ОПТИМАЛЬНОЕ УПРАВЛЕНИЕ И МОДЕЛИРОВАНИЕ В СИСТЕМАХ МАРКОВСКОГО ТИПА

Трушков Александр Сергеевич, trushkov_as@mail.ru

Государственный социально-гуманитарный университет, г. Коломна

Задачи оптимального управления сложны для формализации и аналитического исследования. Системы марковского типа представляют собой простейший пример управляемого объекта. Ограничения, накладываемые на исследуемый процесс в этом случае достаточно жесткие и для большинства реальных систем не выполняются. Тем не менее выводы, которые следуют из теории оптимального управления в марковских системах имеют большое значение для теории оптимального управления в целом, так как являются основой разработки вычислительных алгоритмов управления для более сложных объектов и позволяют определить основные тенденции оптимального управления конкретной системой.

С матрицей переходных вероятностей **P** будем связывать матрицу дохода $\mathbf{R} = \{\mathbf{r_{ij}}\}$. При этом считается, что элементу $\mathbf{p_{ij}}$ матрицы **P** соответствует элемент $\mathbf{r_{ij}}$ матрицы **R**. Таким образом, каждому переходу $\mathbf{i} \to \mathbf{j}$ в марковской системе соответствует доход (или эффективнсть) $\mathbf{r_{ij}}$. Изменяя матрицу переходных вероятностей **P**, изменяем и соответствующую ей матрицу дохода **R**. Система мероприятий, при которой изменяются матрицы переходных вероятностей и дохода, называется управлением в марковской системе. Целью управления в марковском процессе является максимизация среднего ожидаемого дохода.

Будем называть управленческой альтернативой совокупность матрицы переходных вероятностей и соответствующей ей матрицы дохода. Пусть имеется несколько управленческих альтернатив. Правило, в соответствии с которым на каждом шаге марковского процесса выбирается управленческая альтернатива в зависимости от состояния системы называется стратегией. Если стратегия не зависит от номера шага марковского процесса, то она называется стационарной. Для каждой стационарной стратегии можно сформировать свои матрицы переходных вероятностей и дохода, полученные комбинацией строк из соответствующих матриц управленческих альтернатив.

Поведение марковского процесса на долгосрочном горизонте планирования характеризуется его независимостью от начального состояния системы. В этом случае говорят, что система достигла стационарного состояния. Будем рассматривать стратегии, для которых соответствующие марковские цепи допускают существование установившегося решения. Простейшим методом определения оптимального управления в марковском процессе является метод полного перебора.

Если число стратегий велико, то метод полного перебора реализовать трудно. В этом случае используется метод итераций по стратегиям [1]. Суммарный доход за этапы с **n**-го по **N**-й определяется рекуррентным уравнением: $\mathbf{z_n(i) = \nu_i + \sum_{j=1}^{m} p_{ij} z_{n+1}(j)}$, $\mathbf{i = \overline{1,m}}$. Для сокращения записи индекс **s** - номер стратегии - опущен. Здесь **n** - номер этапа управления. Пусть η - число оставшихся этапов. Тогда рекуррентное уравнение динамического программирования примет вид: $\mathbf{z_\eta(i) = \nu_i + \sum_{j=1}^{m} p_{ij} z_{\eta-1}(j)}$, $\mathbf{i = \overline{1,m}}$, здесь $\mathbf{z_\eta}$ - суммарный ожидаемый доход за η оставшихся этапов.

Пусть $\vec{\pi} = (\pi_1, \dots, \pi_m)$ - вектор установившихся вероятностей матрицы переходных вероятностей **P** и $\mathbf{E = \sum_{i=1}^{m} \pi_i \nu_i}$ - ожидаемый доход за этап.

