Управляемые марковские процессы поддержки принятия решений

Введение

Случайный процесс с дискретными значениями x_n называют *марковским*, если он обладает свойством: для любого $n \geq 1$ ($n = 1, 2, \ldots$) и любых возможных значений $i_0, i_1, \ldots, i_{n-1}$ должно выполняться следующее требование для условных вероятностей:
$p(x_n = i_n / x_0 = i_0, x_1 = i_1, \ldots, x_{n-1} = i_{n-1}) = p(x_n = i_n / x_{n-1} = i_{n-1})$.

Значения i_k, которые принимает марковский процесс, можно назвать его внутренними состояниями. Они ни в коем случае не являются его «состояниями» в терминологии динамических систем. Однако далее для краткости изложения вместо термина «внутреннее состояние» в некоторых случаях будем говорить просто о состоянии, опуская слово «внутреннее».

Если вероятность $p(x_n = i_n / x_{n-1} = i_{n-1})$ перехода из состояния i_{n-1} в состояние i_n не зависит от момента времени n, *марковский процесс* называется *стационарным*. В последнем случае случайный процесс перехода из одного состояния в другое на каждом шаге описывается одной и той же стохастической матрицей $P = \left(p_{ij}\right)$, элементы которой p_{ij} являются условными вероятностями того, что следующим состоянием будет состояние j, если текущим состоянием является состояние i. Эти вероятности удовлетворяют двум условиям: $p_{ij} \geq 0$ и $\sum_{j=1}^{m} p_{ij} = 1$, (если число возможных состояний равно m). Марковский процесс с конечным числом внутренних состояний называют *конечной марковской цепью*.

Рассмотрим марковскую цепь с m внутренними состояниями, вероятности нахождения в которых в момент времени n заданы вектором-строкой $p(n) = \left(p_1(n), p_2(n), \ldots, p_m(n)\right)$. Вектор $p(n)$ описывает текущее вероятностное распределение марковской цепи по ее внутренним состояниям. В силу однозначности определения $p(n+1)$ по $p(n)$ вектор вероятностного распределения будет являться состоянием марковской цепи как динамической системы. Оператор изменения этих вероятностей задается стохастической матрицей P: $p(n+1) = p(n)P$.

Сделаем теперь цепь управляемой за счет того, что матрица вероятностей переходов P будет зависеть от некоторой стратегии-управления k ($P^{(k)}$). Предположим, что при каждом внутреннем состоянии цепи мы имеем возможность выбирать одну из K стратегий, задаваемых стохастическими матрицами $P^{(k)}$, $k=1,...,K$. Каждой матрице $P^{(k)}$ сопоставим матрицу доходов $D^{(k)}$ так, что при выборе стратегии k математическое ожидание дохода $q_i^{(k)}$, связанного с попаданием во внутреннее состояние i за один шаг, будет равно

$q_i^{(k)} = p_{i1}^{(k)} d_{i1}^{(k)} + p_{i2}^{(k)} d_{i2}^{(k)} + ... + p_{im}^{(k)} d_{im}^{(k)}$.

Обозначим через $V_n(i)$ максимально возможное математическое ожидание дохода за n шагов, если начальное внутреннее состояние системы было i. Тогда в соответствии с принципом оптимальности мы получим рекуррентное соотношение, являющееся аналогом уравнения Р.Беллмана

$$V_n(i) = \max\left\{ q_i^{(k)} + \sum_{j=1}^{m} p_{ij}^{(k)} V_{n-1}(j) \colon \, k \in \{1,...,K\} \right\}. \qquad (3.4)$$

Функция $V_n(i)$ играет роль функции Беллмана.

Пример 2.1. Задача об игрушечных дел мастере [7]

Игрушечных дел мастер в течение недели изготавливает игрушки, а в воскресенье выходит на рынок, чтобы их продать. Вероятности успешной или неуспешной продажи, а также величины доходов в зависимости от результата предыдущего раунда заданы матрицами

$$P^{(1)} = \begin{pmatrix} 0,5 & 0,5 \\ 0,4 & 0,6 \end{pmatrix},\ P^{(2)} = \begin{pmatrix} 0,6 & 0,4 \\ 0,5 & 0,5 \end{pmatrix},\ P^{(3)} = \begin{pmatrix} 0,7 & 0,3 \\ 0,6 & 0,4 \end{pmatrix};$$

$$D^{(1)} = \begin{pmatrix} 9 & 3 \\ 3 & -7 \end{pmatrix},\ D^{(2)} = \begin{pmatrix} 8 & 2 \\ 1 & -8 \end{pmatrix},\ D^{(3)} = \begin{pmatrix} 6 & 1 \\ 0 & -10 \end{pmatrix}.$$

