

Customer Lifetime Value Modeling and Its Use for Customer Retention Planning

Saharon Rosset

Einat Neumann

Uri Eick

Nurit Vatnik

Yizhak Idan

Amdocs Ltd.

8 Hapnina St.

Ra'anana 43000, Israel

{saharonr, einatn, urieick, nuritv, yizhaki}@amdocs.com

ABSTRACT

We present and discuss the important business problem of estimating the effect of retention efforts on the Lifetime Value of a customer in the Telecommunications industry. We discuss the components of this problem, in particular customer value and length of service (or tenure) modeling, and present a novel segment-based approach, motivated by the segment-level view marketing analysts usually employ. We then describe how we build on this approach to estimate the effects of retention on Lifetime Value. Our solution has been successfully implemented in Amdocs' Business Insight (BI) platform, and we illustrate its usefulness in real-world scenarios.

Keywords

Lifetime Value, Length of Service, Churn Modeling, Retention Campaign, Incentive Allocation.

1. INTRODUCTION

Customer Lifetime Value is usually defined as the total net income a company can expect from a customer (Novo 2001). The exact mathematical definition and its calculation method depend on many factors, such as whether customers are "subscribers" (as in most telecommunications products) or "visitors" (as in direct marketing or e-business).

In this paper we discuss the calculation and business uses of Customer Lifetime Value (LTV) in the communication industry, in particular in cellular telephony.

The Business Intelligence unit of the CRM division at Amdocs tailors analytical solutions to business problems, which are a high priority of Amdocs' customers in the communication industry: Churn and retention analysis, Fraud analysis (Murad and Pinkas 1999, Rosset et al 1999), Campaign management (Rosset et al 2001), Credit and Collection Risk management and more. LTV plays a major role in several of these applications, in particular Churn analysis and retention campaign management. In the context of churn analysis, the LTV of a customer or a segment is important complementary information to their churn probability,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SIGKDD '02, July 23-26, 2002, Edmonton, Alberta, Canada.

Copyright 2002 ACM 1-58113-567-X/02/0007...\$5.00.

as it gives a sense of how much is really being lost due to churn and how much effort should be concentrated on this segment. In the context of retention campaigns, the main business issue is the relation between the resources invested in retention and the corresponding change in LTV of the target segments.

In general, an LTV model has three components: customer's value over time, customer's length of service and a discounting factor. Each component can be calculated or estimated separately or their modeling can be combined. When modeling LTV in the context of a retention campaign, there is an additional issue, which is the need to calculate a customer's LTV before and after the retention effort. In other words, we would need to calculate several LTV's for each customer or segment, corresponding to each possible retention campaign we may want to run (i.e. the different incentives we may want to suggest). Being able to estimate these different LTV's is the key to a *successful and useful* LTV application.

The structure of this paper is as follows:

- In section 2 we introduce the general mathematical formulation of the LTV calculation problem.
- Section 3 discusses practical approaches to LTV calculations from the literature, and presents our preferred approach.
- The practical implementation of our LTV calculation, with some examples, is presented in section 4.
- In section 5 we turn to the business problem of estimating LTV given incentives and using these calculations to guide retention campaigns.
- Section 6 presents our LTV-based solution to the incentive allocation challenge and illustrates its use for real-life applications.

2. THEORETICAL LTV CALCULATIONS

Given a customer, there are three factors we have to determine in order to calculate LTV:

1. The customer's value over time: $v(t)$ for $t \geq 0$, where t is time and $t=0$ is the present. In practice, the customer's future value has to be estimated from current data, using business knowledge and analytical tools.
2. A length of service (LOS) model, describing the customer's churn probability over time. This is usually described by a "survival" function $S(t)$ for $t \geq 0$, which describes the probability that the customer will still be active at time t . We can then define $f(t)$ as the customer's "instantaneous" probability of churn at time t : $f(t) \equiv -dS/dt$. The quantity most commonly modeled, however is the hazard function $h(t) = f(t)/S(t)$. Helsén and Schmittlein (1993) discuss why $h(t)$ is a more appropriate quantity to estimate than $f(t)$. The LOS

model has to be estimated from current and historical data as well.

3. A discounting factor $D(t)$, which describes how much each \$1 gained in some future time t is worth for us right now. This function is usually given based on business knowledge. Two popular choices are:
 - Exponential decay: $D(t) = \exp(-\alpha t)$ for some $\alpha \geq 0$ ($\alpha=0$ means no discounting)
 - Threshold function: $D(t) = I\{t \leq T\}$ for some $T > 0$ (where I is the indicator function).

Given these three components, we can write the explicit formula for a customer's LTV as follows:

$$\text{LTV} = \int_0^{\infty} S(t)v(t)D(t)dt \quad (2.1)$$

In other words, the total value to be gained while the customer is still active. While this formula is attractive and straight-forward, the essence of the challenge lies, of course, in estimating the $v(t)$ and $S(t)$ components in a reasonable way.

