3. Markov Decision Process

3.1 Introduction

Consider a process that is observed at discrete time points to be in any one of M possible states, which we number by $1, 2, \ldots, M$. After observing the state of the process, an action must be chosen, and we let A, assumed finite, denote the set of all possible actions.

- If we let X_n denote the state of the process at time n and an the action chosen at time n, then the preceding is equivalent to stating that

$$P(X_{n+1} = j | X_0, a_0, X_1, a_1, \ldots X_n = i, a_n = a) = P_{ij}(a)$$

Thus, the transition probabilities are functions only of the present state and the subsequent action.

Policy

- By a *policy*, we mean a rule for choosing actions.

- We shall restrict ourselves to policies that are of the form that the action they prescribe at any time depends only on the state of the process at that time (and not on any information concerning prior states and actions).

- However, we shall allow the policy to be "randomized" in that its instructions may be to choose actions according to a probability distribution. In other words, a policy $\boldsymbol{\beta}$ is a set of numbers $\boldsymbol{\beta} = \{\beta_i(a), a \in A, i = 1, \ldots, M\}$ with the interpretation that if the process is in state i, then action a is to be chosen with probability $\beta_i(a)$. Of course, we need have

$$0 \leq \beta_i(a) \leq 1, \text{ for all } i, a \text{ and } \sum_a \beta_i(a) = 1, \text{ for all } i$$

- Under any given policy $\boldsymbol{\beta}$, the sequence of states $\{X_n, n = 0, 1, \ldots\}$ constitutes a Markov chain with transition probabilities $P_{ij}(\boldsymbol{\beta})$ given by

$$P_{ij} = P_{\boldsymbol{\beta}}\{X_{n+1} = j | X_n = i\} = \sum_a P_{ij}(a)\beta_i(a)$$

- Let us suppose that for every choice of a policy $\boldsymbol{\beta}$, the resultant Markov chain $\{X_n, n = 0, 1, \ldots\}$ is ergodic (An irreducible, positive recurrent, aperiodic Markov chain) .

- For any policy $\boldsymbol{\beta}$, let π_{ia} denote the limiting (or steady-state) probability that the process will be in state i and action a will be chosen if policy $\boldsymbol{\beta}$ is employed. That is,

$$\pi_{ia} = \lim_{n\to\infty} P_{\boldsymbol{\beta}}\{X_n = i, a_n = a\}$$

The vector $\pi = (\pi_{ia})$ must satisfy

(i) $\pi_{ia} \geq 0$ for all i, a,
(ii) $\sum_i \sum_a \pi_{ia} = 1$,
(iii) $\sum_a \pi_{ja} = \sum_i \sum_a \pi_{ia} P_{ij}(a)$ for all j.

- Notice
$$\beta_i(a) = P(\boldsymbol{\beta} \text{ choose } a | \text{state is } i) = \frac{\pi_{ia}}{\sum_a \pi_{ia}}$$

- Thus for any policy $\boldsymbol{\beta}$, there is a vector $\boldsymbol{\pi} = (\pi_{ia})$ that satisfies (i)–(iii) and with the interpretation that π_{ia} is equal to the steady-state probability of being in state i and choosing action a when policy $\boldsymbol{\beta}$ is employed.

- Moreover, it turns out that the reverse is also true. Namely, for any vector $\boldsymbol{\pi} = (\pi_{ia})$ that satisfies (i)–(iii), there exists a policy $\boldsymbol{\beta}$ such that if $\boldsymbol{\beta}$ is used, then the steady-state probability of being in i and choosing action a equals π_{ia}.

Reward

- The preceding is quite important in the determination of "optimal" policies. For instance, suppose that a reward $R(i, a)$ is earned whenever action a is chosen in state i. Since $R(X_i, a_i)$ would then represent the reward earned at time i, the expected average reward per unit time under policy $\boldsymbol{\beta}$ can be expressed as

$$\text{expected average reward under } \boldsymbol{\beta} = \lim_{n\to\infty} \mathrm{E}_{\boldsymbol{\beta}}\left[\frac{\sum_{i=1}^{n} R(X_i, a_i)}{n}\right]$$

- Now, if π_{ia} denotes the steady-state probability of being in state i and choosing action a, it follows that the limiting expected reward at time n equals

$$\lim_{n\to\infty} \mathrm{E}[R(X_n, a_n)] = \sum_i \sum_a \pi_{ia} R(i, a) = \text{expected average reward under } \boldsymbol{\beta}$$

Maximizing Reward and Optimal Policy

- Hence, the problem of determining the policy that maximizes the expected average reward is

$$\max_{\pi=(\pi_{ia})} \sum_i \sum_a \pi_{ia} R(i,a)$$

subject to $\pi_{ia} \geq 0$, for all i, a,

$$\sum_i \sum_a \pi_{ia} = 1, \text{and } \sum_a \pi_{ia} = \sum_i \sum_a \pi_{ia} P_{ij}(a) \text{ for all } j$$

- linear program and can be solved by a standard linear programming algorithm known as the simplex algorithm.

- If $\boldsymbol{\pi}^* = (\pi^*_{ia})$ maximizes the preceding, then the optimal policy will be given by $\boldsymbol{\beta}^*$ where

$$\beta^*_i(a) = \frac{\pi^*_{ia}}{\sum_a \pi^*_{ia}}$$

Remarks

(i) It can be shown that there is a π^* maximizing the above equation that has the property that for each i,π^*_{ia} is zero for all but one value of a, which implies that the optimal policy is nonrandomized. That is, the action it prescribes when in state i is a deterministic function of i.

