Association for Information Systems
AIS Electronic Library (AISeL)

ICIS 2007 Proceedings

International Conference on Information Systems (ICIS)

December 2007

Sequential Decision Making for Profit Maximization Under the Defection Probability Constraint in Direct Marketing

Young Ae Kim
KAIST

Soung Kim
KAIST

Hee-Seok Song
Hannam university

SEQUENTIAL DECISION MAKING FOR PROFIT MAXIMIZATION UNDER THE DEFECTION PROBABILITY CONSTRAINT IN DIRECT MARKETING

Young Ae Kim
Korea Advanced Institute of Science and Technology (KAIST)
87 Hoegiro Dongdaemoon-gu Seoul, KOREA
kya1030@business.kaist.ac.kr

Hee Seok Song
Department of Management Information Systems, Hannam University
133 Ojung-dong, Daeduk-gu, Daejon, KOREA
hssong@hannam.ac.kr

Soung Hie Kim
Korea Advanced Institute of Science and Technology (KAIST)
87 Hoegiro Dongdaemoon-gu Seoul, KOREA
seekim@business.kaist.ac.kr

Abstract

Direct marketing is one of the most effective marketing methods with an aim to maximize the customer's lifetime value. Many cost-sensitive learning methods which identify valuable customers to maximize expected profit have been proposed. However, current cost-sensitive methods for profit maximization do not identify how to control the defection probability while maximizing total profits over the customer's lifetime. Unfortunately, optimal marketing actions to maximize profits often perform poorly in minimizing the defection probability due to a conflict between these two objectives. In this paper, we propose the sequential decision making method for profit maximization under the given defection probability in direct marketing. We adopt a Reinforcement Learning algorithm to determine the sequential optimal marketing actions. With this finding, we design a marketing strategy map which helps a marketing manager identify sequential optimal campaigns and the shortest paths toward desirable states. Ultimately, this strategy leads to the ideal design for more effective campaigns

Keywords: Sequential decision making, Reinforcement Learning, Direct marketing strategy; Customer Relationship Management, Marketing strategy map

Introduction

Direct marketing is one of the most effective marketing methods with an aim to maximize the expected profits (Wang et al. 2005). A key to direct marketing is to offer *optimized* campaigns to the most *valuable target* customers to maximize the profit return while minimizing cost. In recent years, a number of cost-sensitive learning methods focusing on predicting profitable customers have been proposed for direct marketing (Domingos 1999; Fan et al. 1999; Wang et al. 2005; Zadrozny et al. 2001). Domingos (1999) proposed the MetaCost framework to convert error-based classifiers to cost-sensitive classifiers by incorporating a cost matrix *C(i,j)* for misclassifying true class *j* into class *i*. MetaCost learns the classifier that predicts a customer's optimal class to minimize the expected cost. Fan et al. (1999) proposed AdaCost, a misclassification cost-sensitive boosting method. The principle of the proposed model is to assign high weights to expensive examples and comparably lower weights to inexpensive examples in order to reduce cumulative misclassification costs by introducing a misclassification cost. This method adjusts the function into weight updated rules. Zadrozny et al. (2001) provided an even more general cost-sensitive decision making method than MetaCost, when misclassification costs and probability are unknown and different from the examples. Wang et al. (2005) further proposed a data mining model for identifying potential customers who are likely to respond in the current campaign. They consider the customer value in selecting association rules and maximize the sum of the net profit over the contacted customers. However, a common objective of these methods is to only maximize the short-term profit associated with each marketing campaign. They ignore the interactions among decision outcomes when sequences of marketing decisions are made over time. These independent decision-making strategies cannot guarantee the maximization of total profits generated over a customer's lifetime because they often inundate profitable customers with frequent marketing campaigns, encourage radical changes in customer behavior, or neglect potential defectors. This approach can decrease customer profitability because of the annoyance factor or their budgetary limits per unit time (Pednault et al. 2002). For some customers, such as potential defectors or those who have already spent their allotted budget, marketing actions which sacrifice short-term profit by sending thank-you letters or discount coupons, may be optimal actions to strengthen customer loyalty for the companies and to ultimately increase total profits over the customer's lifetime.

Some researchers have recognized the importance of sequential decision making to overcome the limitations of isolated decision making. For example, Pednault et al. (2002) and Abe et al. (2004) proposed sequential cost-sensitive learning methods for direct marketing. Pednault et al. (2002) first suggested a novel approach to sequential decision making for direct marketing based on Reinforcement Learning with the goal of maximizing total profits over a customer's lifetime. They provide batch Reinforcement Learning methods with a regression model as a function approximation, which attempts to estimate long-term profits in each customer's state and marketing action pair. Their experimental results shown by simulation reveal that the sequential decision making approach with Reinforcement Learning outperforms the usual target marketing methods for optimizing individual campaigns. Abe et al. (2004), based on their earlier work (Pednault et al. 2002), expanded their research to optimize sequential marketing actions in the respective marketing channels for maximization of long-term profits across different channels. However, current sequential cost-sensitive methods for maximizing long-term profit do not indicate how to control the probability of customer defection while maximizing total profits over the customer's lifetime. Although a primary objective of direct marketing is to maximize total profits, it is also important to control the probability of customer defection, keeping it under a desirable or acceptable level. Customer defections produce both tangible and intangible losses, e.g., increasing the acquisition cost of a new customer, loss of word-of-mouth effects, and loss of future cash flows and profits from a customer, even though the probability of a customer defection is very low. However, the current sequential decision making approach for maximizing long-term total profit can only consider a tangible factor such as the loss of future cash flow and profits from potential defectors. In recent years, customer switching costs are much lower in e-commerce marketplaces, so a company needs to pay more attention to customer defections regardless of customer loyalty.

Unfortunately, optimal marketing actions designed to maximize profits often perform poorly in minimizing the probability of customer defection due to a conflict between profit-maximization and defection probability-minimization. For example, an optimal marketing action for profit maximization is liable to give up unprofitable customers who are most likely to defect but are profitable from a long-term perspective. In contrast, an optimal marketing action for the minimization of defection probability is apt to unnecessarily sacrifice loyal customers' profit with excessive marketing costs.

