УДК 519.21

П.Р. ШПАК, Я.И. ЕЛЕЙКО

ОПТИМАЛЬНЫЕ СТРАТЕГИИ И ОЦЕНКА ПОЛУНЕПРЕРЫВНОГО ОБРЫВНОГО УПРАВЛЯЕМОГО МАРКОВСКОГО ПРОЦЕССА

Аннотация. Рассмотрены обрывные управляемые марковские процессы с несчетными множествами состояний и управлений на конечном промежутке времени. Приведены определения обрывного управляемого марковского процесса, оценки пути и оптимальной стратегии, а также доказано фундаментальное уравнение в случае, когда множествами состояний и управлений являются измеримые пространства. Предложен метод построения оптимальной стратегии и доказано существование равномерно оптимальной стратегии в случае, когда множествами состояний и управлений есть сепарабельные метрические пространства.

Ключевые слова: обрывной управляемый марковский процесс, оптимальная стратегия, равномерно оптимальная стратегия, оценка пути, фундаментальное уравнение.

ВВЕДЕНИЕ

Определение управляемого марковского процесса впервые введено в книге Беллмана [1], в которой автор применяет принципы динамического программирования к этим стохастическим процессам. Управляемые марковские процессы подробно описаны в работе [2], где даны определения процесса и его оценки, а также оптимальной и ε-оптимальной стратегий и их оценки. Однако здесь модели не учитывают фактора риска, т.е. вероятности банкротства в какой-то определенный момент времени. Некоторые основные идеи обрывных управляемых марковских процессов рассмотрены в [3], а процессы с конечными или счетными множествами состояний и наборов действий описаны в [4, 5].

ОБЩЕЕ ОПРЕДЕЛЕНИЕ МОДЕЛИ

Пусть $X = \bigcup\limits_{t=m}^{n} X_t$ и $A = \bigcup\limits_{t=m+1}^{n} A_t$ — измеримые пространства.

Определение 1. Траекторию $l = x_m a_{m+1} x_{m+1} \dots a_n x_n$ будем называть путем и пространство всех возможных путей обозначать $L = X \times (A \times X)^{n-m}$.

Определение 2. Обрывным управляемым марковским процессом на конечном промежутке времени $[m, n]$ называется набор $(X, A, j, p, q, r, c, \mu) \equiv Z_\mu$, где:

1) множество состояний $X = \bigcup\limits_{t=m}^{n} X_t$ является измеримым пространством и подмножество обрывных состояний $X^* \subset X$ измеримо, а также $X_m, X_{m+1}, \dots, X_n$ — непересекающиеся подмножества X;

2) множество всех пар xx $(x \in X_t)$ принадлежит $\sigma(X_t \times X_t)$ $(m \le t \le n)$;

3) множество управлений $A = \bigcup\limits_{t=m+1}^{n} A_t$ является измеримым пространством и $A_{m+1}, A_{m+2}, \dots, A_n$ — непересекающиеся подмножества A;

4) $j : A \to X$ — соответствия проекции, $j(A_{t+1}) = X_t$;

5) $p(\cdot \mid a) = \mathbb{P}(x_t = x \mid a_t = a, x_{t-1})$ — распределения вероятностей на X_t;

6) $q : A \to \mathbb{R}$ — функция на множестве управлений (текущая плата);

7) $r : X_n \to \mathbb{R}$ — функция на множестве финальных состояний (финальная плата);

8) $c : X_t \cap X^* \to \mathbb{R}$ — функция на множестве обрывных состояний

$$c(x) = -\sum_{t=m+1}^{k} \sup_{a \in A_t} q(a), \ x \in X_t \cap X^*;$$

9) μ — распределение вероятностей на X_m (начальное распределение), причем если начальное распределение μ сосредоточено в одной точке x, то вместо Z_μ будем писать Z_x.

Если удовлетворяются только условия 1–8, то будем называть этот объект моделью и обозначать Z.

Определение 3. Функцию $I : L \to \mathbb{R}$ будем называть оценкой пути l, если она удовлетворяет следующим условиям:

$$I(l) = \sum_{t=m+1}^{n} q(a_t) + r(x_n), \ x_t \notin X^* \quad (t = m, m+1, \dots, n),$$

$$I(l) = \sum_{t=m+1}^{k} q(a_t) + c(x_k), \ x_k \in X^* \wedge x_t \notin X^* \quad (t = m, m+1, \dots, k).$$

Определение 4. Если подмножество $A(x) \subset A$ является подмножеством всех возможных управлений в состоянии $x \in X$, то $\varphi : X \to A(x)$ называется простой стратегией при $\varphi(x_{t-1}) = a_t, \ t = (m+1, \dots n)$.

