CSP
07.13 | EB 6801

ANALYZE YOUR BIG DATA AND REDUCE CUSTOMER CHURN

NEW TOOLS HELP SERVICE PROVIDERS RETAIN CUSTOMERS AND INCREASE REVENUE

TERADATA.

By Stephen Butler,
Managing Partner, and
Charles Fournier,
Senior Industry Consultant,
Communications,
Media and Entertainment,
and Utilities Industries,
Teradata

CSP
07.13 | EB 6801

EXECUTIVE SUMMARY

For subscription-based service providers, such as wireless carriers, satellite providers, and cable companies, understanding customer turnover can have a large and meaningful impact on the bottom line. In fact, cutting customer churn by just one basis point can improve revenue to the tune of millions of dollars. In the past, analyzing the reasons why customers didn't renew their subscriptions was a straightforward exercise because the data was limited, but with the advent of big data, that task has become much more challenging.

Subscription-based service providers that can effectively use big data, which includes both structured (e.g., set-top box remote clicks) and unstructured (e.g., customer online activities) data, can gain deeper customer insight and apply human capital more efficiently.

This paper discusses existing methods of predictive churn modeling and the importance of big data in new types of discovery analysis. It explains the benefit of investing in a new platform and visualization tools that enable quick understanding of event sequences to reduce churn and deliver better customer service and significantly more revenue.

DELIVERING SERIOUS SAVINGS

To understand why churn analytics can greatly increase revenue, let's look at U.S. wireless carriers. They have a massive installed base of customers on contract who provide a wealth of data. Running advanced analysis on this data enables carriers to predict customer churn and, often, reduce it, saving millions of dollars.

Here's a hypothetical example: A wireless provider with an installed base of 20 million customers reports ARPU (average revenue per user) at $55 a month and customer churn at 2.5 percent. If a subscriber's average tenure is 24 months (the length of most contracts), the provider could address $660 million over the average tenure of the subscriber base by applying customer churn analysis to its installed customer base—and that doesn't include the sunk cost of acquisition to replace lost subscribers.

This is not an exaggeration. It sounds complicated, but it's not as complex as other modeling exercises because service providers have the advantage of modeling in a customer-data-rich environment. And all large service providers invest in infrastructure that supports the use of customer transactional data for predictive analytics, along with other business intelligence functions.

Transaction-based data available for variable input selection enables advanced methodologies and transformations of these data inputs. An example of transaction-based data for telecommunications companies is a customer's "minutes of use" (MOU). The transformation process entails taking various treatments of this transactional data to determine what yields the best correlation to the dependent variable—in this case, customer churn.

Larger service providers use enterprise-scale statistical modeling tools in the variable transformation process. Regardless of industry, churn modeling exercises traditionally apply the following standard procedures:

- Variable selection and transformation based on correlations to a subscriber churn event
- Model development that uses linear or nonlinear predictive analytics (in most cases, logistic regression)
- Hold-out for model validation
- Scoring for customer base, to understand customers who are at high risk of churn, usually grouping into quintiles or deciles
- Model refreshes and model revamps on a recurring basis, typically monthly
- Use of output and insight of churn modeling: proactively in outbound campaigns and reactively at the "save desk" inside customer care

EMPLOYING A VARIETY OF METHODOLOGIES

While service providers may employ the basic process detailed previously, they often customize it depending on data availability. Often the modeling process is unique to each technique employed. And, typically, each statistician and model uses distinct methods, which can be considered the "art" in the marriage of art and science of statistical analysis. Certainly, the amount of data-infrastructure investment differentiates one service provider from another. Some invest more, some less.

But there are similarities as well. With a suite of tools and applications for model development and customer scoring, along with a large enterprise-wide data warehouse and a team of smart statisticians, companies employing a variety of methodologies in predictive churn analytics can all arrive at a similar baseline.

Another important point to consider is that the vast majority of carrier churn is based on customer experience. For example, the experience of wireless customers is heavily dependent on their perception of the carrier's network performance. And at the end of the day, a carrier cannot expect that marketing campaigns and customer treatment will solve a long-term churn problem.

DEFINING BIG DATA

Anyone who has ever been involved in using customer data to drive insight—whether in financial reporting, inventory and supply-chain management, compensation design, marketing modeling, or a number of other processes—has heard of big data. The term enjoys the same hype today that "dot-com" did in the late '90s. Like dot-com years ago, big data is now burdened with misconceptions of what it means in the marketplace, and everyone is trying to understand how to use it in business.