To overcome this conflict, we regard the customer defection probability as a constraint and attempt to control it under the given threshold because, in general, controlling defection probability under the threshold is more cost effective than completely avoiding customer defection with zero percent. In this paper, we have developed a *sequential* decision-making methodology for profit maximization under the given defection probability constraint. To suggest simpler and more practical business intelligence, we have designed a framework for segmentation marketing, instead of focusing on individualized marketing. We define the possible customer states (i.e., segments) using a Self-Organizing Map (SOM) and monitor the shift of customer behavior states, marketing actions, and resulting profits and defection probability over time. Based on the movement among all these states and marketing actions, we have adopted the Reinforcement Learning algorithm to determine an optimal marketing action in each state (i.e., segment). With these results, we have also suggested the concept of a marketing strategy map which visualizes the results of the learning including an optimal marketing action and customer behavior dynamics in each state. This marketing strategy map can help a company identify sequential optimal campaigns and determine the shortest paths toward desirable states. Ultimately, this strategy leads to the ideal design for more effective campaigns.

Background

Self-Organizing Map(SOM)

The SOM (Kohonen, 1995) is a sophisticated clustering algorithm in terms of the visualization of its clustering results. It not only clusters high-dimensional data points into groups, but also represents the relationships between the clusters in a much lower-dimensional space. In the SOM model, the input vectors are connected to an array of neurons in the output layer. Each neuron in the output layer is represented by an n-dimensional weight vector $\mathbf{m} = [m_1, m_2, ..., m_n]$, where n is equal to the dimension of the input vectors. In each training step with each input vector, the distances between an input vector and all the weight vectors of the output layer are calculated. Then, the neuron whose weight vector is closest to the input vector is selected as the Best-Matching Unit (BMU). After finding the BMU, the weight vector of the BMU is updated so that the BMU is moved closer to the input vector. The topological neighbors of the BMU are also treated in a similar way. This adaptation procedure stretches the BMU and its neighbors toward the sample vector. This mapping tends to preserve the topological relationship of the input data points so the similar input data points are mapped onto nearby output map units.

In our method below, we define the possible customer states onto two-dimensional map using SOM, because the output map of SOM promote visual understanding of analysis such as monitoring customers' state shift and identifying sequential optimal campaigns.

Reinforcement Learning

Reinforcement Learning (Sutton et al. 1998) is characterized by goal-directed learning from interaction with its environment. At each discrete time t, the learning agent observes the current *state* $s_t \in S$, where S is the set of possible states in a system and selects an action $a_t \in \mathrm{A}(s_t)$, where $\mathrm{A}(s_t)$ is the set of actions available in state s_t. As a consequence of its action a_t in state s_t, the agent receives an immediate *reward* r_{t+1}, and next state s_{t+1}. Based on these interactions, the agent attempts to learn a policy $\pi : S \to A$ which is a function of mapping states to actions to maximize the expected sum of its immediate rewards, $R = \sum_{t=0}^{\infty} \gamma^t r_t$ [where γ (i.e., $0 \le \gamma < 1$) is *a discount rate*]. Thus, Reinforcement Learning is particularly well suited to multi-step decision problems where the decision criteria can be represented in a recursive way as a function of the immediate numerical value.

Domain and Data Collection

To evaluate our sequential decision making method's feasibility in direct marketing, we experimented with a part of KDD-CUP-98 datasets which concerns direct-mail promotions soliciting donations.

The dataset for experiments consists of 95,412 records. Each record contains each donor's direct-mail promotion pattern for 22 campaigns conducted monthly for about two years (e.g. which direct-mail was sent, when it was sent, etc.) and response behavior against each promotion (whether a donor responded or not by each promotion, how much was donated, etc). It also includes summary data reflecting donor's promotion and response history such as total number of promotions received to date, total amount of donations to date and RFA (Recency-Frequency-Amount) status at each promotion. For effective experiments, we classified the original dataset into two donor groups. The first group included donors who often responded to campaigns except for the last (22nd) campaign. We collected data from the "active donors" group by excluding the data from the last two campaigns from each donor in this group. The second group included donors who had previously actively donated, but stopped donations long before the last (22nd) campaign. By definition of Paralyzed Veterans of America (a donor of KDD-CUP-98 datasets), the second group included "the lapsed donors" who had not made a donation within the last 12 months. Out of this group, we prepared data from the defector's group by collecting campaigns and response history until the donors became lapsed donors. We defined the lapsed donors as a fatal state (i.e., a defection state). The original dataset had some fields showing whether a donor would become a lapsed donor or not at each promotion. We sampled 10,000 records which consisted of 50 percent active donors and 50 percent defectors to equally ascertain information from both groups. Although the proportion of "active donors" and "lapsed donors" is equal, the probability that a customer defects next period by single campaign is only just 5%.

Method

In this section, we will define state, action and immediate rewards and explain how to learn optimal marketing actions for profit maximization under the control of defection probability in direct marketing. We also provide some examples and experimental results with prepared datasets to help understand our method.

Definition of States and Actions

States are representations of the environment that the agent observes and are the basis on which agent's decisions are made. In this method, states would be represented as customer segments which have similar purchase patterns and response behaviors against promotion (e.g., recency, frequency, and monetary value) at the time of each campaign. Thus,

$$S = \{s1\ ,s2,\ldots\ldots,sN\}$$

where S is the set of states, N is the total number of states.

The actions are defined as all of the marketing campaigns conducted in a company. Thus,

$$A = \{a1\ ,a2,\ldots\ldots,aM\}$$

where A is the set of actions, M is the total number of actions.

In our experiments, we conducted customer segmentation using a Self-Organizing Map (SOM) in order to determine a set of possible states S. Table1 shows the 14 input features of the SOM which were collected at the time of each campaign. Since the SOM was given no information about the optimal number of states, we had to experiment with the number of states of the SOM. Figure 1 illustrates the possible states (i.e. 6x8 SOM) as a result of SOM learning. Each state was assigned its number to distinguish the states (e.g., $s1, s2, \ldots, s48$).