We can build models of varying structural and computational complexity for these two quantities. For example, for LOS we can use a highly simplistic model assuming constant churn rate – so if we observe 5% churn rate in the current month, we can set $S(t) = 0.95^t$. This model ignores the different factors that can affect churn – a customer's individual characteristics, contracts and commitments, etc. On the other hand we can build a complex proportional hazards model, using dozens of customer properties as predictors. Such a model can turn out to be too complex and elaborate, either because it is modeling “local” effects relevant for the present only and not for the future, or because there is not enough data to estimate it properly. So to build practical and useful analytical models we have to find the “golden path” which makes effective and relevant use of the data available to us. We attempt to answer this challenge in the next sections.

3. PRACTICAL LTV APPROACHES

In this section we review some of the approaches to modeling the various components of LTV from the literature and present the segment-based approach, which follows naturally from the way analyses and campaigns are usually conducted in marketing departments. The segment-based approach helps in simplifying calculations and justifies the use of relatively simple methods for estimating the functions.

To model LTV we would naturally want to make use of the most recent data available. Therefore let us assume that we are only going to use churn data from the last available month for modeling LOS. So for the rest of this paper we assume we have a set of n customers, with covariates vectors x_1, \dots, x_n representing their “current” state and churn indicators c_1, \dots, c_n . The customers' tenure with the company is an important churn predictor since LOS frequently shows a strong dependency on customer “age”, in particular when contracts prevent customers from disconnecting during a specific period. Let us denote these tenures by t_1, \dots, t_n . Additional covariates are customer details, usage history, payment history, etc. Some of the covariates may be based on time-dependent accumulated attributes (e.g. averages over time, trends).

Our discussion is going to view time as discrete (measured in months), and thus the t_i 's will be integers and $f(t)$ will be a probability function, rather than a distribution function.

3.1 LOS Modeling Approaches

We now present a brief description of common Survival Analysis approaches and their possible use in LOS modeling. Detailed discussion of prevalent Survival Analysis approaches can be found in the literature, e.g. Venables and Ripley 1999, chapter 12. Pure parametric approaches assume $S(t)$ has a parametric form (Exponential, Weibull etc.) with the parameters depending on the covariates, including t . As Mani et al (1999) mention, such approaches are generally not appropriate for LTV modeling, since the survival function tends to be “spiky” and non-smooth, with spikes at the contract end dates.

Semi-parametric approaches, such as the Cox proportional hazards (PH) model (Cox 1972), are somewhat more flexible. The Cox PH model assumes a model for the hazard function $h(t)$ of the form:

$$h_i(t) = \frac{f_i(t)}{S_i(t)} = \lambda(t) \exp(\beta' x_i) \quad (3.1)$$

or alternatively:

$$\log(h_i(t)) = \log(\lambda(t)) + \beta' x_i \quad (3.2)$$

So there is a fixed parametric linear effect (in the exponent) for all covariates, except time, which is accounted for in the time-varying “baseline” risk $\lambda(t)$. Mani et al (1999) build a Neural Network semi-parametric model, where each possible tenure t has its own output node (the tenure is discretized to the monthly level). They illustrate that the more elaborate NN model performs better than the PH model on their data.

The data as described above, makes LOS modeling a special case of survival analysis where each subject is observed only once in time, and customers who disconnected before this month are “left censored”. Consequently we can approach it either as a survival analysis problem or a standard supervised learning problem where the time (i.e. customer's tenure with the company) is one of the predictors and churn is the response. To include a “baseline hazard” effect, time can be treated as being factorial rather than numerical, thus allowing a different effect for each tenure value. In this setting, a log-linear regression model for churn prediction using left-censored data would be equivalent in representation to a Cox proportional hazards survival analysis model. To see this point, consider that a customer's churn risk is in fact his $h(t)$ value (since if the customer already left we would not observe him). Thus a model of the form:

$$\log(P(c_i = 1)) = \alpha(t_i) + \beta' x_i \quad (3.3)$$

is obviously equivalent to (3.2).

The Kaplan-Meier estimator (Kaplan and Meier 1958) offers a fully non-parametric estimate for $S(t)$ by averaging over the data:

$$S(t) = \frac{\sum_i I(t_i \geq t)}{\sum_i I(t_i \geq t) + C_i} \quad (3.4)$$

Where:

- $I(t_i \geq t)$ equals 1 if customer i 's tenure is at least t months

- $\sum_i I(t_i \geq t)$ is the number of customers whose tenure is at least t months
- C_t is the number of customers who should have been at least t months old at the current date but have already left.

The data as described above is “left censored” and does not include C_t . However it can often be calculated based on historical information found in customer databases, which are typically used for LTV calculations.

3.2 The Segment-Based LOS Approach

When we are considering the use of analytical models for marketing applications, we should take into account the way they are going to be used. An important concept in marketing is that of a “segment”, representing a set of customers who are to be treated as one unit for the purpose of planning, carrying out and inspecting the results of marketing campaigns. A segment is usually implicitly considered to be “homogeneous” in the sense that the customers in it are “similar”, at least for the property examined (e.g. propensity to churn) or the campaign planned.