Тогда при большом η (т.е. при $\mathbf{n} \to \infty$) имеет место соотношение: $\mathbf{z_\eta(i) = \eta E + z(i)}$, где $\mathbf{z(i)}$ - постоянный член, не зависящий от η и описывающий асимптотическое поведение $\mathbf{z(i)}$ при заданном состоянии $\mathbf{i}$. Тогда рекуррентное уравнение перепишется в виде:

$$\mathbf{\eta E + z(i) = \nu_i + \sum_{j=1}^{m} p_{ij}\left((\eta - 1)E + z(j)\right)}.$$

После преобразование данного уравнения получим:

$$\mathbf{E = \nu_i + \sum_{j=1}^{m} p_{ij} z(j) - z(i),\ \ i = \overline{1,m}}$$

Таким образом, имеем систему из $\mathbf{m}$ уравнений с $\mathbf{m+1}$ неизвестными: $\mathbf{z(1), \ldots, z(m), E}$. Целью задачи является определение оптимальной стратегии, которой соответствует $\mathbf{\max\{E\}}$. Оптимальное значение $\mathbf{E}$ из полученной системы уравнений определить за один шаг нельзя. Поэтому выбирается произвольная стратегия и определяются соответствующие ей значения $\mathbf{E}$. Затем определяется новая стратегия, дающая лучшее значение $\mathbf{E}$. Процесс продолжается до тех пор, пока $\mathbf{E}$ не перестанет улучшаться. Итеративный процесс состоит из двух шагов.

Шаг оценивания параметров. Выбрать произвольную стратегию $\mathbf{s}$. Используя матрицы $\mathbf{P^S}$, $\mathbf{R^S}$ и произвольно полагая $\mathbf{z^s(m) = 0}$ решить систему уравнений: $\mathbf{E^s = \nu_i^s + \sum_{j=1}^{m} p_{ij}^s z^s(j) - z^s(i),\ \ i = \overline{1,m}}$ относительно переменных $\mathbf{E^S, z^s(1), \ldots, z^s(m-1)}$.

Шаг улучшения стратегии. Для каждого состояния $\mathbf{i}$ определить управленческую альтернативу $\mathbf{k}$, обеспечивающую:

$$\mathbf{\max_k \left\{ \nu_i^k + \sum_{j=1}^{m} p_{ij} z^s(j) \right\},\ \ i = \overline{1,m}}$$

В качестве величин $\mathbf{z^s(i)}$ принимаются значения, полученные на шаге оценивания. Результирующая оптимальная стратегия $\mathbf{k(1), \ldots, k(m)}$ формирует новую стратегию $\mathbf{t}$. Если $\mathbf{s}$ и $\mathbf{t}$ совпадают, то вычисления заканчиваются, в противном случае принимается, что **s = t,** и производится переход к новому шагу оценивания.

Задача, решаемая на шаге улучшения стратегии: $\max\limits_{k}\left\{\nu_i^k + \sum\limits_{j=1}^{m} p_{ij} z^s(j)\right\}$

эквивалентна задаче определения **max{E}**, так как: $E = \nu_i + \sum\limits_{j=1}^{m} p_{ij} z(j) - z(i)$

но **z(i)** не зависят от выбора стратегии **k**, поэтому данные задачи эквивалентны.

Для расчета результатов работы марковской системы на бесконечном горизонте планирования разработана программа “Имитационное моделирование марковского процесса на бесконечном горизонте планирования”. Программа написана на алгоритмическом языке Visual Basic for Applications (VBA) и оформлена в виде модуля табличного процессора MS Excel. Программа позволяет вычислять оптимальные стратегии управления марковским процессом при его планировании на бесконечное число этапов, а также вычислять средний одношаговый доход при выбранной стратегии.

При конечном числе этапов моделирования средний одношаговый доход может не совпадать с теоретическим прогнозом. Поэтому интересным представляется вопрос об определении фактической величины среднего одношагового дохода по результатам заданного числа этапов марковского процесса. Данная задача также решается с помощью программы “Имитационное моделирование марковского процесса на бесконечном горизонте планирования”. При этом проводится имитационное моделирование марковского процесса при заданной стратегии управления для оценивания параметров распределения вероятностей среднего одношагового дохода.

