Data Mining and Knowledge Discovery, 7, 321–339, 2003
© 2003 Kluwer Academic Publishers. Manufactured in The Netherlands.

Customer Lifetime Value Models for Decision Support

SAHARON ROSSET* saharonr@amdocs.com
EINAT NEUMANN einatn@amdocs.com
URI EICK urieick@amdocs.com
NURIT VATNIK nuritv@amdocs.com
Amdocs Ltd. 8 Hapnina St., Ra'anana, 43000 Israel

Editors: Hand, Keim, Ng

Received October 1, 2002

Abstract. We present and discuss the important business problem of estimating the effect of marketing activities 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 describe in detail how we build on this approach to estimate the effects of retention campaigns on Lifetime Value, and also discuss its application in other situations. Our solution has been successfully implemented by the Business Insight (BI) Professional Services.

Keywords: lifetime value, length of service, churn modeling, retention

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, as it gives a sense of how much is really being lost due to churn and how much effort should be

*Also of Stanford University, Department of Statistics.

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 present and discuss the LTV calculation problem. The practical implementation of our LTV calculation, with some examples, is presented in Section 3. In Section 3 we turn to the business problem of estimating LTV given incentives and using these calculations to guide retention campaigns. Section 5 presents our segment-based solution to the incentive allocation challenge and illustrates its use for real-life applications. We conclude with a brief discussion of LTV calculation solutions to some other business problems in Sections 6 and 7.

2. Calculating current customer LTV

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: $f(t) = -dS/dt$ The quantity most commonly modeled, however is the hazard function $h(t) = f(t)/S(t)$. Helsen 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: $D(t) = \exp(-at)$ for some $a \geq 0$ ($a = 0$ 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:

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

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 hundreds 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.

2.1. 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 $x_1, \ldots, x_N$ representing their "current" state and churn indicators $c_1, \ldots, c_N$. 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 $t_1, \ldots, t_N$. 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.

2.1.1. 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, Chap. 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:

$$h_i(t) = f_i(t)/S_i(t) = \lambda(t)\exp\beta' x_i \tag{2.2}$$

or alternatively:

$$\log(h_i(t)) = log(\lambda(t)) + \beta' x_i \tag{2.3}$$

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 $\lambda(t)$. A different semi-parametric approach is taken by Mani et al. (1999), who 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 categorical 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:

$$\log(\mathbb{P}(c_i = 1)) = \alpha(t_i) + \beta' x_i \tag{2.4}$$

is obviously equivalent to (2.3).

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

$$S(t) = \frac{\sum_i I(t_i \geq t)}{\sum_i I(t_i \geq t) + C_t} \tag{2.5}$$

Where

- $\sum_i I(t_i \geq t)$ 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.

The Kaplan-Meier approach is obviously inappropriate for our purpose, as it ignores the covariates completely and assumes stationarity of churn behavior along time. However our segment-based approach, presented next, employs a slightly modified non-parametric estimate.

2.1.2. 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

$$h(t) = \frac{\sum_i I(t_i = t) I(c_i = 1)}{\sum_i I(t_i = t)} \tag{2.6}$$

Where $\sum_i I(t_i = t) I(c_i = 1)$ 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:

$$S(t) = \prod_{u<t} \frac{S(u+1)}{S(u)} = \prod_{u<t} \frac{S(u) - f(u)}{S(u)} = \prod_{u<t} (1 - h(u)) \tag{2.7}$$

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

In Section 3 we describe Amdocs' LTV platform, which utilizes this approach and illustrate it on real data.

2.1.3. Discussion of a segment-based 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, 1997). 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 previously, the bias of our segment-based non-parametric approach is 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.

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 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.

2.1.4. Practical value calculations. Calculating a customer's current value is usually a straightforward 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.

3. LTV implementation

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 their predictive power.