Amdocs Business Insight tools assist marketing experts in automatically discovering, examining, manually defining and manipulating segments for specific business problems. We assume in our LTV implementation that:

- the marketing analyst is interested in examining segments, not individual customers
- these segments have been pre-defined using Amdocs CMS or some other tool
- they are “homogeneous” in terms of churn (and hence LOS) behavior
- they are reasonably large

Based upon these assumptions, estimating LOS for a segment is reasonable and relatively simple. Under these assumptions we can dispense completely with the covariate vectors x (since all customers within the segment are similar) and adopt a non-parametric approach to estimating LOS in the segment by averaging over customers in the segment.

The Kaplan-Meier approach is reasonable here, but as we discussed before it requires the use of left-censored data referring to customers who have churned in the past. While this data is usually available it refers to churn events from the (potentially distant) past, and so may not represent the current tendencies in this segment, which may well be related to recent trends in the market, offers by competitors etc. So an alternative approach could be to calculate a non-parametric estimate of the hazards rate:

$$h(t) = \frac{\sum_i I(t_i = t) I(c_i = 1)}{\sum_i I(t_i = t)} \quad (3.5)$$

Where:

- $I(t_i = t)$ equals 1 if customer i 's current tenure is t months
- $I(c_i = 1)$ equals 1 if customer i churned in the current month
- $\sum_i I(t_i = t) I(c_i = 1)$ is the number of customers whose current tenure is t months and churned in the current month.

This approach relies heavily on having a sufficient number of examples for each discrete time point t (usually taken in months), but has the advantage of using only current data to estimate the

function. We can obtain an estimate for $S(t)$ through the simple calculation:

$$S(t) = \prod_{u < t} S(u+1) / S(u) = \prod_{u < t} (S(u) - f(u)) / S(u) = \prod_{u < t} (1 - h(u)) \quad (3.6)$$

Where $S(0) = 1$, of course.

In section 4 we describe Amdocs' LTV platform, which utilizes this approach and illustrate it on real data.

3.2.1 Theoretical Discussion of a Segment Approach

When examining the adequacy of a modeling approach, we generally have to consider two statistical concepts:

- Bias / Consistency: if we had infinite data, would our estimate converge to the correct value? How far would it end up being?
- Variance: how much uncertainty do we have in the estimates we are calculating for the unknown value?

These concepts have concrete mathematical definitions for the case of squared error loss regression only (although many suggestions exist for generalized formulations for other cases – see, for example - Friedman 97). However the principles they describe apply to any problem:

- The more flexible and/or adequate the model is, the smaller the bias.
- The more data one has, and the more efficiently one uses it, the smaller the uncertainty.

Under the segment-homogeneity assumption mentioned in the previous section, the bias of our segment-based approach should be close to zero. Furthermore, even without this assumption, if we assume that the marketing expert planning the campaign is only interested in the segment as a whole, then the quantities we want to estimate are indeed segment averages and not individual values. Hence the segment-based estimates are unbiased in this scenario as well.

As for variance, this is obviously a function of segment size. Parametric estimators will tend to have smaller variance. It is an interesting research question to investigate this bias-variance tradeoff between non-parametric and parametric estimates in this case. Under the assumption that segments are “large” (as are indeed the segments in most real-life segments encountered in the communication industry), and that there is a reasonable amount of churn in each segment, we can safely assume that the segment based non-parametric estimates will also have low variance, and hence that our approach is reasonable.

3.3 Practical Value Calculations

Calculating a customer's current value is usually a straight forward calculation based on the customer's current or recent information: usage, price plan, payments, collection efforts, call center contacts, etc. In section 4 we give illustrated examples.

The statistical techniques for modeling customer value along time include forecasting, trend analysis and time series modeling. However the complexity of modeling and predicting the various factors that affect future value: seasonality, business cycles, economic situation, competitors, personal profiles and more, make future value prediction a highly complex problem. The solution in LTV applications is usually to concentrate on modeling LOS, while either leaving the whole value issue to the experts (Mani et

al 1999), or considering customers' current value as their future value (Novo 2001).

Working at the segment level also makes the value calculation task easier, since it implies we do not need to have an exact estimate of individual customers' future value, but can rather average the estimates over all customers in the segment. This does not solve the fundamental problem of predicting *future* value, but it allows us to get a reliable *average current* value estimate at the segment level.

4. LTV WITHIN THE CMS

One of the Amdocs' BI platform systems is the Churn Management System (CMS). The key outputs of the system are churn and loyal segments, as well as scores for each individual in the target population, which represent the individual's likelihood to churn. The first step we take in the process of churn analysis is defining and creating a customer data mart that provides a single consolidated view of the customer data to be analyzed. It includes various attributes that reflect customers' profile and behavior changes: customer data, usage summaries, billing data, accounts receivable information, and social demographic data. Relevant, trends and moving averages are calculated, to account for time-variability in the data and exploit its predictive power. The preparation of the data for the exact needs of the data mining process includes Extracting Transforming and Loading (ETL) the necessary data.

