Вісник Харківського національного університету
Серія «Математичне моделювання. Інформаційні технології. Автоматизовані системи управління»
УДК 519.217:681.3
№ 629, 2004, с.25-32

Оптимизация периода наблюдений в марковском процессе принятия решений

Н. С. Подцыкин
Харьковский национальный университет им. В.Н. Каразина, Украина

The case in point of correlation the transitional matrix of conducted Mark circuit with profits from the controllable process has been considered. An optimizational method of a controllable period and a strategy of conduct regarding the unlimited interval of time.

1. Общая постановка задачи и её актуальность

Задача поиска оптимального периода наблюдений за управляемым случайным процессом является достаточно важной в прикладной теории оптимального управления. При построении модели, в которой решения по применению управлений принимаются в дискретные моменты времени, основное внимание уделяется разработке методов выбора оптимальных управлений. Период принятия решений считается заданным. В то же время оптимизация периода принятия решений может значительно повысить эффективность разработанного метода. В статье рассмотрена задача оптимизации периода принятия решений и стратегии управления.

2. Истоки исследования автора

Настоящая работа выполнена в рамках теории марковских процессов принятия решений [1]. При разработке оптимизационного алгоритма была использована, полученная в [3], оценка близости выбранной стратегии к оптимальной по критерию среднего дохода в единицу времени.

3. Нерешенные проблемы и цели работы

Класс стохастических систем, описываемых марковским процессом принятия решений, объединяет достаточно большую часть всех управляемых систем. Учитывая, что на практике часто возникает необходимость управления целой группой однотипных или разнотипных систем из одного центра, приобретает важное значение задача выбора оптимального периода контроля и управления такой группой систем. Решение этой задачи должно проводиться в рамках условий и ограничений на функционирование рассматриваемой группы, а также влияний на нее внешних воздействий. Целью настоящей работы является разработка оптимизационных алгоритмов, позволяющих отыскивать ε-оптимальные период и стратегию управления одной системой, эволюция которой может быть описана марковским процессом принятия решений. Очевидно, что результаты этой работы могут быть использованы при решении указанных выше задач для сложных систем.

4. Построение математической модели

Рассмотрим управляемую дискретную систему с конечным числом состояний $E = \{x\}$ на неограниченном интервале времени. Система контролируется через равные промежутки времени $\tau \in T$, и в эти же моменты времени принимаются решения о выборе одного из возможных действий $g \in G$. Множества G и T считаем конечными. Действие g определяет распределение вероятностей q на множестве E в следующий момент контроля. Естественно считать, что это распределение зависит от периода τ и наблюденного состояния x. Поэтому обозначим его через $q(\cdot|x, g, \tau)$. Введем в рассмотрение множество управлений $Y = G \times T$ и определим решающую функцию $f : E \to Y$. Обозначим $\tau_f(x)$ - период, планируемый решающей функцией f в состоянии $x \in E$. В этом случае распределение вероятностей $q(\cdot|x, g, \tau)$ удобнее обозначить через $q(\cdot|x, f)$, так как f однозначно определяет и период контроля и воздействие. Если $\tau_f(x)=\tau_0$ для любого $x \in E$, то соответствующую решающую функцию f будем обозначать также и $f^{(\tau_0)}$.

Каждой решающей функции f соответствует матрица переходных вероятностей $Q(f)$ с элементами $q(z|x, f(x))$, $x, z \in E$.

Применение управления $y \in Y$ в состоянии $x \in E$ определяет непосредственный доход $W(x, y)$, который приносит в этих условиях система в единицу времени на рассматриваемом периоде.

Решающей функции f поставим в соответствие вектор непосредственных доходов $W(f)$, с компонентами $W(x, f(x))$, $x \in E$.

Последовательность решающих функций $\pi = \{f_1, f_2, \ldots\}$ назовем стратегией управления. Стратегия $\pi = \{f, f, \ldots\}$ - стационарная. Таким образом, $f(x) \in Y$ - управление, применяемое в состоянии x.