(ii) The linear programming formulation also often works when there are restrictions placed on the class of allowable policies. For instance, suppose there is a restriction on the fraction of time the process spends in some state, say, state 1. Specifically, suppose that we are allowed to consider only policies having the property that their use results in the process being in state 1 less than 100α percent of time. To determine the optimal policy subject to this requirement, we add to the linear programming problem the additional constraint

$$\sum_a \pi_{1a} \leq \alpha$$

since $\sum_a \pi_{1a}$ represents the proportion of time that the process is in state 1.

3.2 Markov Decision Processes for Customer Lifetime Value

For more details in the practice, the process of Markov Decision Process can be also summarized as follows:

(i) At time t,a certain state i of the Markov chain is observed.

(ii) After the observation of the state, an action, let us say k, is taken from a set of possible decisions A_i. Different states may have different sets of possible actions.

(iii) An immediate gain (or loss) $q_i^{(k)}$ is then incurred according to the current state i and the action k taken.

(iv) The transition probabilities $p_{ji}^{(k)}$ are then affected by the action k.

(v) When the time parameter t increases, transition occurs again and the above steps (i)–(iv) repeat.

- A policy D is a rule of taking action. It prescribes all the decisions that should be made throughout the process. Given the current state i, the value of an optimal policy $v_i(t)$ is defined as the total expected gain obtained with t decisions or transitions remaining.

- For the case of one-period remaining, i.e. $t = 1$, the value of an optimal policy is given by

$$v_i(1) = \max_{k \in A_i} \{q_i^{(k)}\}$$

- For the case of two periods remaining, we have

$$v_i(2) = \max_{k \in A_i} \left\{ q_i^{(k)} + \alpha \sum_j p_{ji}^{(k)} v_j(1) \right\}$$

where α is the so called *discount factor*. Since the subsequent gain is associated with the transition probabilities which are affected by the action taken, an optimal policy should consider both the immediate and subsequent gain.

- The model can be easily extended to a more general situation, the process having n transitions remaining.

$$v_i(n) = \max_{k \in A_i} \left\{ q_i^{(k)} + \alpha \sum_j p_{ji}^{(k)} v_j(n-1) \right\}$$

From the above equation, the subsequent gain of $v_i(n)$ is defined as the expected value of $v_j(n-1)$.

- Since the number of transitions remaining is countable or finite, the process is called the discounted finite horizon MDP. For the infinite horizon MDP, the value of an optimal policy can be expressed as

$$v_i = \max_{k \in A_i} \left\{ q_i^{(k)} + \alpha \sum_j p_{ji}^{(k)} v_j \right\}$$

The finite horizon MDP is a dynamic programming problem and the infinite horizon MDP can be transformed into a linear programming problem.

Stationary Policy

A *stationary policy* is a policy where the decision depends only on the state the system is in and is independent of n.

- For instance, a stationary policy D prescribes the action $D(i)$ when the current state is i. Define $\bar{D}$ as the associated one-step- removed policy, then the value of policy $w_i(D)$ is defined as

$$w_i(D) = q_i^{D(i)} + \alpha \sum_j p_{ji}^{D(i)} w_j(\bar{D}).$$

- Given a Markov decision process with an infinite horizon and a discount factor $\alpha, 0 < \alpha < 1$, choose, for each i, an alternative k_i such that

$$\max_{k \in A_i} \left\{ q_i^{(k)} + \alpha \sum_j p_{ji}^{(k)} v_j \right\} = q_i^{(k_i)} + \alpha \sum_j p_{ji}^{k_i} v_j.$$

- Define the stationary policy D by $D(i) = k_i$. Then for each i, $w_i(D) = v_i$, i.e. the stationary policy is an optimal policy.

Example. We consider a nonine game company that plans to stay in business for 4 more years and then it will be closed without any salvage value. Each year, the volume of players only depends on the volume in the last year, and it is classified as either high or low. If a high volume of players occurs, the expected profit for the company will be 8 million dollars; but the profit drops to 4 million dollars when a low volume of players is encountered. At the end of every year, the profit of this year is collected, and then the company has the option to take certain actions that influence the performance of their service and hence the volume of players in the future may be altered. But some of these actions are costly so they reduce instant profit. To be more specific, the company can choose to: take no action, which costs nothing; perform only regular maintenance to the service system, which costs 1 million; or fully upgrade the service system, which costs 3 million. When the volume of players in the last year was high, it stays in the high state in the coming year with probability 0.4 if no action is taken; this probability is 0.8 if only regular maintenance is performed; and the probability rises to 1 if the system is fully upgraded. When the volume of players in the last year was low, then the probability that the player volume stays low is 0.9 with no action taken, 0.6 with regular maintenance, and 0.2 when the service system is fully upgraded. Assume the discount factor is 0.9 and that the company experienced a low volume of players last year. Determine the optimal (profit maximizing) strategy for the company, and Determine the optimal policy and the values for the Markov decision process. The parameters of this problem can be summarized in Table 1.1.