The churn analysis process within the CMS combines automatic knowledge discovery and interactive analyst sessions. The automatic algorithm is a decision tree algorithm followed by a rule extraction mechanism. The analyst can then view and manipulate the automatically generated predictive segments (or patterns), and add to them based on his marketing expertise.

The automatic and interactive tools which the CMS utilizes to discover and analyse patterns, and to perform predictive modelling, have proven themselves as highly successful when compared to the state of the art data-mining techniques (Rosset and Inger 2000, Inger et al 2000, Neumann et al 2000).

The analysis tool includes an easy-to-use graphical user interface. Figure 1 is a capture of one of the system analysis tool's screens which provides the analyst with insight into various customer population segments automatically identified by their attributes, churn likelihood and related value. These segments (or rules) are characterized by several attributes accompanied by statistical measures that describe the significance of the segments and their coverage. Additional graphical capabilities of the CMS include analyzing the distribution of each variable per churn/loyal groups or in comparison to the entire population and an interactive visual data analysis, which provides the ability to further investigate attributes to provide additional insight and support the design of retention actions.

The data is extracted on a monthly basis and accordingly the scoring process is performed once a month. The churn score is one of the main components of the LOS; thus each customer will have a new LTV every month.

As shown in Figure 1 the system produces segments that characterize churn and loyal populations. Thus, the segment level and not the customer level is the basis for the interaction with the analyst. That is the level on which retention campaigns are planned and therefore the level on which the analyst is interested in viewing LTV.

The LOS solution implemented in Amdocs CMS is the segment-based calculation described in section 3.2. For value definition the CMS allows flexibility and it calculates the value individually per customer. It can be a constant value, an existing attribute within the data-mart, or a function of several existing attributes. An example for the customer value can be 'The financial value of a customer to the organization'. This value can be calculated from 'received payments' minus the 'cost of supplying products and services' to the customer.

To effectively use the data mining algorithms in the CMS, the input is usually a biased sample of the population. Often the churn rate in the population is very small but in the sample the two classes (churn and loyal) are much more balanced. The difference in churn rate between the sample and the population is accounted

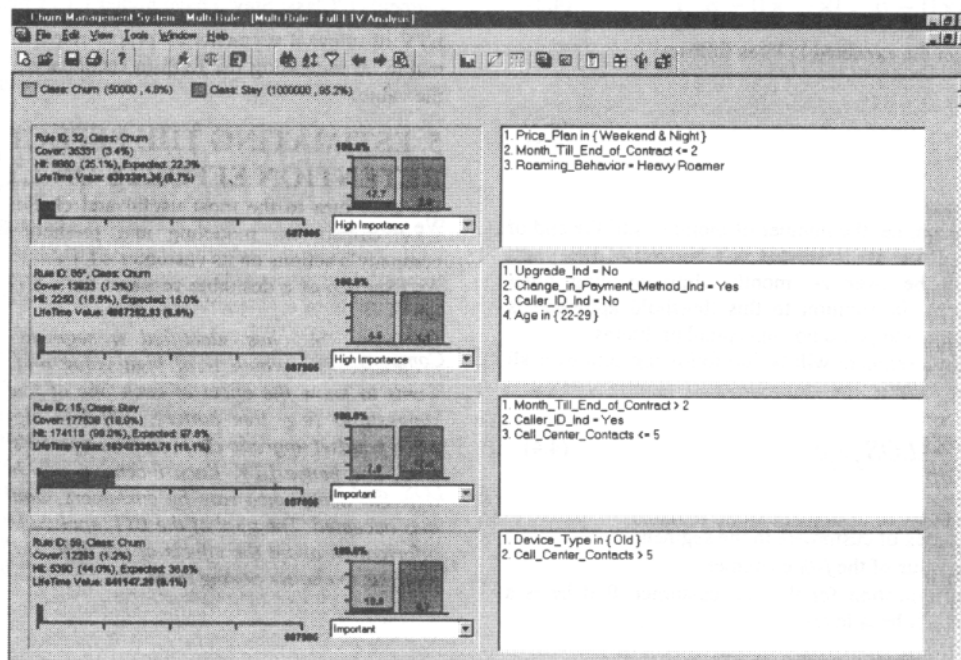


Figure 1. Churn and loyal patterns discovered automatically by the CMS

for in the LOS model, as we describe below. Rosset et. al. 2001 provide a detailed explanation about the relevant inverse transformation.

