Paper 7500-2016

Customer Lifetime Value Modeling

Vadim Pliner, Verizon

ABSTRACT

Customer lifetime value (LTV) estimation involves two parts: the “survival” probabilities and profit margins. This article describes the estimation of those probabilities using discrete-time logistic hazard models and that of profit margins is based on linear regression. In the scenario when outliers are present among margins, we suggest applying robust regression with PROC ROBUSTREG.

INTRODUCTION

According to Wikipedia, “in marketing, customer lifetime value (CLV) (or often CLTV), lifetime customer value (LCV), or life-time value (LTV) is a prediction of the net profit attributed to the entire future relationship with a customer.” Let us assume the time horizon is divided by N periods, e.g. months or weeks. LTV is typically defined as

$$\sum_{t=1}^{N} p_t * M_t, \qquad (1)$$

where p_t is the estimated probability to remain the company’s customer for the next t months/weeks/… and M_t is the predicted profit margin during period t. p_t are estimated using survival analysis (we will cover this component of LTV in the next section), while M_t are usually assumed to be constant for each customer and equal to either the latest profit margin or its average over several most recent time periods. Sometimes $M_t = M*r^t$, where the discount factor r accounts for the way the value of money is discounted over time. The concept is based on the premise that a dollar today is worth more than a dollar tomorrow. We won’t make those simplistic assumptions, but will suggest a regression type of modeling to actually predict future values of M_t (section 3).

2. “SURVIVAL” COMPONENT OF LTV

Survival analysis, also known as failure time analysis and event history analysis, is used to analyze data on the length of time it takes a specific event to occur. Typical examples of such events include death, the onset of a disease, failure of a manufactured item, and customer or employee turnover. The main distinct feature of data used in survival analysis is that some subjects have not experienced the event during the observation period. In that case, the event time is considered (right) censored at the point when the subject (patient, customer, employee, item, etc.) was last observed.

The probability that an event (in our case customer churn/turnover) occurs during time interval t is known as the hazard function $h(t) = P(T = t \mid T \geq t)$. Another important function is the survival distribution function (SDF), which is $S(t) = P(T > t)$.

Most widely used survival analysis techniques for LTV modeling include parametric regression models (a.k.a. accelerated failure time models), Cox proportional hazards regression, and discrete-time logistic hazard models. The discrete-time logistic hazard models were first introduced in biostatistics (Brown 1975), but found more demand in the social sciences (Allison 1982) and business applications (Potts 2003 & 2005). They are better suited for handling events that are observed in discrete units and are able to deal with various complexities of real world applications such as time-dependent covariates, time-varying effects of the covariates, competing risks, and nonlinear hazards.

We will consider a special case of discrete-time logistic hazard models when the model can be expressed by this formula:

$$\text{Log}\,[h(t)/(1-h(t))] = \psi(t,\alpha) + \sum_{j=1}^{m} \beta_j x_j\,, \qquad (2)$$

where h(t) is the hazard function, x_j are covariates, α, β are vectors of parameters to be estimated, and ψ(t,α) is a cubic spline. Let us note that x_j are not time-dependent covariates in (2). In real world applications time-dependent $x_j(t)$ pose a serious challenge in implementation, because we would need to predict future values of $x_j(t)$ to be able to estimate the hazard function going forward. However, since t is a part of ψ(t,α), it is a predictor variable in (2), and therefore strictly speaking our special case still deals with time-dependent covariates. The use of cubic splines allows to track rather arbitrary patterns that are changing at specified points in time (knots). This is especially important for contract customers whose propensities to churn (i.e. hazard) while in contract vs. out-of-contract can be quite different due to their contractual financial obligations.

It is known (Brown 1975) that the likelihood function for the discrete-time logistic hazard models is equivalent to the likelihood function of the logistic regression model for expanded data (with dependent observations) looking like this (shown for the i^{th} subject):

t	event	covariates
0	0	$\mathbf{x}_i$
1	0	$\mathbf{x}_i$
. . .		
. . .		
T_{i-1}	0	$\mathbf{x}_i$
T_i	C_i	$\mathbf{x}_i$,

where T_i is the time of the last observation; C_i = 0 means censoring and C_i=1 indicates that the event of interest occurred at time T_i.

The expanded data set can become very large. There are several ways to reduce the computational burden caused by the amount of data (Potts 2003). We used sub-sampling non-events in macro %LOGHAZ (Pliner 2005). While keeping all event records (with C_i=1), the non-events are randomly sampled with the selection probability P_s specified by the user. The bias caused by sampling can be corrected by adding Log(P_s) to the intercept of the logistic regression equation, and as of SAS 9.1, this can be done automatically using either PRIOREVENT= or PRIOR= options of the SCORE statement in PROC LOGISTIC.