Table 1: A summary of the policy parameters

State i	Alternative k	$q_i^{(k)}$	$p_{i1}^{(k)}$	$p_{i2}^{(k)}$
1 (high volume)	1 (No action)	8	0.4	0.6
	2 (Regular Maintenance)	7	0.8	0.2
	3 (Fully Update)	5	1	0
2(low volume)	1(No action)	4	0.1	0.9
	2 (Regular Maintenance)	3	0.4	0.6
	3 (Fully Update)	1	0.8	0.2

Customer Life Time

- The customer equity should be measured in making the promotion plan so as to achieve an acceptable and reasonable budget. A popular approach is the Customer Life Time (CLV).

- A profitable customer is defined as "a person, household, or company whose revenues over time exceeds, by an acceptable amount, the company costs of attracting, selling, and servicing that customer." This excess is called the CLV. In some literature, CLV is also referred to as "customer equity"

- In this Lecute note, CLV is defined as the present value of the projected net cash flow that a firm expects to receive from the customer over time.

- Recognizing its importance in decision making, CLV has been successfully applied to the problems of pricing strategy, media selection and setting optimal promotion budget.

- To calculate the CLV, a company should estimate the expected net cash flow that they expect to receive from the customer over time. The CLV is the present value of that stream of cash flow.

- However, it is a difficult task to estimate the net cash flow to be received from the customer. In fact, one needs to answer, the following questions:

  1. How many customers one can attract given a specific advertising budget?
  2. What is the probability that the customer will stay with the company?
  3. How does this probability change with respect to the promotion budget?

  * To answer the first question, there are a number of advertising models found.

  * The second and the third questions give rise to an important concept, the *retention rate*. The retention rate is defined as "the chance that the account will remain with the vendor for the next purchase, provided that the customer has bought from the vendor on each previous purchase".

A formula for calculation of CLV

The model is simple and deterministic. The CLV is the sum of two net present values: the return from acquisition spending and the return from retention spending. In the model, CLV is defined as

$$\begin{aligned} \text{CLV} &= \underbrace{am - A}_{\text{acquisition}} + \underbrace{\sum_{k=1}^{\infty} a(m - \frac{R}{r})[r(1+d)^{-1}]^k}_{retention} \\ &= am - A + a(m - \frac{R}{r}) \times \frac{r}{1+d-r} \end{aligned}$$

where a is the acquisition rate, A is the level of acquisition spending, m is the margin on a transaction, R is the retention spending per customer per year, r is the yearly retention rate (a proportion), and d is the yearly discount rate appropriate for marketing investment. The acquisition rate a and retention rate r are functions of A and R respectively, and are given by

$$a(A) = a_0(1 - e^{-K_1 A}) \quad \text{and} \quad r(R) = r_0(1 - e^{-K_2 R})$$

where a_0 and r_0 are the estimated ceiling rates and K_1 and K_2 are two positive constants.

The CLV of the customers for three different scenarios are considered in this lecture note:

1. infinite horizon without constraint (without limit in the number of promotions),

2. finite horizon (with a limited number of promotions), and

3. infinite horizon with constraints (with a limited number of promotions).

Markov Chain Models for Customer Behavior

- According to the usage of the customer, a company customer can be classified into N possible states

$$\{0, 1, 2, \ldots, N-1\}$$

  For example, customers can be classified into four states (N D 4): low-volume user (state 1), medium-volume user (state 2) and high-volume user (state 3) and in order to classify all customers in the market, state 0 is introduced.

- A customer is said to be in state 0 if they are either a customer of the competitor company or they did not purchase the service during the period of observation.

- Therefore at any time, a customer in the market belongs to exactly one of the states in $\{0, 1, 2, \ldots, N\}$. With this notation, a Markov chain is a good choice to model the transitions of customers among the states in the market.

- A markov chain model is characterized by an $N \times N$ transition probability matrix P. Here $P_{ij}(i, j = 0, 1, 2 \ldots, N-1)$ is the transition probability that a customer will move to state i in the next period given that currently they are in state j .

- The retention probability of a customer in state $i(i = 0, 1, 2 \ldots, N-1)$ is given by P_{ii}.

- If the underlying Markov chain is assumed to be irreducible then the stationary distribution $\mathbf{p}$ exists. This means that there is a unique

$$\mathbf{p} = (p_0, p_1, \ldots, P_{N-1})^T$$

such that

$$\mathbf{p} = P\mathbf{p}, \quad \sum_{i=0}^{N-1} p_i = 1, \quad p_i \geq 0.$$

- By making use of the stationary distribution $\mathbf{p}$, one can compute the retention probability of a customer as follows

$$\sum_{i=1}^{N-1}\left(\frac{p_i}{\sum_{j=1}^{N-1}p_j}\right)(1-P_{i0}) = 1-\frac{1}{1-p_0}\sum_{i=1}^{N-1}p_iP_{0i} = 1-\frac{p_0(1-P_{00})}{1-p_0}.$$

  This is the probability that a customer will purchase service with the company in the next period.

- Apart from the retention probability, the Markov model can also help us in computing the CLV. In this case c_i is defined to be the revenue obtained from a customer in state i . Then the expected revenue is given by

$$\sum_{i=0}^{N-1} c_ip_i.$$

- The above retention probability and the expected revenue are computed under the assumption that the company makes no promotion (in a non-competitive environment) throughout the period. The transition probability matrix P can be significantly different when there is a promotion made by the company.

- When promotions are allowed, what is the best promotion strategy such that the expected revenue is maximized? Similarly, what is the best strategy when there is a fixed budget for the promotions, e.g. the number of promotions is fixed?