But for the purposes of this discussion, big data is defined as behavioral unstructured data. One example is a customer's online activities, which are stored in Web logs.

Most large service providers, regardless of industry, make data infrastructure investments to enable statistical modeling that is specifically geared toward using structured data. Structured data can be organized in databases and indexed, as is typically done in enterprise data warehouses or smaller operational databases. This indexed data, which can be used in a variety of enterprise-scale modeling applications, enables the process of variable selection and transformation. Data categories such as transactional (e.g., set-top box remote clicks), key performance indicators (e.g., average videos purchased), and enhanced data (e.g., geocodes) are generally structured data and can easily be stored in relational databases.

How does behavioral unstructured data fit into this structured process, and what does it mean for predictive analytics? Let's start with the basic process of modeling outlined previously—specifically, variable selection and transformation.

"Traditional" variables are the drivers of a model. They are the independent variables driving your outcome—in this case, customer churn. These variables are structured data and are relatively easy to obtain, measure, and combine because they are usually indexed and organized in relational databases.

Unstructured data, on the other hand, is generally not measured, not easy to obtain and store economically, and not easy to combine with structured data, which means it wouldn't be indexed and stored in relational databases. The behavioral aspect of big data is critical because it focuses on the interactions—not necessarily on the transactions. Big data, then, cannot be captured and stored in the same way as traditional transactional data.

For most data modelers and statisticians, accessing data for model preparation requires advanced knowledge of SQL, the standard language used to query data from a relational database. Using SQL permits an analyst to obtain large volumes of data and determine the best set of data to predict an outcome. Because unstructured data cannot be stored in a "traditional" relational database, the first step in the process of model development is incredibly complex—unless an analyst uses solutions and platforms that take advantage of existing skill sets and applies them to behavioral unstructured data. Recently, however, with emerging technologies such as Teradata®, data scientists (and now data analysts) can access unstructured data using SQL.

DISCOVERING NEW VARIABLES

We've discussed existing methods of predictive churn modeling, and we've discussed the importance of big data in new types of discovery analysis. Now let's "visualize" big data analytics in action.

Those who have worked closely with a modeling team know firsthand the impressive benefit of using visualization for path and pattern analysis. In a fraction of the time it would traditionally take, emerging technologies now enable a data scientist to visually demonstrate several different channel event tables, including Web log data. This visualization permits the discovery of new patterns of events that lead to churn, and it does so via a path function, originally used by Web companies for Web-pathing analysis.

Standard methodologies that leverage logistic regression analysis can provide the right mix of variables to predict churn. But in visualization and path/pattern analysis, there are new and exciting methodologies being employed. These include the following:

- Rapid ability to sessionize Web log data
- Ability to use traditional SQL to access unstructured data
- "In-database" prepackaged analytics that can visualize new event patterns

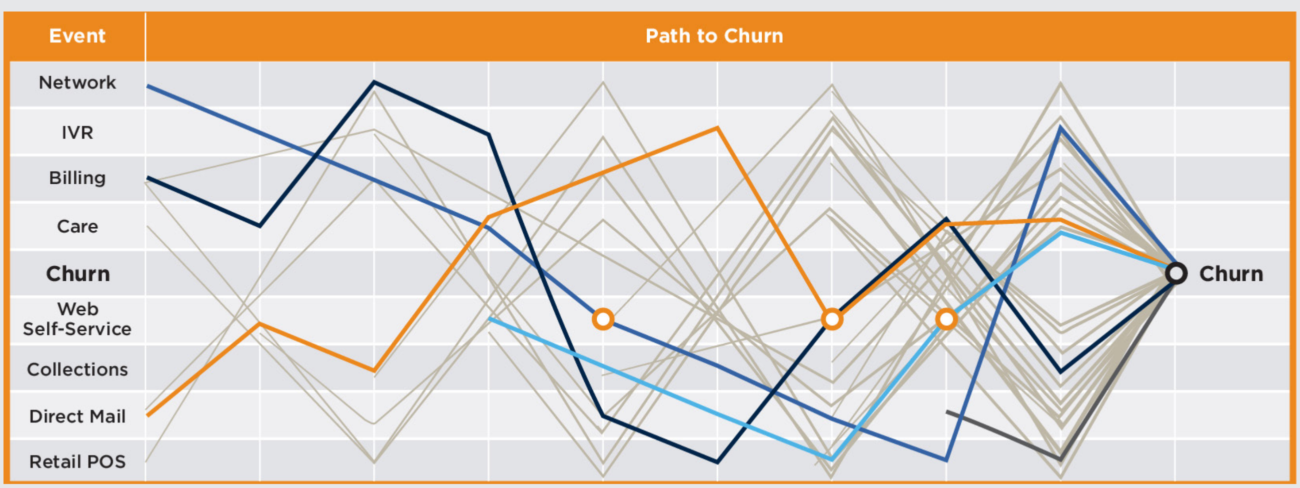
Statistical modeling has always followed multiple phases of development for variable vetting and model calculation, but any analyst with strong data-manipulation and critical-thinking skills can use a visualization tool to identify events or event sequences as churn variables. This boosts a team's productivity by freeing statistical modelers with more advanced skills to focus on downstream tasks in the model-development process. It also gives junior team members a chance to develop new in-demand skills.

