УДК 519.857.3: (519.24+62.50)

АДАПТИВНЕ КЕРУВАННЯ СЛАБКОКЕРОВАНИМИ МАРКОВСЬКИМИ ТА НАПІВМАРКОВСЬКИМИ МОДЕЛЯМИ В ДИСКРЕТНОМУ ЧАСІ

М.В. АНДРЄЄВ

Досліджується байєсів підхід до проблеми марковських процесів рішень [1] в умовах стохастичної невизначеності, коли невідомі перехідні ймовірності слабко збурені, і тільки збурення залежать від стратегії рішень. Процес рішень припускається стаціонарним, розглядається в дискретному часі з скінченним, зчисленним або вимірним фазовим простором і ґрунтується на принципі розділення задач оцінювання та оптимізації.

ВСТУП

Процеси, що відбуваються в природі, житті, сферах людської діяльності характеризуються певною стійкістю, коли йдеться про природу, та деяким усталеним режимом функціонування, якщо маються на увазі процеси, які є результатом людської діяльності. Як приклади наведемо процеси фотосинтезу, що протікають звичайно у природних умовах, процеси функціонування технічних пристроїв або систем (в нормальних умовах їх експлуатації), економічні та фінансові процеси вирівнювання попиту та пропонування в ринкових умовах, тощо.

Природа більшості із наведених процесів носить стохастичний характер, і еволюцію кожного процесу із цієї більшості можна описати матрицею перехідних ймовірностей P_0 у випадку, коли процес характеризується марковською властивістю, яка символічно виражається у незалежності в еволюційному плані майбутнього від минулого при відомому теперішньому стані процесу. Такі поняття як стійкість процесу або усталений режим функціонування системи, що описується цим процесом, накладає деякі умови на матрицю P_0, зокрема, вимагається наявність існування у неї стаціонарного або ергодичного розподілу ρ_0.

Однак на такий процес можуть впливати деякі зовнішні фактори, що призводять до збурення матриці P_0. Будь-яку стратегію збурень можна охарактеризувати деякою стаціонарною нерандомізованою стратегією f, якій відповідає матриця збурень P_1^f. Будемо вважати, що марковський процес, який розглядається, є слабкокерованим, якщо його еволюція описується матрицею ймовірностей переходу

$$P_{\varepsilon}^{f} = P_0 - \varepsilon P_1^{f} , \qquad (1)$$

де ε — малий параметр, тобто $0 \le \varepsilon \le 1$.

Звідси випливає, що слабкокерований процес описується матрицею P_{ε}^{f} (1), яка обумовлена малим збуренням εP_1^{f} матриці P_0, що задає незбурений некерований процес.

Щодо інтерпретацій матриць P_{ε}^{f} та P_1^{f} в (1), то на прикладі технічної системи обслуговування матриця P_{ε}^{f} характеризує потенційну можливість виходу системи із ладу — можливість її неполадки [2], а матриця збурень P_1^{f}, що відповідає стратегії f, характеризує план або обсяг профілактичних та ремонтних заходів, необхідних для нормального функціонування системи. Інші інтерпретації процесів, що описуються матрицями P_0, P_1^{f} та P_{ε}^{f} можна знайти в роботах [3, 4].

Будь-якій стратегії керування можна надати ціну (оцінку) або охарактеризувати її деяким функціоналом якості. Наша задача полягає в тому, щоб знайти оптимальну стратегію, тобто таку стратегію, якій можна дати найвищу оцінку або охарактеризувати її екстремальними значеннями функціонала якості.

З огляду на задання слабкокерованого процесу формулою (1) та наявності деякого функціонала якості стратегії керування слід зазначити, що відповідна стратегії f ціна u^{f} задовольняє деяке лінійне функціональне рівняння, а ціна u^{*} оптимальної стратегії f^{*} задовольняє нелінійне рівняння оптимальності типу Беллмана.

Щодо практичного використання задання (1), то, як правило, в цьому заданні матриці P_0 і P_1^{f} є невідомими, і в зв'язку з цим першочерговою задачею постає статистична задача ідентифікації елементів матриць P_0 і P_1^{f} на основі неспостережуваної але оцінюваної реалізації незбуреного процесу X, що відповідає матриці P_0, та спостережуваної реалізації процесу y_{ε}^{f}, динаміка якого описується матрицею P_{ε}^{f}, що відповідає стратегії керування f.

Оскільки завдяки стратегії керування f збуреним процесом забезпечується відповідна з точки зору критерію «витримування» усталеного режиму функціонування, який задається ідентифікованою оцінкою матриці P_0, яка описує динаміку реального незбуреного процесу, то ця стратегія уже наразі за визначенням є стратегією адаптивного керування. Таким чином, з точки зору статистичного аналізу, ми маємо справу з двохкомпонентною послідовністю з неповною інформацією або частково спостережуваною марковською послідовністю (x, y^f), яку за допомогою підходу, викладеного в роботі [5], можна звести до двохкомпонентної послідовності з повною інформацією або до повністю спостережуваної узагальненої марковської послідовності (v^f, y^f).