Набор объектов $\{E, Y, q, \pi, W\}$ задает марковский процесс принятия решений. Заметим, что в нашем случае q и W зависят от периода. В данной работе рассматривается задача оптимизации периода наблюдений и выбора оптимальной стратегии в марковском процессе принятия решений.

Конкретизируем модель и уточним определения некоторых из приведенных объектов. Пусть работающая система приносит доход v в единицу времени. Будем допускать, что реализация управления требует определенного интервала времени, в течение которого система простаивает и интенсивность дохода, связанного с процессом, нулевая. Обозначим через t_y длительность такого интервала времени, в течение которого реализуется управление y и система простаивает.

Пусть τ - планируемый период между соседними моментами контроля процесса и пусть в начале периода наблюдалось состояние x, в котором было применено управление y. Стоимость управления y, вообще говоря, зависящую от состояния, обозначим через $r(x, y)$. Тогда положим, что значение функции непосредственных доходов равно $W(x, y) = \frac{1}{\tau}\left(\chi\left(\tau - t_y\right)v\left(\tau - t_y\right) - r(x, y)\right)$, где $\chi(t) = \begin{cases} 1, & t \geq 0, \\ 0, & t < 0 \end{cases}$ - функция Хевисайда.

В дальнейшем мы ограничимся рассмотрением только тех процессов, множество состояний которых образуют один эргодический класс при любой допустимой стратегии управления [1].

Каждой стационарной стратегии $\pi = \{f, f, \ldots\}$ поставим в соответствие математическое ожидание среднего в единицу времени дохода, связанного с рассматриваемым процессом, на неограниченном промежутке времени:

$$\varphi_\tau(\pi)(x) = \lim_{N \to \infty} \frac{1}{N} \mathrm{M}_{x,\tau}^{\pi} \sum_{k=0}^{N-1} W(x_k, y_k), \qquad x \in E.$$

Индекс k здесь указывает на номер периода, $\mathrm{M}_{x,\tau}^{\pi}$ - знак математического ожидания, взятого по мере, индуцированной стратегией $\pi = \{f, f, \ldots\}$, x - начальное состояние, τ - вектор периодов контроля, x_k - состояние в начале k-го периода, $y_k = f(x_k)$.

Задача состоит в обосновании оптимизационной процедуры, позволяющей отыскивать ε- оптимальные стратегию управления π^* и период контроля τ^*, удовлетворяющие следующему условию:

для заданного $\varepsilon > 0$ выполнено

$$\left|\varphi_{\tau^*}\left(\pi^*\right)(x) - \varphi_\tau(\pi)(x)\right| < \varepsilon$$

для всех стратегий π, всех $x \in E$ и периодов наблюдений $\tau \in T$.

Если период фиксирован, то для простоты в дальнейшем индекс τ может быть опущен.

Приведем ниже некоторые вспомогательные сведения и результаты.

Пусть V - линейное пространство векторов размерности n. Каждому вектору $v \in V$ поставим в соответствие величину $\Delta v = \max\limits_x v(x) - \min\limits_x v(x)$. Она определяет полунорму в V. Пусть $S = \{c \in V : \Delta c = 0\}$. Из определения Δ следует, что S - подпространство в V такое, что все его элементы – это векторы с равными компонентами. Каждому вектору $v \in V$ поставим в соответствие класс эквивалентности $\tilde{v}$ пространства V по S, содержащий v, по правилу

$$\lambda(v) = v + S = \{v + c : c \in S\} = \tilde{v}.$$

Множество $\tilde{V} = \{\tilde{v}\}$ всех классов эквивалентности образует векторное пространство, называемое фактор-пространством. Определим на $\tilde{V}$ функцию

p, положив $p(\tilde{v}) = p(\lambda(v)) = \Delta v$. Как известно [2], она определяет норму $\|\tilde{v}\| = \Delta v$ в $\tilde{V}$.

Далее нам будет удобно (при построении оптимизационного алгоритма) прообразом класса эквивалентности $\tilde{v}$ при отображении λ считать элемент $v \in V$, имеющий нулевую первую компоненту.