The churn analysis process within the CMS combines automatic knowledge discovery and interactive analyst sessions. The automatic algorithm is a decision tree followed by a rule extraction mechanism. The analyst can then view and manipulate the automatically generated predictive segments, and enhance them based on marketing expertise. The automatic and interactive tools which the CMS utilizes to discover and analyze patterns, and to perform predictive modeling, have proven to be 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's analysis tool's screens which provides the analyst with insight into various customer population segments automatically identified by their churn likelihood. These segments are characterized by several conditions accompanied by statistical measures that describe the significance of the segments and their coverage. Additional graphical capabilities of the CMS include an interactive data visualization tool and a graphical display of dependence plots between single predictors and the response.

We now describe a LTV solution tailored for one of our customers based on its specific requirements. The CMS was customized to support this implementation. LTV within this implementation is calculated and presented at the segment level only. The LOS solution implemented is the segment-based calculation described in Section 2.1.2. 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 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 for each tenure group t within the segment, i.e. each group of customers with the same tenure in the segment will have the same LOS. The basis for this calculation is p_t, the proportion of churners for each tenure t as defined in the following formula (this is (2.6), adjusted for sampling)

$$p_t = \frac{\sum_i I(t_i = t)I(c_i = 1)}{\text{factor} \cdot \sum_i I(t_i = t)I(c_i = 0) + \sum_i I(t_i = t)I(c_i = 1)} \tag{3.1}$$

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

Where factor = (churn to loyal sample ratio)/(churn to loyal population ratio). Now, given a customer who is currently at tenure t_0, we can use (3.1) to get the $S(t)$—the probability of a customer to reach tenure t.

$$S(t) = (1 - p_{t-1}) \times (1 - p_{t-2}) \times \cdots \times \left(1 - p_{t_0}\right) \tag{3.2}$$

And then we can get the expected LOS as follows:

$$\text{ELOS} = \sum_{t=0}^{h} S(t) \tag{3.3}$$

where h is the horizon, i.e. the number of months until the end of the period of interests. Implementing other discounting functions, in addition to this threshold approach is planned for the future, and poses no conceptual problems.

The value definition is flexible and it is calculated individually per customer. It can be a constant value, an existing attribute within the data-mart, or a function of several existing attributes (e.g. total revenues-total costs). Figure 2 is a screen capture of the window for calculating LTV. It is necessary to select the customer's age (tenure) field, 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 customer 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

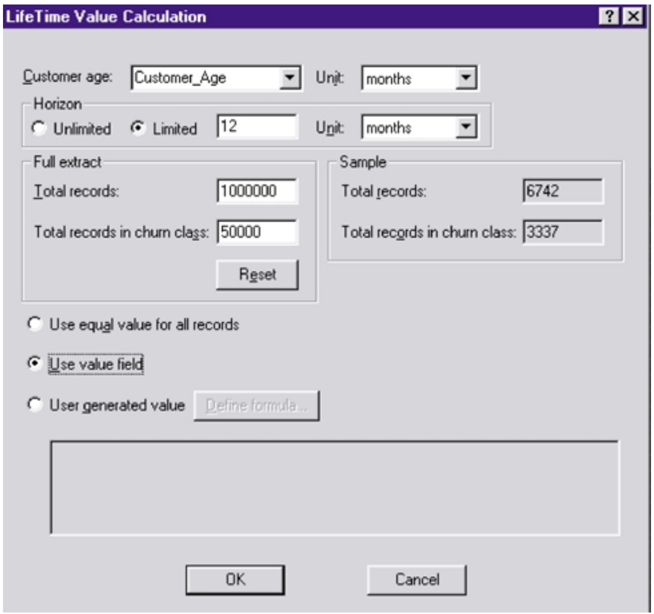

Figure 2. LTV Calculation CMS screen.

“average bill” was previously selected), or a derived value function. LTV of a segment is the following sum over all customers in the segment:

$$\mathrm{LTV} = \mathrm{ratio} \times \sum_{j \in \mathrm{segment}} \mathrm{ELOS}_j v_j c_j \tag{3.4}$$

v_j, c_j are the value and churn indicator for the j-th customer, respectively, and *ratio* is the population to sample ratio of loyal customers. The result of the LTV calculation can also be seen in figure 1. 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 LOS (a secondary goal is increasing value).

