Paper 073-30

Predicting Customer Value

Will Potts, Data Miners, Inc.

ABSTRACT

The future value of a contractual customer depends on the remaining lifetime of their products and the path of their future cash flows. Predicting customer value involves modeling the churn hazard as a function of tenure and other customer attributes. The mean, restricted mean, and median lifetime value are computed by scoring future time intervals with the hazard model.

REMAINING CUSTOMER VALUE

Customer lifetime value (CLV) is used for allocating discretionary marketing investments (Malthouse and Blattberg 2005) and for corporate valuation (Bauer et al. 2003).

Pfeifer et al. (2005) distinguish the finance usage of *value* from the accounting usage of *profit*. A customer's lifetime value is the present value of their future cash inflows minus cash outflows. In contrast, a customer's profitability is their accrual-based revenue minus costs over a fixed, usually past, time.

The remaining value of a customer from the current time *t* forward depends on their residual life

$$R_t = 0,1,2,\ldots .$$

The residual life is the time remaining until churn, the end of the customer relationship. It is a discrete random variable having non-negative integer values. Zero means that the customer churned at the current time *t*. If the current time is zero—the customer just started, then their residual life is their entire life.

The remaining CLV from the current time *t* forward is the sum of the discounted cash flows over the residual life

$$val(R_t) = \sum_{r=0}^{R_t} \delta_r (1+d)^{-r} .$$

The cash flows δ_r linked to the customer relationship can vary in time. The churn time is defined to be the last time with non-zero cash flows attributed to the customer. Thus, all customers active at *t* are worth at least δ_0, even those who churned at $R_t = 0$.

The present value of future cash flows is discounted for not having the opportunity to invest the money at time *t*. The discount factor $(1+d)^{-1}$ is the amount you would pay now for one monetary unit received at one time later. The discount rate per time unit *d* is compounded once every time unit.

CLV is a random variable. Business decisions are based on the center of its probability distribution. The general formula for the expected CLV is an infinite sum of an ever-expanding product

$$\begin{aligned} E(val(R_t)) &= \sum_{r=0}^{\infty} val(r)\Pr(R_t = r) \\ &= \sum_{r=0}^{\infty}\sum_{j=0}^{r} \delta_j (1+d)^{-j} \Pr(R_t = r) \\ &= \delta_0 + \sum_{r=1}^{\infty} \delta_r (1+d)^{-r} \prod_{j=t}^{t+r-1} (1 - h(j \mid \mathbf{x}(j))) \end{aligned} .$$

For discrete-time data, the hazard function $h(t)$ is the conditional probability of churn at time *t* given that the customer has not churned yet

$$\begin{aligned} h(t \mid \mathbf{x}(j)) &= \Pr(T = t \mid T \ge t, \mathbf{x}(j)) \\ &= \Pr(R_t = 0 \mid \mathbf{x}(j)) \end{aligned} .$$

The random variable *T* is the duration until churn (customer tenure). The hazards depend on a vector of customer covariates $\mathbf{x}(t)$.

Note that the mean CLV does not equal the value function evaluated at the mean residual life

$$E(val(R_t)) \neq \sum_{r=0}^{E(R_t)} \delta_r (1+d)^{-r}$$

(unless the cash flows are constant and the discount rate is zero). CLV is a nonlinear function of the residual life, so simply plugging in the mean of R_t gives only a rough first-order approximation.

The formula for the mean CLV depends on three unknowns:

- The value function needs to be specified.
- The churn hazard needs to be modeled as a function of tenure and other customer covariates.
- The summation goes to infinity, but the data does not. Either we can make assumptions about the shape of the hazard outside the range of the data or we can reformulate the problem.

SPECIFYING THE VALUE FUNCTION

If we assume that the δ_r are time-constant, then they can be estimated using the current revenue and expenses associated with the customer's products and services. This accounting could be done for each individual or for customer segments.

When the δ_r are time-constant (equal to δ), the value function is a geometric series

$$val(R_t) = \sum_{r=0}^{R_t} \delta(1+d)^{-r} = \delta\left(\frac{1+d}{d}\right)\left(1-(1+d)^{-(R_t+1)}\right).$$

CLV follows a confined exponential growth curve, increasing at a decreasing rate from δ to an upper asymptote

$$\lim_{R_t \to \infty}(val(R_t)) = \delta\left(\frac{1+d}{d}\right).$$

As the discount rate *d* approaches zero the value function becomes linear

$$\lim_{d \to 0}(val(R_t)) = \delta(R_t+1).$$

Reichheld (1996, chapter 2) argues that the δ_r increase with customer tenure because of increasing revenue growth, operating cost savings, referrals, and a price premium. Berger and Nasr (1998, p. 24) give an example of credit card customers where profit per time unit increases sigmoidally. The form of the functions can be estimated using regression on historical customer cash flows.