Estimation of the Transition Probabilities: An example in the computer service company

- In the captured database of customers, each customer has four important attributes (A, B, C, D): A is the "Customer Number", each customer has a unique identity number. B is the "Week", the time (week) when the data was captured. C is the "Revenue" which is the total amount of money the customer spent in the captured week. D is the "Hour", the number of hours that the customer consumed in the captured week.

- The total number of weeks of data available is 20. Among these 20 weeks, the company has a promotion for 8 consecutive weeks and no promotion for the other 12 consecutive weeks.

- For each week, all the customers are classified into four states 0,1,2,3, according to the amount of "hours" consumed, see Table 2. We recall that a customer is said to be in state 0, if they are a customer of a competitor company or they did not use the service for the whole week.

Table 2: The four classes of customer

State	0	1	2	3
Hours	0.00	1-20	21-40	>40

- From the data, one can estimate two transition probability matrices, one for the promotion period (8 consecutive weeks) and the other one for the no-promotion period (12 consecutive weeks). For each period, the number of customers switching from state i to state j is recorded. Then, divide this number the total number of customers in the state i, and one obtains the estimates for the one-step transition probabilities.

- Hence the transition probability matrices under the promotion period $P^{(2)}$ and the no-promotion period $P^{(2)}$ are given respectively below:

$$P^{(1)} = \begin{pmatrix} 0.8054 & 0.4163 & 0.2285 & 0.1372 \\ 0.1489 & 0.4230 & 0.3458 & 0.2147 \\ 0.0266 & 0.0992 & 0.2109 & 0.2034 \\ 0.0191 & 0.0615 & 0.2148 & 0.4447 \end{pmatrix}$$

and

$$P^{(2)} = \begin{pmatrix} 0.8762 & 0.4964 & 0.3261 & 0.2380 \\ 0.1064 & 0.4146 & 0.3837 & 0.2742 \\ 0.0121 & 0.0623 & 0.1744 & 0.2079 \\ 0.0053 & 0.0267 & 0.1158 & 0.2809 \end{pmatrix}$$

Retention Probability and CLV

- The stationary distributions of the two Markov chains having transition probability matrices $\mathbf{p}^{(1)}$ and $\mathbf{p}^{(1)}$ are given respectively by

$$\mathbf{p}^{(1)} = (0.2306, 0.0691, 0.0738, 0.6265)^T$$

and

$$\mathbf{p}^{(2)} = (0.1692, 0.0285, 0.0167, 0.7856)^T$$

- The retention probabilities in the promotion period and no-promotion period are given respectively by 0.6736 and 0.5461.

- the expected revenue from a customer in the promotion period (assume that the only promotion cost is the discount rate) and no-promotion period are given by 2.42 and 17.09 respectively

Table 3: The average revenue of the four classes of customers

State	0	1	2	3
Promotion	0.00	6.97	18.09	43.75
No-promotion	0.00	14.03	51.72	139.20

- Although one can obtain the CLVs of the customers in the promotion period and the no-promotion period, one would expect to calculate the CLV in a mixture of promotion and no-promotion periods.

- This is especially true when the promotion budget is limited (the number of promotions is fixed) and one would like to obtain the optimal promotion strategy.

- Stochastic dynamic programming with Markov process provides a good approach for solving the above problems.

- Moreover, the optimal stationary strategy for the customers in different states can also be obtained by solving the stochastic dynamic programming problem.

Stochastic Dynamic Programming Models

In this section, stochastic dynamic programming models are presented for maximizing the CLV under an optimal promotion strategy. The notations of the model are given as follows:

1. N, the total number of states (indexed by $i = 0, 1, \ldots, N-1$)
2. A_i, the set containing all the actions in state i (indexed by k)
3. T, number of months remaining in the plan horizon (indexed by $t = 1, \ldots, T$)
4. d_k, the resource required for carrying out the action k in each period.
5. $c_i^{(k)}$, the revenue obtained from a customer in state i with the action k in each period.
6. $p_{ij}^{(k)}$, the transition probability for a customer moving from state j to state i under the action k in each period.
7. α, discount rate.

The recursive relation for maximizing the revenue is given as follows:

$$v_i(t) = \max_{k \in A_i} \left\{ c_i^{(k)} - d_k + \sum_{j=0}^{N-1} p_{ji}^{(k)} v_j(t-1) \right\}.$$

Infinite Horizon Without Constraints

From the standard results in stochastic dynamic programming, for each i, the optimal values v_i for the discounted infinite horizon Markov decision process satisfy the relationship

$$v_i = \max_{k \in A_i} \left\{ c_i^{(k)} - d_k + \sum_{j=0}^{N-1} p_{ji}^{(k)} v_j \right\}.$$

Therefore we have

$$v_i \geq c_i^{(k)} - d_k + \alpha \sum_{j=0}^{N-1} p_{ji}^{(k)} v_j$$

for each i. This suggests that the problem of determining the v_i value can be transformed into the following linear programming problem

$$\begin{cases} \min x_0 = \sum_{i=0}^{N-1} v_i \\ \text{subject to} \\ v_i \geq c_i^{(k)} - d_k + \alpha \sum_{j=0}^{N-1} p_{ji}^{(k)} v_j, \quad \text{for } i = 0, \ldots, N-1; \\ v_i \geq 0, \quad \text{for } i = 0, \ldots, N-1. \end{cases}$$