Первая стратегия соответствует отсутствию рекламы, вторая стратегия соответствует рекламе по радио, третья стратегия соответствует рекламе по телевидению. Требуется определить оптимальную стратегию в смысле максимума математического ожидания дохода на несколько шагов вперед. Пусть $V_0(1) = V_0(2) = 0$. Тогда рекуррентное соотношение (3.4) позволяет нам найти оптимальную стратегию поведения $\left(k_1(1), k_2(1)\right)$ в расчете на один шаг:

$$V_1(1) = \max\begin{Bmatrix} 0,5\cdot 9 + 0,5\cdot 3 = 6 \\ 0,6\cdot 8 + 0,4\cdot 2 = 5,6 \\ 0,7\cdot 6 + 0,3\cdot 1 = 4,5 \end{Bmatrix},\ V_1(2) = \max\begin{Bmatrix} 0,4\cdot 3 + 0,6\cdot(-7) = -3 \\ 0,5\cdot 1 + 0,5\cdot(-8) = -3,5 \\ 0,6\cdot 0 + 0,4\cdot(-10) = -4 \end{Bmatrix}.$$

Итак, оптимальная стратегия поведения ($k_1(1), k_2(1)$) = (1;1) в расчете на один шаг, при этом $V_1(1)=6,\ V_1(2)=-3$. Теперь подсчитаем оптимальную стратегию поведения ($k_1(2), k_2(2)$) в расчете на два шага:

$$V_2(1)=\max\begin{Bmatrix}6+0,5\cdot 6+0,5\cdot(-3)=7,5\\ 5,6+0,6\cdot 6+0,4\cdot(-3)=8\\ 4,5+0,7\cdot 6+0,3\cdot(-3)=7,8\end{Bmatrix},$$
$$V_2(2)=\max\begin{Bmatrix}-3+0,4\cdot 6+0,6\cdot(-3)=-2,4\\ -3,5+0,5\cdot 6+0,5\cdot(-3)=-2\\ -4+0,6\cdot 6+0,4\cdot(-3)=-1,6\end{Bmatrix}.$$

Итак, в расчете на два шага оптимальная стратегия поведения ($k_1(2), k_2(2)$) = (2;3) , при этом $V_2(1)=8,\ V_2(2)=-1,6$. Теперь подсчитаем оптимальную стратегию поведения ($k_1(3), k_2(3)$) в расчете на три шага:

$$V_3(1)=\max\begin{Bmatrix}6+0,5\cdot 8+0,5\cdot(-1,6)=9,2\\ 5,6+0,6\cdot 8+0,4\cdot(-1,6)=9,76\\ 4,5+0,7\cdot 8+0,3\cdot(-1,6)=9,62\end{Bmatrix},$$
$$V_3(2)=\max\begin{Bmatrix}-3+0,4\cdot 8+0,6\cdot(-1,6)=-0,76\\ -3,5+0,5\cdot 8+0,5\cdot(-1,6)=-0,3\\ -4+0,6\cdot 8+0,4\cdot(-1,6)=0,16\end{Bmatrix}.$$

Итак, в расчете на три шага оптимальная стратегия поведения ($k_1(3), k_2(3)$) = (2;3) , при этом $V_3(1)=9,76,\ V_3(2)=0,16$. Можно сделать предположение, что стратегия (2;3) останется оптимальной и на большее число шагов. Рассмотрим марковский процесс, соответствующий этой стратегии:

$$p(n+1)=p(n)\cdot\begin{pmatrix}0,6 & 0,4\\ 0,6 & 0,4\end{pmatrix}.$$

Известно, что если все элементы матрицы вероятностей переходов строго положительны, то вне зависимости от начального распределения $p(0)$ существует $\lim\limits_{n\to\infty} p(n)=p^*$. В этом случае марковскую цепь называют *эргодической* [4, 5]. Переходя к пределу в записанном соотношении, получим систему двух зависимых уравнений относительно вектора p^*. Дополняя ее условием нормировки $p_1^*+p_2^*=1$, находим соответствующий нашей задаче вектор предельных вероятностей p^* = (0.6;0.4). Таким образом, предполагая процесс достаточно длительным, мы можем подсчитать средний доход M за один шаг при соблюдении стратегии (2;3):

M = 0.6· (0.6·8 + 0.4· 2) + 0.4· (0.6·0 + 0.4· (−10)) =1.76.