Определим на V ряд операторов.

$$Q(y)v(x) = \sum_{z \in E} v(z) q(z|x, y), \quad x \in E, y \in Y.$$

Если $y \in Y$ определяется решающей функцией f, то этот оператор будем записывать $Q(f)$. Так же будем обозначать матрицу этого линейного оператора. Введем далее оператор

$$F(y)v(x) = W(x, y) + Q(y)v(x).$$

Если здесь $y \in Y$ определяется с помощью решающей функции f, то этот оператор обозначим $F(f)$ и запишем в векторном виде

$$F(f)v = W(f) + Q(f)v.$$

Определим, наконец, оператор U следующим образом

$$Uv = \max_f F(f)v.$$

Заметим, что введенные операторы можно считать заданными и на фактор-пространстве $\tilde{V}$. Свойства этих операторов можно найти в [1].

Лемма [1]. Пусть $\sigma = \{f, f, \ldots\}$ - стационарная стратегия. Тогда

1. Матрица $Q(f)$ сходится по Чезаро к матрице $Q^*(f)$, т.е. существует предел
$$Q^*(f) = \lim_k \frac{1}{k} \sum_{i=0}^{k-1} Q^i(f).$$
2. $\varphi(\sigma) = Q^*(f)W(f)$ и вектор $\varphi(\sigma)$ - имеет одинаковые компоненты.
3. Вектор $\varphi(\sigma)$ является решением уравнения $\varphi(\sigma) = F(f)v - v$, где одна из компонент неизвестного вектора v выбирается произвольно.

Теорема. Пусть $\sigma = \{f, f, \ldots\}$ - стационарная стратегия, $\sigma^* = \{f^*, f^*, \ldots\}$ - оптимальная стратегия и $v \in V$ - произвольный вектор. Обозначим

$$\Phi(\sigma) = F(f)v - v, \quad \overline{\Phi}(\sigma) = \max_x \Phi(\sigma)(x), \quad \underline{\Phi}(\sigma) = \min_x \Phi(\sigma)(x).$$

Тогда справедливы оценки:

$$\underline{\Phi}(\sigma) \le \varphi(\sigma) \le \overline{\Phi}(\sigma).$$

Обозначим далее

$$\Phi = Uv - v, \quad \overline{\Phi} = \max_x \Phi(x), \quad \underline{\Phi} = \min_x \Phi(x).$$

Тогда для оптимального значения φ справедливы оценки:

$$\underline{\Phi} \le \varphi(\sigma^*) \le \overline{\Phi}.$$

Доказательство. Докажем первую часть утверждения. Умножим обе части равенства $\Phi(\sigma) = W(f) + Q(f)v - v$ на $Q^*(f)$:

$Q^*(f)\Phi(\sigma) = Q^*(f)W(f) + Q^*(f)v - Q^*(f)v$ или $Q^*(f)\Phi(\sigma) = \varphi(\sigma)$.

Отсюда следует, что $\underline{\Phi}(\sigma) \le \varphi(\sigma) \le \overline{\Phi}(\sigma)$.

Докажем теперь вторую часть. Если $Uv = F\left(f^*\right)v$, то требуемое доказано в первой части. Пусть $Uv = F(f)v$, где $f \ne f^*$. Тогда

$\Phi(\sigma) = F_t(f)v - v \le Uv - v = \Phi$.

Отсюда следует, что $\overline{\Phi}(\sigma) \le \overline{\Phi}$. По доказанному $\varphi(\sigma) \le \overline{\Phi}(\sigma)$. Поэтому получаем:

$$\varphi(\sigma) \le \overline{\Phi}(\sigma) \le \overline{\Phi}. \qquad (1)$$

С другой стороны, по определению оптимальной стратегии $\varphi(f) \le \varphi(\sigma^*)$. В соответствии с доказанным выше, получаем

$\underline{\Phi} = \underline{\Phi}(f) \le \varphi(\sigma)$.

С учетом (1) получаем

$\underline{\Phi} \le \varphi(f) \le \varphi(\sigma) \le \overline{\Phi}(\sigma) \le \overline{\Phi}$.