Нижняя **a** и верхняя **b** границы матрицы дохода определяют диапазон возможных значений элементов матриц дохода при их автоматической генерации с помощью программы. Принимается, что элементы матриц являются реализациями равномерно распределенной на интервале **(a, b)** случайной величины. При генерации матриц дохода используется следующая формула: $\mathbf{a_{ij} = a + r \cdot (b - a)}$, где **r** - равномерно распределенная на интервале **(0, 1)** случайная величина, генерируемая датчиком случайных чисел.

При генерации матрицы переходных вероятностей используется тот факт, что при заданном состоянии **i** в начале текущего шага процесса вероятности $\mathbf{p_{ij}}$ перехода в состояние **j** в конце текущего шага являются вероятностями полной группы несовместных событий и удовлетворяют условию: $\sum_{j=1}^{m} \mathbf{p_{ij}} = \mathbf{1}$.

Для определения строки матрицы переходных вероятностей, соответствующей выбранной управленческой альтернативе **k** и состоянию системы **i** в начале текущего шага, генерируются **m-1** случайное число из диапазона **(0, 1)**, которые упорядочиваются в порядке возрастания в массиве $\mathbf{r_j}$, $\mathbf{j=1, \ldots, m-1}$. Тогда для заданного состояния системы **i** вероятности $\mathbf{p_{ij}}$ определяются по формуле: $\mathbf{p_{ij} = r_j - r_{j-1}}$. Принимается, что $\mathbf{r_0 = 0}$. Аналогично генерируется вектор начальных вероятностей состояний системы. Использование сгенерированных матриц переходных вероятностей и дохода сокращает время ввода исходных данных и может применяться при изучении марковских процессов и свойств их решения.

Если необходимо использовать заданные матрицы переходных вероятностей и дохода, то границы **a** и **b** целесообразно принять равными 0. Тогда все элементы матрицы дохода при формировании таблицы для ее ввода будут равны 0, и в соответствующие клетки вносятся фактические значения. Элементы матриц переходных вероятностей будут по-прежнему генерироваться программой и при использовании заданных управленческих альтернатив необходимо исправить предложенные программой значения.

При считывании исходных данных производится контроль корректности ввода. Контролируются значения **m** - количество состояний системы, **n** - число управленческих альтернатив и число **k** этапов моделирования, которые должны удовлетворять следующим условиям: $\mathbf{1 \le k \le 10000}$, $\mathbf{1 \le m \le 10}$, $\mathbf{1 \le n \le 10}$. Число **k** этапов моделирования при расчете оптимальной стратегии и ожидаемого одношагового дохода не учитывается.

Вид рабочего листа **Расчет** после расчета оптимальной стратегии и моделировании марковского процесса на 1000 шагов приведен на рисунке. Рассчитанное теоретически методом итераций по стратегиям значение ожидаемого одношагового дохода подтверждено в численном моделировании марковского процесса по рассчитанной оптимальной стратегии. Расхождение не превышает 0,1 (относительная погрешность ~ 1%, тогда как среднее отклонение составляет более 25% от одношагового дохода).

Данный алгоритм и программа используются в компьютерном практикуме при изучении дисциплины «Компьютерное моделирование» [2].

Литература

1.Таха, Х. Введение в исследование операций: в 2 кн. / Х. Таха. – М.: Мир, 1985.

2. Трушков, А.С. Стохастическое моделирование. Основы теории и компьютерный практикум. Часть 2. Имитационный эксперимент в задачах оптимального управления и принятия решений. Учебное пособие./ А.С. Трушков - МГОСГИ, г. Коломна, 2012.

