

LTV PREDICTION USING PROBABILISTIC MODELS

CONTENTS

1. Introduction	1
2. Customer Basis Analysis:	1
2.1. Probabilistic models for Customer Basis Analysis:	2
2.2. Possible scenarios:	2
2.3. Our case : Model LTV prediction	2
3. Dataset description	3
4. Model : Shifted Geometrical Beta distribution for retention prediction	3
4.1. Model Assumptions	4
4.2. Prior and Likelihood	5
4.3. Maximum Likelihood	7
References	7

1. INTRODUCTION

In this project we want to perform an estimation of customer life-time value (LTV) as a case of a Customer-Base Analysis using what literature in this domain calls probabilistic models (which is in reality no other thing than using Bayesian approach), using data from the company where we work .

This work will be divided in the following parts: First discussion Customer-Basis Analysis and possible scenarios and characterization of our scenario, second description of the model we will use (assumptions, ...), third discussion about implementation and finally presentation of results and accuracy measurement.

We found that this would make a proper project for a "Bayesian Statistics" course, as it is in between a Data-oriented and an Application-oriented project type.

2. CUSTOMER BASIS ANALYSIS:

Customer Basis Analysis is a concept that is becoming more and more critical as Companies have access (or generate) to more customer's transaction data (increase in complexity and volume). This availability of these datasets has led to a general change to transaction based strategies to more customer-centric strategies. In other words, the first strategies lead typically to descriptive analysis (for example summary statistics like rates, average number of subscriptions ...), while the more customer-centric approach is more focused on analysis on activities that are

more predictive, i.e. activities and facts that by past information (found in those databases) a forecast is made. This is why a Bayesian Probabilistic approach can be convenient.

2.1. Probabilistic models for Customer Basis Analysis:

How can we model this analysis with Probabilistic Models ? As usual in this situations, problem will be modeled in a way that observed behavior has underlying a random process, thus a past behavior is not a true mirror of future actions. With this interpretation, the past, and the "future" are function of latent characteristics denoted by θ .

While this approach may not be new (in other fields), the novelty resides in the fact that classically for Customer Basis Analysis, the approaches have been regression-like (which also happened in the company we work) and the use of probabilistic approaches is relatively new.

2.2. Possible scenarios:

Commonly in literature we will find the following classification for scenarios in Customer-Base Analysis :

By Opportunities of Transactions:

It can be discrete and continuous.

By Relation with the Customer:

The key distinction between contractual and noncontractual is that in contractual scenario, you observe when a customer leaves, while in a noncontractual scenario, customer departure is unobservable and therefore must be inferred. For example in a contractual scenario, a user might pay for a service only when needed, so that if during a long time this user didn't made any transaction either it should be considered that is not any more a client or it's just a hiatus, we don't know it directly.

Correctly knowing in which scenario our problem is in matters, especially while selecting a model.

The company in which we work, has a few subscription based websites, where users are subscribed in order to have access to contents and services. After each cycle (30 days) the user is automatically rebilled and in order to drop out, the user should cancel. As a consequence this scenario corresponds to a contractual scenario where transactions are done periodically.

In concrete we will work with data from a concrete product.

2.3. Our case : Model LTV prediction.

For the scope of this project, we will model a the Customer Life Time Value (or LTV) as a particular case of Customer Basis Analysis, using Bayesian Probabilistic model.

Opportunities for Transactions	Continuous	Grocery purchases Doctor visits Hotel stays	Credit card Student mealplan Mobile phone usage
	Discrete	Event attendance Prescription refills Charity fund drives	Magazine subs Insurance policy Health club m'ship
		Noncontractual	Contractual
Type of Relationship With Customers			

FIGURE 1. Examples for the different scenarios. From **fader09**

In marketing, customer lifetime value (CLV or often CLTV), lifetime customer value (LCV), or life-time value (LTV) is a prediction of the net profit attributed to the entire future relationship with a customer. The prediction model can have varying levels of sophistication and accuracy, ranging from a crude heuristic to the use of complex predictive analytics techniques.