4. Estimating the effect of retention efforts on LTV

We now turn to a useful and challenging application of LTV calculations: modeling the effects of a company’s actions on its customers’ LTV. An example of a desirable scenario for an LTV application:

> 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, reduced price handset upgrade etc.), on the

segment's value 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 various incentives, and help analysts 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 their effect on LTV:

- 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.
- The cost incurred if the customer accepts the incentive (e.g. the cost of the battery offered). This figure is either known or can be reliably estimated based on business knowledge. Denote the offer cost by G
- 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 quantity has to be estimated from past experience, or simply guessed (in which case different values for it can be tried, to see how each would affect the outcome). Denote the acceptance probability by P
- 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 generate additional revenue. The change in value has to be estimated from past data. Denote the new value function by $v^{(i)}(t)$.
- 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 (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 these, calculating the expected change in LTV of a customer from suggesting an incentive is a straight forward ROI calculation:

$$\mathrm{LTV}^{(i)} - \mathrm{LTV} = P \cdot \left(\int_0^\infty \left[S^{(i)}(t) v^{(i)}(t) - S(t) v(t) \right] D(t)\, dt - G \right) - C$$

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.

4.1. 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, while assuming that the segment is "homogeneous".

To expand the segment-level approach described in Section 2.1.2 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 we assume 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:

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

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 (3.2) and (3.3) now gives us a post-incentive expected LOS estimate of:

$$\mathrm{ELOS}^{(i)} = cmt^{(i)} + \frac{1}{N} \sum_{j=1}^{N} \sum_{t=cmt^{(i)}}^{h} \prod_{u=cmt^{(i)}}^{t} \left(1 - c(a+u)rc^{(i)}\right)$$

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 valued 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: $v^{(i)} = v(1 + \text{change}^{(i)})$, where $\text{change}^{(i)}$ is the percentage 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:

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

Where $\text{avLTV}^{(i)}$ 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 profitable.

4.2. A simpler solution with further 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 give that for each customer, $S(t) = (1 - p)^t$, where p is the customer's churn probability 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:

$$\text{LTV}_{\text{old}} = v \sum_{t=0}^{h-1} (1 - p)^t \cong v \sum_{t=0}^{h-1} (1 - pt) = vh(1 - p(h - 1)/2)$$

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

$$\text{LTV}_{\text{new}} = P(hv - G) + (1 - P)\text{LTV}_{\text{old}} - C$$

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

$$\text{LTV}_{\text{new}} - \text{LTV}_{\text{old}} \cong P \frac{h(h-1)}{2} vp - PG - C$$

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

$$\text{LTV}_{\text{new}} - \text{LTV}_{\text{old}} > 0 \Leftrightarrow vp > 2(PG + C)/(Ph(h - 1))$$

Figure 3. Churn segment.

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** vp is big enough.

5. Retention LTV implementation

We now illustrate how the concepts of the previous section drive the incentive LTV within this implementation, by following the details of the steps in the application and the calculation for a couple of real-life examples. 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. The first step is to calculate the current LTV of this segment. We follow the definitions of the LTV parameters from Section 3. 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 is \$4,967,202.

The next step is to define the possible incentives. An example of how this is done is illustrated in figure 4. On the left we have the incentive definition screen and on the right the incentive allocation screen, where 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. 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 respectively). 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