LOS is calculated on the segment level. It is calculated for each “age” group t within the segment, i.e. for each group of customers with the same tenure in the segment, there will be the same LOS. This calculation is based on a large amount of data (customers with the same tenure in the segment). The base for this value is the proportion of churners for each age $t - p_t$ as defined in the following formula (this is an extended version of the calculation from equation (3.5) in section 3.3)

$$p_t = \frac{\sum_i I(t_i = t)I(c_i = 1)}{\text{factor} \times \sum_i I(t_i = t)I(c_i = 0) + \sum_i I(t_i = t)I(c_i = 1)} \quad (4.1)$$

Where:

- $\sum_i I(t_i = t)I(c_i = 1)$ is the number of churners at tenure t
- $\sum_i I(t_i = t)I(c_i = 0)$ is the number loyal customers at tenure t
- $\text{factor} = (\text{churn to loyal sample ratio}) / (\text{churn to loyal population ratio})$

This quantity is calculated for each tenure t in the segment. There are several assumptions underlying this calculation. First, that the current churn probabilities for customers at tenure t represent the future ones at tenure t ; second, that the customers come from a “homogeneous” population and third, that both $\sum_i I(t_i = t)I(c_i = 1)$ and $\sum_i I(t_i = t)I(c_i = 0)$ are large enough to give reliable estimates of p_t .

Now, given a customer who is currently at tenure t_0 , we can use (4.1) to get the ‘Probability of a customer to reach age $t' - S(t)$

$$S(t) = (1 - p_{t-1}) \times (1 - p_{t-2}) \times \dots \times (1 - p_{t_0}) \quad (4.2)$$

And then we can get the expected LOS as follows-

$$LOS = \sum_{t=0}^h S(t) \quad (4.3)$$

Where h is the horizon, i.e. the number of months until the end of the interest period. If we are interested in a horizon of two years then the sum will be over 24 months. Implementing other discounting functions, in addition to this threshold approach is planned for the future, and poses no conceptual problems.

Finally, LTV within a segment will be the following sum over all customers in the segment is

$$LTV = \text{ratio} \times \sum_j LOS_j v_j c_j \quad (4.4)$$

where

- j is the index of customers in the segment
- v_j is the value of the j -th customer
- c_j is an indication for the j -th customer. 0 if he is a churner, 1 if he is loyal

- LOS_j is the Expected Length of Service for the j customer
- ratio is the population to sample ratio of loyal customers

Figure 2. LTV calculation CMS screen

Figure 2 is a screen capture of the CMS window for calculating LTV. It is necessary to select the customer's age (tenure), enter the horizon and enter the full population churn rate (the sample churn rate is already derived). It is also necessary to select/define the value, which may be one of three options: an equal value for all customers, a field that was previously selected as the value (in this case the “average bill” was previously selected), or a new value function.

The result of the LTV calculation can also be seen in Figure 1. The statistic measures (including LTV) of the identified segments are already transformed to the full population. In general, Loyal segments (“Class: Stay”) have higher LTVs than Churn segments, since the LOS of churners is 0. The aim is to try to increase the LTV of relevant segments by proper retention efforts, which aim mainly at increasing the LOS (a secondary purpose is increasing the value).

5. ESTIMATING THE EFFECT OF RETENTION EFFORTS ON LTV

We now turn to the most useful and challenging application of LTV calculation: modeling and predicting the effects of a company's actions on its customer's LTV.

An example of a desirable scenario for a LTV application would be:

Company “A” has identified a segment of “City dwelling professionals”, which is of high value and high churn rate. It wants to know the effect of each one of five possible incentives suggestions (e.g. free battery, 200 free night minutes, reduced price handset upgrade etc.), on the segment's value over time and LOS, and hence LTV. Each incentive may have a different cost, different acceptance rate by customers, and different effect if it gets accepted. The goal of the LTV application is to supply useful information about the effects of the different incentives, and help analysts to choose among them.

From the definition of the problem it is clear that there is some information about the incentives which we must know (or estimate) before we can calculate its effect on LTV:

1. The cost of the effort involved in suggesting the incentive. This figure is usually known and depends for example on the channel utilized (e.g. proactive phone contact, letter, comment on written bill). Denote the suggestion (or contact channel) cost by C .
2. The cost to the company if the customer accepts the incentive (e.g. the cost of the battery offered). Again this figure is either known or can be reliably estimated based on business knowledge. Denote the offer cost by G .
3. The probability that a customer in the approached segment will agree to accept the incentive (which can be around 100% if the incentive is completely free, but that is rarely the case). This is a more problematic quantity to figure out and it has to be estimated from past experience, or simply guessed (in which case many different values for it can be tried, to see how each would affect the outcome). Denote the acceptance probability by P .
4. Change in the value function if the incentive is accepted. For example, if the incentive is free voice-mail, the customer's calls to the voice-mail can still generate additional revenue. Similarly to P , the change in value has to be assumed or approximated from past data. Denote the new value function by $v^{(i)}(t)$.
5. The effect on customer's LOS if the incentive is accepted. The most obvious way for the incentive to affect LOS is if it includes a commitment by the customer (i.e. the customer commits not to leave the company in the next X months). Denote the new survival function by $S^{(i)}(t)$.