If the δ_r increases linearly $\delta_r = \alpha_0 + \alpha_1 r$ (for $\alpha_1 > 0$), then the value function is an arithmetic-geometric series

$$\begin{aligned} val(R_t) &= \sum_{r=0}^{R_t} (\alpha_0 + \alpha_1 r)(1+d)^{-r} \\ &= \alpha_0 \frac{1}{d}\left(1+d-(1+d)^{-R_t}\right) + \alpha_1 \sum_{j=1}^{R_t}\sum_{r=j}^{R_t}(1+d)^{-r} \\ &= \frac{1+d}{d}\left(\alpha_0 + \frac{\alpha_1}{d}\right) - \frac{(1+d)^{-R_t}}{d}\left(\alpha_0 + \frac{\alpha_1(1+d)}{d} + \alpha_1 R_t\right) \end{aligned}$$

CLV increases sigmoidally from α_0 to an upper asymptote at

$$\lim_{R_t \to \infty}(val(R_t)) = \frac{1+d}{d}\left(\alpha_0 + \frac{\alpha_1}{d}\right).$$

As the discount rate *d* approaches zero, the inflection point moves toward infinity and value function becomes an increasing quadratic curve

$$\lim_{d \to 0}(val(R_t)) = (2\alpha_0 + R_t\alpha_1)(R_t+1)/2.$$

Abrupt falls in the individual cash flows are caused by dropping products or services. Banking customers can close their money market account but keep their checking account. Insurance customers can drop their vehicle policy but keep their homeowners policy. When a company offers several detached products or services, you can avoid these abrupt changes by modeling product churn instead of customer churn. The cash flows are separated by product and separate churn models are built for each product. CLV is the sum of the product-level lifetime values within a customer.

Abrupt rises in the individual cash flows are caused by adding products or services. When the probability of these events is small, incorporating predictive models for add-on buying can cause greater error than assuming no product additions.

HAZARD MODELING

The churn hazard can be estimated using customer history data extracted from company databases. The outcome is the duration from inception to churn. Customers who are still active at the observation date are censored. The available data is often left-truncated to exclude customers who churned before some date. The data also contains customer covariates such as the product changes, usage, marketing channel, and demographics. Many of the covariates are time-dependent.

Hazard models are used to estimate the shape of the hazard function (the time effect) and how the shape is affected by the covariates. Predictive hazard models can be used to score customers. The inputs are the current customer tenure and the current values of the other covariates. The output is the churn hazard at that time—the conditional probability of churn.

The discrete-time logistic-hazard model is well suited to customer history data. The model is fit using logistic regression on data expanded longitudinally to one row for each time that each customer was at risk. The nonlinear shapes of the churn hazard can be handled in PROC LOGISTIC by including regression spline terms or other transformations of time. Neural networks are a more flexible but more troublesome alternative. PROC NEURAL can fit the logistic-hazard model.

The semi-parametric Cox model can be adapted for prediction by coupling it with profile likelihood estimates of the baseline hazard. The most effective way of doing this in PROC PHREG is with the counting-process style input and the BASELINE statement. If the TIES=DISCRETE option is used on the MODEL statement, the form of the Cox model is equivalent to the logistic-hazard model with a saturated (over-parameterized) time effect.

Parametric accelerated failure time (AFT) models assume that the outcome time follows a named continuous distribution such as the Weibull or log-logistic distribution. PROC LIFEREG and PROC RELIABILITY support several probability distributions but none for discrete time. Parametric AFT models can make convenient scoring models—though this is seldom their purpose. Their chief disadvantage is their inflexibility. The shape of the baseline hazard and the way the covariates affect the hazard is constrained by the distributional assumptions. Another disadvantage is that they cannot handle time-dependent covariates.

CONSTANT HAZARD EXTRAPOLATION

Plugging-in the maximum likelihood estimate of the hazard gives the maximum likelihood estimate of the mean CLV

$$\delta_0 + \sum_{r=1}^{\infty} \delta_r (1+d)^{-r} \prod_{j=t}^{t+r-1} (1-\hat{h}(j \mid \mathbf{x}(j))) .$$

You cannot write a DO loop to compute the formula directly. But this infinite series converges for certain value functions and certain assumptions about the hazard.

