Predicting Customer Lifetime Value --- 2008 DMEF Customer Lifetime Value Modeling Competition (Task 2)

Zainab Jamal, Alex Zhang

HP Laboratories
HPL-2009-224

Keyword(s):
Lifetime value (LTV); Logit models; Linear models; Regression Tree; Recency, Frequency, and Monetary (RFM).

Abstract:
We describe the 2008 DMEF (Direct Marketing Educational Foundation) Customer Lifetime Value Modeling Competition, which the HP Labs team consisting of Zainab Jamal and Alex Zhang won Task 2 (out of 3 tasks). We present our approach for predicting individual donor's total gift amount over a two-year target period. We divide the donors into 8 segments; for each segment, we fit a logit model for predicting the probability of giving, and a log-linear model for predicting the amount of gifts conditional on a donor giving. We found that recency, frequency, and first gift amount are good predictors of the probability of giving, while time-weighted total gift amount in the past years is a good predictor for future gift amount.

External Posting Date: September 6, 2009 [Fulltext] Approved for External Publication
Internal Posting Date: September 6, 2009 [Fulltext]

To be published in Journal of Interactive Marketing, Vol 23, 2009.

Predicting Customer Lifetime Value --- 2008 DMEF Customer Lifetime Value Modeling Competition (Task 2)

Zainab Jamal and Alex Zhang
zainab.jamal@hp.com; alex.zhang@hp.com
HP Labs
1501 Page Mill Road, M/S 1140
Palo Alto, CA 94304

August 27, 2009

Abstract

We describe the 2008 DMEF (Direct Marketing Educational Foundation) Customer Lifetime Value Modeling Competition, which the HP Labs team consisting of Zainab Jamal and Alex Zhang won Task 2 (out of 3 tasks). We present our approach for predicting individual donor's total gift amount over a two-year target period. We divide the donors into 8 segments; for each segment, we fit a logit model for predicting the probability of giving, and a log-linear model for predicting the amount of gifts conditional on a donor giving. We found that recency, frequency, and first gift amount are good predictors of the probability of giving, while time-weighted total gift amount in the past years is a good predictor for future gift amount.

Keywords:
Lifetime value (LTV); Logit models; Linear models; Regression Tree; Recency, Frequency, and Monetary (RFM).

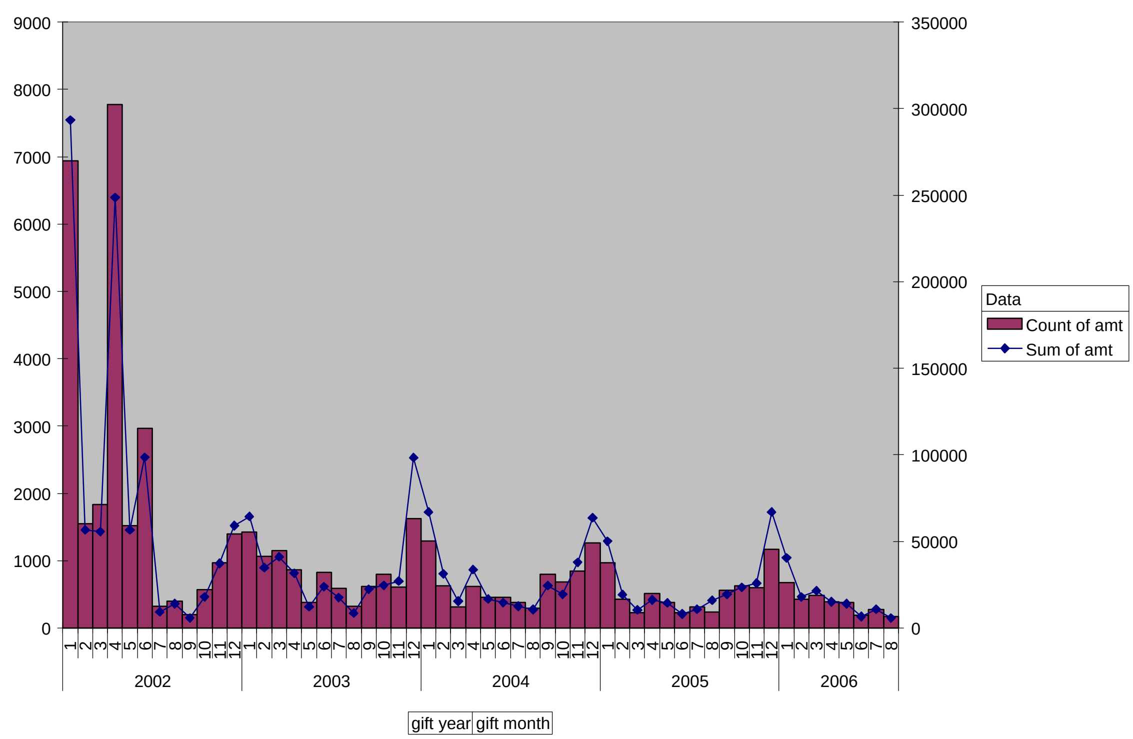
1. Introduction: The 2008 DMEF Customer Lifetime Value Modeling Competition