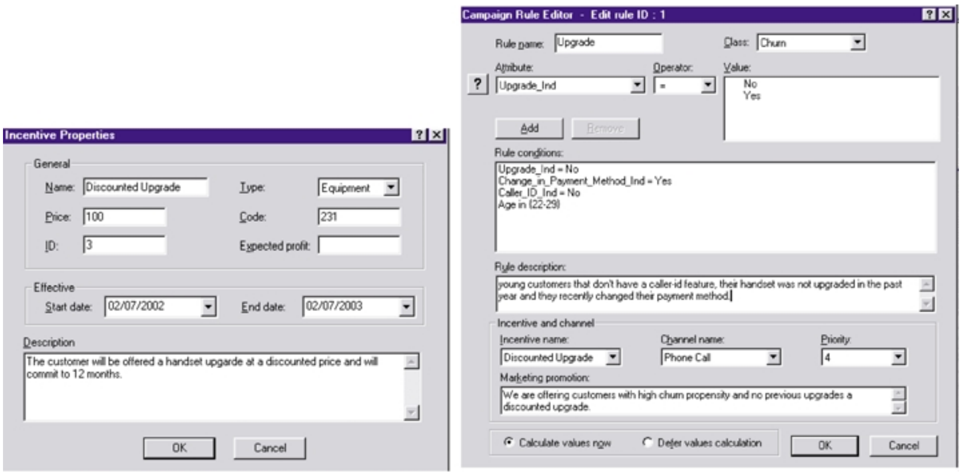

Figure 4. Incentive CMS screens: definition (left) and allocation (right).

month the incentive may lengthen their LOS. The new LTV calculation takes into account all these parameters and the result 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. Suppose we wanted to examine the same two incentives for a different segment, as shown in figure 5. This is a segment with many loyal customers, comprised of older customers with stable usage and medium average bill amounts. 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 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 the upgrade incentive was negative: -\$485,450. On the other hand, the caller-id incentive yielded an estimated LTV increase of \$1,422,540.

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

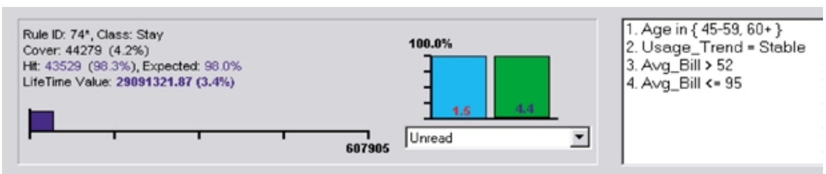

Figure 5. Loyal segment.

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.

6. Segment-based LTV models for other problems

We have described the segment-based approach to estimating the current LTV of a customer (Section 2) and to estimating the effect on LTV of retention by incentive in the communication industry (Section 4). This approach can be utilized to build LTV models for supporting other marketing campaigns and business decisions in general. The components and calculation of current customer LTV are quite standard. The main differences lie in estimating the effect of business activities on LTV, and the information we require in order to make these estimates. We now discuss some examples of extending the segment-based approach to other business problems in telecom and to other industries.

Cross- and up-sell campaigns in telecom. The problem of estimating the effect on LTV of cross-sell (adding new products to the customer) and up-sell (adding services to existing products) is quite similar to that of retention. The pre-marketing activity LTV calculation is exactly the same, of course. The list of components for post-campaign LTV estimation (Section 4) is valid here too:

– the costs of running the campaign
– the costs associated with the customer accepting the offer (in the case of cross/up-sell the cost of the promotional offer made, if any)
– the probability of acceptance by the customer
– the effect of acceptance on customer LOS
– the effect of acceptance on customer value

However the emphasis in cross-sell would usually be quite different than in the case of retention. Cross-sell campaigns are much less likely to include commitments (and hence affect the LOS) and much more likely to affect customer usage patterns (and hence customer value). In addition, the offered product or service is usually available to the general public, and thus the propensity to buy it can be modeled from historical data. This is different than the case with incentives for retention, which often are not offered outside the realm of a specific retention campaign.

Thus, the segment-based approach we discussed (and its implementation) can be applied, almost as-is, to the problem of estimating the effect of cross-sell campaigns on customer LTV, and hence can be used for guiding the decisions on running these campaigns. The methodological and statistical considerations brought up in the discussion of the segment-based approach to retention are all valid here as well.

Direct marketing campaigns. As a representative of non-telecom marketing problems, let us consider direct marketing campaigns, soliciting purchases or donations.

