COMPUTERWORLD An IDG company

Survival of the fittest

Survival analysis can now be used to forecast customer attrition

News Story by Gary H. Anthes

SEPTEMBER 06, 2004 (COMPUTERWORLD) - Survival analysis could help you predict that one of your best customers is about to jump ship for a competitor. Or it could help you decide whether that costly promotion is really going to be worth it. Or it could help you tailor that next catalog mailing and double your return.

The aptly named analytic technique, also called survival data mining, has been used by doctors for decades to predict the life expectancy of heart-transplant patients and by biologists to assess the probability that a cell invaded by a virus will die within 24 hours. Engineers have long used it to estimate the mean time to failure of a disk drive or a robotic welder. More recently, sociologists and psychologist have started using it to predict when certain types of people will divorce or seek help for depression.

But until recently, attempts to apply survival analysis to business problems have been mostly university projects, says Edward Malthouse, a marketing communications professor at Northwestern University. "Now it's really taking off in the database marketing world—for credit cards, hotels, airlines, catalogs and so on," he says.

Survival analysis refers to a family of "time to event" prediction techniques mathematically geared to problems with the following characteristics:

• **They deal with discrete events that will occur to some but not all members of a given population.** Certain patients will die, some disk drives will fail, a certain number of prescriptions for Valium will be written, and some of your best customers will desert you.

• **They involve time-dependent outcomes.** Is that key customer going to cut up his charge card tomorrow (better call him today), next quarter (send him mail) or not in the next five years (leave him be)?

• **The outcomes of interest, or "dependent variables," aren't continuous—like income, height or IQ—but are dichotomous.** Either the patient will die within six months or he won't. Your customer will leave this year or he won't.

• **Outcomes often can be anticipated by trigger events, such as customer complaints.**

Randy Collica, a senior analyst at Hewlett-Packard Co., says use of survival mining to understand and predict customer behavior has sprung up in the past couple of years. He says it's really just an extension of older practices in which a company would take its most recent customer data, or data from a time slice such as a quarter or year, and try to predict attrition based on that.

But survival mining puts time itself into the analysis as a variable. "It is a superset technique," Collica says. "Including time as an element is adding that much more information."

Choosing a Technique

Software for survival analysis most often uses some kind of regression technique, notably logistic regression or a kind of regression called Cox proportional hazards modeling. In addition, some vendors sell products that do time-to-event predicting using neural networks or genetic algorithms.

For example, SAS Institute Inc. has survival prediction capabilities in several of its products. According to Anne Milley, manager of analytic strategy at SAS, deciding which product or technique is appropriate depends on how the problem is formulated, the type of output desired and users' ability to code and prepare data.

New York-based investment firm The Dreyfus Corp. used SAS's logistic regression routines to beat down customer fund redemption rates from more than 20% a year to less than 7%. "Assets were going out of the complex, and we didn't know why," says Prasanna Dhore, an executive vice president.

First Dreyfus used survival analysis to help it understand the factors that were leading customers to cash out. The company organized terabytes of customer history along three dimensions: customer life stage (e.g., young married, peak earnings age, retired), past investor behavior and basic demographics, such as education.

"Then we asked if we could predict behavior," Dhore says. "What's the probability that Customer A is going to take the money out in the next two years, for example? Then, what does it take to extend his life with us?"

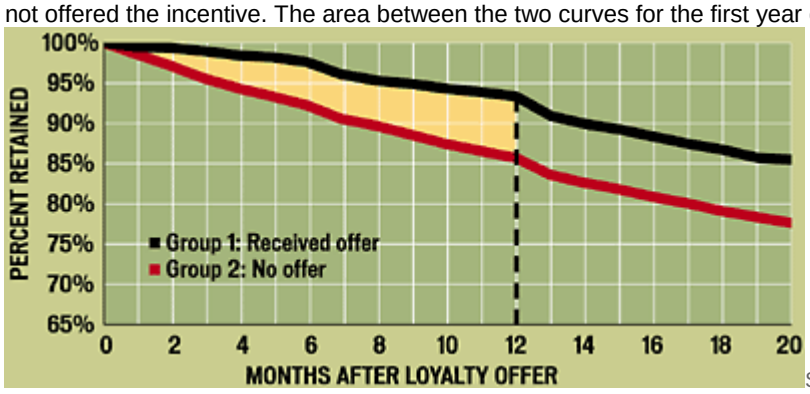
The answers to those questions varied by customer type. For example, the analysis showed that customers identified as "street savants" would redeem within three months of a fund's performance slump. "So if you can call him up and give him other opportunities, he can be kept in the Dreyfus family and we can extend his life another six months, a year or two years," Dhore says.

Catholic Relief Services uses software from Genalytics Inc. in Newburyport, Mass., to "find needles in a haystack," says Kevin Whorton, director of direct-response fundraising. Those are donors who will not simply make one-time gifts in response to an emergency such as a hurricane, but who will keep on giving thereafter. The software uses genetic algorithms that learn over time in order to produce "self-adjusting models," Whorton says.

For example, one model showed that Hispanic donors have a far higher attrition rate than non-Hispanic donors, a special concern, Whorton says, because by 2025, Hispanics will account for half of all Catholics in the U.S. "So now we can do more with these people at the four-to-six-month point in the relationship," he says.

LOYALTY PROGRAM:

The graph below compares customer survival (retention) after acceptance of a loyalty incentive (top curve) with survival of customers not offered the incentive. The area between the two curves for the first year can be computed to find the value of the offer.



Source: Data Miners Inc., Boston