Customer lifetime value and customer equity have attracted widespread attention from marketing researchers and practitioners over the past 15 years. For a review of research methodologies and models in customer lifetime value, we refer to Gupta et al (2006) and Jain and Singh (2002). In 2008, DMEF (Direct Marketing Educational Foundation) held a Customer Lifetime Value Modeling Competition. According to the DMEF organizers, the purpose of this contest was to compare and improve the estimation methods and applications for these concepts.

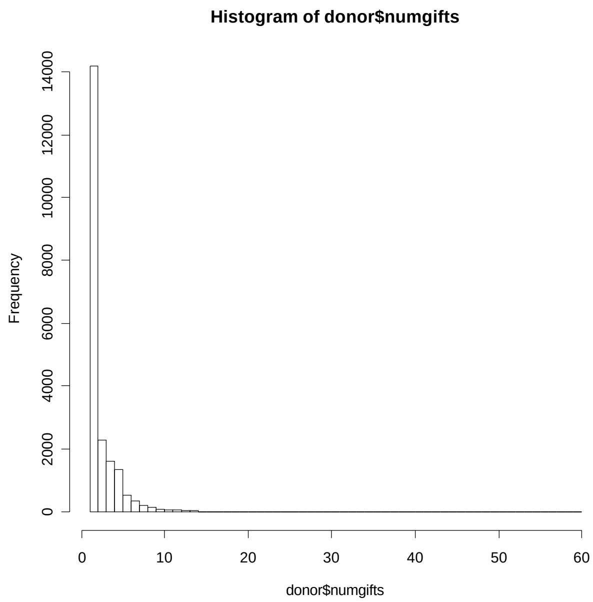
The setting and the data

The setting is to predict the gift dollar amount of 21,166 donors in a 2-year "target period", given the donation history of these (same) donors over the past 4.5 years.

A leading US nonprofit organization has agreed to release detailed transaction and contact (appeal) history of the 21,166 donors it acquired during the first half of 2002. The transaction and contact history extend through August 31, 2006. The donation history is a "log" (transactions) file that records the donor ID, gift amount, date of gift, and the "source" (code) of the campaign which the gift corresponded to. The following chart shows the aggregated number of gifts (histogram, scale on left) and the total gift amount in dollars (line chart, scale on right) over time:



The following histogram shows the statistics of the frequency of giving of these 21,166 donors; for example, 50.2% of the 21,666 donors gave exactly one gift (which happened in the first half of 2002); 16.7% gave 2 gifts; 10.8% gave 3 gifts; one donor gave 57 gifts.



The contact history is another log file that records donor ID, contact date, and "source" (code) of the campaign.

No demographic descriptions about the donors are available to us. The "source" of a campaign is described only by its "cost", and no other attributes about the appeal's content or how an appeal was executed are given.