Given all of the above, calculating the change in LTV of a customer from a retention campaign, in which a given incentive is suggested is a straight forward ROI calculation:

$$LTV_{\text{new}} - LTV_{\text{old}} = P \cdot \left(\int_0^{\infty} S^{(i)}(t) v^{(i)}(t) D(t) dt - \int_0^{\infty} S(t) v(t) D(t) dt - G \right) - C \quad (5.1)$$

As for the basic LTV calculation described in section 2, and even more so, the main challenge is in obtaining reasonable and usable estimates for the above quantities, in particular the functions $v^{(i)}$, $S^{(i)}$.

We now describe two approaches to this problem: one that builds on our segment-level LTV calculation approach presented above, and another that makes further simplifying assumptions, negating the need to predict the future.

5.1 Segment-level calculation

As was mentioned before, working at the segment level allows us to "average" our information over the whole segment and avoid parametric assumptions, at the price of assuming that the segment population is "homogeneous".

To expand the segment-level approach described in section 3.2 to estimate the effect of incentives on a segment's LTV, we need to describe how we change the LOS model per segment, and how we adjust customer value for the incentive effects.

We define two possible effects of an incentive on LOS: *commitment* and *percentage decrease*. If an incentive includes a commitment period of X months (usually with a penalty for commitment violation that makes it unprofitable to leave during this period), then obviously any customer who accepts the incentive will not leave during this period. On the other hand, incentives that do not include a commitment also cause the churn probability to decrease. Our model allows a percentage decrease in the monthly churn rate. This percentage is presumed to be constant in all months and for all customers within the segment.

Thus, to estimate post-incentive LOS for a specific segment and a specific incentive, we need to know:

- Commitment period included in incentive, denote by $cmt^{(i)}$
- Reduction in churn probability from incentive, denote by $rc^{(i)}$

Which gives us for a specific customer:

$$S^{(i)}(t) = I\{t < cmt^{(i)}\} + I\{t \geq cmt^{(i)}\} \prod_{u=cmt^{(i)}}^t (1 - c(a+u) \cdot rc^{(i)}) \quad (5.2)$$

where a is the customer's current "age", and $c(a+u)$ is the churn probability estimate for age $a+u$ as estimated for the whole segment.

Then a similar calculation to the expected LOS calculation in equations (4.2) and (4.3) now gives us a post-incentive expected LOS estimate of:

$$ELOS^{(i)} = cmt^{(i)} + \frac{1}{n} \sum_{j=1}^n \sum_{t=cmt^{(i)}}^h \prod_{u=cmt^{(i)}}^t (1 - c(a_j + u) \cdot rc^{(i)}) \quad (5.3)$$

where the index j runs over the customers in the segment.

So we are using the homogeneity assumption to average the effect of the incentive on LOS over all customers in the segment, and we are assuming again that the "age" effect is the only differentiating factor of individual behavior within the sample. We also assume that the probability of accepting the incentive is constant across the segment and independent of all customer properties (including age), not used for the segment definition.

We also assume that once the commitment period is over, customers will "on average" return to the churn behavior that would characterize them at their "age" have they not churned for other reasons (rather than the commitment from the incentive).

The incentive's effect on customer value is assumed to be as a percentage change in the customer value. This change should reflect both the reduced value to the company due to the incentive cost and the increased value due to the increase in the relevant customer's usage. For example, when offering a free voicemail incentive the reduced value would be the voicemail cost and the increased value would be derived from the increase in billed incoming calls and the increase in outgoing calls due to the customer's calls to the voicemail box. Thus, we get that for every customer: $v^{(i)} = v \cdot (1 + change^{(i)})$, where $change^{(i)}$ is the change in value due to the incentive, assumed constant for all customers.

We can now combine all of the above into an estimate of the average change in LTV in the segment due to the incentive:

$$avLTV^{(i)} - avLTV =$$

$$P \cdot \left(\frac{1}{n} \sum_{j=1}^n [ELOS^{(i)} \cdot v^{(i)}(j) - ELOS \cdot v(j)] - G \right) - C \quad (5.4)$$

Where $avLTV^{(i)}$ is the estimated average LTV per customer in the segment after the retention campaign and $avLTV$ is the estimated current average LTV per customer in the segment. If this difference is positive it means we expect the retention campaign to be beneficial to the company.

5.2 Simplified Calculation Based on Constant Churn Assumptions

Let us now assume the following:

1. $D(t)$ is a threshold function with horizon h .
2. The churn risk is constant for each customer in the segment for any horizon. This would translate to assuming that for each customer, $S(t) = 1 - p^t$, where p is the churn probability of the customer for the next month.
3. p is small
4. The incentive includes a commitment for h months at least.
5. Customer value is constant over time, $v(t) = v$, and is not affected by the incentive's acceptance.