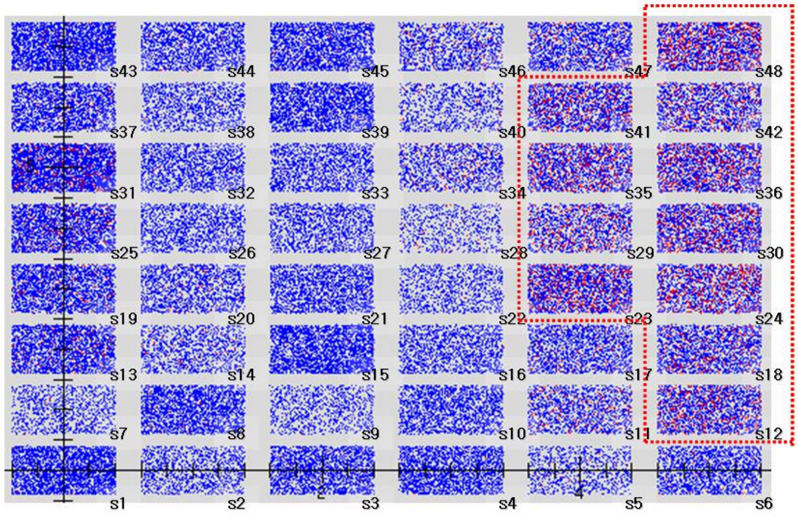

Figure 1. A Set of States on the Output Map of SOM (6x8 SOM)

Table 1. Input Features of SOM

Category		Features	Descriptions
Promotion Pattern		tot_num_pro	Total number of promotions to date
		tot_num_pro_6m	Total number of promotions in the last 6 months
Response Behavior	History	tot_amt_don	Total amount of donations to date
		tot_num_don	Total number of donations to date
		amt_per_don	Average amount per donation to date *(tot_amt_don / tot_num_don)*
		frequency	Response rate to date *(tot_num_don / tot_num_pro)*
		amt_per_pro	Average amount per promotion to date *(tot_amt_don / tot_num_pro)*
	Recent (6months)	tot_amt_don_6m	Total amount of donations in the last 6 months
		tot_num_don_6m	Total number of donations in the last 6 months
		amt_per_don_6m	Average amount per donation in the last 6 months *(tot_amt_don_6m / tot_num_don_6m)*
		frequency_6m	Response rate in the last 6 months *(tot_num_don_6m / tot_num_pro_6m)*
		amt_per_pro_6m	Average amount per promotion in the last 6 months *(tot_amt_don_6m / tot_num_pro_6m)*
	Last	recency	Number of months since the last donation
		last_amt	Amount of the last donation

In the original dataset, the set of actions, A had 11 types of action, and we gave a number to each action from $a1$ to $a11$. Table 2 shows the descriptions of marketing actions conducted by Paralyzed Veterans of America.

Table 2. Marketing Actions

Action	Description	Action	Description
a1	Mailings are general greeting cards (an assortment of birthday, sympathy & get well) with labels	a7	Mailings are Christmas cards with labels
a2	Mailings have labels and a notepad	a8	Mailings have labels and a notepad
a3	Mailings have labels only	a9	Mailings have labels only
a4	Mailings have thank you printed on the outside with labels	a10	Mailings are calendars with stickers but do not have labels
a5	Mailings are blank cards with labels	a11	Mailings are blank cards with labels
a6	Mailings are blank cards that fold into thirds with labels	-	-

Definition of Profit and Defection Probability

The agent achieves both profit and defection probability as immediate rewards at each transition. An immediate profit P is the net profit which is computed as the donation amount minus the cost of the campaign ($ 0.68) in our experiment. An immediate defection probability D is computed as the probability of falling into a fatal state (i.e., defection state). The concept of fatal state was first introduced by Geibel et al. (2001) who noted that processes, in general, have a dangerous state which the agent wants to avoid by the optimal policy. For example, a chemical plant may explode when temperature or pressure exceeds some threshold. The optimal strategy of operating a plant is not to completely avoid the fatal state when considering the related control costs, but to control the probability of entering a fatal state (i.e., an exploration) under a threshold.

In this method, a fatal state means the status of customer defection such as being a lapsed donor in our experiment. Like an exploration in a chemical plant, customer defection is fatal to a company and brings about tangible and intangible loss. However, it is difficult to reflect both the tangible and intangible loss from defection to the reward of profit. It is also impossible and cost-ineffective to completely avoid customer defection, but customer defection could be controlled under the threshold. The immediate defection probability D on transition from s to s' under action a is defined by:

$$D(s,a,s') = \begin{cases} 1 & \text{if } s \text{ is a non-fatal state, } s' \text{ is a fatal state} \\ 0 & \text{else} \end{cases} \qquad (1)$$

If the agent enters a fatal state (i.e. a lapsed donor) from a non-fatal state (i.e. an active donor), the immediate defection probability is 1 and the immediate profit is zero.

Learning Strategy

The objective of the proposed method is to maximize the total profit while controlling the defection probability under the given threshold for all states, as follows:

$$V_P^{\pi}(s) = \mathrm{E}\left(\sum_{t=0}^{\infty} \gamma_P^t P_t\right) \to \max$$

$$V_D^{\pi}(s) = \mathrm{E}\left(\sum_{t=0}^{\infty} \gamma_D^t D_t\right) \le \theta \qquad (2)$$

where $V_P^{\pi}(s)$ is the cumulative profits and $V_D^{\pi}(s)$ is the probability that an agent ends in a defection state, when it starts in state s. An immediate profit P and an immediate defection probability D are discounted by discount rates

γ_P ($0 < \gamma_P < 1$) and $\gamma_D = 1$, respectively. Since γ_P is lower than 1, the agent will try to reach a more profitable state as quickly as possible, controlling the defection probability under the threshold. In addition, since γ_D is 1 and D is defined by (1), a value of $\sum_{t=o}^{k} \gamma_D^t D_t$ is 1, if and only if a customer in state s enters a defection state and ends his relationship with the company at time k (Geibel et al. 2001). In order to construct an optimal policy π^*, the state-action value function $Q^\pi(s,a)$, which is a value of taking action a in state s under policy π, is computed by Watkin's Q-learning algorithm (Watkins 1992). The state-action value function $Q_P(s,a)$ and $Q_D(s,a)$ can be defined by:

$$Q_P(s,a) = \mathrm{E}\left[P(s,a) + \gamma_P V_P^*(s')\right]$$

$$Q_D(s,a) = \mathrm{E}\left[D(s,a) + \gamma_D V_D^*(s')\right] \quad (3)$$

Where $P(s,a)$ and $D(s,a)$ are an immediate profit and defection probability of taking action a in state s, respectively, $V_P^*(s')$ and $V_D^*(s')$ are optimal values of the next state s' under the optimal policy π^*. To optimize the total profit under the given defection probability, the optimal policy is selected by *a reverse-1st lexicographic ordering:*

$$\pi^*(s) = \arg\max_a \succeq \left[\max\left(Q_D^\pi(s,a), \theta\right), Q_P^\pi(s,a)\right] \quad (4)$$

where $\max\left(Q_D^\pi(s,a),\theta\right)$ is higher value between $Q_D^\pi(s,a)$ and θ.