The tasks

The competition requires completion of any of the following three tasks:

Task 1: Aggregate estimate of total donations from the donors provided during the two-year "target" period. This evaluates different estimates of the "customer equity" of the nonprofit organization.

Task 2: Best estimates of individual-level donations, as measured by the mean squared error of the logged donations:

$$\text{MSE} = \sum_i \left(\log(y_i + 1) - \log(\hat{y}_i + 1)\right)^2 / 21{,}166.$$

where y_i is the true future value for customer i over the "target period" (September 1, 2006 through August 30, 2008) and $\hat{y}_i$ is the prediction. This task is meant to test how well models can predict the future behavior of individual donors.

Task 3: Best predictions, as measured by the highest classification rate, of which 2-year lapsed donors (i.e., those who donated most recently between September 1, 2004 and August 30, 2006) will make at least one donation during the target period.

We have won Task 2 by achieving the least mean-squared-error among the contestants. For an account of the approaches and a comparison of modeling techniques among the contestants, we refer to Malthouse (2009). In the next section, we will describe our approach in detail.

2. Our Modeling Approach

Our approach for Task 2 involves several sub-tasks: (1) Re-labeling the years to determine a training period and a test period; (2) Segmenting the 21,166 donors and building a separate model for each segment; (3) Looking for predictive attributes and determining the model form; (4) Applying the model to the target period. Below, we describe each of these components in more detail. These technique details are presented also in Jamal and Zhang (2009).

Training period and test period

The observation period is from 2002-01-01 (here we use yyyy-mm-dd format) through 2006-08-31; the "target period" for prediction is from 2006-09-01 through 2008-08-31. Since the target period is 2 years, we use the last 2-year period from 2004-09-01 through 2006-08-31 as our test period; hence, our dependent variable y_i is the total gift amount (in dollars) of donor i in the test period 2004-09-01 through 2006-08-31.