Определение 5. Соответствия $\pi : H \to \pi(\cdot \mid h \in H)$, где $\pi(\cdot \mid h \in H)$ — распределение вероятностей на $A(x_t)$ и H — пространство историй $(h \in H \Leftrightarrow h = x_m a_{m+1}, \dots, a_t x_t)$, называются стратегией.

Если заданы переходная функция $p(\cdot \mid a)$ и стратегия $\pi(\cdot \mid h)$, то каждому начальному распределению μ соответствуют распределения вероятностей P на пространстве L, определенные формулой:

$$P(dx_m da_{m+1} dx_{m+1} da_{m+2} \dots dx_{n-1} da_n dx_n) = \mu(dx_m)\pi(da_{m+1} \mid x_m) \times$$

$$\times\, p(dx_{m+1} \mid a_{m+1}) \dots p(dx_{n-1} \mid a_{n-1})\pi(da_n \mid x_m a_{m+1} \dots x_{n-1}) p(dx_n \mid a_n).$$

Для каждой функции f, определенной на пространстве L, математическое ожидание f имеет вид

$$Ef = \int_{X_m} \mu(dx_m) \int_{A(x_m)} \pi(da_{m+1} \mid x_m) \int_{X_{m+1}} p(dx_{m+1} \mid a_{m+1}) \dots \int_{X_{n-1}} p(dx_{n-1} \mid a_{n-1}) \times$$

$$\times \int_{A(x_{n-1})} \pi(da_{m+1} \mid x_m a_{m+1} \dots x_{n-1}) \times$$

$$\times \int_{X_n} p(dx_n \mid a_n) f(x_m a_{m+1} x_{m+1} a_{m+2} \dots x_{n-1} a_n x_n). \qquad (1)$$

Примером такой функции есть оценка пути l, ее математическое ожидание обозначим ω:

$$\omega = EI(l). \qquad (2)$$

Определение 6. Величину ω из (2) будем называть оценкой стратегии π $(\omega = \omega(\pi))$.

Определение 7. Величину $v = \sup\limits_{\pi} \omega(\pi)$ будем называть оценкой обрывного управляемого марковского процесса Z_μ или оценкой начального распределения μ.

Определение 8. Стратегия π называется оптимальной, если $\omega(\pi) = v$.

Определение 9. Стратегия π называется равномерно оптимальной, если π оптимальна для каждого начального распределения μ.

Определение 10. Модель $(X', A', j, p, q, r, c) \equiv Z'$, где $X' = \bigcup\limits_{t=m+1}^{n} X_t$ и $A' = \bigcup\limits_{t=m+2}^{n} A_t$, называется производной.

Если f — оценка пути, то для существования математического ожидания (1) необходимо, чтобы функции q, r и c были измеримы и ограничены сверху, а также $p(\cdot \mid a)$ должна быть измерима по a и $\pi(\cdot \mid h)$ — по h.

Утверждение 1. Справедливо следующее уравнение:

$$\omega(x, \pi) = \int\limits_{A(x)} \pi(da \mid x)(q(a) + \omega'(p_a, \pi_a)), \tag{3}$$

где $p_a = p(\cdot \mid a)$, $\pi_a(\cdot \mid h') = \pi(\cdot \mid yah')$, $a \in A_{m+1}$, $y = j(a)$, h' — история в производной модели Z'.

Уравнение (3) называется фундаментальным и выражает оценку ω стратегии π в модели Z через оценку ω' стратегий в Z'.

Доказательство. Из определения P, (1) и (2) следует, что

$$\omega(\mu, \pi) = \int\limits_{X_m} \omega(x, \pi)\mu(dx),$$

следовательно,

$$\omega'(p_a, \pi_a) = \int\limits_{X_{m+1}} \omega'(y, \pi_a) p(dy \mid a).$$

Из определения (2) получаем следующее равенство:

$$Ef(x_{m+1}a_{m+1} \dots x_n) = \int\limits_{A(x)} E_a f(x_{m+1}a_{m+1} \dots x_n)\pi(da \mid x),$$

где E — начальное состояние x и стратегии π в модели Z, а E_a — начальное распределение p_a и стратегии π_a в производной модели Z'. Поскольку $I(xal') = q(a) + I(l')$, где l' — путь в производной модели Z', получаем следующее равенство:

$$\omega(x, \pi) = \int\limits_{A(x)} \pi(da \mid x)(q(a) + \omega'(p_a, \pi_a)).$$

В конечном и счетном случаях [4, 5] для построения рекурсивного метода нахождения оптимальной стратегии использовались операторы U и V. В общем случае эти операторы будут иметь вид

$$Uf(a) = q(a) + \int\limits_{X} f(y)\, p(dy \mid a),$$

$$Vg(x) = \sup_{a \in A(x)} g(a).$$

В общем случае оператор может переводить измеримые функции в неизмеримые. Одним из способов решения этой проблемы является использование только измеримых функций из некоторого класса $\mathcal{L}$, инвариантного относительно операторов U и V.