So how does this apply to churn insight? The following graphic shows a concentration of activity from a store visit or a call center, followed by activity at the "account status" area of the Web channel. Further analysis of the Web logs shows most of the online activity is from verification of a contract end date or a visit to the FAQ page that contains "how to cancel" information.

Seems obvious, right? And it raises the question of why a company would need to invest in data scientists and a big data analytics platform. But remember this: One of the most complex steps in model development is variable selection and transformation. Subscription-based service providers have terabytes of data, and the task of drilling through it all, trying to normalize across the same system types, is a significant effort. Just imagine having three different billing systems, each with its own termination code for calls. Simply getting through the data can add months to the model-development process.

A new platform and visualization tools enable a modeler to quickly understand event sequences that are causing churn and that can be "variable-ized" for use in the company's existing logistic regression model.

But once you understand that a combination of these new capabilities boosts the model-development process, how do you determine the return on investment (ROI), which can be significant?



Identify critical paths and events for canceled subscriptions using visualization of big data on a discovery platform

UNDERSTANDING INCREMENTAL ANALYSIS

Because the costs associated with an investment in big data analytics tend to be as straightforward as any non-frontline technology investment, let's focus on the revenue and margin impact end of the equation. How does a company measure incremental value in churn identification?

First, it's important to remember that not all variables in advanced analysis exercises are equally useful. The same is true of the variables discovered using path and pattern analysis. While determining the most powerful inputs to churn modeling efforts isn't new, what is new is the predictive power of fresh data. With a relatively modest investment, the ROI on big data analytics can grow very large, very fast.

It helps to understand the new variables that path analysis makes possible and how incremental analysis can complement traditional churn-modeling efforts.

Let's first address how to define the variable types of event sequences, or paths, you can discover. We assign a variable to one of three categories: point, stage, and perpetual. All three bring value, but the amount of value can vary significantly.

The most valuable variable type is perpetual because it's an event sequence that isn't limited to a single point in time or stage. These events can have the largest impact in a recurring model calculation and scoring process for churn. They are, in short, the gift that keeps on giving.

Whether point or stage variables should be incorporated into a company's premier churn model is an open question. It needs to be reviewed on a case-by-case basis. But regardless of the outcome, these variables have value in secondary models and in proactive retention campaigns.

Table 1 provides definitions of the three types of variables and an example of each from the wireless industry, as well as the appropriate valuation method.

The formula that is consistently applied is a straightforward NPV (net present value) formula, but it varies by what is included.

The identification of incremental forecasted churners may vary by churn or profitability decile. For example, a company may want to include only the top decile of ARPU that scores in the top three deciles of churn probability. Or it may want to include the incremental forecasted churners in the top two deciles of ARPU that score in the top decile of churn. Since most finance organizations prefer conservative forecasting, a company may consider the incremental churners in the top ARPU decile (e.g., $200) who scored in the top churn decile.

