$H_{\text{макс}}(t) = H_k(t)$. Теперь сформулируем достаточное условие восстановимости управления.

Теорема 2. *По известному на интервале* $[0, T]$ *выходу* $y(t)$ *и начальному вектору* $x(0) = x_0$ *можно однозначно восстановить управление* $u(t)$, *если* $\operatorname{rg}(H_{\text{макс}}(t) B(t)) = p$.

Доказательство теоремы совпадает с доказательством достаточности для стационарного случая.

Список литературы: 1. *Маринич А. П.* Об относительно идеально наблюдаемых системах.— Дифференц. уравнения, 1976, 12, № 7, с. 1204—1210. 2. *Уонем М.* Линейные многомерные системы управления: Пер. с англ.— М.: Наука, 1980.— 326 с.

Поступила в редколлегию 25.10.83.

УДК 62-50

Л. М. ЛЮБЧИК, канд. техн. наук.

ГРАДИЕНТНЫЕ АЛГОРИТМЫ ОПТИМИЗАЦИИ МАРКОВСКИХ ПРОЦЕССОВ ПРИ НЕПОЛНОЙ ИНФОРМАЦИИ

Управляемые марковские цепи с доходами широко используются в качестве математических моделей процессов принятия решений. Существуют эффективные конечно-сходящиеся алгоритмы оптимизации решающих правил [1], основанные на методах динамического и линейного программирования. Практическая реализация указанных методов в условиях неполной информации сопровождается идентификацией набора переходных матриц управляемой марковской цепи для каждого из возможных решающих правил. При этом возникает задача статистического оценивания большого числа параметров, что затрудняет расчеты в реальном масштабе времени и отвлекает большие вычислительные ресурсы.

В работе [2] на основе адаптивного подхода предложены градиентные алгоритмы типа стохастической аппроксимации для оптимизации решающих правил в условиях неопределенности. Соответствующие условия оптимальности получены в работе [3]. Однако при таком подходе нужно решать уравнение чувствительности для финального вектора вероятностей состояний, использующее оценки переходных матриц.

Нами разработаны адаптивные алгоритмы оптимизации решающих правил, относящиеся к классу многошаговых алгоритмов с накоплением и не требующие предварительного восстановления переходных матриц. Рассмотрим управляемую марковскую цепь с конечными множествами состояний $x \in X = \{1, \ldots, N\}$ и управлений $u \in U = \{1, \ldots, M\}$.

Пусть смена состояний цепи осуществляется в дискретные моменты времени $n = 1, 2, \ldots$, тогда вероятности перехода из предыдущего состояния в последующее образуют условную переходную матрицу $\Pi = \|\pi_{ij}^k\|$, причем

$$\pi_{ij}^k = \boldsymbol{P}\{x[n+1] = j \mid x[n] = i,\ u[n] = k\};$$

$$\sum_{j=1}^{N} \sum_{k=1}^{M} \pi_{ij}^k = 1;\ \pi_{ij}^k \geqslant 0;\ i, j = 1, \ldots, N;\ k = 1, \ldots, M. \qquad (1)$$

Выбор управлений производится в соответствии с матрицей решающих правил $\boldsymbol{D}[n] = \| d_{ik}[n] \|$, элементы которой обусловливают вероятность применения соответствующего управления:

$$d_{ik}[n] = \boldsymbol{P}\{u[n] = k \mid x[n] = i\};$$

$$\sum_{k=1}^{M} d_{ik}[n] = 1;\ d_{ik}[n] \geqslant 0;\ i = 1, \ldots, N;\ k = 1, \ldots, M. \quad (2)$$

В качестве критерия оптимальности примем средний доход в единицу времени

$$J = \lim_{n \to \infty} \frac{1}{n} \sum_{m=1}^{n} \xi[m], \quad (3)$$

где $\xi[m]$ — случайный доход, определяемый текущим состоянием и управлением.