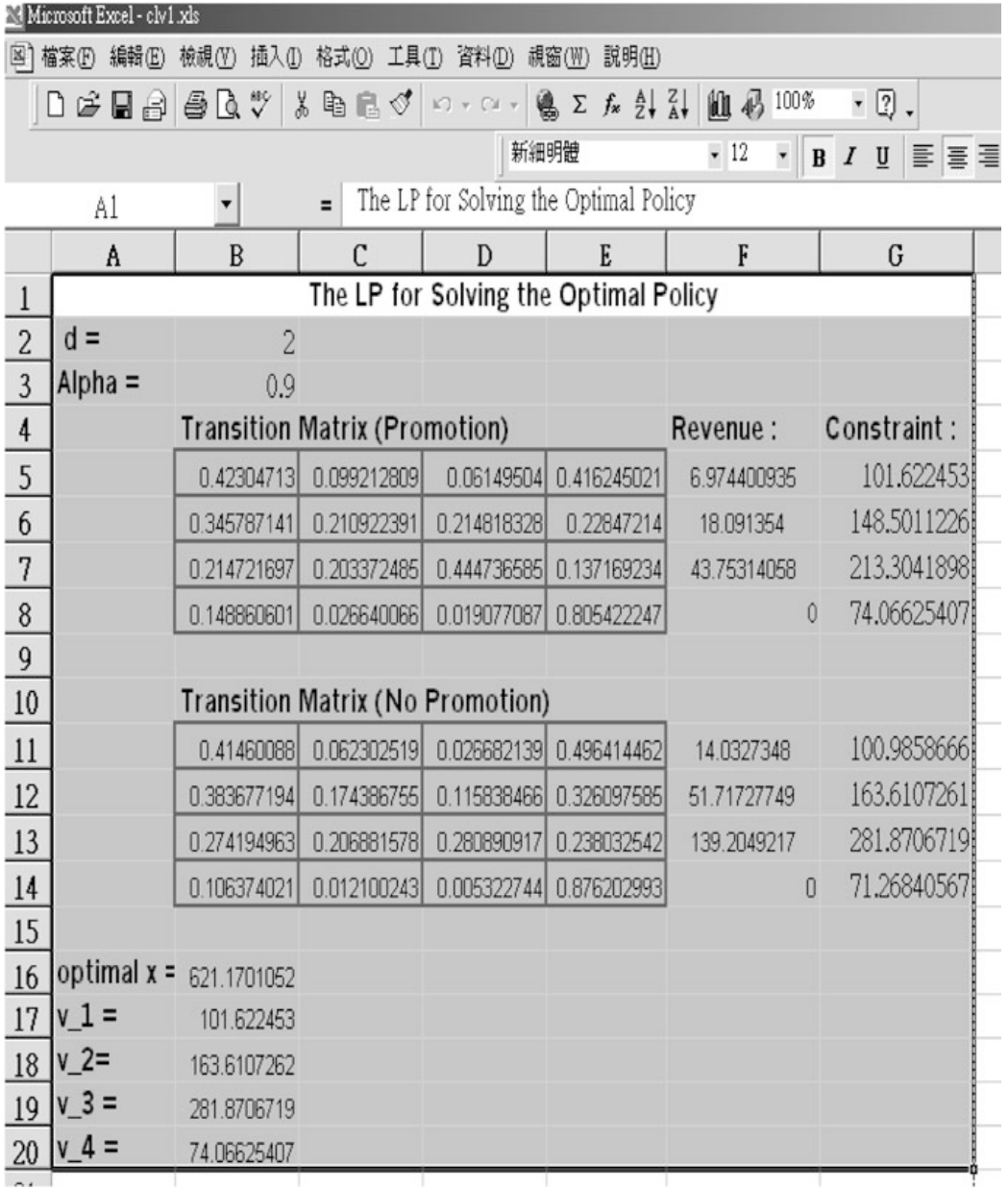
- The above linear programming problem can be solved easily by using speadsheet software such as EXCEL.

- Returning to the model for the computer service company, there are 2 actions available (either (P) promotion or $(NP\ )$ no-promotion) for all possible states. Thus $A_{=}\{P, NP\}$ for all $i = 0, \ldots, N-1$. Moreover, customers are classified into 4 clusters, thus $N = 4$ (possible states of a customer are 0,1,2,3).

- Since no promotion cost is incurred for the action (NP), therefore $d_{NP} = 0$. For simplification, d is used to denote the only promotion cost instead of d_P in the application.

- Table 4 presents the optimal stationary policies(i.e.,to have promotion of $D_i = P$ or no-promotion $D_i = NP$ depends on the state i of the customer) and the corresponding revenues for different discount factors α and fixed promotion costs d .