ПОЛУНЕПРЕРЫВНЫЙ СЛУЧАЙ

Определение 11. Функция f определена на метрическом пространстве E и называется полунепрерывной сверху, если множество $\{x : f(x) \geq c\}$ замкнуто. Множество всех полунепрерывных сверху функций на E будем обозначать $\mathcal{L}(E)$.

Определение 12. Модель Z называется полунепрерывной, если:

— множества состояний X и управлений A — сепарабельные метрические пространства;

— множества $X_m \cap X^*$, $X_m \setminus X^*$, $X_{m+1} \cap X^*$, $X_{m+1} \setminus X^* \ldots X_n \cap X^*$, $X_n \setminus X^*$ — замкнутые подмножества X и $A_{m+1}, A_{m+2}, \ldots, A_n$ — замкнутые подмножества A;

— соответствие $A(x)$ квазинепрерывное (если $x_k \to x \in X$ и $a_k \in A(x_k)$, то $\{a_k\}$ имеет предельную точку, принадлежащую $A(x)$);

— если $f \in \mathcal{L}(X_t)$ и $g(a) = \int\limits_{X_t} p(dx \mid a) f(x)$ $(a \in A_t)$, то $g \in \mathcal{L}(A_t)$ $(t = m+1, \ldots, n)$;

— функция q, определенная на A_t, принадлежит $\mathcal{L}(A_t)$ и q, определенная на $X_t \cap X^*$, принадлежит $\mathcal{L}(X_t \cap X^*)$, а r принадлежит $\mathcal{L}(X_n)$.

Теорема 1. Пусть E и E' — сепарабельные метрические пространства, $Q(x)$ — квазинепрерывное соответствие из E в E'. Если $f \in \mathcal{L}(E')$, то функция $g(x) = \sup\limits_{y \in Q(x)} f(y)$ $(x \in E)$ принадлежит $\mathcal{L}(E)$, множества $\overline{Q}(x) = \{y : y \in Q(x), f(y) = g(x)\}$ $(x \in E)$ непустые и соответствие $\overline{Q}(x)$ допускает измеримый выбор.

Утверждение 2. Для полунепрерывной модели имеют место следующие свойства:

1) оценка v принадлежит $\mathcal{L}(X_m)$;

2) оценка $v(\mu) = \mu v$ для любого начального распределения μ;

3) существует равномерно оптимальная стратегия.

Доказательство. Допустим, что свойства 1–3 справедливы для производной модели Z'. Покажем, что следующие условия справедливы для модели Z:

а) оценка v модели Z выражается через оценку v' производной модели Z' уравнениями $v = Vu$, $u = Uv'$, где операторы U и V заданы формулами

$$Uf(a) = q(a) + \int\limits_X f(y) p(dy \mid a) \quad (a \in A),$$

$$Vg(x) = \sup_{a \in A(x)} g(a) \quad (x \in X \setminus (X_n \cup X^*));$$

б) существует измеримый селектор ψ соответствия $A(x)$ из X_m в A_{m+1} такой, что $u(\psi(x)) = v(x)$;

в) если π' — оптимальная стратегия Z' и ψ — селектор из условия б), то $\psi\pi'$ — оптимальная стратегия для модели Z;

г) свойства 1–3 справедливы для Z.

В случае, когда пространство состояний состоит из одного множества X_n, свойства 1–3 тривиальны.

Из фундаментального уравнения получаем следующее:

$$\omega(x,\pi) \le Vu(x) \quad (x \in X_m), \tag{4}$$

где π — любая стратегия и $u(a) = q(a) + v'(p_a)$ $(a \in A_{m+1})$.

Поскольку свойства 1, 2 справедливы, получаем, что $v' \in \mathcal{L}(X_{m+1})$ и $v' = \int\limits_{X_{m+1}} v'(y)p(dy \mid a)$, следовательно, $u \in \mathcal{L}(A_{m+1})$ и $u = Uv'$.