МАРКОВСКИЙ ПРОЦЕСС

БЕСКОНЕЧНЫЙ ГОРИЗОНТ ПЛАНИРОВАНИЯ

Генерация массивов

Бесконечный марковский процесс

Одношаговый доход

Моделирование марковского процесса

Исходные данные

№	Название параметра	Значение
1	Число альтернатив управления	3
2	Число состояний системы	4
3	Нижняя граница матриц дохода	0
4	Верхняя граница матриц дохода	10
5	Число этапов моделирования	1000

Начальные вероятности состояний системы

Состояние	1	2	3	4
Вероятность	0.110	0.751	0.115	0.024

Матрица 1 переходных вероятностей

	Сост.1	Сост.2	Сост.3	Сост.4
Сост.1	0.349	0.536	0.097	0.018
Сост.2	0.054	0.096	0.090	0.759
Сост.3	0.269	0.162	0.542	0.027
Сост.4	0.240	0.292	0.244	0.224

Матрица 1 дохода

	Сост.1	Сост.2	Сост.3	Сост.4
Сост.1	8.944	0.108	9.533	9.381
Сост.2	4.839	5.336	7.610	6.365
Сост.3	2.747	7.531	5.766	0.180
Сост.4	8.950	8.476	6.771	8.145

Матрица 2 переходных вероятностей

	Сост.1	Сост.2	Сост.3	Сост.4
Сост.1	0.129	0.019	0.229	0.623
Сост.2	0.018	0.491	0.465	0.026
Сост.3	0.240	0.052	0.314	0.394
Сост.4	0.505	0.062	0.050	0.382

Матрица 2 дохода

	Сост.1	Сост.2	Сост.3	Сост.4
Сост.1	4.619	5.279	7.426	5.477
Сост.2	7.226	6.153	1.225	7.325
Сост.3	4.447	8.691	3.504	3.303
Сост.4	0.570	0.657	9.103	4.516

Матрица 3 переходных вероятностей

	Сост.1	Сост.2	Сост.3	Сост.4
Сост.1	0.281	0.160	0.239	0.321
Сост.2	0.160	0.049	0.074	0.717
Сост.3	0.145	0.282	0.050	0.523
Сост.4	0.028	0.182	0.054	0.736

Матрица 3 дохода

	Сост.1	Сост.2	Сост.3	Сост.4
Сост.1	5.400	3.259	2.520	3.096
Сост.2	5.666	6.935	0.566	4.233
Сост.3	8.928	9.633	7.525	0.703
Сост.4	0.011	7.590	2.807	9.901

Оптимальная стратегия

Состояние	1	2	3	4
Стратегия	2	1	3	3

Ожидаемый одношаговый доход 8.186

Результаты моделирования

Номер этапа	1000
Средний доход	8.093
Среднее отклонение	2.372

Расчет окончен

Матрица переходных вероятностей

	Сост.1	Сост.2	Сост.3	Сост.4
Сост.1	0.129	0.019	0.229	0.623
Сост.2	0.054	0.096	0.090	0.759
Сост.3	0.145	0.282	0.050	0.523
Сост.4	0.028	0.182	0.054	0.736

Матрица дохода

	Сост.1	Сост.2	Сост.3	Сост.4
Сост.1	4.619	5.279	7.426	5.477
Сост.2	4.839	5.336	7.610	6.365
Сост.3	8.928	9.633	7.525	0.703
Сост.4	9.981	7.590	2.807	9.901

Количество переходов

	Сост.1	Сост.2	Сост.3	Сост.4
Сост.1	3	0	14	32
Сост.2	13	20	15	142
Сост.3	9	25	1	33
Сост.4	24	144	38	487

Матрица частот

	Сост.1	Сост.2	Сост.3	Сост.4
Сост.1	0.061	0.000	0.286	0.653
Сост.2	0.068	0.105	0.079	0.747
Сост.3	0.132	0.368	0.015	0.485
Сост.4	0.035	0.208	0.055	0.703