It's important to recognize that we're identifying *incremental* churners missed by the existing churn model. After all, if the existing model catches and scores

Table 1: Three Categories of Variables with Examples from the Wireless Industry

	POINT	STAGE	PERPETUAL
Definition	A variable that's singular, a one-time event	A variable that happens in recurring fashion over a defined period of time	A variable that happens on a recurring basis in perpetuity
Example	A subscriber following up on a holiday promotion and searching for more information online	A subscriber migrating from an older phone to a newly released smartphone and engaging with customer service	A subscriber migrating from Apple iOS to Android, or checking a contract end date online, or searching the FAQs for "cancel service"
Valuation Method	NPV of ARPU applied to the segment of subscriber lines associated with the event, correlated with the survival curve of the segment	NPV of ARPU applied to the segment of subscriber lines, correlated with the survival curve of the segment, carrying forward through the window where the segment is meaningful	NPV of ARPU over extended life of the subscriber line, applying decay rate (try monthly churn multiplied by average months of tenure) for each month in perpetuity

the same churners without the new variable, little has been gained other than the possibility of more insight into "why" the subscriber churned. This process should include some statistical method that scores the churners and then quantifies them with a simple two-dimensional view by decile.

Table 2 shows a company that has identified 600 subscribers spending $200 a month who were not identified before. And it should be able to identify a similar number each month, assuming the variable was categorized as a perpetual or stage. But there are other factors to consider before the company can calculate ROI. After all, not all of these customers will be saved. So let's apply three more factors:

- **Average tenure:** 27 months
- **Decay rate:** 0.3 percent per month (or the non–pay rate of the base)
- **Conversion rate:** 25 percent (this should be a blend of both the outbound retention campaigns and inbound saves at the call centers but is optional to apply when considering "value" in churn identification)

Table 2: Company Increases Churner Identification by 12% with Big Data Analysis

		SUBSCRIBERS IDENTIFIED BY EXISTING CHURN MODEL	
		Yes	No
SUBSCRIBERS IDENTIFIED WITH BEHAVIORAL DATA	Yes	5,000	600
	No	200	300

Using all these factors and a discount rate of 10 percent in the NPV formula, the *monthly* value delivered by this new capability is $736,364, or a little more than $8.8 million per year. When applying the decay rate to the average tenure of the base, this new analytic would bring more than $21 million in saved revenue over that same time frame. And this is delivered by just the *first* analytic within 90 days after investment in and implementation of big data analytics. Assuming a fully loaded investment of $1.4 million, the ROI would be 629 percent and would multiply as additional analytics are delivered every 30 to 60 days, at hundreds of thousands of dollars in monthly NPV per analytic. Even the stingiest comptroller would consider that investment in platform, tools, and a couple of data scientists worthwhile.

How did the original churn model miss the identification of so many incremental churners? It did not address the behavioral markers that aren't reflected by the transactions associated with churn. Most churn models used by subscription-based service providers are built on transactional data from operational systems and, occasionally, network elements. They don't take into account behavioral interactions that the new generation of churn modeling and analysis can provide.

CONCLUSION

For years Web companies have understood and tapped into the power of big data, and now companies in telecommunications, cable, satellite, financial services, and other industries are beginning to realize its potential.

First, they are using event sequences as new inputs within standard modeling efforts, effectively using the right infrastructure to access and transform big data. This ability to use customer behavioral data is new, and most companies are still figuring out how to maximize their analytics programs with insights gained from big data.

Second, they are incrementally changing standard methods of churn modeling. Layering in additional insight gained from new technologies can be tested and tuned with current "business as usual" practices.

Once you have a platform that can quickly access and transform data into actionable inputs to drive churn insight, why not just replace your old model with the new one? The answer is simple: Employ both efforts, which run parallel paths.

You should continue with your current modeling efforts while testing new, independent variables in the model. As you refine your churn model over time, you can continue to test new event sequences to learn more from your customers' online behavior. Moreover, you can use visualization tools to drive insight on customer churn, insight that can then be tested in customer treatments. Don't limit yourself to one approach, especially to an approach that everyone recognizes is early in its maturity.

We aren't recommending the best methodology, but we are recommending you invest in the correct tools and test a variety of methodologies. Discovery analysis is the best way to describe your exploration and refinement process. To drive big data's maximum impact, you can't simply use technology alone. You need to keep these goals in mind:

- Gain deeper customer insights, with a complete view, using their behavior and their transaction perspective.
- Learn to apply limited human capital most efficiently.
- Implement new processes (and alter current processes), especially those that turn data into action, such as churn analytics that supports a discover-validate-infuse approach to uncover everything that big data has to offer.

For more information on how discovery analytics can increase the effectiveness of your churn identification and how Teradata can help you maximize the benefits of your big data investment, visit us at **www.Teradata.com/communications**.

10000 Innovation, Drive Dayton, OH 45342 teradata.com

TERADATA. THE BEST DECISION POSSIBLE™

EB-6801 > 0713