For this we will first implement the proposed model for Customer Retention Prediction described in **fader07** on a dataset of one of the products of the company.

3. DATASET DESCRIPTION

Our case is of a Discrete Contractual scenario (as we see in 1) , with a 30 days period cycle of rebill. The users can cancel their relation with the company at any time. The data we use has been collected from July the15th and July the 15th, and it is composed of number of clients that rebills after each cycle, for one of the company products. We have added two additional columns that may be useful , the number of clients as a percentage and the number of users that cancels during each cycle.

4. MODEL : SHIFTED GEOMETRICAL BETA DISTRIBUTION FOR RETENTION PREDICTION

Our first idea to approach this problem was using a Inverse Binomial with a Beta distribution as prior. We first tried this implementation with a MCMC implementation in order to estimate the probability of cancel of the users. But at the end we rejected this model as it just considered the users that cancelled and

numRebill	Count	CountNorm	Diff
0	41747	100.00000	0
1	16026	38.38839	25721
2	10960	26.25338	5066
3	8578	20.54758	2382
4	7027	16.83235	1551
5	5878	14.08005	1149

FIGURE 2. Number of user rebilling at each cycle

not the remaining users. Then we found the model presented in **feder07** that also considers the users that doesn't cancels after the observation period.

The model we want to test is also based in using a negative binomial distribution (more precisely a geometric distribution as the failure number r is set to 1), with some other considerations that enables to consider the surviving users. Let's see the model in more detail.

4.1. Model Assumptions.

First the assumptions for the model :

- At the end of each cycle, the user renews its subscription or cancels with some probability
- That probability remains constant
- There is the assumption of heterogeneity between clients

Those assumptions are translated in the following probabilities :

The probability of duration for a given user(we will call it retention function), in terms of counts of rebills (success) before cancel the contract (failure), thus a classical Negative Binomial (with $r = 1$, also called Geometric)

$$(1) \quad P(T = t|\theta) = \theta(1 - \theta)^{t-1}$$

The other is the probability of remaining as client in for a given cycle (survival function), which is a Negative Binomial with $r=0$:

$$(2) \quad S(t|\theta) = (1 - \theta)^t$$

Where T is the number of cycles that the user has rebilled (the number of counts), and θ the probability of cancel.

So in this model, the parameter that is not directly measurable and has to be estimated is θ .

4.2. Prior and Likelihood.

For this setting it is not a bad assumption to consider as a prior distribution for θ a Beta distribution, as it can be fitted to our data, with a proper choice of α and β parameters.

Prior Distribution :

$$(3) \quad f(\theta; \alpha, \beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)}$$

As for the Likelihood function, we should not forget what we want to model. In a classic scenario of counting successes/failures, we would only take into consideration the retention probability part i.e. the users that cancels after a number of rebills, (thus with a Negative Binomial distribution the problem). In our case, we want to model either the behavior of the users that cancelled after some cycles and the users still in contract (the idea on doing this approach is that with data of few cycles we can predict behavior of the following cycles). That's the reason a survival (equation 2) function is considered.

Let's assume that we have data for a limited number of cycles, that we call T_{max} . Assuming independence between users, we will have as a likelihood distribution :

$$(4) \quad \prod_{t=1}^{T_{max}} P(T = t|\theta)^{n_t} S(T_{max}|\theta)^{N - \sum_{t=1}^{T_{max}} n_t}$$

where :

- n_t is the number of users that cancels on cycle t
- N is the number of total subscriptions (i.e. users that made transaction rebill 0)
- T_{max} is the maximum number of cycles

Note that this equation has two distinguishable parts : The first term, $\prod_{t=1}^{T_{max}} P(T = t|\theta)^{n_t}$ models the probability of users that had cancelled before T_{max} cycle and the second term $S(T_{max}|\theta)^{N - \sum_{t=1}^{T_{max}} n_t}$ the probability of the users still "alive". So we consider a term that is not in a classical Negative Binomial scenario.