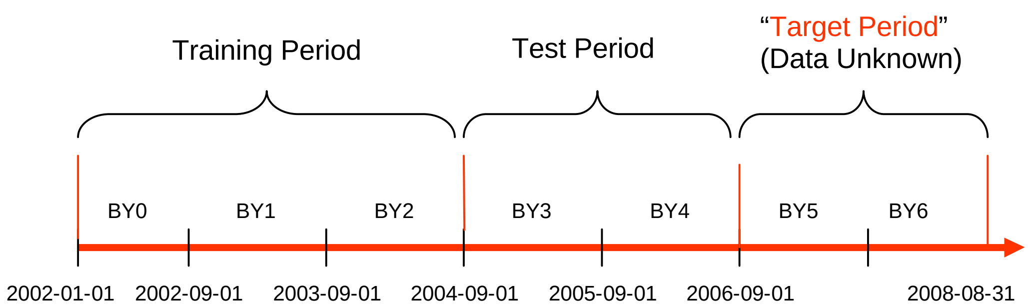


Figure 1: Training, test, and target periods

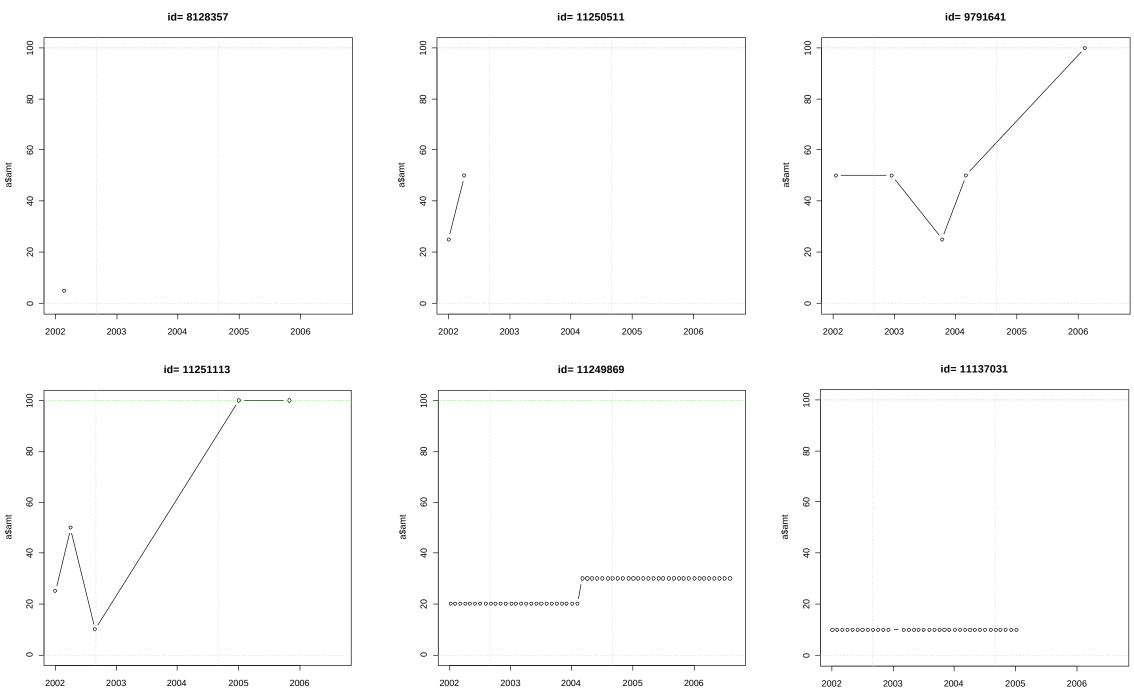
Shown in Figure 1, we re-label the years as follows:

BY0 = 2002-01-01 through 2002-08-31 (8 months);
BY1 = 2002-09-01 through 2003-08-31 (1 year);
BY2 = 2003-09-01 through 2004-08-31 (1 year);
BY3 = 2004-09-01 through 2005-08-31 (1 year);
BY4 = 2004-09-01 through 2005-08-31 (1 year).

Observations for donor *i* prior to 2004-09-01 are used to form potential predictors. In effect, we set 2004-09-01 as our time 0. Recency attribute values will then be computed relative to time 0; for example, a recency of 61 days means that the donor’s last gift had occurred 61 days prior to 2004-09-01 (i.e. on 2004-07-01).

Segmenting the 21,166 donors

We divide the donors into one-time donors and multi-time donors (who gave at least twice). The one-time donors gave exactly one gift between the first half of 2002 and 2004-09-01. As an initial investigation, we examine the time pattern of giving of individual donors, see below for some (more prominent) cases, where each chart shows the gift amount (the vertical axis) and the date of each gift (the horizontal axis) of an individual donor. We note that 54.2% of the 21,166 donors gave only once – the pattern typified in the first of the 12 charts below.