Для эргодических марковских цепей [1, 2] существует стационарное решающее правило $\boldsymbol{D}$, удовлетворяющее ограничениям (2) и обеспечивающее максимальное значение критерия (3). При этом

$$J(\boldsymbol{D}) = \sum_{i=1}^{N} \sum_{k=1}^{M} r_{ik} d_{ik} p_i(\boldsymbol{D}). \quad (4)$$

Здесь $r_{ik} = \boldsymbol{M}\{\xi[n] \mid x[n] = i, u[n] = k\}$, а стационарное распределение вероятностей состояний (финальный вектор) $\boldsymbol{p}(\boldsymbol{D}) = (p_1(\boldsymbol{D}), p_2(\boldsymbol{D}), \ldots, p_N(\boldsymbol{D}))^T$ удовлетворяет условиям

$$p^T(\boldsymbol{D}) = \boldsymbol{p}^T = (\boldsymbol{D})\,\Pi(\boldsymbol{D});\ \boldsymbol{p}^T(\boldsymbol{D})\,\boldsymbol{e} = 1;\ p^T(\boldsymbol{D}) \geqslant 0, \quad (5)$$

где $\Pi(\boldsymbol{D}) = \| \pi_{ij}(\boldsymbol{D}) \|$ — безусловная переходная матрица управляемой марковской цепи

$$\pi_{ij}(\boldsymbol{D}) = \sum_{k=1}^{M} \pi_{ij}^{k} d_{ik};\ \sum_{j=1}^{N} \pi_{ij}(\boldsymbol{D}) = 1;\ \pi_{ij}(\boldsymbol{D}) \geqslant 0;$$

$\boldsymbol{e} = (1, 1, \ldots, 1)^T$; T — знак транспонирования.

Задача оптимизации управляемой марковской цепи в условиях неполной информации состоит в нахождении оптимальной матрицы решающих правил $\boldsymbol{D}^*$, максимизирующей критерий оптимальности (4), причем элементы условных переходных матриц π_{ij}^{k} (1) и условные доходы r_{ik} предполагаются неизвестными, а наблюдению доступны лишь случайные последовательности состояний $x[n]$, управлений $u[n]$ и доходов $\xi[n]$. Для построения алгоритмов оптимизации решающих правил воспользуемся адаптивным подходом.

Введем матрицу $\boldsymbol{Q}(\boldsymbol{D}) = \| q_{ij}(\boldsymbol{D}) \|$, полученную путем замены на вектор $\boldsymbol{e}$ первого столбца матрицы $\boldsymbol{E} - \Pi(\boldsymbol{D})$:

$$q_{ij}(\boldsymbol{D}) = \left(\Delta_{ij} - \sum_{k=1}^{M} \pi_{ij}^{k} d_{ik}\right)(1 - \Delta_{1j}) + \Delta_{1j}. \quad (6)$$

Здесь Δ_{ij} — символ Кронекера, $\Delta_{ij} = 1$ при $i = j$, $\Delta_{ij} = 0$ при $i \neq j$; $\boldsymbol{E}$ — единичная матрица.

Поскольку матрица $Q(D)$ невырожденная, из условий (5) следует, что

$$p^T(D) = e_1^T Q^{-1}(D);\ e_1 = (1, 0, \ldots, 0)^T;$$

$$J(D) = e_1^T Q^{-1}(D) r(D);\ r_i(D) = \sum_{k=1}^{M} r_{ik} d_{ik}.$$

Вычислим компоненты градиента критерия оптимальности по элементам матрицы решающих правил:

$$\frac{\partial J(D)}{\partial d_{\alpha\beta}} = r_{\alpha\beta} p_\alpha(D) + \sum_{i=1}^{N} \sum_{k=1}^{M} r_{ik} d_{ik} \frac{\partial p_i(D)}{\partial d_{\alpha\beta}}, \qquad (7)$$

$$\alpha = 1, \ldots, N, \quad \beta = 1, \ldots, M.$$

При этом функции чувствительности стационарного распределения вероятностей состояний удовлетворяют условиям

$$\frac{\partial p_j(D)}{\partial d_{\alpha\beta}} = \pi_{\alpha j}^{\beta} p_\alpha(D) + \sum_{i=1}^{N} \sum_{k=1}^{M} \pi_{ij}^{k} d_{ik} \frac{\partial p_i(D)}{\partial d_{\alpha\beta}};$$

$$\sum_{i=1}^{N} \frac{\partial p_i(D)}{\partial d_{\alpha\beta}} = 0;\ \alpha = 1, \ldots, N;\ \beta = 1, \ldots, M.$$