The agent prefers an action a to a' if $\max\left(Q_D^\pi(s,a),\theta\right) < \max\left(Q_D^\pi(s,a'),\theta\right)$ or if $\max\left(Q_D^\pi(s,a),\theta\right) = \max\left(Q_D^\pi(s,a'),\theta\right)$ and $Q_P^\pi(s,a) \geq Q_P^\pi(s,a')$. If several marketing actions would have $Q_D^\pi(s,a)$ less than threshold θ, they have the same value θ compared to the first component, $\max\left(Q_D^\pi(s,a),\theta\right)$. Then, the agent compares the second component $Q_P^\pi(s,a)$ and selects the action with the highest profit value as an optimal action.

Figure 2 shows the algorithm for learning $Q_P(s,a)$ and $Q_D(s,a)$ and achieving the optimal policy $\pi^*(s)$. Input training data E is a set of episodes where each episode is a sequence of events, and each event consists of a state, an action, profit, and defection probability. Each episode represents the campaign interactions between a customer and a company as time goes on. Note that we introduce a dummy state s_{def} and a dummy action a_{def} for a defection state and its action for technical reasons. If a customer falls into a defection state, we construct his last event with $< s_{def}, a_{def} >$. We also compute both $Q_P\left(s_{def}, a_{def}\right)$ and $Q_D\left(s_{def}, a_{def}\right)$ at zero, because customers entering a defection state have no permanent rewards. $Q_P(s,a)$ and $Q_D(s,a)$ are updated with each episode from line 2 to 8. At line 5, α is a step-size-parameter which affects the rate of convergence to $Q^*(s,a)$. To ensure convergence, α is set up to be a decreasing function of time t. If each state-action set is visited infinitely often, the above training rule assures convergence to $Q^*(s,a)$ as $t \to \infty$, for all (s,a) (Sutton et al. 1998; Watkins 1992). After the learning for all episodes, the agent can achieve the optimal policy in each state at line 9 and 10. However, the agent changes the optimal action a^* into "no action" if the cumulative profit (i.e., $Q_P(s,a^*)$) is negative regardless of the $Q_D(s,a^*)$ value. If companies determine that "no action," is needed for the state, they have to give up customers in the state or develop new campaigns which are especially effective for the state.

```
Input: a set of episodes $E = \{E_1, E_2, \ldots, E_i\}$
        where $E_i = \{< s_{i,1}, a_{i,1}, P_{i,1}, D_{i,1} >, < s_{i,2}, a_{i,2}, P_{i,2}, D_{i,2} >, \ldots, < s_{i,j}, a_{i,j}, P_{i,j}, D_{i,j} >\}$,
    thrshold $\theta$, $\gamma_P (0 < \gamma_P < 1)$, $\gamma_D = 1$.
Output: state-action value function $Q_P(s,a), Q_D(s,a)$,
    optimal policy(action) $\pi^*(s)$ in each state.
$Q_P(s,a) = 0; Q_D(s,a) = 0;$
For each episode $E_i$
    For $l = 1$ to $j$ -1
        $t = visit(s,a)$;  //total number of times that $(s,a)$ has been visited up to
        $\alpha_t = \frac{1}{1+t}$;
        $Q_P^t(s_{i,l}, a_{i,l}) = (1-\alpha_t) Q_P^{t-1}(s_{i,l}, a_{i,l}) + \alpha_t \left(P_{i,l} + \gamma_P Q_P^{t-1}(s_{i,l+1}, a^*)\right)$;
        $Q_D^t(s_{i,l}, a_{i,l}) = (1-\alpha_t) Q_D^{t-1}(s_{i,l}, a_{i,l}) + \alpha_t \left(D_{i,l} + \gamma_D Q_D^{t-1}(s_{i,l+1}, a^*)\right)$;
            where $a^* = \arg\max_a \succeq \left[\max\left(Q_D^{t-1}(s_{i,l+1}, a), \theta\right), Q_P^{t-1}(s_{i,l+1}, a)\right]$
For each state $s$
    Find $\pi^*(s) = \arg\max_a \succeq \left[\max\left(Q_D(s,a), \theta\right), Q_P(s,a)\right]$;
```

Figure 2. The Learning Strategy with Profit Maximization under the Given Defection Probability

Experiments

In this section, we discuss how to determine the optimal number states (i.e. segments) of SOM in our experiments. We also report the tradeoffs between two conflicting objectives such as profit and defection probability and show the potential benefit of our methodology in direct marketing strategy. In particular, we compare our method with single objective models such as a model for profit maximization and a model for defection probability minimization based on sequential decision making. However, we don't compare with existing isolated (i.e. single) decision model which selects an optimal action based on response rate and short-term profit or risk for a given action. Because sequential decision model with Reinforcement Learning has shown to be superior to single decision model when its objective is to maximize total profit (Pednault et al. 2002). To the best of our knowledge this is the first study on suggesting *sequential* optimal marketing actions for maximizing *total profit* while the *defection probability* is kept below the threshold (an acceptable level). Our work differs from a Pareto-Genetic algorithm proposed by Bhattacharyya (2000) in that they suggested multiple optimal actions based on short-term profit and defection probability for a given *single* action, i.e. based on isolated decision making. Our work also differs from LTV model (Life Time Value) suggested by Hwang (2004) in that they just selected target customers with high lifetime value through an LTV model considering expected future cash flow and defection probability, but could not suggest which actions are optimal to maximize profit and control defection probability.

The Optimal Number of states of SOM

We experiment on varied number of states of SOM to choose optimal number of states of SOM. Figure 3 shows performance of SOM models with varied number of states when learning to maximize total profit and to minimize defection probability respectively. The 6x8 SOM outperformed the others by achieving the highest average total profit over all the states and was optimal for profit maximization. The 6x7 SOM outperformed the others by achieving the lowest average defection probability over all the states and was optimal for defection probability minimization. We can also observe that there is no relationship (i.e. positive or negative) between the number of states and the performance of SOM. For experiments to test our method, we chose two SOM models, 6x8 SOM (48 states) and 6x7 SOM (42 states). It is noted that these single objective models of each SOM also learned with

Reinforcement Learning algorithm. And all experiments of the SOM model training were conducted with Clementine, a SPSS Data Mining tool.