Then we get the following value for customer LTV without retention:

$$LTV_{old} = v \cdot \sum_{t=0}^{h-1} (1 - p)^t \cong v \cdot \sum_{t=0}^{h-1} 1 - pt = vh(1 - p(h-1)) \quad (5.5)$$

where the approximation relies on p and h being reasonably small. And adding retention we get:

$$LTV_{new} = P(hv - G) + (1 - P)LTV_{old} - C \quad (5.6)$$

Since if we succeed in giving the incentive we are guaranteed loyalty for the full relevant period of h months. So the difference in LTV due to retention is:

$$LTV_{new} - LTV_{old} \cong P \cdot h(h-1) \cdot v \cdot p - P \cdot G - C \quad (5.7)$$

which, given P, G and C and ignoring the inaccuracy in our calculation gives us the elegant result that:

$$LTV_{new} - LTV_{old} > 0 \Leftrightarrow v \cdot p > (P \cdot G - C) / (P \cdot h(h-1)) \quad (5.8)$$

In other words, we get the intuitive conclusion, that if we have a reasonable model for v and p , we should suggest the incentive only to customers whose *value weighted risk* $v \cdot p$ is big enough.

6. RETENTION LTV IMPLEMENTATION IN THE CMS

We now illustrate how the concepts of the previous section drive the incentive LTV implementation within the CMS, by following the details of the steps in the application and the calculation for a couple of real-life examples.

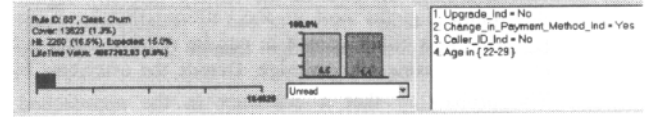


Figure 3. Churn segment

Figure 3 demonstrates one of the churn segments selected for a retention campaign. The segment consists of young customers who don't have a caller-id feature, whose handset was not upgraded in the past year and who have recently changed their payment method (for example from direct debit to check).

A marketing analyst came up with two possible incentives for this segment: an upgrade at a discounted price (lets assume for simplicity that all will be offered the same new handset) or a free caller-id feature. Both incentives will involve a 12 months commitment period.

Figure 4. Incentive definition CMS Screen

The first step is to calculate the current LTV of this segment. As displayed in section 4 we defined the LTV parameters. Recall, that the field selected as value was the monthly average bill, the selected horizon is 12 months and the population churn rate is 5% (in the sample it's about 50%). The LTV of this segment, which is already displayed in Figure 3 is \$4,967,202.

The next step is to define the possible incentives. An example of how this is done in the CMS is illustrated in Figures 4 & 5. In Figure 4 the incentive is defined and in Figure 5 the incentive is attached to a specific segment. At this stage it is also possible to refine the segment definition. Note that the same incentive may be allocated to different segments.

Figure 5. Incentive allocation CMS screen

Finally, we compare the change in LTV related to each of the incentives. The cost of giving a discounted handset upgrade is much higher than the cost of a free caller id (in this example we used \$100 and \$10 correspondingly). On the other hand, the acceptance rate will be higher since it's a more attractive offer (caller-id - 10% of the churners and 20% of the loyals, upgrade - 20% of the churners and 30% of the loyals). Actually, churners often switch providers in order to receive an improved handset promised by the competitor. So, the result of the upgrade incentive will be a higher retention rate than the caller-id incentive. Additionally, a more sophisticated handset will probably increase the usage and thus the added value, while adding a caller-id will have very little or no impact on the usage (the relative value increase for the upgrade is 10% in this example and none for the caller-id). Note that the added value affects both potential churners who accept the offer and loyal customer who will accept the offer. Furthermore, loyal customers will also be committed to 12 more months, so even though they weren't about to churn in the next month the incentive may lengthen their LOS.

The new LTV calculation takes into account all these parameters and the result as can be seen in Figure 6 is that the estimated increase in LTV due to offering a discounted upgrade is \$2,413,338 and due to offering a free caller-id is \$1,982,294.

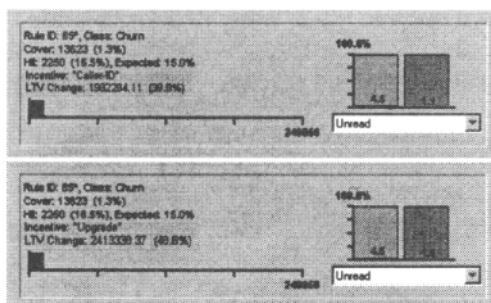


Figure 6. Estimated LTV change due to a free caller-id and due to a discounted upgrade offer

Suppose we wanted to examine the same two incentives for a different segment, as shown in Figure 7. This is a segment with many loyal customers, comprised of older customers with stable usage and medium bill average amounts.