A good example of such a campaign can be found in the data of KDD-Cup 1998, also used for the knowledge discovery contest in KDD-Cup 1999. The modeling problem was to select a sub-population to which to send donation solicitations. In Rosset and Inger (2000), we have described several solutions to this problem, among them a segment-based solution, selecting 11 "most profitable" segments to send the solicitations to. As we discuss, this model was less accurate on this dataset than the models which target individual customers according to propensity scores (see Section 7 below), however it has the advantage of being highly interpretable, easy to implement and easy to modify. It also has the advantage of following the "marketing paradigm" of selecting customers by segments. The quality of the model is, essentially, as good as the quality of the segmentation—if the segments are large, homogeneous and statistically correlated with the property being modeled, then the models are bound to do well.

Applying the segment-based approach to this problem is more straightforward than the cases discussed previously. If we limit the definition of "lifetime value" to the outcome of the current campaign only, ignoring issues such as loyalty and referrals (as direct marketers usually do), then our problem is essentially reduced to predicting, at the segment level, the total expected donation in this campaign, and choosing the most profitable segments.

Credit and collection in telecom. Debt prevention and debt collection are major concerns of communication companies. Though not a marketing domain, it is a problem where LTV considerations are apt to be a major driving factor. Our experience indicates that a segment-based approach is common in this area too. The area of bad debt prevention is complex and contains many different decisions during the "delinquency cycle". Let us concentrate on a limited example, of deciding whether to reduce the credit limit of customers who are likely to become delinquent (accumulate debt) and possibly end up being written off (as lost debt or debt transferred to external collection agencies).

Once we bring potential debt considerations into the equation, we may have to adjust our "current LTV" calculation to take into account the potential to lose future revenue from the customer due to non-payment. The simplest model would just add another argument to the product in the LTV formula (2.1), say $np(t)$, to denote the non-payment probability for age t, to give us $\mathrm{LTV} = \int_0^\infty S(t)v(t)np(t)D(t)\,dt$.

As for the effect of reducing credit limit on a customer's (or a segment's) LTV, we need to consider a somewhat different set of quantities than in the marketing problems. This activity may certainly affect a customer's LOS (an effect we would have to model); it will affect customer value; it can also affect the non-payment probability np, since a customer with reduced credit limit is more likely to be able to afford his bill. However, there are now no acceptance probabilities or offer costs associated with our activity. In other words, if we are able to estimate (following a similar methodology to the one we described for retention) the effect of our activity on segment-level average LOS, average customer value, and average non-payment probability, we can calculate and utilize a segment-level estimate of the net LTV effect of reducing a segment's credit limit.

7. Value-weighted propensity scores as LTV models

In Section 4.2 we presented an alternative to the segment-based approach for modeling the effect of retention on LTV. Given some assumptions on the components of LTV we concluded that an optimal decision rule for determining which customers to suggest the incentive to was based on their value-weighted propensity for churn, and had the form: **suggest incentive** $\Leftrightarrow$ **vp** $>$ **const**, where v is the customers value and p is a churn propensity score. This criterion (value-weighted propensity scores) for including customers in campaigns—for retention as well as other tasks—turns out to be quite popular among marketing people. And as in the case of retention, if we want to justify it as an optimal policy we need to make some simplifying, sometimes problematic, assumptions. We now briefly illustrate this approach and discuss assumptions to justify it on the same three examples as in Section 6.

Cross- and up-sell. The setup presented in Section 4.2 also applies to the case of cross-/up-sell models, however its simplifying assumptions seem even less realistic here. Since cross-sell is more concerned with customer value and less with length of service, we can formulate an alternative value-weighted propensity approach, where the relevant propensity score would be the customer's propensity to accept the offer and the relevant value indicator would be the change in customer value from accepting the incentive. An appropriate set of assumptions would be:

– That the discounting model is a threshold function
– That no churn occurs within this period
– That the change in customer value $v^{(c)}$is fixed (and "known")

Then if we denote the propensity for acceptance by $p^{(c)}$ we easily see that an optimal decision rule for profitability of presenting a cross-sell offer to a customer has the form: $v^{(c)}p^{(c)} - p^{(c)}G > C$ (where as before C is the channel cost and G is the acceptance cost).