Отсюда окончательно следует

$\underline{\Phi} \le \varphi(\sigma^*) \le \overline{\Phi}$.

Вторая часть теоремы доказана.

Следствие 1. Пусть $v_k \in V$ и σ_{k+1} определяется из условия $Uv_k = F(\sigma_{k+1})v$. Обозначим $\Phi_k = Uv_k - v_k$, $\overline{\Phi}_k = \max_x \Phi_k(x)$, $\underline{\Phi}_k = \min_x \Phi_k(x)$. Тогда для $\varphi(\sigma_{k+1})$ верно

$$\underline{\Phi}_k \le \varphi(\sigma_{k+1}) \le \overline{\Phi}_k. \qquad (2)$$

Пусть σ^* - оптимальная стратегия. Тогда для $\varphi(\sigma^*)$ верно

$$\underline{\Phi}_k \le \varphi(\sigma^*) \le \overline{\Phi}_k. \qquad (3)$$

Доказательство следует из теоремы 1, если в качестве v взять v_k.

Следствие 2. Пусть σ^* - оптимальная стратегия, v_k, v_{k+1} и σ_{k+1} удовлетворяют условиям: $Uv_k = F(\sigma_{k+1})v_k$ и $v_{k+1} = Uv_k$. Обозначим $\varepsilon = \left\| v_{k+1} - v_k \right\|$. Тогда σ_{k+1} является ε-оптимальной стратегией.

Доказательство. Из (2) и (3) следует

$\left| \varphi(\sigma_{k+1}) - \varphi(\sigma^*) \right| \le \overline{\Phi}_k - \underline{\Phi}_k$.

Требуемое получаем из того, что $\overline{\Phi}_k - \underline{\Phi}_k = \left\| \Phi_k \right\| = \left\| Uv_k - v_k \right\| = \left\| v_{k+1} - v_k \right\|$.

Приведем далее известный алгоритм вычисления ε- оптимальной стратегии $[1]$, дополнив его правилом остановки, полученном в теореме. Этот алгоритм основан на принципе сжатых отображений. В связи с этим будем предполагать, что оператор $\widetilde{U}$ является сжимающим в $\widetilde{V}$. Заметим, что для этого достаточно потребовать регулярность матрицы $Q(f)$ для любой решающей функции f [1]. Далее считаем, что это условие выполнено.

Алгоритм вычисления ε-оптимальной стратегии состоит в выполнении следующих действий.

1. Выбрать $\varepsilon > 0$.
2. Положить $v_0 = \bar{0}$, $k = 1$.

   k - я итерация:
3. Вычислить вектор v_k фактор-пространства $\widetilde{V}$ и решающую функцию f_k из системы линейных уравнений
   $v_k = Uv_{k-1} = F_t(f_k)v_{k-1}$.
4. Проверить условие: $\|v_k - v_{k-1}\| < \varepsilon$.
5. Если оно выполнено, то $\pi_k = \{f_k, f_k, \ldots\}$ - ε - оптимальная стратегия. Иначе увеличить k на 1 и перейти в 3.

За конечное число итераций гарантируется достижение ε - оптимальной стратегии.

В пункте 3 алгоритма достаточно вычислить прообраз вектора v_k.

5. Разработка оптимизационных алгоритмов

Практический интерес представляют две различные постановки задачи оптимизации периода контроля. В первой – оптимальный период контроля находится для каждого состояния отдельно, во второй – общий оптимальный период контроля для всех состояний. При этом, как в первой постановке, так и во второй, предполагается, что вместе с оптимизацией периода контроля, оптимизируется правило применения управляющих воздействий.

Решение задачи поиска ε - оптимальной стратегии в первой постановке реализуется применением приведенного выше алгоритма. В момент остановки алгоритма будет получена ε - оптимальная стационарная стратегия управления $\pi_k = \{f_k, f_k, ...\}$, где решающая функция f_k определяет оптимальное воздействие для каждого состояния и вектор ε - оптимальных периодов $\tau_{f_k}(x)$.