After we specify ψ(t,α) in (2), running PROC LOGISTIC on the expanded data will estimate parameters in (2), which was implemented in macro %LOGHAZ for a cubic spline ψ(t,α) having this form:

$$\psi(t,\alpha) = \sum_{j=1}^{J} \alpha_j\, b(t,k_j),$$

where J is the number of knots k_j and $b(t, k_j) = I\{t > k_j\}(t- k_j)^3 - t^3 + 3 k_j t^2 - 3 k_j^2 t$. Such a selection of ψ(t,α) makes the hazard function constant over time when t > k_J (assuming the knots are sorted in the ascending order).

The selection of the number of knots and their placement is equivalent in this context to a variable selection, where $b(t,k_j)$ can be treated as potential predictors similarly to covariates x_j. Knots can be added, dropped, or moved, then PROC LOGISTIC run again. To select the best model traditional approaches can be applied, such as, for example, comparing AIC among competing models.

For scoring, macro %LOGHAZ writes a code similarly to (Potts 2003). There is a convenient alternative to having a scoring text file in recent versions of SAS: using PROC LOGISTIC with the SCORE statement.

After hazards are estimated, the survival probabilities can be calculated using the following formula expressing the relationship between the hazard function and SDF:
$S(t)=S(t-1)*[1-h(t)]$, with $S(0)$ being equal to 1. And finally p_t in (1) are estimated as $S(t+t_0)/S(t_0)$, where t_0 is the customer's tenure with the company at the time these calculations are carried out. Here is the logic behind this:

$$p_t = P(T > t+t_0) | T > t_0) = P(T > t+t_0 \text{ and } T > t_0) / P(T > t_0) = P(T > t+t_0) / P(T > t_0) = S(t+t_0)/S(t_0)$$

It is worth noting that the above approach of survival analysis has been also implemented in the SAS Enterprise Miner starting from version 7.1 (Schubert et al 2012).

3. PREDICTING PROFIT MARGINS

Instead of assuming that margins M_t in (1) are constant for each customer we applied linear regression:

$$M_t = \sum_{j=1}^{k} a_j M_{t-j} + \sum_{i=1}^{m} b_i z_i, +ct + d, \qquad (3)$$

using as predictors latest historic margins in addition to *t* and other variables (z_i) that could potentially improve the model's predictability.

If one does not want to discard relatively new customers for whom some historic margins in (3) may not be available yet, those values should be imputed.

The parameters in (3) can be easily estimated with either PROC GLM or PROC REG, but one may encounter a problem if outliers are present in data. In fact profit margins can have extremely high positive and even negative (in case of refunds) values especially if we deal with business customers.

To handle outliers we suggest using robust regression which is capable of providing stable results in the presence of outliers. Four methods of robust regression are implemented in PROC ROBUSTREG (Chen 2002): the Huber M estimation, Least Trimmed Squares (LTS), S estimation, and MM estimation. The last three (LTS, S, and MM) are suited for scenarios with outliers in both dependent and independent variables, which is the case when we deal with outliers in profit margins. The method is specified on the PROC ROBUSTREG statement (METHOD=LTS or METHOD=S or METHOD=MM). A few additional options can be used with those methods. For example, for METHOD=LTS, the option h= specifies the quantile for the LTS estimate. LTS estimates are based on minimizing the sum of h lowest linear regression residuals. You can specify any integer h between $[n/2]+1$ and $[(3n+p+1)/4]$, where n is the number of data points and p is the number of independent variables ($p= m+k+1$ in our case). By default, $h = [(3n+p+1)/4]$.

To select the best performing method one can randomly split available data into two data sets: the first one will be used for model building (estimating parameters) and the second set – for validation. If the amount of available data is too limited for such a split, a cross-validation procedure can be applied.

After the parameters in (3) and M_t are estimated, this equation is used to estimate all future margins in (1) going forward by shifting historic margins by one period at a time so that just estimated M_t serves as M_{t-1} at the next step.

REFERENCES

Allison P.D. 1982. "Discrete-time methods for the analysis of event histories." In Samuel Leinhardt (ed.),

Sociological Methodology, Jossey-Bass.
Brown C.C. 1975. "On the use of indicator variables for studying the time dependence of parameters in a response-time model." *Biometrics*, 31: 863-872.
Chen C. 2002. "Robust regression and outlier detection with the ROBUSTREG procedure." *Proceedings of the twenty seventh annual SAS users group international conference (SUGI 27),* Orlando, FL.
Pliner V 2005. "Predictive Modeling Using Survival Analysis." *Proceedings of NESUG 2005 (18th Annual NorthEast SAS Users Group Conference)*; Portland, ME.
Potts, W. 2003. "Survival data mining: predictive hazard modeling for customer history data." Will Potts, Data Miners Inc.
Potts, W. 2005. "Predicting Customer Value.*" Proceedings of the SAS Global Forum 2005 Conference*, Philadelphia, PA
Schubert, S, Haller, S, and Lee T. 2012. "It's About Time: Discrete Time Survival Analysis Using SAS® Enterprise Miner™." *Proceedings of the SAS Global Forum 2012 Conference.* Orlando, FL.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

vadim.pliner@verizonwireless.com