Дополнение к расчётам по управляемым марковским процессам

Шаг первый:

$V_1(1)=6$, $V_2(1)=-3$

Оптимальная стратегия: $K_1(1)=1$, $K_2(1)=1$

Шаг второй:

$$\begin{Bmatrix} 6{,}0+0{,}5*6+0{,}5*(-3{,}0)=7{,}5 \\ 5{,}6+0{,}6*6+0{,}4*(-3{,}0)=8{,}0 \\ 4{,}5+0{,}7*6+0{,}3*(-3{,}0)=7{,}8 \end{Bmatrix} \quad V_1(2)=8{,}0$$

$$\begin{Bmatrix} -3{,}0+0{,}4*6+0{,}6*(-3{,}0)=-2{,}4 \\ -3{,}5+0{,}5*6+0{,}5*(-3{,}0)=-2{,}0 \\ -4{,}0+0{,}6*6+0{,}4*(-3{,}0)=-1{,}6 \end{Bmatrix} \quad V_2(2)=-1{,}6$$

Оптимальная стратегия: $K_1(2)=2$, $K_2(2)=3$

Шаг третий:

$$\begin{Bmatrix} 7{,}5+0{,}5*8+0{,}5*(-1{,}6)=10{,}70 \\ 8{,}0+0{,}6*8+0{,}4*(-1{,}6)=12{,}16 \\ 7{,}8+0{,}7*8+0{,}3*(-1{,}6)=12{,}92 \end{Bmatrix} \quad V_1(3)=12{,}92$$

$$\begin{Bmatrix} -2{,}4+0{,}4*8+0{,}6*(-1{,}6)=-0{,}16 \\ -2{,}0+0{,}5*8+0{,}5*(-1{,}6)=+1{,}20 \\ -1{,}6+0{,}6*8+0{,}4*(-1{,}6)=+2{,}56 \end{Bmatrix} \quad V_2(3)=2{,}56$$

Оптимальная стратегия: $K_1(3)=3$, $K_2(3)=3$

Домашние задания

1. Задача об экзаменационной сессии

Студент уже сдал один экзамен на 4, но ему предстоит сдать еще три экзамена. При подготовке к экзаменам он из-за недостатка времени может выбрать одну из следующих двух стратегий: либо выучить часть материала довольно хорошо, либо пройтись быстро по всему материалу. Определить оптимальную в смысле набранных баллов стратегию поведения студента на оставшиеся три экзамена, если матрицы вероятностей получения оценок 5, 4, 3, 2 в зависимости от предыдущей оценки для двух стратегий имеют вид:

$$P^{(1)} = \begin{pmatrix} 0,2 & 0,3 & 0,3 & 0,2 \\ 0,1 & 0,3 & 0,4 & 0,2 \\ 0,0 & 0,3 & 0,4 & 0,3 \\ 0,0 & 0,3 & 0,3 & 0,4 \end{pmatrix}, \ P^{(2)} = \begin{pmatrix} 0,1 & 0,3 & 0,5 & 0,1 \\ 0,0 & 0,3 & 0,6 & 0,1 \\ 0,0 & 0,2 & 0,7 & 0,1 \\ 0,0 & 0,1 & 0,8 & 0,1 \end{pmatrix}.$$

2. Задача об экзаменационной сессии

Решить предыдущую задачу №1 для следующих исходных данных

$$P^{(1)} = \begin{pmatrix} 0,2 & 0,0 & 0,4 & 0,4 \\ 0,1 & 0,3 & 0,4 & 0,2 \\ 0,0 & 0,3 & 0,4 & 0,3 \\ 0,0 & 0,3 & 0,3 & 0,4 \end{pmatrix}, \ P^{(2)} = \begin{pmatrix} 0,1 & 0,3 & 0,5 & 0,1 \\ 0,2 & 0,3 & 0,4 & 0,1 \\ 0,0 & 0,2 & 0,7 & 0,1 \\ 0,0 & 0,1 & 0,8 & 0,1 \end{pmatrix}.$$

3. Задача об экзаменационной сессии

Решить предыдущую задачу №1 для следующих исходных данных

$$P^{(1)} = \begin{pmatrix} 0,2 & 0,0 & 0,4 & 0,4 \\ 0,1 & 0,3 & 0,4 & 0,2 \\ 0,2 & 0,3 & 0,2 & 0,3 \\ 0,0 & 0,3 & 0,3 & 0,4 \end{pmatrix}, \ P^{(2)} = \begin{pmatrix} 0,1 & 0,3 & 0,5 & 0,1 \\ 0,2 & 0,3 & 0,4 & 0,1 \\ 0,0 & 0,2 & 0,7 & 0,1 \\ 0,2 & 0,1 & 0,6 & 0,1 \end{pmatrix}.$$