Рассмотрим теперь решение задачи оптимизации стратегии управления во второй постановке. Оно будет основано на результатах леммы и теоремы. Далее будем пользоваться обозначениями: оператор U обозначим через $U^{(\tau)}$, если U максимизирует вектор $(W + Qv)$ по решающим функциям $f^{(\tau)}$ с постоянным вектором периодов, и $U^{(g)}$, - если по решающим функциям с заданными фиксированными в каждом состоянии воздействиями.

Пусть множество периодов T представлено m возможными значениями $\tau_1, ..., \tau_m$. Выберем начальное значение вектора $v \in V$. Для каждого τ_i, $i = 1, ..., m$, определим начальную стратегию $f^{(\tau_i)}$, выбрав произвольно допустимое воздействие из G для каждого состояния $x \in E$. Затем вычислим интервалы $\left(\underline{\varphi}^{(\tau_i)}, \overline{\varphi}^{(\tau_i)}\right)$, где $\underline{\varphi}^{(\tau_i)} = \min_x \left(U^{(\tau_i)} v - v\right)$, $\overline{\varphi}^{(\tau_i)} = \max_x \left(U^{(\tau_i)} v - v\right)$, $i = 1,...,m$. Обозначим $v_i = U^{(\tau_i)} v$.

По доказанной теореме интервал $\left(\underline{\varphi}^{(\tau_i)}, \overline{\varphi}^{(\tau_i)}\right)$ содержит максимальное значение критерия φ, соответствующее оптимальной стратегии управления при фиксированном значении периода наблюдения τ_i общего для всех состояний. Если окажется, что для некоторого k- ого интервала найдется l- ый интервал такой, что $\overline{\varphi}^{(\tau_k)} < \underline{\varphi}^{(\tau_l)}$, то период контроля τ_k не может быть оптимальным и должен быть исключен из дальнейшего процесса оптимизации. Обозначим через T_1 подмножество множества T периодов контроля, оставшихся после исключения тех из них, которые удовлетворяют приведенному выше условию.

Далее повторим операцию исключения периодов из множества T_1, вычислив интервалы $\left(\underline{\varphi}^{(\tau_i)}, \overline{\varphi}^{(\tau_i)}\right)$ для всех $\tau_i \in T_1$, где $\underline{\varphi}^{(\tau_i)} = \min_x \left(U^{(\tau_i)} v_1 - v_1\right)$, $\overline{\varphi}^{(\tau_i)} = \max_x \left(U^{(\tau_i)} v_1 - v_1\right)$. Полученное множество периодов обозначим через T_2.

Аналогично продолжим процесс построения множеств $T_3, T_4, ...$. При этом, в силу сжатия оператора U, длины интервалов $\left(\underline{\varphi}^{(\tau_i)}, \overline{\varphi}^{(\tau_i)}\right)$, будут стремиться к нулю, а множества допустимых периодов контроля, вообще говоря, сужаться.

Пусть рассматриваемый процесс остановлен после получения множества T_k. Обозначим $\delta_k = \bigcup_{\tau_i \in T_k} \left(\underline{\varphi}^{(\tau_i)}, \overline{\varphi}^{(\tau_i)}\right)$ и его длину через $s(\delta_k)$. Можно утверждать, что любая стационарная стратегия, определяемая решающей функцией $f_k^{(\tau_i)}$, $\tau_i \in T_k$, является $s(\delta_k)$- оптимальной. Действительно, обозначим $\varphi(\tau_i)$, $\tau_i \in T_k$, значения средних доходов в единицу времени, соответствующие оптимальным стратегиям с фиксированными значениями периодов контроля. Так как $\varphi(\tau_i) \in \left(\underline{\varphi}^{(\tau_i)}, \overline{\varphi}^{(\tau_i)}\right)$ и оптимальное значение $\varphi^*(\tau^*) \in \delta_k$, то $\left|\varphi^*(\tau^*) - \varphi(\tau_i)\right| < s(\delta_k)$. Из того, что длины интервалов $\left(\underline{\varphi}^{(\tau_i)}, \overline{\varphi}^{(\tau_i)}\right)$, составляющие интервал δ_k, стремятся к нулю, следует, что для любого $\varepsilon > 0$ найдется такой номер k_0, что для всех $k > k_0$ будет выполняться неравенство $s(\delta_k) < \varepsilon$.