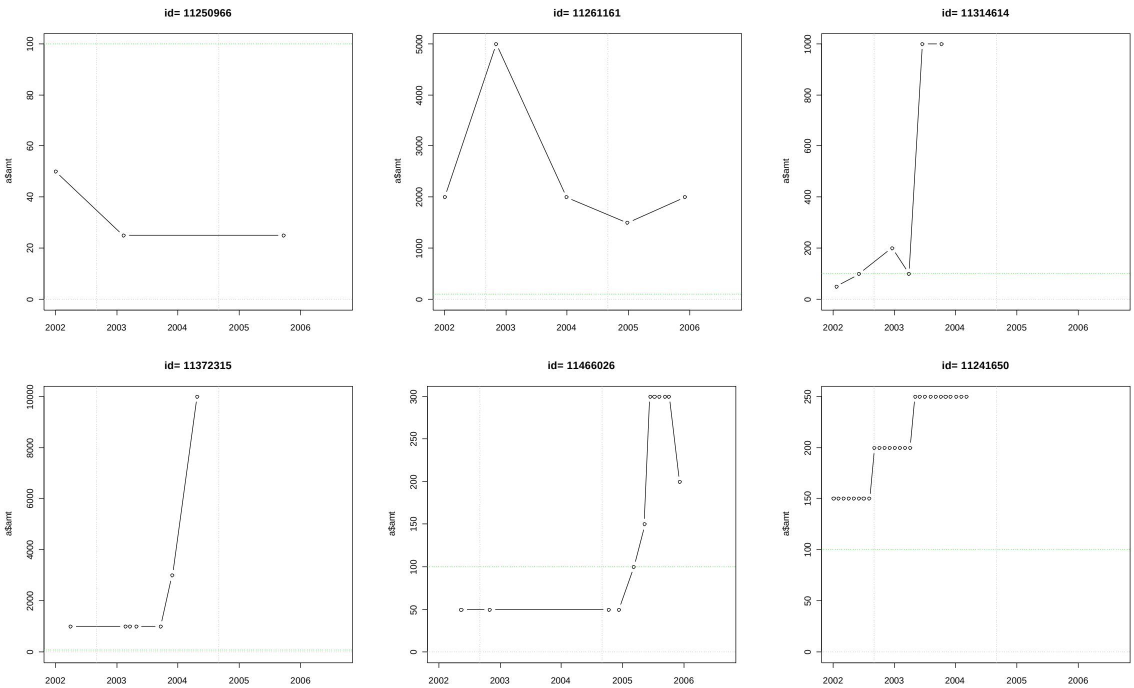


Figure 2: Some patterns of individual donations

We then use a CART tree to segment the 21,166 donors. Combined with our own intuition and judgment, we eventually divide the 21,166 donors into 8 segments, depending on the each donor's first gift amount, recency (last gift date prior to 2004-09-01, hence "the last two years" refers to 2002-09-1 through 2004-08-31, and "the past year" refers to 2003-09-01 through 2004-08-31), and frequency (number of gifts):

– One-time donors: 11,474 donors. All one-time donors are lapsed donors. Depending the amount of the first gift (attribute "firstgiftamt"), we further have:

  - 1a: firstgiftamt =$1 or >=$1,000: 56 donors.

  - 1b: $1 < firstgiftamt <= $47.5: 9,086 donors.

  - 1c: $47.5 < firstgiftamt < $1,000: 2,332 donors.

– Multi-time donors (at least twice): 9,692 donors:

  - 2a: Frequent donors (recency • 61 and number of gifts in the past two years > 6): probability of donating in BY3 and BY4 is 0.9912. 114 donors are "frequent donors".

  - 2b: Lapsed multi-time donors (gave no gifts in the last two years), with average gift amount of less than $7.25 or maximum gift amount • $1000: 47 donors.

- 2c: Lapsed multi-time donors (gave no gifts in the last two years), minus 2b: 525 donors.

- 2d: The regular donors: made at least one donation in the past two years, is not a frequent donor (2a), and has not donated more than $10,000 in the past year (sumgiftsBY2 < 10000): 9,004 donors.

- 2e: Gave $10,000 or more in the past year (sumgiftsBY2 >= 10000): 2 donors.

Predictive models

We first predict the probability $\Pr\{y_i > 0\}$, then predict $E[\log(1+ y_i) \mid y_i > 0]$. The desired $E[\log(1+ y_i)] = E[\log(1+ y_i) \mid y_i > 0] \Pr\{y_i > 0\}$, as $E[\log(1+ y_i) \mid y_i = 0] = E[\log(1+0)] = 0$.

For each of the eight segments, we use a logit model to predict the probability $\Pr\{y_i > 0\}$. To predict $E[\log(1+ y_i) \mid y_i > 0]$, we use an ordinary least squares (OLS) linear regression of $z_i = \log(1+ y_i)$ on the predictor variables using only cases where $y_i > 0$. Figure 3 below shows that y_i (given $y_i > 0$) is a long-tailed distribution while $z_i = \log(1+ y_i)$ given $y_i > 0$ has a bell-shaped histogram, more amenable to the normality assumption required for OLS. The log transformation also serves as variance stabilizing device which also reduces the influence of outliers (large gifts). Furthermore, OLS regression minimizes the squared error $\sum_i [\log(1+y_i) - \log(1+\hat{y}_i)]^2$.