Для оцінки елементів матриць P_0 та P_ε^f використовується підхід, аналогічний тому, суть якого стосовно керованих марковських гауссових послідовностей викладена в роботі [6]. Цей підхід характеризується двома етапами: на першому — перевіряється низка статистичних гіпотез щодо адекватної приналежності спостережуваних реалізацій реальних процесів у дискретному часі до марковських гауссових послідовностей, на другому — оцінюються функції перехідних ймовірностей для цих послідовностей.

Щодо задання (1) динаміка процесу x у фазовому просторі X описується матрицею P_0, а динаміка процесу y^f у фазовому просторі Y описується матрицею P_ε^f.

ЗВЕДЕННЯ КЕРОВАНИХ МАРКОВСЬКИХ ПОСЛІДОВНОСТЕЙ З НЕПОВНОЮ ІНФОРМАЦІЄЮ ДО КЕРОВАНИХ МАРКОВСЬКИХ ПОСЛІДОВНОСТЕЙ З ПОВНОЮ ІНФОРМАЦІЄЮ

За кожною послідовністю з неповною інформацією будується деяка послідовність з повною інформацією так, щоб оцінки відповідних стратегій співпадали.

Ідея полягає в тому, що вводяться нові фазові простори станів, розглядаючи як стан в момент t всю суттєву для подальшого керування інформацію, якою ми володіємо в цей час. У початковий момент m ця інформація описується спостережуваним станом y_m і заздалегідь заданим апріорним розподілом v_m для неспостережуваного стану x_m. У будь-який момент $t > m$ її природно описати парою $v_t y_t$, де v_t — апостеріорний розподіл ймовірностей для стану x_t, обчислений з урахуванням всієї історії $\{y_m\}$, яку спостерігали до цього моменту.

Розглядається випадок, коли простори X_t, Y_t — скінченні. Тоді ймовірність траєкторії $l = a_m x_m y_m^f,\ a_{m+1} x_{m+1} y_{m+1}^f \dots a_n x_n y_n^f$ за початковим розподілом μ і довільної стаціонарної рандомізованої стратегії π визначається як

$$\begin{aligned} P_\mu^\pi(l) = \mu(x_m y_m)\,\pi\left(a_{m+1}\middle|v_m y_m\right) P\left(x_{m+1} y_{m+1}\middle|x_m a_{m+1}\right)\dots \\ \dots\pi\left(a_n\middle|v_m y_m a_{m+1} v_{m+1}\dots a_{n-1} y_{n-1}\right) P\left(x_n y_n\middle|x_{n-1} a_n\right), \end{aligned} \tag{2}$$

де розподіл v_m обчислюється за формулою

$$v_m = v_m\left(x_m\middle|y_m\right) = \frac{\mu\left(x_m y_m\right)}{\sum\limits_{z \in X_m} \mu\left(z, y_m\right)}. \tag{3}$$

Якщо знаменник дорівнює 0, то можна прийняти за $v_m\left(\cdot|y_m\right)$ будь-яку ймовірносну міру на X_m, наприклад, раз і назавжди вибрану міру v_m^0.

Побудова допоміжної послідовності з повною інформацією починається з вибору просторів станів $\widetilde{Y}_t$. Покладають $\widetilde{Y}_t = Y_t \times N_t$, де N_t — сукупність всіх ймовірносних мір на множині X_t (із N_t приймають значення розподіли v_t).

Керування у новій послідовності залишаються такими, якими були і раніше. Одне і те ж керування a_t можливе тепер при різних станах $\widetilde{y}_{t-1} = v_{t-1}y_{t-1}$, що відрізняються розподілами v_{t-1}.

Щоб задати нову перехідну функцію $\widetilde{p}$, необхідно співставити кожній парі $v_{t-1}a_t$ розподіл ймовірностей у просторі $Y_t \times N_t$. Вихідна перехідна функція задає розподіл у просторі $X_t \times Y_t$ як функцію від $x_{t-1}a_t$. Зафіксуємо $v_{t-1}a_t$. Природно співставити цій парі розподіл у просторі $X_t \times Y_t$, що визначається формулою

$$\widetilde{p}\left(x_t y_t|v_{t-1}a_t\right) = \sum_{v_{t-1}N_{t-1}} p\left(x_t y_t|x_{t-1}a_t\right)v_{t-1}\left(x_{t-1}\right). \tag{4}$$

За теоремою множення ймовірносних розподілів цей розподіл подається у вигляді добутку розподілу в X_t на розподіл у Y_t

$$\widetilde{p}\left(x_t y_t|v_{t-1}a_t\right) = v_t\left(x_t|v_{t-1}a_t y_t\right)\widetilde{p}\left(y_t|v_{t-1}a_t\right). \tag{5}$$