Assume that after some time point ***z*** the hazard becomes constant at it’s current value

$$h(j) = h(z) \quad \text{for } j \geq z .$$

The time point ***z*** could be the maximum tenure in the data or longer if you trust the ability of the hazard model to extrapolate. Under this assumption, the mean converges for the value function with constant cash flows $\delta_r = \delta$

$$E(val(R_t)) = \delta + \delta \sum_{r=1}^{\infty} (1+d)^{-r} \prod_{j=t}^{t+r-1} (1 - h(j|\mathbf{x}(j)))$$
$$= \delta + \delta \sum_{r=1}^{z-t} \left((1+d)^{-r} \prod_{j=t}^{t+r-1} (1 - h(j|\mathbf{x}(j))) \right) + \delta \sum_{r=z-t+1}^{\infty} (1+d)^{-r} \prod_{j=t}^{z-1} (1 - h(j|\mathbf{x}(j))) \prod_{j=z}^{t+r-1} (1 - h(z|\mathbf{x}(z))) .$$
$$= \delta + \delta \sum_{r=1}^{z-t} \left((1+d)^{-r} \prod_{j=t}^{t+r-1} (1 - h(j|\mathbf{x}(j))) \right) + \delta (1+d)^{-(z-t)} \frac{1 - h(z|\mathbf{x}(z))}{d - h(z|\mathbf{x}(z))} \prod_{j=t}^{z-1} (1 - h(j|\mathbf{x}(j)))$$

The formula is not pretty, but all the sums and the products are finite. The value function with linear cash flows is tractable as well.

RESTRICTED CLV

For many business problems, it is more sensible to base CLV on the restricted residual life

$$\min(R_t, \gamma + 1)$$

the remaining time until churn or an upper bound—whichever comes first. This new random variable replaces R_t with an upper bound $\gamma + 1$ whenever R_t exceeds γ.

The value of a customer over the next $\gamma + 1$ time units is a function of the restricted residual life

$$val(\min(R_t, \gamma + 1)) = \sum_{r=0}^{\min(R_t, \gamma+1)} \delta_r (1+d)^{-r} .$$

The mean restricted CLV is the finite sum of a finitely expanding product

$$E(val(\min(R_t, \gamma + 1))) = \delta_0 + \sum_{r=1}^{\gamma+1} \delta_r (1+d)^{-r} \prod_{j=t}^{t+r-1} (1 - h(j|\mathbf{x}(j))) .$$

This is the expected value of a customer over the next $\gamma + 1$ time units. The mean restricted CLV is more sensibly interpreted as the expected *long-term* value. The upper bound $\gamma + 1$ might represent the limit on a meaningful remaining lifetime. So that customers whose residual life exceeds this time horizon are considered equal.

The discount factor causes diminishing returns for longer lifetimes. The upper bound could be chosen as the time when the value function is within *c* units of the upper asymptote. For the value function with constant cash flow, choosing γ to be the smallest integer greater than

$$\frac{\ln(\delta / c) - \ln(d)}{\ln(1+d)}$$

ensures that CLV is within *c* units of its maximum.

The restricted mean CLV can be computed for any given value function. Computing the sum of an expanding product

$$\begin{aligned}
&\delta_0 \\
&+\delta_1 (1+d)^{-1} (1 - h(t|\mathbf{x}(t))) \\
&+\delta_2 (1+d)^{-2} (1 - h(t|\mathbf{x}(t)))(1 - h(t+1|\mathbf{x}(t+1))) \\
&+\ldots \\
&+\delta_{\gamma+1} (1+d)^{-(\gamma+1)} (1 - h(t|\mathbf{x}(t)))(1 - h(t+1|\mathbf{x}(t+1))) \cdots (1 - h(t+\gamma|\mathbf{x}(t+\gamma)))
\end{aligned}$$

using two nested DO loops is inefficient, because it would involve repeatedly calculating the hazard for the same customer at the same time. A more efficient method is to rearrange sum of the products so that the elements of the product only appear once. The general form can be factored

$$v_1 a_0 + v_2 a_0 a_1 + v_3 a_0 a_1 a_2 + v_4 a_0 a_1 a_2 a_3 = a_0 (v_1 + a_1 (v_2 + a_2 (v_3 + a_3 v_4)))$$

and computed with a single DO loop. The a_j represent the elements of the product $1 - h(j)$ and the v_r represent the coefficients of the products $\delta_r (1+d)^{-r}$. On the right-hand-side, each a_j only appears once.

The following program demonstrates the method. For simplicity, the cash flows are assumed constant and the hazard function has no covariates other than time.

```
%let gamma=11;
%let delta=100;
%let d=.0125;
%let hf=1/(1+exp(3-0.7*tanh(0.3-0.09*t)-1*tanh(0.2-0.4*t)
            +1*tanh(1-0.5*t)));

data a;
set a(rename=(t=t0));
rclv=&delta/(1+&d)**(&gamma+1);
do r=&gamma to 0 by -1;
   t=t0+r;
   hf=&hf;
   rclv=rclv*(1-hf)+&delta/(1+&d)**r;
end;
run;
```

The dataset contains the current tenure of a set of customers. The time variable is renamed so that the hazard can be evaluated at a range of future times. The program creates a new variable RCLV for the expected value over the next 12 time units. The restricted CLV is calculated in the factored form starting from the inner most term. Thus, the DO loop counts down, starting at the last time.

