

Efficient Allocation of Marketing Resources using Dynamic Programming

Giuliano Tirenni*
tir@zurich.ibm.com

Abderrahim Labbi
abl@zurich.ibm.com

André Elisseeff
ael@zurich.ibm.com

Cèsar Berrospi
ceb@zurich.ibm.com

Abstract

In this paper we address the following question: how to estimate a Markov Decision Process modeling the dynamics of customer relationships. Once the model is estimated, we discuss how to efficiently allocate marketing resources and instruments in order to maximize the long-term value generated by customers in a given future time horizon using dynamic programming. Our methodology allows us both to predict and to optimize the future value generated by customers. We show our approach using a case study involving a major European airline.

1 Introduction

In the last years there has been an increasing interest in the allocation of marketing resources both in the marketing (e.g. [17], [6], [19]) and in the data mining (e.g. [16], [13], [4]) communities. There is common agreement that marketing initiatives should be evaluated by measuring their impact on the customer lifetime value [9], i.e. the long-term value generated by a relationship with a customer.

In this paper we discuss how to model the dynamic relationships between the customers and the company using Markov Decision Processes [15] and how to allocate efficiently the marketing budget by finding marketing actions that maximize the long-term value generated by the customers.

We model the customer behavior in time taking into account the marketing actions executed by the company. Customers are segmented into a finite number of states, then the transition probabilities from one state to another and the expected rewards generated when applying a given marketing action in a state are estimated from the transactional data. Once all the states, the transition probabilities, and the expected rewards are known, it is possible to find the marketing policy which maximizes the expected long-term value generated by the relationship with the customers. A *marketing policy* is a mapping from customer states to marketing actions.

The use of dynamic programming techniques to

maximize customer future long-term value and the concept of Markov Decision Processes itself originated from the catalog industry in the 1950s [8]. A main issue which has not been addressed in the literature is the estimation of robust Markov Decision Processes modeling the customer relationship and the long-term effects of marketing actions. To the best of our knowledge, with the exception of [18] where customer states are built using a supervised clustering algorithm, all the models found in the literature (e.g. [1], [5], [14], [3]) assume a given state representation *ad hoc*, without providing any theoretical justification. Most of the models are based on the recency, the frequency, and the monetary value (RFM) segmentation. While the RFM segmentation is very popular in the marketing practice [11], there is not a theoretical motivation that justifies its use in modeling customer dynamics. As discussed in [18], the issue of estimating robust Markov Decision Processes is relevant because a non-reliable model can lead to a non-optimal policy, which in some cases could even perform worse than the historical, i.e. the current, policy.

The remainder of the paper is organized as follows. In section 2 we briefly review Markov Decision Processes and dynamic programming. In sections 3 and 4 we describe how to estimate robust Markov Decision Processes from the customer transactional data. In section 5 we describe a case study. Finally, the conclusions follow in section 6.

2 Markov Decision Processes

A Markov Decision Process (MDP) [15] can be defined as a set of decision epochs $\mathcal{T} \subset \mathcal{N}$, a finite set of states S , a finite set of actions A , a transition probability $p_t(s'|s, a)$ modeling the probability of moving from state $s \in S$ to state $s' \in S$ if action $a \in A$ is applied at decision epoch $t \in \mathcal{T}$, and a reward function $r_t(s, a)$ modeling the expected reward obtained in state $s \in S$ if action $a \in A$ is applied at decision epoch $t \in \mathcal{T}$.

If the transition probabilities and the rewards do not depend on the decision epoch, the process is said to be *stationary*. We consider stationary Markov Decision Processes when modeling the dynamics of customer behavior.

*Computer Science Department, IBM Zurich Research Laboratory, Saeumerstrasse 4, CH-8803 Rueschlikon, Switzerland.

A deterministic¹ *policy* π defines, for each decision epoch $t \in \mathcal{T}$, a mapping from states to actions. The expected *value* of state s at time step i , given a policy π , and a finite horizon of length T , is defined as

$$(2.1) \quad \nu_T^\pi(s_i) = \mathbb{E}_{s_i}^\pi \left[\sum_{t=i}^{T-1} r_t(s_t, a_t) + r_T(s_T) \right],$$

r_t , s_t , and a_t are, respectively, the expected reward, the state, and the action executed at time step t . $r_T(s_T)$ is the terminal reward obtained at the last epoch T which depends only on the state s_T .