Direct marketing. As mentioned in Section 6, if we only consider the current campaign as the horizon, then the propensity to respond, multiplied by expected purchase or donation, is the quantity of interest. In (Rosset and Inger, 00) we discussed in detail some of the prediction models that can be used to model these quantities. It seems that for this problem, the value-weighted propensity scores approach is justified and appropriate.

Credit and collection. Continuing the credit limit determination example from Section 6, let us now assume that the non payment probability $np(v)$ is a function of the customer's bill size. Consider a customer with current bill size v and consider limiting this customer's credit to $v_0 < v$.

If we assume no churn and a threshold discounting function we get that the expected monthly income from this customer with no credit limit is $v \cdot (1 - np(v))$ while if we limit the customer's credit we expect $v_0 \cdot (1 - np(v_0))$. So under these assumptions an optimal decision would be to limit the customer's credit based on the relation between these two value-weighted propensity scores.

8. Summary: LTV models and data mining

In this paper we have tried to illustrate the usefulness of combining business knowledge and analytical expertise to build practical solutions to practical problems. We have tackled the use of analytical models for estimating the effect of various activities 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. 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 business and marketing activities on customers' LTV.

We believe that problems 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 they usually get in the data mining community.

References

Cox, D.R. 1972. Regression models and life tables. Journal of the Royal Statistical Society, B34, 187–220.

Friedman, J.H. 1997. On bias, variance, 0/1-loss and the curse-of-dimensionality. Data Mining and Knowledge Discovery, 1(1):55–77.

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.

Inger, A., Vatnik, N., Rosset, S., and Neumann, E. 2000. KDD-cup 2000 Question 1 Winner's report. SIGKDD Explorations, 2(2):94.

Kaplan, E.L. and Meier, R. 1958. "Non-parametric estimation from incomplete observations. Journal of the American Statistical Association, 53:457–481.

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.

Murad, U. and Pinkas, G. 1999. Unsupervised Profiling for Identifying Superimposed Fraud. PKDD-99, 251–261.

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.

Novo, J. 2001. Maximizing Marketing ROI with Customer Behavior Analysis, http://www.drilling-down.com.

Rosset, S. and Inger, A. 2000. KDD-cup 99: Knowledge discovery in a charitable organization's donor database. SIGKDD Explorations, 1(2):85–90.

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.

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.

Venables, W.N. and Ripley, B.D. 1999. Modern Applied Statistics with S-PLUS, 3rd edition. Springer-Verlag.

Saharon Rosset has been researching applications of Data Mining for Communications since 1997. His interests are in solving modeling problems while taking into account business considerations and in developing modeling tools for large data sets. He is currently completing his Ph.D. in the Department of Statistics at Stanford University.

Einat Neumann has been developing Data Mining and Business Insight applications and solutions for Communications Service Providers since 1998. Her research and development areas include Fraud Detection, Churn and Retention Analysis, Bad Debt Analysis, Product Analysis and more. Einat is working as a senior Data Mining Analyst and Business Consultant at Amdocs Ltd. She is an Information Systems Engineer and is currently completing her M.Sc degree in the Department of Statistics at Tel-Aviv University.

Uri Eick is a senior Data Mining Analyst at the Business Insight unit in the CRM Division of Amdocs Ltd. In this position Uri has worked with leading Communications Service Providers providers in Europe, and North America. Uri has experience as an applied statistician and worked in the Statistical Laboratory of Israel's High Institute for Science and Engineering, both as a researcher and as a consultant. He was involved in data analyses

in fields ranging from Medicine to Behavioral Sciences. Uri received his M.Sc in Applied Statistics from Israel's High Institute for Science and Engineering.

Nurit Vatnik has been implementing high-scale data mining solutions for Communication Service Providers since 1999. Her main interests are multi-dimensional prediction models, and their deployment in real-world business scenarios. Nurit has been Researching Financial Time-series as a graduate at the Tel Aviv University. She is currently the Manager of Analytical Professional Services in the CRM Division of Amdocs Ltd.