Построим стратегию π, которая превращает неравенство (4) в равенство. Пусть π' — оптимальная стратегия для производной модели Z', тогда для любой стратегии $\gamma\pi'$ справедливы следующие уравнения:

$$\omega(x,\gamma\pi') = \int\limits_{A(x)} \gamma(da \mid x)(q(a) + \omega'(p_a,\pi')) =$$

$$= \int\limits_{A(x)} \gamma(da \mid x)(q(a) + v'(p_a)) = \int\limits_{A(x)} u(a)\gamma(da \mid x).$$

Если $\gamma(\cdot \mid x)$ сосредоточено в одной точке с $\overline{A}(x) = \{a : a \in A(x), u(a) = Vu(x)\}$, то

$$Vu(x) = \sup\limits_{A(x)} u(a) = \int\limits_{A(x)} u(a)\gamma(da \mid x).$$

Произведение $\psi\pi'$ будет стратегией, если ψ — измеримый селектор $u \in \mathcal{L}(A_{m+1})$. Согласно теореме 1 измеримый селектор ψ существует. Из равенства $\omega(x,\psi\pi') = Vu(x)$ и (4) следует, что $v = Vu$, значит, условие а) справедливо.

Очевидно, что селектор ψ соответствия $A(x)$ удовлетворяет условию б) тогда и только тогда, когда ψ — измеримый селектор соответствия $\overline{A}(x)$. В результате показано, что условия б) и в) справедливы.

Покажем, что свойства 1–3 справедливы для модели Z. Свойство 3 справедливо в силу построения стратегии $\psi\pi'$. Поскольку $u \in \mathcal{L}(A_{m+1})$ и $v = Vu$, из теоремы 1 следует, что свойство 1 справедливо. Если π равномерно оптимальна для модели Z, то

$$v(\mu) = \omega(\mu,\pi) = \int\limits_{X_m} \omega(x,\pi)\mu(dx) = \int\limits_{X_m} v(x)\mu(dx) = \mu v,$$

следовательно, свойство 2 справедливо.

ЗАКЛЮЧЕНИЕ

Таким образом, в работе доказано, что имеет место фундаментальное уравнение в случае, когда множествами состояний и управлений являются измеримые пространства. Также предложен метод построения оптимальной стратегии и доказано существование равномерно оптимальной стратегии в случае, когда множествами состояний и управлений есть сепарабельные метрические пространства.

СПИСОК ЛИТЕРАТУРЫ

1. Bellman R.E. Dynamic programming. — Princeton (NJ): Princeton University Press, 1957. — 400 p.
2. Дынкин Е.Б., Юшкевич А.А. Управляемые марковские процессы и их приложения. — М.: Наука, 1975. — 334 с.
3. Pakes A.G. Killing and resurrection of Markov processes // Stochastic Models. — 1997. — **13**, N 2 — P. 255–269.

4. Пароля Н.Р., Єлейко Я.І. Обривні керовані марковські процеси на скінченному інтервалі часу для скінченних моделей // Вісник Львівського університету. Серія механіко-математична. — 2010. — № 72. — С. 243–254.
5. Parolya N.R., Yeleyko Y.I. Killed Markov decision processes on finite time interval for countable models // Transactions of NAS of Azerbaijan. — 2010. — **30**, N 4. — P. 141–152.

Надійшла до редакції 12.11.2015

П.Р. Шпак, Я.І. Єлейко

ОПТИМАЛЬНІ СТРАТЕГІЇ ТА ОЦІНКА НАПІВНЕПЕРЕВНИХ ОБРИВНИХ КЕРОВАНИХ МАРКОВСЬКИХ ПРОЦЕСІВ

Анотація. Розглянуто обривні керовані марковські процеси з незліченними множинами станів та керувань на скінченному часовому інтервалі. Наведено означення обривного керованого марковського процесу, оцінки шляху та оптимальної стратегії, а також доведено істинність фундаментального рівняння за умов, коли множини станів та керувань є вимірними просторами. Наведено метод побудови рівномірно оптимальної стратегії у випадку, коли множини станів та керувань являють собою сепарабельні метричні простори.

Ключові слова: обривний керований марковський процес, оптимальна стратегія, рівномірно оптимальна стратегія, оцінка шляху, фундаментальне рівняння.

P.R. Shpak, Y.I. Yeleyko

ASSESSMENT AND OPTIMAL POLICIES OF SEMI-CONTINUOUS KILLED MARKOV DECISION PROCESSES

Abstract. In the paper, we consider killed Markov decision processes with uncountable sets of states and controls on a finite time interval. Definitions of killed Markov decision process and assessment of the way and optimal policy are given, as well as fundamental equation is proved in the case where the set of states and set of controls are measurable spaces. We also proposed a method to construct the optimal strategy and proved the existence of a uniformly optimal policy in case where the set of states and set of controls are separable metric spaces.

Keywords: killed Markov decision process, optimal policy, uniformly optimal policy, assessment of the way, fundamental equation.

Шпак Павел Романович,
аспирант Львовского национального университета имени Ивана Франко,
e-mail: prshpak@gmail.com.

Елейко Ярослав Иванович,
доктор физ.-мат. наук, профессор Львовского национального университета имени Ивана Франко,
e-mail: yikts@yahoo.com.