiatives to boost average demand can indirectly increase peak demand to beyond capacity.

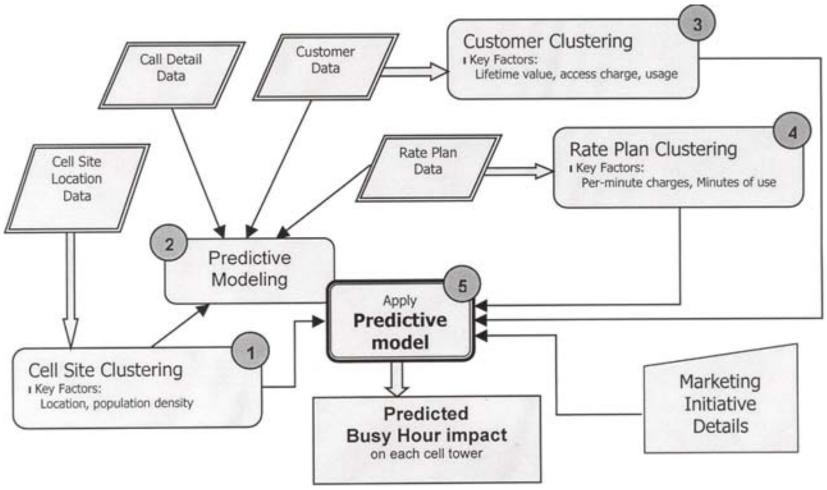


BACKGROUND

In this business context, applying data mining (Abramowicz & Zurada, 2000; Berry & Linoff, 2004; Han & Kamber, 2000) to capital efficient marketing is illustrated here by a study from wireless telephony[1] (Green, 2000), where marketing plans[2] introduced to utilize excess off-peak network capacity[3] (see Figure 1) potentially could result in requiring fresh capital outlays by indirectly driving peak demand to levels beyond current capacity.

We specifically consider marketing initiatives (e.g., rate plans with free nights and weekends) that are aimed at stimulating off-peak usage. Given a rate plan with a fixed peak minute allowance, availability of extra off-peak min-

Figure 2. Flowchart describing data sources and data mining operations used in predicting busy-hour impact of marketing initiatives



utes could potentially increase peak usage. The quantification of this effect is complicated by the corporate reality of myriad rate plans and geographically extensive and complicated peak usage patterns. In this study, we use data mining methods to analyze customer, call detail, rate plan, and cell-site location data to predict the effect of marketing initiatives on busy-hour[4] network utilization. This will enable forecasting network cost of service for marketing initiatives, thereby leading to optimization of capital outlay.

MAIN THRUST

Ideally, the capital cost of a marketing initiative is obtained by determining the existing capacity, the increased capacity required under the new initiative, and then factoring the cost of the additional capital; data for a study like this would come from a corporate data warehouse (Berson & Smith, 1997) that integrates data from relevant sources. Unfortunately, such detailed cost data are not available in most corporations and businesses. In fact, in many situations, the connection between promotional marketing initiatives and capital cost is not even recognized. In this case study, we therefore need to assemble relevant data from different and disparate sources in order to predict the busy-hour impact of marketing initiatives.

Data

The parallelograms in the flowchart in Figure 2 indicate essential data sources for linking marketing initiatives to busy-hour usage. Customer data characterize the customer by indicating a customer's mobile phone number(s), lifetime value[5], access charge, subscribed rate plan, and peak and off-peak minutes used. Rate plan data provide details for a given rate plan, including monthly charges, allowed peak, off-peak, weekend minutes of use, per-minute charges for excess use, long distance, roaming charges, and so forth. Call detail data, for every call placed in a given time period, provide the originating and terminating phone numbers (and, hence, originating and terminating customers), cell sites used in handling the call, call duration, and other call details. Cell site location data indicate the geographic location of cell sites, capacity of each cell site, and details about the physical and electromagnetic configuration of the radio towers.

Data Mining Process

Figure 2 provides an overview of the analysis and data-mining process. The numbered processes are described in more detail to illustrate how the various components are integrated into an exploratory tool that allows marketers and network engineers to evaluate the effect of proposed initiatives.

1. **Cell-Site Clustering:** Clustering cell sites using the geographic location (latitude, longitude) results in cell site clusters that capture the underlying population density, with cluster area generally inversely proportional to population. This is a natural consequence of the fact that heavily populated urban areas tend to have more cell towers to cover the large call volumes and provide good signal coverage. The flowchart for cell site clustering is included in Figure 3 with results of *k*-means clustering (Hastie, Tibshirani & Friedman, 2001) for the San

Related Content

Distributed Association Rule Mining
Mafruz Zaman Ashrafi, David Taniarand Kate A. Smith (2005). *Encyclopedia of Data Warehousing and Mining (pp. 403-407).*
www.irma-international.org/chapter/distributed-association-rule-mining/10631

A Tutorial on Hierarchical Classification with Applications in Bioinformatics
Alex Freitasand Andre´ C.P.L.F. de Carvalho (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications (pp. 119-145).*
www.irma-international.org/chapter/tutorial-hierarchical-classification-applications-bioinformatics/7637

Exception Rules in Data Mining
Olena Dalyand David Taniar (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications (pp. 336-342).*
www.irma-international.org/chapter/exception-rules-data-mining/7647

Neural Network-Based Stock Market Return Forecasting Using Data Mining for Variable Reduction
David Encke (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications (pp. 2476-2493).*
www.irma-international.org/chapter/neural-network-based-stock-market/7777

An Algebraic Approach to Data Quality Metrics for Entity Resolution over Large Datasets
John Talburt, Richard Wang, Kimberly Hessand Emily Kuo (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications (pp. 3067-3084).*
www.irma-international.org/chapter/algebraic-approach-data-quality-metrics/7822