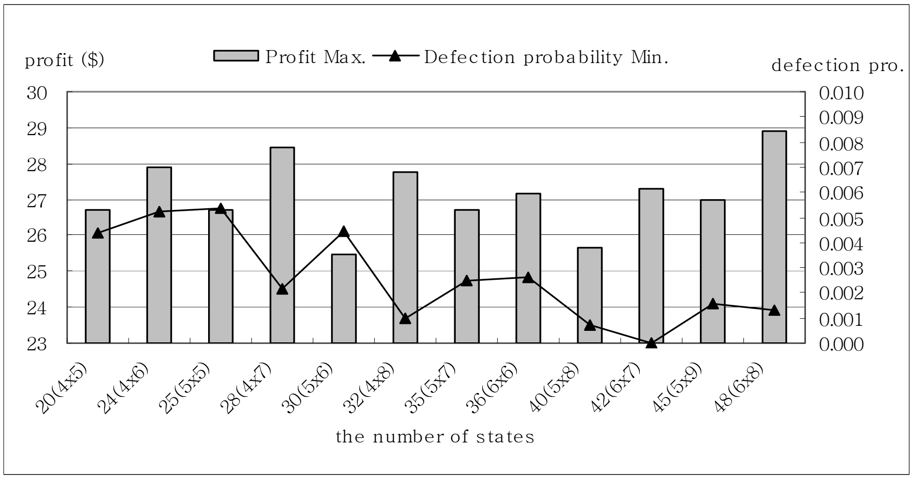

Figure 3 Performance Comparison of SOM Model with Single Objectives

Results

In our method, the threshold reflects a desirable or acceptable level of customer defections for each company. The decision of the threshold is invariably dependent on several factors including the conditions of the market and the characteristics and goals of each company. Therefore, we experimented with our method using various levels of thresholds and the selected SOM models (i.e., 6x8 SOM and 6x7 SOM). We then observed the change of the average Q_P and Q_D values. For finite experiments on the threshold, we increased the value of the threshold by 0.05 (5%) within a meaningful range of thresholds.

Table 3 shows the results of the single objective models which were learned for defection probability minimization or profit maximization, respectively. The highest defection probability of each single objective model provided the lower and upper bound on experiment thresholds. In the case of the 6x8 SOM model, if the agent learns to minimize the defection probability, the agent is able to control the defection probability under 0.0307 over all states. This value of 0.0307 is the lowest threshold which the agent is able to achieve with the 6x8 SOM. In contrast, if the agent learns to maximize the total profit, the agent does not consider the defection probability and, therefore, retains the defection probability under 0.3025 over all states. The value of 0.3025 is the upper bound of the threshold. Values over this threshold are meaningless in the 6x8 SOM because the agent is not able to achieve more than the average total profit in the total profit maximization model (i.e., $28.91) even though the threshold is increased over 0.3025. Based on the results in Table 3, we changed the thresholds by 0.05 between 0.0307 and 0.3025 in the 6x8 SOM model and between 0 and 0.2846 in the 6x7 SOM model.

Table 3. Results of Single Objective Models for the Decision of a Meaningful Range of Thresholds

	Min. of Defection Probability			Max. of Total Profit		
	Average total profit	Average defection pro.	The highest defection pro.	Average total profit	Average defection pro.	The highest defection pro.
6x8 SOM	7.29	0.0013	**0.0307**	28.91	0.1333	**0.3025**
6x7 SOM	7.41	0	**0**	27.28	0.1400	**0.2846**

Figure 4 shows the performance comparison of our methods with different thresholds. We compared the average Q_P and Q_D values over all starting states assuming equal distribution of donors into all states. Note that the first bar and triangular point correspond to the single objective model for minimization of the defection probability, and the last bar and triangular point are the single objective model for maximization of the total profits. As mentioned earlier, we were able to observe the conflict between the two marketing objectives: maximization of the total profit and minimization of the defection probability. As we achieved more total profit by alleviating the constraints of the defection probability, the number of donors who were apt to defect increased more. Figure 4 also shows that the single objective model for minimization of the defection probability achieved very poor performance of the total profits, \$7.29. In addition, the single objective model for maximization of the total profits achieved poor performance of the defection probability, 0.133 (13.3%), and the threshold, 0.3025 (30.25%), because it disregarded the other objective. However, when learning with each threshold in our model, we could achieve far more satisfactory results of both the total profit and the defection probability.