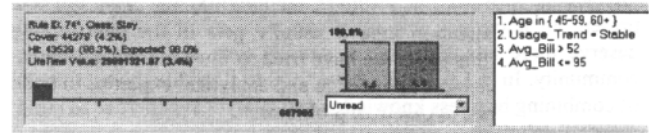


Figure 7. Loyal segment

In addition to the purpose of retaining the churners in this segment, offering an incentive to this segment is done also to increase the usage / value of the loyal customers and lengthen their LOS. The original LTV as displayed on Figure 6.5 is \$29,091,321. The same cost and acceptance rates were applied for the caller-id and upgrade incentives. Note that since the acceptance rate is higher for loyal customers the overall acceptance rate of this segment will be higher than in the previous churn segment.

The result was that the increase in value and LOS wasn't large enough to cover the high cost of the upgrade offers. Thus, the estimated change in LTV due to that incentive is negative: -\$485,450. On the other hand, the caller-id incentive yielded an estimated LTV increase of \$1,422,540 (Figure 8).

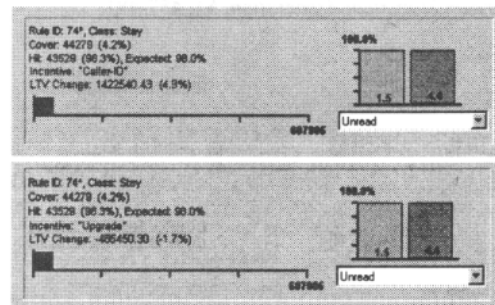


Figure 8. Estimated LTV change due to a free caller-id and due to a discounted upgrade offer

The examples illustrate that different incentives may have different impacts on LTV of the same segment, and the same incentive may have different impacts on LTV of different segments. The calculations involved are complex enough that the differential effect of different incentives on different segments cannot be easily guessed even when all the incentive's parameters are known. Using the application's mechanism for estimating that impact, it is possible to fit the appropriate incentive (out of the given options) to selected segments.

7. SUMMARY AND FUTURE WORK

In this paper we have tackled the practical use of analytical models for estimating the effect of retention measures on customers' lifetime value. This issue has been somewhat ignored in the data mining and marketing literature. We have described our approach and illustrated its usefulness in practical situations. The approach presented here to LTV calculation is not necessarily the best approach. However our emphasis is on practical and usable solutions, which will enable us to reach our ultimate goal -

to get useful and actionable information about the effects of different incentives. As our approach is modular, additional LOS and value models can certainly be integrated into the solution we presented.

We believe that this problem, like many others that arise from the interaction between the business community and data miners, present an important and significant data mining challenge and deserve more attention than it usually gets in the data mining community. In this paper we have tried to illustrate the usefulness of combining business knowledge and analytical expertise to build practical solutions to practical problems.

8. REFERENCES

- [1] Cox, D.R. (1972), "Regression Models and Life Tables," *Journal of the Royal Statistical Society*, B34, 187-220.
- [2] Friedman, J.H. (1997), "On Bias, Variance, 0/1-Loss and the Curse-of-Dimensionality", *Data Mining and Knowledge Discovery* 1(1), 55-77.
- [3] Helsen, K. and Schmittlein, D.C. (1993), "Analyzing Duration Times in Marketing: Evidence for the Effectiveness of Hazard Rate Models," *Marketing Science*, 11, 395-414.
- [4] Inger, A., Vatnik, N., Rosset, S. and Neumann, E. (2000), "KDD-Cup 2000 Question 1 Winner's Report," *SIGKDD Explorations*, 2(2), 94.
- [5] Kaplan, E.L. and Meier, R. (1958), "Non-parametric Estimation from Incomplete Observations," *Journal of the American Statistical Association*, 53, 457-481.
- [6] Mani, D.R., Drew, J., Betz, A. and Datta, P. (1999), "Statistics and Data Mining Techniques for Lifetime Value Modeling," *Proceedings of KDD-99*, 94-103.
- [7] Murad, U. and Pinkas, G. (1999), "Unsupervised Profiling for Identifying Superimposed Fraud," *PKDD-99*, 251-261.
- [8] Neumann, E., Vatnik, N., Rosset, S., Duenias, M., Sassoon, I. and Inger, A. (2000), "KDD-Cup 2000 Question 5 Winner's Report," *SIGKDD Explorations*, 2(2), 98.
- [9] Novo, J. (2001), "Maximizing Marketing ROI with Customer Behavior Analysis," <http://www.drilling-down.com>.
- [10] Rosset, S. and Inger, A. (2000), "KDD-Cup 99: Knowledge Discovery In a Charitable Organization's Donor Database," *SIGKDD Explorations*, 1(2), 85-90.
- [11] Rosset, S., Murad, U., Neumann, E., Idan, I. and Pinkas, G. (1999), "Discovery of Fraud Rules for Telecommunications - Challenges and Solutions," *Proceedings of KDD-99*, 409-413.
- [12] Rosset, S., Neumann, E., Eick, U., Vatnik, N. and Idan, I. (2001), "Evaluation of prediction models for marketing campaigns," *Proceedings of KDD-2001*, 456-461.
- [13] Venables, W.N. and Ripley, B.D. (1999), *Modern Applied Statistics with S-PLUS*, 3rd edition. Springer-Verlag.