Figure 1: For solving infinite horizon without constraint



Microsoft Excel - clv1.xls

檔案(F) 編輯(E) 檢視(V) 插入(I) 格式(O) 工具(T) 資料(D) 視窗(W) 說明(H)

新細明體 12 B I U

A1 = The LP for Solving the Optimal Policy

	A	B	C	D	E	F	G
1	The LP for Solving the Optimal Policy						
2	d =	2					
3	Alpha =	0.9					
4		Transition Matrix (Promotion)				Revenue :	Constraint :
5		0.42304713	0.099212809	0.06149504	0.416245021	6.974400935	101.622453
6		0.345787141	0.210922391	0.214818328	0.22847214	18.091354	148.5011226
7		0.214721697	0.203372485	0.444736585	0.137169234	43.75314058	213.3041898
8		0.148860601	0.026640066	0.019077087	0.805422247	0	74.06625407
9							
10		Transition Matrix (No Promotion)					
11		0.41460088	0.062302519	0.026682139	0.496414462	14.0327348	100.9858666
12		0.383677194	0.174386755	0.115838466	0.326097585	51.71727749	163.6107261
13		0.274194963	0.206881578	0.280890917	0.238032542	139.2049217	281.8706719
14		0.106374021	0.012100243	0.005322744	0.876202993	0	71.26840567
15							
16	optimal x =	621.1701052					
17	v_1 =	101.622453					
18	v_2=	163.6107262					
19	v_3 =	281.8706719					
20	v_4 =	74.06625407					

Table 4: Optimal stationary policies and their CLVs

	$d=0$			$d=1$			$d=2$		
α	0.99	0.95	0.90	0.99	0.95	0.90	0.99	0.95	0.90
x_0	4791	1149	687	4437	1080	654	4083	1012	621
v_0	1112	204	92	1023	186	83	934	168	74
v_1	1144	234	119	1054	216	110	965	198	101
v_2	1206	295	179	1118	278	171	1030	261	163
v_3	1328	415	296	1240	399	289	1153	382	281
D_0	P	P	P	P	P	P	P	P	P
D_1	P	P	P	P	P	P	P	P	P
D_2	NP	NP	NP	NP	NP	NP	NP	NP	NP
D_3	NP	NP	NP	NP	NP	NP	NP	NP	NP
	$d=3$			$d=4$			$d=5$		
α	0.99	0.95	0.90	0.99	0.95	0.90	0.99	0.95	0.90
x_0	3729	943	590	3375	879	566	3056	827	541
v_0	845	151	65	755	134	58	675	119	51
v_1	877	181	94	788	164	88	707	151	82
v_2	942	245	156	854	230	151	775	217	145
v_3	1066	366	275	978	351	269	899	339	264
D_0	P	P	P	P	P	P	P	P	P
D_1	P	P	NP	P	NP	NP	NP	NP	NP
D_2	NP	NP	NP	NP	NP	NP	NP	NP	NP
D_3	NP	NP	NP	NP	NP	NP	NP	NP	NP

From the numerical examples, the following conclusions are drawn:

1. When the fixed promotion cost d is large, the optimal strategy is that the company should not conduct any promotion on the active customers and should only conduct the promotion scheme to inactive (purchase no service) customers and customers of the competitor company. However, when d is small, the company should take care of the low-volume customers to prevent this group of customers from switching to the competitor companies.

2. It is also clear that the CLV of a high-volume user is larger than the CLV of other groups.

3. The CLVs of each group depend on the discount rate significantly. Here the discount rate can be viewed as the technology depreciation of the computer services in the company. Therefore, in order to generate the revenue of the company, new technology and services should be provided.

Finite Horizon with Hard Constrains

In the computer service and telecommunication industry, the product life cycle is short, e.g., it is usually one year. Therefore, the case of finite horizon with limited budget constraint is considered. The model?s parameters are defined as follows:

- $n =$ number of weeks remaining
- $p =$ number of possible promotions remaining

The recursive relation for the problem is given as follows

$$v_i(n,p) = \max \quad \{ \quad c_i^{(p)} - d_p + \alpha \sum_{j=0}^{N-1} p_{ji}^{(p)} v_j(n-1, p-1),$$

$$c_i^{(NP)} - d_{NP} + \alpha \sum_{j=0}^{N-1} p_{ji}^{(NP)} v_j(n-1, p)\}$$

for $n = 1, \ldots, n_{\max}$ and $p = 1, \ldots, p_{\max}$ and

$$v_{n,0} = c_i^{(NP)} - d_{NP} + \alpha \sum_{j=0}^{N-1} p_{ji}^{(NP)} v_j(n-1, 0)$$

for $n = 1, \ldots, n_{\max}$.

The above dynamic programming problem can be solved easily by using spreadsheet software such as EXCEL.

In the numerical experiment of the computer service company,the length of the planning period is set to be $n_{\max} = 52$ and the maximum number of promotions is $p_{\max} = 4$.By solving the dynamic programming problem, the optimal values and promotion strategies are listed in Table 4. The optimal solution in the table is presented as follows:

$$(t_1, t_2, t_3, t_4, r^*)$$

where r^* is the optimal expected revenue, and t_i is the promotion week of optimal promotion strategy and "-" means no promotion. Findings are summarized as follows

- For different values of the fixed promotion cost d, the optimal strategy for the customers in states 2 and 3 is to conduct no promotion.
- For those in state 0, the optimal strategy is to conduct all four promotions as early as possible.
- In state 1, the optimal strategy depends on the value of d . If d is large, then no promotion will be conducted. However, if d is small, promotions are carried out and the strategy is to conduct the promotions as late as possible.