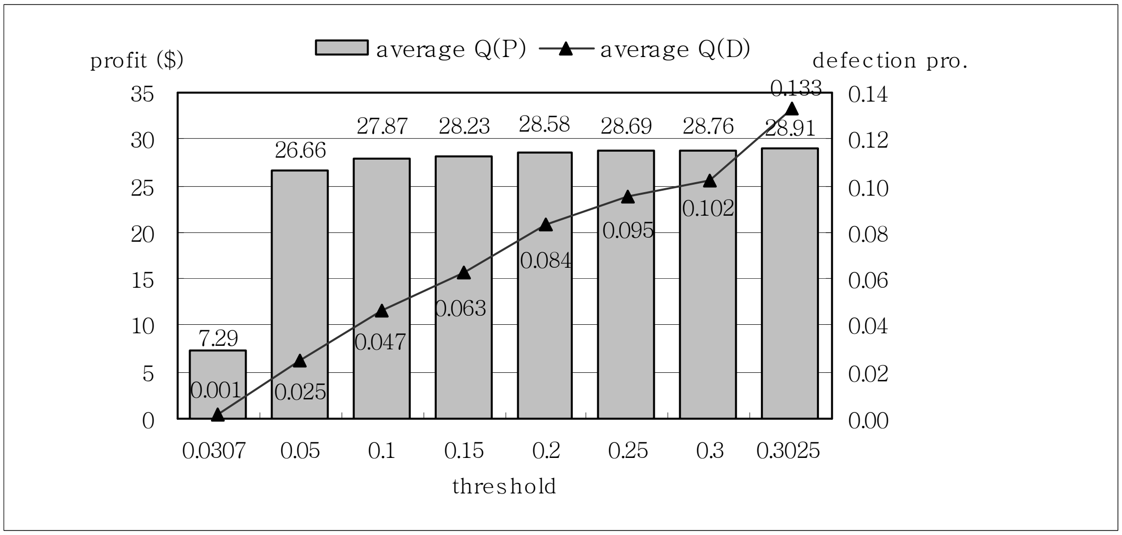

Figure 4 (a) Performance Comparison of Different Thresholds in the 6x8 SOM Model

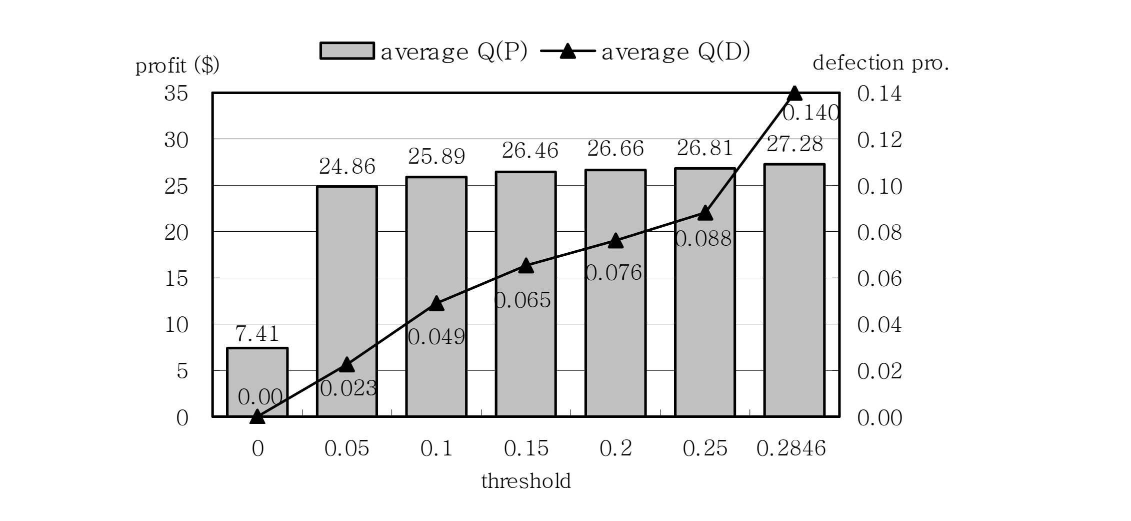

Figure 4 (b) Performance Comparison of Different Thresholds in the 6x7 SOM Model

At this point, marketing experts in each company could decide the threshold based on observed tradeoffs and acquired knowledge. For further analysis, we selected 0.05 (5%) based on Duncan's test (significance level=0.05) of the average Q_P and Q_D values. When we select 0.05 as the threshold in both the 6x8 SOM and 6x7 SOM, we could achieve a significantly lower average Q_D value than the models with thresholds of 0.1 or 0.15. We could also achieve the same average Q_P value as models with thresholds of 0.1 or 0.15 achieved.

To choose a better model between two SOM models (the 6x7 SOM and 6x8 SOM with a threshold of 0.05), we took a T-test (significance level=0.05). The T-test demonstrated that the 6x8 SOM model significantly outperformed the 6x7 SOM model. The 6x8 SOM achieved a significantly higher average Q_P value ($26.66) than the 6x7 SOM ($24.86), but a significant difference in the average Q_D value was not observed between the 6x8 SOM (0.025) and the 6x7 SOM (0.023).

Table 4 shows the performance comparison of our model with a threshold of 0.05 and the single objective models in the 6x8 SOM. The last row in Table 4 gives the results when campaigning without any optimization model. The $Q_P(s,a)$ and $Q_D(s,a)$ values in no optimization model are calculated as follows:

$$\begin{aligned} Q(s,a) &= \mathrm{E}\left[r(s,a)+\gamma V(s')\right] \\ &= \mathrm{E}\left[r(s,a)\right]+\gamma\sum_{s'} p(s' \mid s,a)Q(s',a') \end{aligned} \quad (5)$$

Unlike other optimization models, there is no strategy to select optimal action a^* in transition state s'. It uses the $Q_P(s',a')$ and $Q_D(s',a')$ values observed from a training dataset instead of the optimal values $Q_P^*(s',a^*)$ and $Q_D^*(s',a^*)$.

Table 4. Performance Comparison with Single Objective Models.

	Average Q_P	Average Q_D	Expected revenue	Threshold	Q_P Improvement	(1-Q_D) Improvement	Exp. Revenue Improvement
Max. of Total Profit	28.91	0.1333	25.95	0.3025	4.5 (28.91/6.38)	1.55 (0.8667/0.56)	7.46 (25.95/3.48)
Our model ($\theta = 0.05$)	26.66	0.0248	**26.02**	0.05	4.2 (26.66/6.38)	1.74 (0.9752/0.56)	**7.48** (26.02/3.48)
Min. of Defection Pro.	7.29	0.0013	7.20	0.0307	1.1 (7.29/6.38)	1.78 (0.9987/0.56)	2.07 (7.20/3.48)
No optimization model	6.38	0.44	3.48	0.67	-	-	-

As shown in Table 4, our method significantly outperformed the single objective models in terms of expected revenue (significance level=0.1). The expected revenue is the average expected revenue generated from the surviving customers who do not defect. It is computed by multiplying the average total profit by the rate of surviving customers as follows:

$$\frac{\sum_{S} Q_P(s,a^*)(1-Q_D(s,a^*))}{N} \quad (6)$$

where N is the total number of states.

The last three columns of Table 4 show the improvement over no optimization model. Our model increased the average expected revenue by 7.48 times over no optimization model, while the single objective models increased the average by 7.46 times and 2.07 times, respectively. Our method was able to improve the total profit and prevent the customer defection probability under the given threshold over all states and ultimately, achieve higher expected revenues. It is noted that the objective of the proposed method is to maximize the total profit while controlling the defection probability under the given threshold (e.g. 5%) for all states, not to maximize the expected revenues by (6). Although the single objective model for profit maximization seems to perform as well as our method in terms of the expected revenue, it is ineffective to reduce defection probability in some states. Its average defection probability over all states is 13.3% and it lost 30.25% of customers in some states. As we mentioned before, a company will suffer great intangible loss.

Design of a Marketing Strategy Map

To clearly show an optimal action and customer behavior dynamics in each state, we designed a marketing strategy map on the output map of SOM. Figure 5 (a) illustrates the marketing strategy map of our experiments. To find customers' paths in each state, we exploited the association rules of the form $(state = s \;\&\; action = a^*) \rightarrow (next\ state = s')$, where action a* is the optimal action of state s. We selected the association rules in order of high confidence until the sum of confidences from the selected rules was over 70%.

As expected, most customers shift from a current state to nearby states on the strategy map by the targeted campaign because input behavior patterns between two nearby states are mostly similar according to the topology preserving property of the SOM. However, some customers significantly change their behavior states. We describe these transitions with a direction arrow and a state number on the map. For example, customers in state s2 move into state s3 (24.3%), s9 (21.6%) or s45 (16.2%) or remain in state s2 (21.6%) by action a3 in Figure 5 (a).

