Predictive Analytics & Big Data

Efficiently Using Big Data for Advanced Analytics

By Paul Maiste (President & CEO, Lityx) and Gary Robinson (COO, Lityx)

Marketing and CRM Optimization

In recent years, analytics has brought to the forefront of business decision making the importance of understanding customer churn, value, channel preference, and other behaviors. Providing support for the marketer's understanding of these metrics has become a key deliverable for IT departments, and big data has increasingly become a more integral aspect of the deliverable. However, a bigger challenge is on the horizon: integrating predictive analytic and optimization techniques into the big data infrastructure. This allows marketers to go to the next level - optimal planning and decision making. One of the biggest levers a marketer can pull to improve performance is to improve the targeting of your CRM strategies. The more marketers can predict customer differences and align strategies and tactics around this knowledge, the more efficient they will be able to spend marketing dollars. However, this all adds up to more complexity for the IT environment

BI vs. Predictive Analytics and Marketing Optimization

The proliferation of Business Intelligence (BI) tools demonstrates the demand that exists for historical customer insight. BI tools offer a wonderful way to visualize data and better understand customers and prospects. However, they fall short in their ability to take all of the information at once and determine the most effective way to weave it together in multi-dimensional fashion. Generally BI tools look at one to three pieces of information at a time, but then things start to get too complex for a visual representation. Predictive Analytics uses modeling algorithms to examine data and determine the best combination of data elements and weightings to predict a behavior of interest. Going further, Marketing Optimization can then combine those predictions with hard business constraints, such as constraints on budget and resources, to provide the optimal way to spend marketing budget. The definition of "Optimal" can be up to the marketer, and possibly changes over time. For example, perhaps growing the customer base is a marketing goal for the third quarter, but profitability is next quarter's goal. Optimization can provide the blueprint for making budgeting decisions to satisfy each objective. In the end, it is the optimal use of a combination of all available data that makes Predictive Analytics and Optimization, in conjunction with BI, such a powerful set of complementary tools.

Big Data + Predictive Analytics = Computational Burden

Without a doubt, the era of Big Data has arrived. Big Data brings even more sources of information about customers, prospects, channels, and competitors into the mix. This can only serve to enhance our ability to optimize marketing efforts. However, there are challenges. The data is often unstructured and can require enormous computing power to summarize and evaluate. In addition, advanced algorithms for modeling and optimization require much greater computing resources than BI tools. The IT administrator has to make a decision about infrastructure planning, and often the first thought is to throw the kitchen sink at the problem, thinking that an increase in data size requires a requisite increase in computing power for advanced analytics.

Thankfully, many aspects of the predictive analytics solution can rely on core tenets from the realm of statistics. Specifically, statistical sampling and variable reduction techniques can provide great computational efficiencies while sacrificing very little in terms of the accuracy or power of the results.

Big Data + Efficiency = Fast, Powerful Results

Statistical sampling approaches ensure a random or stratified subset of all available data provides sufficient information to draw essentially the same insight as if the full file were being analyzed. The brute force method is to crunch the entire file, which in the case of big data can be billions of records. Some might argue this is the only way to realize 100% of the information value. Without disagreeing, the fact is that as little as .01% of your data could be used to generate predictive results to within to 99% or greater accuracy (depending on the situation). Closing the gap from, say, 99% to 99.5% accuracy can require exponentially greater computing power with very little gain. The law of diminishing returns is in full force when combining big data with predictive analytics.

In addition, in many cases, what makes big data "Big" is the width of the information. The number of variables, or pieces of information, available to us can often be in the hundreds, if not many thousands. More variables greatly increases our power to build compelling models, but also increases the computational needs for predictive analytics. As we discussed above, there are tried-and-true techniques that can be employed to reduce the number of pieces of information used to develop predictive models with very little effect on accuracy. Variable reduction techniques such as correlation analysis and principal components analysis can determine subsets of variables that contain a high proportion of the interesting variability necessary for predictive modeling. Focusing on a powerful subset can provide significant efficiency gains when working with big data.

Putting it Together for IT

This article has attempted to draw your attention to the power of combining advanced analytic techniques such as Predictive Analytics and Marketing Optimization with Big Data. For the IT practitioner, the key thing to remember is that computational resources can be saved or re-allocated by keeping in mind a few key concepts such as sampling and data reduction techniques. Certainly these techniques do not solve all Big Data issues and cannot always be employed. But for some of the most challenging and resource-intensive problems in advanced analytics, they can help make the supporting IT infrastructure much more streamlined and efficient.