Table 5: Optimal promotion strategies and their CLVs

	α	State 0	State 1	State 2	State 3
	0.9	(1, 2, 3, 4, 67)	(1, 45, 50, 52, 95)	(-,-,-,-,158)	(-,-,-,-,276)
$d = 0$	0.95	(1, 2, 3, 4, 138)	(45, 48, 50, 51, 169)	(-,-,-,-,234)	(-,-,-,-,335)
	0.99	(1, 2, 3, 4, 929)	(47, 49, 50, 51, 963)	(-,-,-,-,1031)	(-,-,-,-,1155)
	0.9	(1, 2, 3, 4, 64)	(47, 49, 51, 52, 92)	(-,-,-,-,155)	(-,-,-,-,274)
$d = 1$	0.95	(1, 2, 3, 4, 133)	(47, 49, 51, 52, 164)	(-,-,-,-,230)	(-,-,-,-,351)
	0.99	(1, 2, 3, 4, 872)	(47, 49, 51, 52, 906)	(-,-,-,-,974)	(-,-,-,-,1098)
	0.9	(1, 2, 3, 4, 60)	(49, 50, 51, 52, 89)	(-,-,-,-,152)	(-,-,-,-,271)
$d = 2$	0.95	(1, 2, 3, 4, 128)	(48, 50, 51, 52, 160)	(-,-,-,-,225)	(-,-,-,-,347)
	0.99	(1, 2, 3, 4, 815)	(48, 49, 51, 52, 849)	(-,-,-,-,917)	(-,-,-,-,1041)
	0.9	(1, 2, 3, 4, 60)	(−, −, −, −, 87)	(-,-,-,-,150)	(-,-,-,-,269)
$d = 3$	0.95	(1, 2, 3, 4, 123)	(49, 50, 51, 52, 155)	(-,-,-,-,221)	(-,-,-,-,342)
	0.99	(1, 2, 3, 4, 758)	(48, 50, 51, 52, 792)	(-,-,-,-,860)	(-,-,-,-,984)
	0.9	(1, 2, 3, 4, 54)	(−, −, −, −, 84)	(-,-,-,-,147)	(-,-,-,-,266)
$d = 4$	0.95	(1, 2, 3, 4, 119)	(−, −, −, −, 151)	(-,-,-,-,217)	(-,-,-,-,338)
	0.99	(1, 2, 3, 4, 701)	(49, 50, 51, 52, 736)	(-,-,-,-,804)	(-,-,-,-,928)
	0.9	(1, 2, 3, 4, 50)	(-,-,-,-,81)	(-,-,-,-,144)	(-,-,-,-,264)
$d = 5$	0.95	(1, 2, 3, 4, 114)	(-,-,-,-,147)	(-,-,-,-,212)	(-,-,-,-,334)
	0.99	(1, 2, 3, 4, 650)	(-,-,-,-,684)	(-,-,-,-,752)	(-,-,-,-,876)

Infinite Horizon with Constrains

For the purpose of comparison, the model is extended to the infinite horizon case. Similar to the previous model, the finite number of promotions available is denoted by $p_{\max}$. The value function $v_i(p)$, which represents the optimal discounted utility starting at state i when there are p promotions remaining, is the unique fixed point of the equations:

$$v_i(p) = \max \left\{ c_i^{(p)} - d_p + \alpha \sum_{j=0}^{N-1} p_{ji}^{(p)} v_j(p-1), c_i^{(NP)} - d_{NP} + \alpha \sum_{j=0}^{N-1} p_{ji}^{(NP)} v_j(p) \right\}$$

for $p = 1, \ldots, p_{\max}$, and

$$v_i(0) = c_i^{(NP)} - d_{NP} + \alpha \sum_{j=0}^{N-1} p_{ji}^{(NP)} v_j(0).$$

It can be solved by using the linear programming approach:

$$
\begin{cases}
\min x_0 = \sum_{i=0}^{N-1} \sum_{p=1}^{p_{\max}} v_i(p) \\
\text{subject to} \\
v_i(p) \geq c_i^{(p)} - d_1 + \alpha \sum_{j=0}^{N-1} p_{ji}^{(p)} v_j(p-1), \\
\text{for } i = 0, \ldots, N-1, p = 1, \ldots, p_{\max}; \\
v_i(p) \geq c_i^{(NP)} - d_2 + \alpha \sum_{j=0}^{N-1} p_{ji}^{(NP)} v_j(p), \\
\text{for } i = 0, \ldots, N-1, p = 1, \ldots, p_{\max}.
\end{cases}
$$

We note that $v_i(0)$ is not included in the linear programming constraints and the objective function; $v_i(0)$ is solved beforehand using the equation for $v_i(0)$.

- Tables 6 and 7 give the optimal values and promotion strategies of the computer service company. For instance, when the promotion cost is 0 and the discount factor is 0.99, then the optimal strategy is that when the current state is 1, 2 or 3, the promotion should be done when there are some available promotions, i.e.,

$$D_1(p) = D_2(p) = D_3(p) = P \text{ for } p = 1, 2, 3, 4$$

  and when the current state is 0, no promotion is required, i.e. $D_0(p) = NP$ for $p = 1, 2, 3, 4$.

- The optimal strategy for the customers in states 1, 2 and 3 is to conduct no promotion.

- Moreover, it is not affected by the promotion cost and the discount factor. These results are slightly different from those for the finite horizon case.

- However, the optimal strategy is to conduct all the four promotions to customers in state 0 as early as possible.