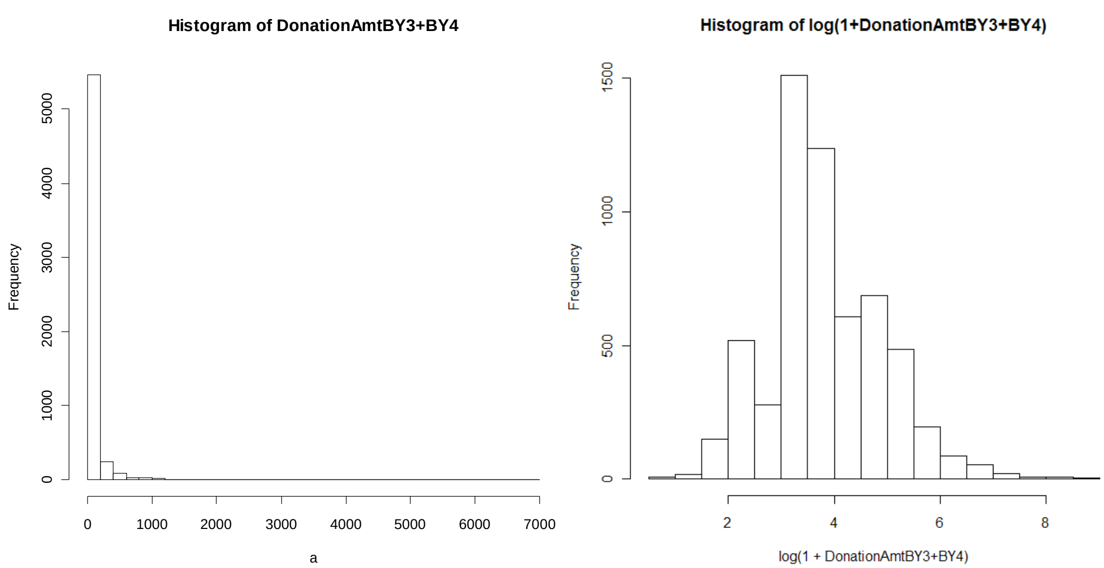


Figure 3: Distribution of y_i (left) and $\log(1+y_i)$ (right)

We use the R statistical software package to fit all our models. In the following, we give the model summary in R for each of the eight segments.

Segment 1a: $\Pr\{y_i > 0\} = 0$ (none of the 56 donors in segment 1a gave in the test period).

Segment 1b: The predictors are the dollar amount of the first (and only) gift (firstgiftamt), and the source code of the first gift (firstgiftsource). The logit model is given below:

```
                                      Estimate Std. Error z value Pr(>|z|)
(Intercept)                             -7.501      1.463   -5.13  3.0e-07 ***
log(1 + firstgiftamt)                    3.738      1.110    3.37  0.00076 ***
I((log(1 + firstgiftamt))^2)            -0.708      0.206   -3.44  0.00059 ***
ifelse(firstgiftsource == "7", 1, 0)     1.261      0.323    3.90  9.7e-05 ***
ifelse(firstgiftsource == "HF", 1, 0)   -0.639      0.325   -1.97  0.04893 *
---
    Null deviance: 4107.0  on 9085  degrees of freedom
Residual deviance: 4076.2  on 9081  degrees of freedom
AIC: 4086
```

Segment 1c: The logit model is given below:

```
                                      Estimate Std. Error z value Pr(>|z|)
(Intercept)                            -1.7541     0.0633  -27.69  < 2e-16 ***
ifelse(firstgiftamt > 125, 1, 0)       -1.0985     0.3167   -3.47  0.00052 ***
ifelse(firstgiftsource == "HF", 1, 0)  -0.6918     0.3073   -2.25  0.02436 *
---
    Null deviance: 1839.2  on 2331  degrees of freedom
Residual deviance: 1814.0  on 2329  degrees of freedom
AIC: 1820
```

Segment 2a: $\Pr\{y_i > 0\} = 0.9912$.