The optimal policy is defined as the policy maximizing the long-term expected value (2.1) in each state and can be found using backward induction [15].

3 Estimation of the MDP

We assume that customers are segmented into different *states* and that customer transactional data is available. For each customer, the transactional data consists of a sequence of events. Each event is defined by a triple composed of a state s , an action a , and a reward r . The next event is defined by the triple s', a', r' where s' is the state resulting from applying a to s and so on. Each customer has an associated sequence of events defined as episode.

Given transactional data D , the state and action spaces are obtained by considering respectively all the states and all the actions that appear in D . In order to estimate the transition probabilities we can simply use the maximum likelihood estimator:

$$(3.2) \quad p(s'|s, a) = \frac{\#(s'|s, a)}{\#(s, a)}$$

where $\#(s'|s, a)$ is the total number of transitions from s to s' if action a is applied and $\#(s, a)$ is the total number of actions a applied to state s . If the quantity $\#(s, a)$ is zero, then the above equation is not defined. Moreover if the quantity $\#(s'|s, a) = 0$ then $p(s'|s, a) = 0$. As we are estimating the transition probabilities from a limited sample of data, we can assume that the absence of particular transitions does not necessarily imply that the real probabilities are undefined or null. To address these two issues we adopt a *Bayesian* approach incorporating the prior transition probability $\hat{p}_{s'|s, a}$ into equation 3.2. This leads to the following estimator [12]:

$$(3.3) \quad p(s'|s, a) = \frac{\#(s'|s, a) + m_1 \hat{p}_{s'|s, a}}{\#(s, a) + m_1}.$$

¹Deterministic policies are a special case of stochastic policies, which associate to each state a probability distribution on the actions.

The quantity m_1 can be interpreted as the number of instances following the prior probability that are injected into the data set D , m_1 acts therefore as a weight defining the relative importance of the prior probability with respect to the probability estimated from the data.

There are two possibilities to model the prior transition probability $\hat{p}_{s'|s, a}$: a) adopt a *state-driven* approach, emphasizing the role of the origin state, or b) adopt an *action-driven* approach, emphasizing the role of the action.

In the first case, the prior is modeled as follows

$$\hat{p}_{s'|s, a} = p(s'|s) = \frac{\#(s'|s) + m_2 \hat{p}_{s'|s}}{\#(s) + m_2},$$

where $\#(s)$ is the number of times state s appears in the data set D and $\#(s'|s)$ is the number of times a transition from state s to state s' is observed. Finally, the nested prior $\hat{p}_{s'|s}$ is estimated as follows

$$\hat{p}_{s'|s} = p(s') = \frac{\#(s') + m_3 \hat{p}}{\sum_{s \in S} \#(s) + m_3}.$$

If we set equiprobable state probabilities, then the prior \hat{p} becomes

$$\hat{p} = \frac{1}{|S|},$$

where $|S|$ is the cardinality, i.e. the number of different states, of the set S . If we set $m_3 = |S|$, we obtain the Laplace estimator:

$$\hat{p}_{s'|s} = p(s') = \frac{\#(s') + 1}{\sum_{s \in S} \#(s) + |S|},$$

we use this estimator to model the prior $\hat{p}_{s'|s}$.

In the action-driven approach, the prior is modeled as follows

$$\hat{p}_{s'|s, a} = p(s'|a) = \frac{\#(s'|a) + m_2 \hat{p}_{s'|a}}{\#(a) + m_2},$$

where $\#(a)$ is the number of times action a appears in the data set D and $\#(s'|a)$ is the number of times a transition to state s' is due to the execution of action a . The nested prior is equal to $\hat{p}_{s'|a} = p(s')$.