Отсюда с помощью выражения (6) получим

$$\frac{\partial p^T(D)}{\partial d_{\alpha\beta}} = p_\alpha(D)\, \pi_\alpha^\beta Q^{-1}(D);\ \pi_\alpha^\beta = \left(0, \pi_{\alpha 2}^{\beta}, \ldots, \pi_{\alpha N}^{\beta}\right). \qquad (8)$$

Из равенств (7), (8) следует, что

$$\frac{\partial J(D)}{\partial d_{\alpha\beta}} = [r_{\alpha\beta} + \pi_\alpha^\beta Q^{-1}(D) r(D)]\, p_\alpha(D). \qquad (9)$$

Для построения адаптивных алгоритмов оптимизации решающих правил, использующих лишь реализацию последовательности состояний марковской цепи, применим рандомизированное представление градиента критерия оптимальности (9). Пусть $\alpha = x[n]$; $\beta = u[n]$; $\delta = x[n+1]$; $\xi(\alpha, \beta) = \xi[n]$. Тогда

$$r_{ik} = M\left\{\frac{\xi(\alpha, \beta)\, \Delta_{i\alpha} \Delta_{k\beta}}{p_i(D)\, d_{ik}}\right\};\quad r_i(D) = M\left\{\frac{\xi(\alpha, \beta)\, \Delta_{i\alpha}}{p_i(D)}\right\};$$

$$\frac{\partial J(D)}{\partial d_{ik}} = M\left\{\xi(\alpha, \beta)\left[d_{ik}^{-1} + (1 - \Delta_{ik})\,(Q^{-1}(D))_{\delta\alpha}\right] \Delta_{i\alpha} \Delta_{k\beta}\right\}. \qquad (10)$$

Обозначим через $B = \|b_{ij}\|$ оценку матрицы $Q^{-1}(D)$ при фиксированной матрице решающих правил D. С учетом рандомизированного представления (10) адаптивные алгоритмы оптимизации проекционного типа приобретают вид

$$x[n] = \alpha;\ x[n+1] = \delta;\ u[n] = \beta;$$

$$D[n+1] = \Omega_{\varepsilon[n]}\{D[n] + \gamma[n]\,\omega[n]\,e_\alpha e_\beta^T\};$$
$$\omega[n] = \xi[n]\left(d_{\alpha\beta}^{-1}[n] + (1-\Delta_{1\beta})\,b_{1\alpha}[n]\right);$$
$$e_\alpha = (\underbrace{0, \ldots, 0}_{\alpha}, 1, 0, \ldots, 0)^T, \qquad (11)$$

где $\Omega_{\varepsilon[n]}\{\cdot\}$ — оператор проектирования на множество матриц D, удовлетворяющих условию $(D-E)\,e = 0$, $D \geqslant \varepsilon[n]\,ee^T$, $0 \leqslant \varepsilon[n] < 1$. Скалярные последовательности $\gamma[n]$, $\varepsilon[n]$ выбираются из условия сходимости алгоритма (11) по методике [2].

Реализация полученных алгоритмов оптимизации сопровождается нахождением матрицы оценок B. Воспользовавшись специфической структурой матрицы $Q(D)$, можно предложить рекуррентные алгоритмы вычисления оценок $B[n]$, не требующие восстановления всего набора условных переходных матриц, и избежать обращения матрицы.

Представим матрицу $Q(D)$ в рандомизированной форме:

$$Q(D) = I - M\{(1-\Delta_{1\beta})\,v_\alpha e_\beta^T\};$$
$$I = E + e_0 e_1^T = \sum_{i=1}^{N} e_i e_i^T + e_0 e_1^T, \quad e_0 = e - e_1; \qquad (12)$$
$$v_\alpha = (0, \ldots, 0, p_\alpha^{-1}, 0, \ldots, 0)^T.$$

С помощью рекуррентного алгоритма оценивания типа текущего среднего получим выражение для оценки матрицы $Q(D)$ (12):

$$\hat{Q}[n+1] = \frac{n}{n+1}\hat{Q}[n] + \frac{1}{n+1}(I - \chi[n]\,v[n+1]\,e^T[n+1]). \qquad (13)$$

Здесь $v[n]$, $e[n]$ — случайные реализации соответствующих векторов; $\chi[n] = 1$ при $x[n+1] \neq 1$; $\chi[n] = 0$ при $x[n+1] = 1$. Введем обозначения

$$S_m = \frac{n}{n+1}\hat{Q}[n] + \frac{1}{n+1}\sum_{i=1}^{m} e_i e_i^T, \quad m = 1, \ldots, N; \qquad (14)$$
$$R = S_N + \frac{1}{n+1} e_0 e_1^T.$$