Segment 2b: $\Pr\{y_i > 0\} = 0$.

Segment 2c: The predictors are the minimum gift amount so far (mingiftamt), and firstgiftsource. The logit model is given below:

```
                                      Estimate Std. Error z value Pr(>|z|)
(Intercept)                              3.767      1.634    2.31   0.0212 *
log(1 + mingiftamt)                     -2.942      1.030   -2.86   0.0043 **
I((log(1 + mingiftamt))^2)               0.401      0.158    2.53   0.0113 *
ifelse(firstgiftsource == "HF", 1, 0)   -1.917      1.031   -1.86   0.0630 .
---
    Null deviance: 541.64  on 524  degrees of freedom
Residual deviance: 524.64  on 521  degrees of freedom
AIC: 532.6
```

Segment 2d: We look at a number of predictors: recency (any recency less than 60 days is rounded up to 60 days); number of gifts in the most recent year and the year before the most recent year, numgiftsBY2 and numgiftsBY1, respectively; the total gift amount in the most recent year, and the firstgiftsource. The logit model is given below:

```
                                          Estimate Std. Error z value Pr(>|z|)
(Intercept)                               1.80e-01   1.82e-01    0.99  0.32205
pmax(recency, 60)                        -1.59e-03   2.80e-04   -5.68  1.3e-08 ***
log(1 + numgiftsBY1)                      1.64e+00   1.51e-01   10.86  < 2e-16 ***
log(1 + numgiftsBY2)                      1.25e+00   2.26e-01    5.53  3.1e-08 ***
ifelse(numgiftsBY2 > 2, 1, 0)             1.00e+00   1.95e-01    5.13  2.9e-07 ***
log(1 + (numgiftsBY1 + numgiftsBY2)/730) -3.49e+02   4.72e+01   -7.39  1.5e-13 ***
ifelse(sumgiftsBY2 < 6.5, 1, 0)          -6.13e-01   1.14e-01   -5.37  7.8e-08 ***
```

```
ifelse(firstgiftsource == "HF", 1, 0)    -5.28e-01   1.51e-01   -3.50  0.00047 ***
---
    Null deviance: 12443  on 9003  degrees of freedom
Residual deviance: 11314  on 8996  degrees of freedom
AIC: 11330
```

Segment 2e: $\Pr\{y_i > 0\} = 0$.

To summarize these models, we make several observations. First, recency is a dominating predictor of future gift probability. However, if recency is greater than 2 years, then recency is not a good predictor. Second, frequency (number of gifts in the most recent year and in the two most recent years) is a good predictor. Third, for those who gave only once, the gift dollar amount is quite predictive; those who gave a very small amount ($1) or a large amount ($1,000 or more) are less likely to donate again. This prompts us to include the quadratic term `(log(1 + firstgiftamt))^2` in the logit model for segment 1b. Finally, the source of the first gift, “HF” (house file), indicates a lower tendency to action than other sources, while we found that the source cost is not predictive.

Next, we give the linear regression models for $z_i = \log(1+ y_i)$ given $y_i > 0$ (we select those data rows where $y_i > 0$ for the linear regression).