The expected reward $r(s, a)$ if action a is applied to state s can be estimated as follows

$$(3.4) \quad r(s, a) = \frac{\sum_{(s, a) \in D} r(s, a)}{\#(s, a)},$$

where $r(s, a)$ is the reward observed in the data when action a is applied to state s . If the quantity $\#(s, a)$ is zero, because action a has never been applied to state s ,

we can estimate the expected reward considering either a state-driven approach or an action-driven approach. The state-driven estimate is

$$r(s, a) = r(s) = \frac{\sum_{(s) \in D} r(s, a)}{\#(s)},$$

while the action-driven estimate is

$$r(s, a) = r(a) = \frac{\sum_{(a) \in D} r(s, a)}{\#(a)}.$$

3.1 Estimation of the historical policy We define as *historical policy* the policy used by the company when targeting customers. The knowledge of the historical policy allows us to simulate the customer dynamics. We learn from the available transactional data a stochastic policy, assuming that it is *stationary*², and estimate the probability of executing action a in state s as follows:

$$\pi(a|s) = \frac{\#(a|s) + m\hat{\pi}_{a|s}}{\#(s) + m},$$

where $\#(a|s)$ are the number of events (i.e. transactions) with state s and action a and $\#(s)$ are the number of events with state s . The quantity $\hat{\pi}_{a|s}$ is the prior probability of executing action a in state s . We define the prior probability using the Laplace estimator as follows

$$\hat{\pi}_{a|s} = p(a) = \frac{\#(a) + 1}{\sum_{a \in A} \#(a) + |A|},$$

where $\#(a)$ is the total number of actions of type a in all the events and $|A|$ is the number of available actions.

4 Model selection

Several parameters influence the estimation of a Markov Decision Process modeling customer relationships, such as the segmentation used to define the customer states, the state-driven or action-driven approach to estimate the transition probabilities and the rewards, the length of the time horizon in the training data, etc. The definition of the state space will probably have the highest impact on the performance of the model because the transition probabilities, the rewards, and the historical policy are estimated based on the states encountered in the training data.

Assuming there is a finite list of possible models $\mathcal{M}_1, \dots, \mathcal{M}_n$, corresponding to different choices of the parameters, we use *cross-validation* [7] to select the model with the best accuracy in predicting the long-term value generated by customers.

²This assumption is realistic if the company is not using any multistage decision model to target the customer base.

Feature	Description
rectrip	elapsed time since last purchase
freqtrip3	number of transactions in the last 3 months
freqtrip12	number of transactions in the last 12 months
value3	value generated in the last 3 months
value3camp	value generated from responding to campaigns in the last 3 months
value12	value generated in the last 12 months
value12camp	value generated from responding to campaigns in the last 12 months
miles3	miles flown in the last 3 months
miles12	miles flown in the last 12 months
longevity	number of days since first transaction

Figure 1: Customer features.

5 Case study

We apply our methodology to the customers of a major European airline and estimate the Markov Decision Process modeling the relationships with the company in order to efficiently allocate the marketing resources.

5.1 Data We use transactional data of customers for a period length of two years (2002, 2003). We do not segment customers into a finite number of states, as this is part of the model definition. At this stage we represent each customer with the set of numeric features defined in Figure 1.

After removing the outliers³, we randomly extract 10,000 customers. The customers are assigned randomly to two independent sets of size 5,000: the *validation set* used in the model selection phase and the *evaluation set* used to predict the future long-term value by simulating the historical and the optimal policy.

5.2 Defining customer states In order to build a MDP we need to discretize the high-dimensional feature space into a finite number of states. We propose a list of segmentation criteria which can be divided into two categories: a) business-based segments, and b) statistical-based segments.

The business-based segments are obtained by using recency, frequency, and monetary value. Each segmentation criterion can have several parameters. The segments are defined as follows.

- *RFM*(n) Scores the customers according to recency, frequency, and monetary value, then divides the so ranked customers into n segments of equal

³We removed marketing actions which have been applied very rarely and customers whose cumulative value is larger than the 99% percentile.

size.

- $ABC(a, b, c)$ Scores the customers according to a value feature, e.g. *value3*, and generates three segments by assigning the first $a\%$ to segment *A*, the next $b\%$ to segment *B*, and the remaining $c\%$ to segment *C*.
- $VD(a, b, c)$ The Value-Defectors (VD) segmentation performs $ABC(a, b, c)$ segmentation both on a value feature and on a loyalty index⁴, 9 segments are then obtained (e.g. *AA*, *AB*, *AC*, etc.).
- $RV(a, b, c)$ Recency-Value performs $ABC(a, b, c)$ segmentation both on a recency feature and on a value feature, there are 9 possible segments.

The following statistical-based segments use all the features defined in Figure 1.