Тогда из выражений (13), (14) следует, что

$$S_{m+1} = S_m + \frac{1}{n+1} e_{m+1} e_{m+1}^T, \quad S_0 = \frac{n}{n+1}\hat{Q}[n],$$
$$m = 0, \ldots, N-1;$$
$$Q[n+1] = R - \frac{1}{n+1}\chi[n+1]\,v[n+1]\,e^T[n+1].$$

Воспользовавшись известным матричным тождеством

$$(A + uv^T)^{-1} = A^{-1} - (1 + v^T A^{-1} u)^{-1} A^{-1} u v^T A^{-1},$$

запишем следующую систему соотношений для рекуррентного вычисления оценки $B[n] = \hat{Q}^{-1}[n]$:

$$B[n+1] = R^{-1} + (1+n-\chi[n+1]\,e^T[n+1]\,R^{-1}v[n+1])^{-1} \times$$
$$\times \chi[n+1]\,R^{-1}v[n+1]\,e^T[n+1]\,R^{-1};$$
$$R^{-1} = S_N^{-1} - (1+n+e_1^T S_N^{-1} e_0)^{-1} S_N^{-1} e_0 e_1^T S_N^{-1};$$
$$S_N^{-1} = S_{N-1}^{-1} - (1+n+e_N^T S_{N-1}^{-1} e_N)^{-1} S_{N-1}^{-1} e_N e_N^T S_{N-1}^{-1};$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$S_1^{-1} = \frac{n+1}{n}\,(B[n] - (n+e_1^T B[n]\,e_1)^{-1} B[n]\,e_1 e_1^T B[n]).$$

Предложенные алгоритмы отличаются простотой вычислительной реализации и могут использоваться в задачах адаптивного управления дискретными стохастическими объектами.

Список литературы: 1. *Майн Х., Осаки С.* Марковские процессы принятия решений.— М.: Наука, 1977.— 176 с. 2. *Любчик Л. М., Позняк А. С.* Обучающиеся автоматы в задачах управления стохастическими объектами.— Автоматика и телемеханика, 1974, № 5, с. 95—109.

3. *Yoursi M. El-Fattah.* Recursive Algorithms for Adaptive Control of Finite Marcov Chains. — IEEE Trans. on Systems, Man and Cybernetics, 1981, **SMC-11**, N 2, p. 135 — 144.

Поступила в редколлегию 10.10.83.

УДК 62-50

А. Н. СИРЕНКО

ИДЕНТИФИКАЦИЯ ПРОЦЕССОВ С АПРИОРНО НЕИЗВЕСТНЫМИ ДЕТЕРМИНИРОВАННЫМИ ОСНОВАМИ

Рассмотрим задачу определения характера и структуры взаимосвязей между анализируемыми показателями, характеризующими состояние или поведение статистически обследуемого процесса. Для простоты ограничимся изучением процесса, в котором результирующий показатель $S_e(x_i, y_j)$ является функцией двух факторов-аргументов: $x_i (i = 1, 2, \ldots, m)$ и $y_j (j = 1, 2, \ldots, n)$. Предположим, что $S_e(x_i, y_j) = S(x_i, y_j) + \varepsilon_{ij}$, где $S(x_i, y_j)$ — истинная модель процесса; ε_{ij} — некоррелированные случайные числа, распределенные по нормальному закону с нулевым средним ($i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$).

Для получения оптимальных оценок $\hat{S}(x_i, y_j)$ (несмещенных, состоятельных, эффективных) при таких предположениях следует решать вариационную задачу

$$\sum_{i=1}^{m}\sum_{j=1}^{n}[S_e(x_i, y_j) - S(x_i, y_j)]^2 \to \min, \qquad (1)$$

которую рассмотрим для случая $S(x_i, y_j) = \lambda f_1(x_i) f_2(y_j)$, когда на искомые функции $f_1(x_i)$, $f_2(y_j)$ наложены ограничения

$$\sum_{i=1}^{m}[f_1(x_i)]^2 = 1; \quad \sum_{j=1}^{n}[f_2(y_j)]^2 = 1. \qquad (2)$$