More complicated hazard functions and time-dependent cash flows could be included (%INCLUDE) from text files instead of using MACRO variables. If the hazard function depends on time-dependent covariates, then forecasting formulas would need to be included as well.

MEDIAN CLV

The median CLV can be computed by simply plugging in the median residual life

$$med(val(R_t)) = val(med(R_t)) \; .$$

This convenient identity does not hold with the mean unless the value function is linear.

For the value function with constant cash flows the median equals

$$med(val(R_t)) = \delta\left(\frac{1+d}{d}\right)\left(1-(1+d)^{-(med(R_t)+1)}\right) .$$

For the value function with linear cash flows the median equals

$$med(val(R_t)) = \frac{1+d}{d}\left(\alpha_0 + \frac{\alpha_1}{d}\right) - \frac{(1+d)^{-med(R_t)}}{d}\left(\alpha_0 + \frac{\alpha_1(1+d)}{d} + \alpha_1 med(R_t)\right).$$

Half of all customers of the same tenure and the same values of the other covariates will be worth more than the median, half will not.

The probability distribution of CLV is usually not symmetric. If the discount rate *d* is relatively small, then it is often positively skewed. If *d* is relatively large, then most of the CLV is close to the upper asymptote and the distribution is negatively skewed. The median CLV is arguably a better measure of centrality than the mean for skewed distributions.

The median residual life does not have a closed form but is easy to compute. It is the solution to the equation

$$\prod_{j=t}^{t+med(R_t)} (1-h(j\,|\,\mathbf{x}(j))) = \tfrac{1}{2} \; .$$

The left hand side is the $\Pr(R_t > med(R_t))$ expressed in terms of the hazards. This equation can be solved by iterating from *t*, updating the product, and stopping when it equals or drops below ½.

The following program finds the median residual life and plugs it into the value function with constant cash flows to estimate the median CLV.

```
data a;
set a(rename=(t=t0));
cp=1;
do t=t0 to 1000 until(cp<=.5);
   hf=&hf;
   cp=cp*(1-hf);
end;
medr=t-t0;
medclv=&delta*((1+&d)/&d)*(1-(1+&d)**(-(medr+1)));
run;
```

It looks like this program computes the restricted median residual life, because there is an upper limit on the DO loop. Choose the upper limit large enough so that it is unlikely that the DO loop will reach the end.

CONCLUSION

Three tractable ways to predict CLV were presented. The first way estimates the mean by assuming that the churn hazard eventually levels off. The second and third ways estimate other measures of centrality: the restricted mean and the median. Each of the three solutions requires the statistical modeling of the churn hazard and the financial analysis of the revenues and expenses associated with the customers products and services.

The foregoing definition of CLV is suitable for businesses with contractual products such as banking, insurance, and telecom. When a contractual customer churns, they are considered to be “lost-for-good,” because the relationship ends when the account is closed, the policy is canceled, or the service is disconnected.

In contrast, businesses such as mail order and online retail have non-contractual products. Their customers are characterized as “always-a-share,” because the customer relationship does not formally end. You could try to fit non-contractual customers into this CLV framework by defining churn as a long period of inactivity. A better fitting framework would be based on the intensity function of future purchases.

REFRENCES

Allison P.D. (1982), “Discrete-Time Methods for the Analysis of Event Histories,” in *Sociological Methodology 1982*, Jossey-Bass.

Allison P.D. (1995), *Survival Analysis using the SAS System: A Practical Guide*, SAS Institute Inc.

Bauer, H.H., Hammerschmidt, M., and Braehler, M. (2003), “The Customer Lifetime Concept and its Contribution to Corporate Valuation,” In: *Yearbook of Marketing and Consumer Research*, 1, 47-67.

Berger, P.D. and Nasr, N.I. (1998), “Customer Lifetime Value: Marketing Models and Applications,” *Journal of Interactive Marketing*, 12, 17-30.

Malthouse E.C. and Blattberg R.C. (2005), “Is it Possible to Predict Customer Long-Term Value?,” *Journal of Interactive Marketing*, to appear.

Pfeifer, P.E., Haskins, M.E., and Conroy, R.M. (2005) “Customer Lifetime Value, Customer Profitability, and the Treatment of Acquisition Spending,” *Journal of Managerial Issues*, to appear.

Reichheld F.F. (1996), *The Loyalty Effect*, Harvard Business School Press.

Singer, J.D. and Willett, J.B. (1993), “It’s About Time: Using Discrete-Time Survival Analysis to Study Duration and the Timing of Events,” *Journal of Educational Statistics*, 18, 155-195.

CONTACT INFORMATION

Email: wjep@data-miners.com