- $Trees(n)$ Regression Trees [2] are used for supervised clustering. A regression tree is trained on an independent data set to predict the immediate reward of each customer. The leaves of the tree correspond to the segments. The parameter n indicates the number of leaves in the training set obtained by acting on the parameters of the algorithm. This is a supervised clustering technique as the leaves are built specifically to minimize the standard deviation of the reward.
- $SOM(n, m)$ Self-Organizing Maps [10] allow us to map the high-dimensional feature space into a two-dimensional rectangular $n \times m$ grid. The features are normalized and Euclidian distance is used.
- $K - means(n)$ K-means clustering [7] finds n clusters which are the centers minimizing the total within-cluster variance. The features are normalized and Euclidian distance is used.

5.3 Model selection We perform model selection using cross-validation. Each model defined in Figure 2 is tested in the case of state-driven and action-driven approach using the validation set.

We compare the mean absolute errors of each model using the state-driven and action-driven approaches. As shown in Figure 3, the state-driven approach outperforms the action-driven approach for each segmentation.

Therefore we focus on the state-driven approach in the remainder of the paper. The best model is the regression tree with 10 leaves (#19), followed by ABC (#10), the regression tree (#20), and ABC (#9).

⁴The loyalty index is a function of the frequency and the longevity of a customers and has been used by IBM in different CRM projects as a measure of the loyalty of customers.

#	Segment	Used features
1	RFM (10)	rectrip,freqtrip3,value3
2	RFM (20)	rectrip,freqtrip3,value3
3	RFM (30)	rectrip,freqtrip3,value3
4	RFM (10)	rectrip,freqtrip12,value12
5	RFM (20)	rectrip,freqtrip12,value12
6	RFM (30)	rectrip,freqtrip12,value12
7	ABC (10,10,80)	value3
8	ABC (10,20,70)	value3
9	ABC (10,10,80)	value12
10	ABC (10,20,70)	value12
11	VD (10,10,80)	value3,freqtrip3,rectrip
12	VD (10,20,70)	value3,freqtrip3,rectrip
13	VD (10,10,80)	value12,freqtrip12,rectrip
14	VD (10,20,70)	value12,freqtrip12,rectrip
15	RV (10,10,80)	value3,rectrip
16	RV (10,20,70)	value3,rectrip
17	RV (10,10,80)	value12,rectrip
18	RV (10,20,70)	value12,rectrip
19	Trees (10)	all
20	Trees (29)	all
21	K-means (10)	all
22	K-means (15)	all
23	K-means (20)	all
24	K-means (30)	all
25	SOM (3 × 3)	all
26	SOM (3 × 5)	all
27	SOM (4 × 5)	all

Figure 2: List of the used segmentations.

5.4 Simulation and optimization of customer dynamics

We use an independent data set (evaluation set) to train the MDP using the tree-based segmentation #19. Then we estimate the historical policy and simulate the Markov Decision Process in order to predict the future value obtained by following the historical policy.

In order to maximize the long-term value, we apply backward induction [15] and find the optimal policy⁵. Figure 4 compares the expected value per state obtained using the historical and the optimal marketing policy, the time horizon is set to 12 months.

6 Conclusions

In this paper we provide a rigorous methodology for the estimation of Markov Decision Processes modeling the dynamic relationship between the customers and the company. Although the use of Markov Decision Processes and dynamic programming is not new in the

⁵For confidentiality reasons we cannot show the historical policy and the optimal policy in terms of marketing actions undertaken by the company.

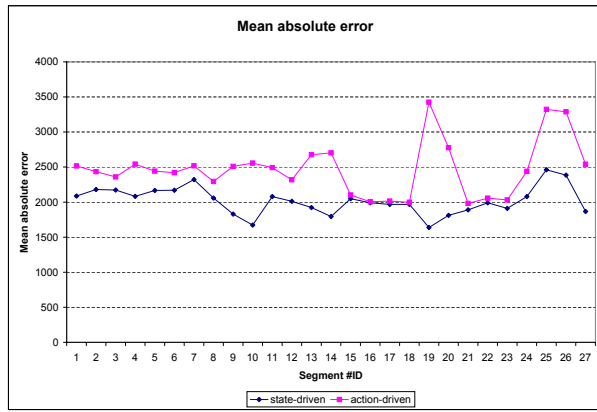


Figure 3: Comparison of the mean absolute error based on the prediction of the value generated in 12 months, for action-driven and state-driven approach.