4. Задача об экзаменационной сессии

Решить предыдущую задачу №1 для следующих исходных данных

$$P^{(1)} = \begin{pmatrix} 0,2 & 0,0 & 0,4 & 0,4 \\ 0,1 & 0,3 & 0,4 & 0,2 \\ 0,2 & 0,3 & 0,2 & 0,3 \\ 0,1 & 0,2 & 0,3 & 0,4 \end{pmatrix}, \ P^{(2)} = \begin{pmatrix} 0,1 & 0,3 & 0,5 & 0,1 \\ 0,2 & 0,3 & 0,4 & 0,1 \\ 0,1 & 0,2 & 0,6 & 0,1 \\ 0,2 & 0,1 & 0,6 & 0,1 \end{pmatrix}.$$

5. Задача о погоне

Догоняющий находится в i -той клетке из 5 клеток, образующих круг. За один такт он с вероятностью $p = 1/2$ перемещается по часовой стрелке в соседнюю клетку, с вероятностью $q = 1/3$ перемещается против часовой стрелки в соседнюю клетку, с вероятностью $r = 1/6$ остается на месте. Убегающий находится в j -той клетке и на каждом такте может выбрать одну из трех стратегий поведения: (a) переместиться по часовой стрелке в соседнюю клетку; (b) остаться на месте; (c) переместиться против часовой стрелки в соседнюю клетку. Расстояние между догоняющим и убегающим определяется по формуле $d = |i - j|$. Определить стратегию убегающего на три такта вперед, максимизирующую сумму расстояний между догоняющим и убегающим.

6. Задача о погоне

Решить задачу №5 при следующих исходных данных

$p = 1/3$, $q = 1/3$, $r = 1/3$.

7. Задача о погоне

Решить задачу №5 при следующих исходных данных

$p = 1/6$, $q = 1/3$, $r = 1/2$.

8. Стохастическая задача о фермере

Состояние продуктивности земли, используемой фермером, может быть (a) хорошим, (b) удовлетворительным, (c) плохим. Вероятности перехода продуктивности земли из одного состояния в другое без проведения агротехнических мероприятий за один сезон заданы матрицей $P^{(1)}$. Однако фермер может провести комплекс агротехнических мероприятий, и тогда вероятности перехода продуктивности земли из одного состояния в другое за один сезон будут заданы матрицей $P^{(2)}$. Матрицы доходов для двух стратегий поведения: $D^{(1)}$, $D^{(2)}$. Найти оптимальную стратегию фермера на 4 сезона.

$$P^{(1)} = \begin{pmatrix} 0,2 & 0,5 & 0,3 \\ 0,0 & 0,5 & 0,5 \\ 0,0 & 0,0 & 1,0 \end{pmatrix}, \; P^{(2)} = \begin{pmatrix} 0,3 & 0,6 & 0,1 \\ 0,2 & 0,6 & 0,2 \\ 0,1 & 0,5 & 0,4 \end{pmatrix};$$

$$D^{(1)} = \begin{pmatrix} 7 & 6 & 3 \\ 0 & 5 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \; D^{(2)} = \begin{pmatrix} 6 & 5 & -1 \\ 5 & 4 & -1 \\ 4 & 3 & -2 \end{pmatrix}.$$

9. Стохастическая задача о фермере

Решить задачу №8 для следующих исходных данных

$$P^{(1)} = \begin{pmatrix} 0,2 & 0,5 & 0,3 \\ 0,1 & 0,4 & 0,5 \\ 0,0 & 0,0 & 1,0 \end{pmatrix}, \; P^{(2)} = \begin{pmatrix} 0,3 & 0,6 & 0,1 \\ 0,2 & 0,5 & 0,3 \\ 0,1 & 0,5 & 0,4 \end{pmatrix};$$

$$D^{(1)} = \begin{pmatrix} 7 & 6 & 3 \\ 1 & 6 & 1 \\ 0 & 2 & 1 \end{pmatrix}, \; D^{(2)} = \begin{pmatrix} 6 & 5 & -2 \\ 5 & 4 & -1 \\ 4 & 3 & -3 \end{pmatrix}.$$

10. Стохастическая задача о фермере

Решить задачу №8 для следующих исходных данных

$$P^{(1)} = \begin{pmatrix} 0,2 & 0,5 & 0,3 \\ 0,1 & 0,4 & 0,5 \\ 0,0 & 0,2 & 0,8 \end{pmatrix}, \; P^{(2)} = \begin{pmatrix} 0,3 & 0,6 & 0,1 \\ 0,2 & 0,5 & 0,3 \\ 0,0 & 0,6 & 0,4 \end{pmatrix};$$

$$D^{(1)} = \begin{pmatrix} 8 & 6 & 3 \\ 2 & 6 & 1 \\ 0 & 2 & 1 \end{pmatrix}, \; D^{(2)} = \begin{pmatrix} 7 & 5 & -2 \\ 6 & 4 & -1 \\ 5 & 3 & -3 \end{pmatrix}.$$