As this likelihood term doesn't correspond exactly to any standard distribution, we opted for a Posterior Maximization strategy to estimate the parameter.

For this the approach that is used at **feder07** is to compute the likelihood function as a function of α and β parameters by averaging the likelihood with the

prior (thus in reality computing the posterior predictive) and maximize for α and β .

The derivations to obtain such expressions (adapted from **feder07**) :

For the retention part :

$$\begin{aligned}
 P(T = t|\alpha, \beta) &= \int_0^1 P(T = t|\Theta = \theta) f(\theta; \alpha, \beta) d\theta \\
 &= \int_0^1 \theta(1 - \theta)^{t-1} \frac{\theta^{\alpha-1}(1 - \theta)^{\beta-1}}{B(\alpha, \beta)} d\theta \\
 &= \frac{1}{B(\alpha, \beta)} \int_0^1 \theta^{\alpha}(1 - \theta)^{\beta+t-2} d\theta \\
 &= \frac{B(\alpha + 1, \beta + t - 1)}{\alpha, \beta}
 \end{aligned}$$

For the survival part :

$$\begin{aligned}
 S(t|\alpha, \beta) &= \int_0^1 S(t|\Theta = \theta) f(\theta; \alpha, \beta) d\theta \\
 &= \int_0^1 (1 - \theta)^t \frac{\theta^{\alpha-1}(1 - \theta)^{\beta-1}}{B(\alpha, \beta)} d\theta \\
 &= \frac{1}{B(\alpha, \beta)} \int_0^1 \theta^{\alpha-1}(1 - \theta)^{\beta+t-1} d\theta \\
 &= \frac{B(\alpha, \beta + t)}{\alpha, \beta}
 \end{aligned}$$

The retention part has been implemented in a recursive way, using the expression proposed in the article:

First we should note that:

$$\frac{\gamma(x+1)}{\gamma(x)} = x$$

$$B(\alpha, \gamma) = \frac{\gamma(\alpha)\gamma(\beta)}{\gamma(\alpha + \beta)}$$

The general recursive expression :

$$P(T = t|\alpha, \beta) = \frac{P(T = t|\alpha, \beta)}{P(T = t-1|\alpha, \beta)} P(T = t-1|\alpha, \beta)$$

First we have to compute the initial expression in which we use the above expressions of the Gamma function and the definition of Beta function in terms of Gamma functiond :

$$P(T = 1|\alpha, \beta) = \frac{B(\alpha + 1, \beta)}{B(\alpha, \beta)} = \frac{\gamma(\alpha + 1)/\gamma(\alpha)}{\gamma(\alpha + \beta + 1)/\gamma(\alpha + \beta)} = \frac{\alpha}{\alpha + \beta}$$

Then we can compute the first term of the recursive expression :

$$\begin{aligned} \frac{P(T = t|\alpha, \beta)}{P(T = t - 1|\alpha, \beta)} &= \frac{B(\alpha + 1, \beta + t - 1)}{B(\alpha + 1, \beta + t - 2)} \\ &= \frac{\gamma(\beta + t - 1)/\gamma(\beta + t - 2)}{\gamma(\alpha + \beta + t)/\gamma(\alpha + \beta + t - 1)} \\ &= \frac{\beta + t - 2}{\alpha + \beta + t - 1} \end{aligned}$$

So at the end the recursive expression is :

$$(5) \quad P(T = t|\alpha, \beta) = \begin{cases} \frac{\alpha}{\alpha + \beta} & \text{if } t = 1 \\ \frac{\beta + t - 2}{\alpha + \beta + t - 1} P(T = t|\alpha, \beta) & \text{if } t > 1 \end{cases}$$

4.3. Maximum Likelihood.

REFERENCES

- [1] Peter S. Fader and Bruce G. S. Hardie *How To Project Customer Retention*, 2007.
- [2] Peter S. Fader and Bruce G. S. Hardie *Probability Models for Customer-Base Analysis*, 2009.