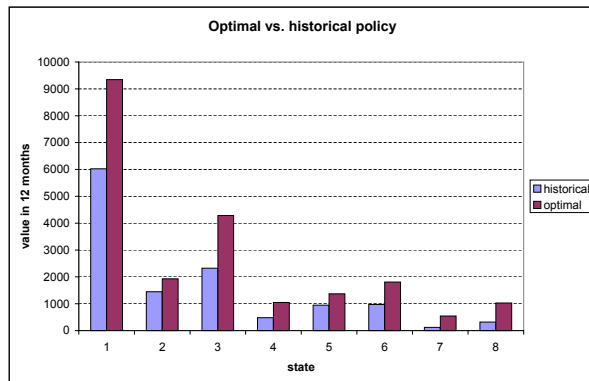


Figure 4: Comparison of the long-term value generated in 12 months when using the optimal and the historical policy.

literature, the issue of estimating robust models from customer relationship transactional data has not been addressed so far.

Using cross-validation for model selection, we are able to build a reliable Markov Decision Process taking into account the impact that the uncertainty in the parameters of the model has on the performance measure, i.e. the mean absolute error on the long-term value prediction. Our methodology enables efficient allocation of marketing resources by optimizing the long-term return on investment using the optimal marketing policy.

References

[1] G. R. Bitran and S. V. Mondschein. Mailing decisions

in the catalog sales industry. *Management Science*, 42(9):1364–1381, September 1996.

[2] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and Regression Trees*. Wadsworth Int. Group, California, USA, 1984.

[3] W. K. Ching, M. K. Ng, K. K. Wong, and E. Altman. Customer lifetime value: stochastic optimization approach. *Journal of the Operational Research Society*, 55:860–868, 2004.

[4] J. H. Drew, D. R. Mani, A. L. Betz, and P. Datta. Targeting customers with statistical and data-mining techniques. *Journal of Service Research*, 3(3):205–219, 2001.

[5] F. Gönül and M. Z. Shi. Optimal mailing of catalogs: A new methodology using estimable structural dynamic programming models. *Management Science*, 44(9):1249–1262, September 1998.

[6] S. Gupta, D. R. Lehmann, and J. A. Stuart. Valuing Customers. *Journal of Marketing Research*, 41(1):7–18, 2004.

[7] T. Hastie, R. Tibshirani, and J. H. Friedman. *The Elements of Statistical Learning*. Springer-Verlag, 2001.

[8] R. A. Howard. Comments on the origin and application of Markov Decision Processes. *Operations Research*, 50(1):100–102, 2002.

[9] D. Jain and S. S. Singh. Customer lifetime value research in marketing: A review and future directions. *Journal of Interactive Marketing*, 16(2):34–46, 2002.

[10] T. Kohonen. *Self-Organizing Maps*. Springer-Verlag, Berlin, 2 edition, 1997.

[11] P. Kotler. *Marketing Management*. Prentice-Hall, 10 edition, 2000.

[12] T. M. Mitchell. *Machine Learning*. McGraw-Hill, 1997.

[13] E. Pednault, N. Abe, B. Zadrozny, H. Wang, W. Fan, and C. Apte. Sequential cost-sensitive decision making with reinforcement learning. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2002.

[14] P. Pfeifer and R. Carraway. Modeling customer relationships as Markov Chains. *Journal of Interactive Marketing*, 14(2):43–55, 2000.

[15] M. L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, 1994.

[16] S. Rosset, E. Neumann, U. Eick, N. Vatnik, and S. Idan. Lifetime Value Models for Decision Support. *Data Mining and Knowledge Discovery Journal*, 7:321–339, 2003.

[17] R. Rust, K. Lemon, and V. Zeithalm. Return on marketing: Using customer equity to focus marketing strategy. *Journal of Marketing*, 68:109–127, 2004.

[18] D. I. Simester, P. Sun, and J. N. Tsitsiklis. Dynamic catalog mailing policies. Submitted, March 2003; revised May 2004.

[19] G. Tirenni. *Allocation of Marketing Resources to Optimize Customer Equity*. PhD thesis, University of St. Gallen, Switzerland, 2004.