The marketing strategy map also shows the desirable states to which a company attempts to drive customers. We selected the top 10% of states in terms of total profits as the desirable states (i.e. s6, s34, s36, s37 and s40). However, we did not consider defection probability, because defection probability is controlled under the given threshold over all states.

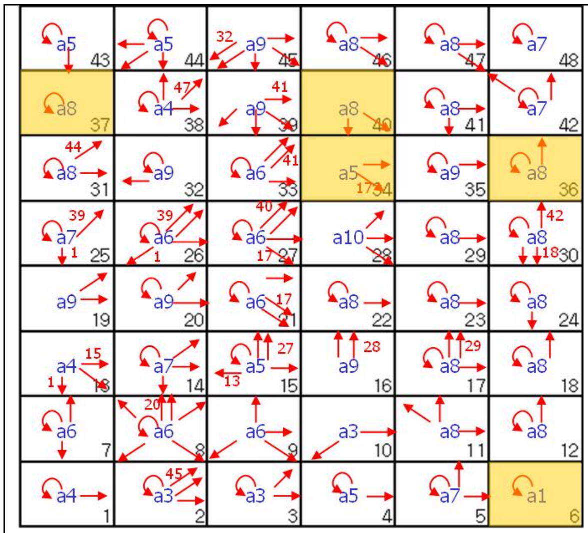

Figure 5 (a) The Marketing Strategy Map: Optimal Actions and Major Customer Paths

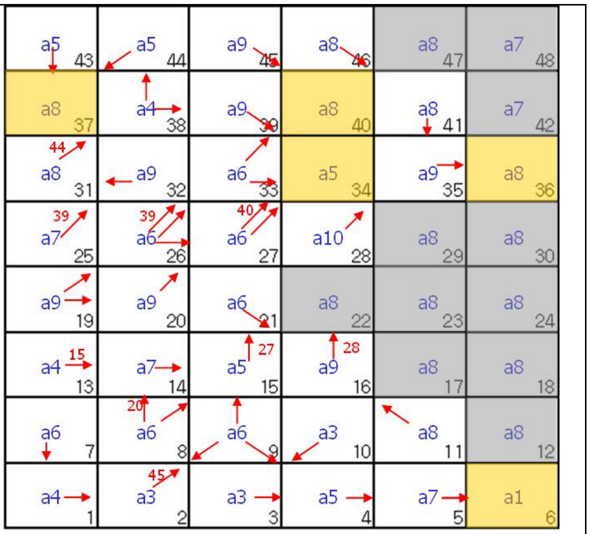

Figure 5 (b) The Shortest Path Map: Customers' Shortest Paths to the Desirable States (s6, s34, s36, s37 and s40)

Marketing Strategy Map Application

The marketing strategy map itself in Figure 5(a) shows multiple sequential optimal campaigns and paths to the desirable states from each state, because each action in each state is the optimal campaign when considering long term objectives based on sequential decision makings. Among the multiple solutions and paths, a marketing manager can find the shortest path to design more effective campaign strategies as well as to identify the shortest paths and its sequential optimal campaigns towards desirable states. To design the shortest path map, we found a set of states with the shortest paths which lead to the desirable states after n period (i.e. D_n) (See Table 5). We first selected all rules of the form $(state = s \in (S - D_0)\ \&\ action = a^*) \rightarrow (next\ state = s' \in D_0)$, where S is a set of all possible states and D_0 is a set of desirable states. With these rules, we found all states (i.e. D_1) and their paths which lead to the desirable states after 1 period. We then selected all rules of the form $(state = s \in (S - \bigcup_{i=0,1} D_i)\ \&\ action = a^*) \rightarrow (next\ state = s' \in D_1)$. With these rules, we found all states (i.e. D_2) and their paths which led to the desirable states after 2 periods via one of the states in D_1 . By repeating this process, we finally found all the states from D_1 to D_4 and their paths. There was no further $D_n (n \geq 5)$. A state in D_n has direct paths to states in D_{n-1} and leads to states in D_0 via states in $D_{n-i} (i = 1, 2, ... n-1)$, sequentially.

Table 5. A Set of States Leading to the Desirable States after n Periods (D_n)

D_n	States
D_0	s6, s34, s36, s37, s40
D_1	s5, s27, s33, s35, s39, s43, s44, s45
D_2	s2, s4, s15, s20, s25, s26, s28, s31, s38, s41
D_3	s1, s3, s8, s9, s13, s14, s16, s19, s32, s46
D_4	s7, s19, s11, s21

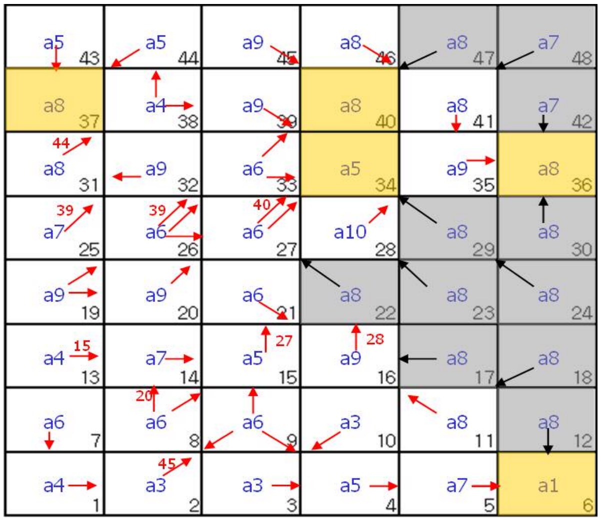

Figure 5 (c) A Marketing Campaign Strategy for the Ineffective States

Figure 5(b) illustrates the shortest path map. A marketing manager can identify the shortest paths and its sequential optimal campaigns toward desirable states in each state. For example, customers in state s38 can go to the desirable states after 2 periods through 2 different paths: s38(by a4)→s44(by a5)→s37 or s38(by a4)→s39(by a9)→s34. The shortest path map can then be used to determine if a current campaign is effective or not by identifying states which have no path to the desirable states. In Figure 5(b), a total of 10 states, gray color states (e.g., s12, s17, s18), cannot lead to desirable states. Among these 10 ineffective states, 8 states (i.e., s12, s18, s23, s24, s29, s30, s42, s48) are risky states in which the probability of being a defector in next period is over 10%.