Полученные результаты позволяют предложить следующий алгоритм поиска ε - оптимальных правила применения воздействий и общего для всех состояний периода контроля.

Задать $\varepsilon > 0$.

Выбрать вектор $v_0 \in V$ (произвольно), положить $T^* = T$, $\varepsilon_1 = \varepsilon/2$.

1. Для каждого $\tau_i \in T^*$ вычислить последовательно векторы $v_1^{(\tau_i)} = U^{(\tau_i)} v_0$, $v_2^{(\tau_i)} = U^{(\tau_i)} v_1^{(\tau_i)}$, …, $v_k^{(\tau_i)} = U^{(\tau_i)} v_{k-1}^{(\tau_i)}$ так, чтобы $\left\| \widetilde{v}_k^{(\tau_i)} - \widetilde{v}_{k-1}^{(\tau_i)} \right\| < \varepsilon_1$.
2. Образовать интервалы $\left(\underline{\varphi}^{(\tau_i)}, \overline{\varphi}^{(\tau_i)} \right)$, где: $\underline{\varphi}^{(\tau_i)} = \min\limits_x \left(U^{(\tau_i)} v_{k-1}^{(\tau_i)} - v_{k-1}^{(\tau_i)} \right)(x)$, $\overline{\varphi}^{(\tau_i)} = \max\limits_x \left(U^{(\tau_i)} v_{k-1}^{(\tau_i)} - v_{k-1}^{(\tau_i)} \right)(x)$.
3. Период τ_i, $i = 1,\dots,n$, исключить из множества T^*, если выполнено условие: $(\exists j) \left\{ \overline{\varphi}^{(\tau_i)} < \underline{\varphi}^{(\tau_j)} \right\}$.
4. Образовать множество $\delta = \bigcup\limits_{\tau_i \in T^*} \left(\underline{\varphi}^{(\tau_i)}, \overline{\varphi}^{(\tau_i)} \right)$ и вычислить $s(\delta) = \max\limits_{\tau_i \in T^*} \overline{\varphi}^{(\tau_i)} - \min\limits_{\tau_i \in T^*} \underline{\varphi}^{(\tau_i)}$.

Любая стратегия $\left(f_k^{(\tau_i)}, f_k^{(\tau_i)}, \dots \right)$, $\tau_i \in T^*$, является ε - оптимальной.

Действительно, это следует из того, что длина $s(\delta)$ интервала δ меньше ε.

6. Выводы по результатам и направления дальнейших исследований

В работе обоснованы два метода оптимизации в марковском процессе принятия решений, позволяющие получить ε-оптимальные стратегию управления и период контроля. Первый метод оптимизирует следующий момент контроля в зависимости от наблюденного состояния в текущий момент времени. Второй метод оптимизирует период контроля общий для всех состояний. При этом оба метода позволяют одновременно оптимизировать и стратегию управления. В марковском процессе принятия решений по умолчанию принят общий период контроля. Однако его величина не оптимизируется.

Использование предложенной методологии в математической теории надежности может служить дополнительным резервом в повышении эффективности и работоспособности технических систем. Особенно перспективным является ее использование при моделировании технических систем, состояния которых определяют вероятность отказа и уровень работоспособности. Кроме того, полученные результаты могут быть использованы для разработки методов оптимизации периода наблюдений и принятия решений в сложных системах.

ЛИТЕРАТУРА

1. Майн Х., Осаки С. Марковские процессы принятия решений, - М.: Наука, 1977. – 175с.
2. Рудин У. Функциональный анализ. М.: Мир, 1975. 443с.
3. Подцыкин Н.С. Метод оптимизации в периодических марковских процессах принятия решений // Кибернетика. – Киев, 1991. – № 2. – С. 91-94, 99.