Segment 1a, 1b, 1c (all one-time donors together): the only predictor is the last gift amount (which is also the first and only gift amount for the one-time donors). The model is given below:

```
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)            0.9699     0.1055    9.19   <2e-16 ***
log(1 + lastgiftamt)   0.7890     0.0307   25.71   <2e-16 ***
---
Residual standard error: 0.708 on 853 degrees of freedom
Multiple R-squared: 0.437,      Adjusted R-squared: 0.436
F-statistic:  661 on 1 and 853 DF,  p-value: <2e-16
```

Segment 2a: The only predictor is the total gift amount in the past two years. The model is:

```
                                   Estimate Std. Error t value Pr(>|t|)
(Intercept)                          -1.249      0.396   -3.16   0.0021 **
log(1 + sumgiftsBY1 + sumgiftsBY2)    1.135      0.072   15.78   <2e-16 ***
---
Residual standard error: 0.771 on 111 degrees of freedom
Multiple R-squared: 0.692,      Adjusted R-squared: 0.689
F-statistic:  249 on 1 and 111 DF,  p-value: <2e-16
```

Segment 2b: *Null*, as $\Pr\{y_i > 0\} = 0$.

Segment 2c: The only predictor is the total gift amount in the first 8 months (sumgiftsBY0). The model is:

```
                     Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)              0.0792     0.3505    0.23     0.82
log(1 + sumgiftsBY0)   0.8914     0.0916    9.74   <2e-16 ***
---
Residual standard error: 0.67 on 109 degrees of freedom
Multiple R-squared: 0.465,      Adjusted R-squared: 0.46
F-statistic: 94.8 on 1 and 109 DF,  p-value: <2e-16
```

Segment 2d: Two predictors, the first is a weighted sum of gift amounts in the past three years where the weights are 0.8^0 (for the most recent year), 0.8, and 0.8^2. The model is:

```
                                         Estimate Std. Error t value Pr(>|t|)
(Intercept)                                0.3657     0.0432    8.46   <2e-16 ***
log(1 + sumgiftsBY0 * 0.64 +
        sumgiftsBY1 * 0.8 + sumgiftsBY2)   0.4924     0.0208   23.70   <2e-16 ***
log(1 + lastgiftamt)                       0.4601     0.0218   21.07   <2e-16 ***
---
Residual standard error: 0.656 on 4795 degrees of freedom
Multiple R-squared: 0.613,      Adjusted R-squared: 0.613
F-statistic: 3.8e+03 on 2 and 4795 DF,  p-value: <2e-16
```

Segment 2e: *Null*, as $\Pr\{y_i > 0\} = 0$.

Applying the model to the target period

Once the model is "learned", we reset "time 0" to the beginning of the target period 2006-09-01, and update the predictor values relative to the new time 0. For example, recency will be number of days since the last gift till 2006-09-01; sumgiftsBY2 will be the total gift amount in the most recent year (now BY4), and sumgiftsBY1 will be the total gift amount in the year BY3. We then apply the learned model on the updated predictor values. Our final MSE (mean squared error) is 1.6913 on the target period.

3. Conclusions

We quote Malthouse (2009) "Success is determined more by the modeler than the model family." Although we have attempted several other approaches, such as Poisson count models, in the end, it is the understanding of the raw data that is available to the modeler and the somehow manual "tuning" or customization of the modeling technique that have enabled us to produce the best model.

4. References

Gupta, Sunil, Dominique Hanssens, Bruce Hardie, William Kahn, V. Kumar, Nathaniel Lin, Nalini Ravishanker and S. Sriram. 2006. Modeling Customer Lifetime Value. *Journal of Service Research*, Vol 9 (2006), No. 2, pp. 139-155. DOI: 10.1177/1094670506293810.

Jain, Dipak, and Siddhartha S. Singh. 2002. Customer Lifetime Value Research in Marketing: A Review and Future Directions. *Journal of Interactive Marketing*, Vol. 16 (2002), No. 2, pp.34-46.

Malthouse, Edward C. 2009. The Results from the Lifetime Value and Customer Equity Modeling Competition. *Journal of Interactive Marketing*, Vol. 23 (2009), pp. 272-275.

Jamal, Zainab, and Alex Zhang. 2009. 2008 DMEF Customer Lifetime Value Competition (Task 2). *Journal of Interactive Marketing*, Vol. 23 (2009), pp. 279-283.