For these ineffective states, a marketing manager needs to develop new campaign strategies for ineffective states. The objective of designing new strategies is to provide the shortest path to get to the desirable states. We adopt a gradual approach which suggests next state with the fastest path to desirable states among immediate neighbors of an ineffective state. This approach is based on the fact that it is very difficult to significantly change customer behavior in such a short period. Figure 5(c) illustrates new campaign strategies for the ineffective states. For example, we designed this strategy to drive customers in state s22 to state s27 (D_1) among its immediate neighbors, (i.e. s15 (D_2), s16 (D_3), s21 (D_4), s27 (D_1) and s28 (D_2)) (See Table 5) because state s27 has the fastest path to desirable states. In the case of state s29, we could select two strategies, (i.e. s34 (D_0) and s36 (D_0)), but we selected state s34 with a higher total profit as the next state. In the case of state s24, it has no immediate neighbor states which have the shortest path to the desirable states. Therefore, after designing new strategies of its neighbors (i.e., s17, s18, s23, s29 and s30), we designed its strategy based on new strategies.

As shown in Figure 5(c), most of the shortest paths to desirable states including the shortest paths designed by new strategies have a trend towards state s34 and s40 because these two states are the top 2 total profit states. Therefore, we can say that our proposed method gives a good performance in terms of suggesting optimal marketing campaigns and designing new campaigns.

Discussions

According to the literature of Reinforcement Learning and Q-learning (Sutton et al. 1998; Watkins 1992), the large volume of input data, which contains sufficient information in a real world such as customer's reaction (i.e. state transition, profit amount, defection probability) to all campaigns in each state, can assures convergence of $Q(s,a)$ value of all state-action sets. In our experiments, we have enough input datasets with 10,000 episodes which consist of a sequence of (state, action) compared to the number of states (i.e. less than 50) and actions (i.e. 11). Therefore, we could observe that the average $Q_P(s,a)$ and $Q_D(s,a)$ values converge as the number of episodes learned increase.

A common problem in performance evaluation of reinforcement learning methods is that it is difficult or impossible to conduct real life experiments in which the learning methods have access to on-line interactions (Abe et al. 2004). Our concern is also to validate our method with historical data sets. Because the past datasets were collected using some policy which is different from our optimal policy generated from our learning method. For further research, we have a plan to develop a sequential decision making method for suggesting a personalized campaign and evaluate it by connecting to a real-world company which often has thousands of campaigns. As the number of actions and states increase, we may need to define each state and action more sophisticatedly. For example, the transition of states during last 3 month such as a set of (s1→s4→s7) can be a new state and a sequence of conducted campaigns such as a set of (a7→a3→a9) can be a new action.

Our methodology is similar to current Multi-Criteria-Decision-Making (MCDM) which is a relatively developed, in terms of dealing with multiple conflicting criteria. However, a simple MCDM approach does not consider interactions among decision outcomes when sequences of marketing decisions are made over time. We think combining MCDM into Reinforcement Learning is one of the interesting research areas.

Conclusion

While direct marketing has garnered a great deal of attention, few studies have addressed the tradeoff between two conflicting objectives such as the profit and defection probability even though these tradeoffs are of great interest to companies. To solve this tradeoff conflict, we have developed a sequential decision-making methodology for profit maximization under the given defection probability constraint. Our method suggests sequential optimal marketing actions for maximizing long-term total profit while controlling the defection probability under the threshold over a customer's lifetime. In addition, the suggested marketing strategy map clearly shows an optimal action and customers' behavior dynamics in each state. It also helps a marketing manager identify sequential optimal campaigns and the shortest paths toward desirable states and, ultimately, a design for more effective campaigns. Our experiments demonstrate the feasibility of our proposed method in direct marketing. The proposed method will be a practical implementation procedure for direct marketing in telecommunications, online shopping malls, and other highly competitive marketplaces suffering from profit loss and customer defections.

Acknowledgements

We thank an associate editor and all reviewers for giving quality comments and recommendations which would develop our work.

References

Abe, N., Verma, N., Apte, C. and Schroko, R. "Cross Channel Optimization Marketing by Reinforcement Learning ", in *Proceedings of the 10th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (SIGKDD '04)*, ACM Press, Seattle WA, August 2004, pp. 767-772

Bhattacharyya S., "Evolutionary Algorithms in Data Mining: Multi-Objective Performance Modeling for Direct Marketing," in *Proceedings of the 6th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (SIGKDD '00)*, ACM Press, Boston, MA, USA, August 2000, pp.465-473.

Domingos, P. "MetaCost: A General Method for Making Classifiers Cost Sensitive", in *Proceedings of the 5th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (SIGKDD '99)*, ACM Press, San Diego, CA, August 1999, pp. 155-164.

Fan, W., Stolfo, S. J., Zhang, J. and Chan, P. K. "AdaCost: Misclassification Cost-Sensitive Boosting", in *Proceedings of the 16th Int'l Conf. Machine Learning (ICML '99)*, Morgan Kaufmann Publishers, Bled, Slovenia, June 1999, pp. 97-105.

Geibel, P. "Reinforcement Learning with Bounded Risk", in *Proceedings of the 18th Int'l Conf. Machine Learning (ICML '01)*, Morgan Kaufmann Publishers, Williamstown, MA, June 2001, pp. 162-169.

Hwang H., Jung T., and Suh E. "An LTV Model and Customer Segmentation based on Customer Value: A Case Study on The Wireless Telecommunication Industry," Expert Systems with Applications (26), 2004, pp.181-188.

Kohonen, T. *Self-Organizing and Associative Memory*, Berlin: Springer-Verlag, 1995.

Pednault, E., Abe, N. and Zadrozny, B. "Sequential Cost-Sensitive Decision Making with Reinforcement Learning", *in Proceedings of the 8th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (SIGKDD '02)*, ACM Press, Edmonton, Canada, July 2002, pp. 259-268.

Sutton, R. S. and Barto, A. G. *Reinforcement Learning: An Introduction*, MIT Press, 1998.

Wang, K., Zhou, S., and Yeung, J. M. S. "Mining Customer Value: From Association Rules to Direct Marketing," Data Mining and Knowledge Discovery (11:1), July 2005, pp.57-79.

Watkins, C. J. C. H., Dayan, P. "Q-learning", Machine Learning (8), 1992, pp. 279-292.

Zadrozny, B. and Elkan, C. "Learning and Making Decisions When Costs and Probabilities are Both Unknown", in *Proceedings of the 7th ACM SIDKDD Int'l Conf. Knowledge Discovery and Data Mining (SIGKDD '01*), ACM Press, San Francisco, CA, August 2001, pp. 204-213.

. Leave the footer untouched.