Table 6: Optimal promotion strategies and their CLVs

	$d=0$			$d=1$			$d=2$		
α	0.99	0.95	0.90	0.99	0.95	0.90	0.99	0.95	0.90
x_0	11355	3378	2306	11320	3344	2277	11277	3310	2248
$v_0(1)$	610	117	55	609	116	54	608	115	53
$v_1(1)$	645	149	85	644	148	84	643	147	84
$v_2(1)$	713	215	149	712	214	148	711	213	147
$v_3(1)$	837	337	267	836	336	267	845	335	266
$v_0(2)$	616	122	60	614	120	58	612	118	56
$v_1(2)$	650	154	89	648	152	87	647	150	86
$v_2(2)$	718	219	152	716	218	151	714	216	149
$v_3(2)$	842	341	271	840	339	269	839	338	268
$v_1(3)$	656	158	92	654	156	90	650	153	88
$v_2(3)$	724	224	155	722	221	153	718	219	151
$v_3(3)$	848	345	273	846	343	271	842	340	270
$v_0(4)$	628	131	67	624	128	63	620	124	60
$v_1(4)$	662	162	95	658	159	92	654	158	89
$v_2(4)$	730	228	157	726	225	155	722	221	152
$v_3(4)$	854	349	276	850	346	273	846	343	271
$D_0(1)$	P	P	P	P	P	P	P	P	P
$D_1(1)$	NP	NP	NP	NP	NP	NP	NP	NP	NP
$D_2(1)$	NP	NP	NP	NP	NP	NP	NP	NP	NP
$D_3(1)$	NP	NP	NP	NP	NP	NP	NP	NP	NP
$D_0(2)$	P	P	P	P	P	P	P	P	P
$D_1(2)$	NP	NP	NP	NP	NP	NP	NP	NP	NP
$D_2(2)$	NP	NP	NP	NP	NP	NP	NP	NP	NP
$D_3(2)$	NP	NP	NP	NP	NP	NP	NP	NP	NP
$D_0(3)$	P	P	P	P	P	P	P	P	P
$D_1(3)$	NP	NP	NP	NP	NP	NP	NP	NP	NP
$D_3(3)$	NP	NP	NP	NP	NP	NP	NP	NP	NP
$D_0(4)$	P	P	P	P	P	P	P	P	P
$D_1(4)$	NP	NP	NP	NP	NP	NP	NP	NP	NP
$D_2(4)$	NP	NP	NP	NP	NP	NP	NP	NP	NP
$D_3(4)$	NP	NP	NP	NP	NP	NP	NP	NP	NP

Table 7: Optimal promotion strategies policies and their CLVs

	$d=3$			$d=4$			$d=5$		
α	0.99	0.95	0.90	0.99	0.95	0.90	0.99	0.95	0.90
x_0	11239	3276	2218	11200	3242	2189	11161	3208	2163
$v_0(1)$	607	114	52	606	113	51	605	112	50
$v_1(1)$	641	146	83	641	146	82	640	145	81
$v_2(1)$	710	212	146	709	211	145	708	211	145
$v_3(1)$	834	334	265	833	333	264	832	332	264
$v_0(2)$	610	116	54	608	114	52	606	112	50
$v_1(2)$	645	149	84	643	147	83	641	145	81
$v_2(2)$	713	214	148	711	213	146	709	211	145
$v_3(2)$	837	336	266	835	334	265	833	333	264
$v_0(3)$	613	119	56	610	116	53	607	113	50
$v_1(3)$	647	151	86	645	148	83	642	146	81
$v_2(3)$	715	216	149	713	214	147	710	211	145
$v_3(3)$	839	338	268	837	336	266	834	333	264
$v_0(4)$	616	121	57	612	117	54	608	113	50
$v_1(4)$	650	152	87	646	149	84	643	146	81
$v_2(4)$	718	218	150	714	215	147	711	212	145
$v_3(4)$	842	340	269	838	337	266	835	334	265
$D_0(1)$	P	P	P	P	P	P	P	P	P
$D_1(1)$	NP	NP	NP	NP	NP	NP	NP	NP	NP
$D_2(1)$	NP	NP	NP	NP	NP	NP	NP	NP	NP
$D_3(1)$	NP	NP	NP	NP	NP	NP	NP	NP	NP
$D_0(2)$	P	P	P	P	P	P	P	P	P
$D_1(2)$	NP	NP	NP	NP	NP	NP	NP	NP	NP
$D_2(2)$	NP	NP	NP	NP	NP	NP	NP	NP	NP
$D_3(2)$	NP	NP	NP	NP	NP	NP	NP	NP	NP
$D_0(3)$	P	P	P	P	P	P	P	P	P
$D_1(3)$	NP	NP	NP	NP	NP	NP	NP	NP	NP
$D_2(3)$	NP	NP	NP	NP	NP	NP	NP	NP	NP
$D_3(3)$	NP	NP	NP	NP	NP	NP	NP	NP	NP
$D_0(4)$	P	P	P	P	P	P	P	P	P
$D_1(4)$	NP	NP	NP	NP	NP	NP	NP	NP	NP
$D_2(4)$	NP	NP	NP	NP	NP	NP	NP	NP	NP
$D_3(4)$	NP	NP	NP	NP	NP	NP	NP	NP	NP

Extensions

- In the previous discussions, the problem under consideration is to decide whether to offer the promotion at the start of each time unit with the assumption that the promotion only lasts for a single time unit. The analysis is extended to consider multi-period promotions. A multi-period promotion refers to a promotion that lasts for $2, 3, \ldots, R$ time units. This encourages more purchases or continuous subscriptions than a single-period promotion does.

- The MDP presented in the previous section is a first-order type, i.e., the transition probabilities depend on the current state only. For the Higher-order Markov Decision Process (HMDP), the transition probabilities depend on the current state and a number of previous states.