Тут перший множник має вигляд

$$v_t\left(x_t|v_{t-1}a_t y_t\right) = \frac{\widetilde{p}\left(x_t y_t|v_{t-1}a_t\right)}{\widetilde{p}\left(y_t|v_{t-1}a_t\right)}, \tag{6}$$

а другий

$$\widetilde{p}\left(y_t|v_{t-1}a_t\right) = \sum_{x_t \in Y_t} p\left(x_t y_t|v_{t-1}a_t\right). \tag{7}$$

Якщо знаменник (6) перетворюється в 0, то за v_t приймається фіксована міра v_t^0 на X_t. Формула (6) визначає відображення $y_t \to v_t$, і тим самим задається розподіл у просторі $Y_t \times N_t$, якщо при цьому приймається для другої компоненти розподіл (7) і вважається, що перша компонента є функцією другої за формулою (6). Таким чином визначається ймовірносний розподіл у просторі $Y_t \times N_t = \widetilde{Y}_t$, який залежить від $v_{t-1}a_t$, тобто перехідна функція $\widetilde{p}$ з $v_{t-1} \times A_t$ в $\widetilde{Y}_t$.

Ймовірносний розподіл v_t має бути апостеріорним розподілом x_t для керованих марковських послідовностей з неповною інформацією $\left(x, y^{\pi}\right)$, тобто розподіл v_t у (6) є апостеріорним розподілом x_t з урахуванням усіх спостережень (y^{π}), здійснених до моменту t. Іншими словами, має виконуватись формула

$$v_t = P_{\mu}^{\pi}\left\{x_t \middle| y_m a_{m+1} y_{m+1} \ldots a_t y_t\right\} =$$

$$= \frac{P_{\mu}^{\pi}\left(x_m a_{m+1} x_{m+1} \ldots a_t x_t y_t\right)}{\sum\limits_{z \in Y_t} P_{\mu}^{\pi}\left(x_m a_{m+1} x_{m+1} \ldots a_t x_t z_t\right)}. \quad (8)$$

При $t = m$ це вірно за формулою (3), при $t > m$ перевіряється за індукцією формулами (2) та (4) — (7).

Оскільки ймовірносна міра v_t на x виражається через y^t, яка індукується матрицею P_{ε}^{f} в заданні (1), то в подальшому для простоти викладок марковську послідовність з повною інформацією позначатимемо через $x_{\varepsilon} = \left\{x_n^{\varepsilon},\ n \geq 0\right\}$.

Слабкокеровані марковські та напівмарковські моделі з дискретним часом в умовах повної інформації

Розглянемо слабкокеровані марковські та напівмарковські системи з скінченним, зчисленним або вимірним фазовим простором станів (ФПС) E та компактним простором керувань або рішень A. Математичними моделями цих систем є слабкокеровані марковські та напівмарковські процеси в дискретному часі.

Нехай F — компактна множина марковських стаціонарних нерандомізованих стратегій f, які представляють собою у загальному випадку функції, що відображають E в A. Для стратегії f вводяться поняття слабкокерованої марковської моделі (СКММ), яка задається набором $\left(E, A, P_{\varepsilon}^{f}, r^{f}\right)$, де P_{ε}^{f} — матриця ймовірностей переходу вигляду (1), причому $P_0 = \left\{p_{kr}^{0};\ k, r \in E\right\}$ — матриця ймовірностей переходу незбуреного ланцюга Маркова; $P_1^{f} = \left\{p_{xr}^{1, f(k)};\ k, r \in E\right\}$ — матриця збурень, елементи якої залежать від керувань; $r^{f} = \left\{r_k^{f(k)};\ k \in E\right\}$ — вектор або функція однокрокового доходу, індукованого СКММ, що відповідає стратегії f.

Набір $\left(E, A, P_{\varepsilon}^{f}, r^{f}\right)$ характеризує також слабкокеровану напівмарковську модель (СКНММ) в дискретному часі, або, що те ж саме, слабкокеровану модель марковського відновлення (СКММВ), яка описується слабкокерованим процесом марковського відновлення з доходом (ПМВД). При цьому P_{ε}^{f} — матриця ймовірностей переходу (МЙП) вкладеного ланцюга Маркова (ВЛМ), а r^{f} — функція однокрокового доходу, що відповідає стратегії f. Таким чином, набір $\left(E, A, P_{\varepsilon}^{f}, r^{f}\right)$ задає СКНММ в дискретному часі або, що те ж саме, СКММВ.

Подання матриці P_{ε}^{f} у вигляді різниці стохастичної матриці P_0 та матриці збурень P_1^f, помноженої на малий параметр ε у формулі (1), дозволяє розглянути два типи моделей: 1) моделі зі станом поглинання у випадку, коли $P_1^f 1 > 0$ (1 — одиниця в E або вектор з одиничними компонентами розмірності E), та напівстохастичною збуреною матрицею P_{ε}^{f}; 2) моделі без станів поглинання, коли $P_1^f 1 = 0$. При цьому в першому випадку критерій якості керування (або ціна стратегії) визначається середнім значенням адитивного функціонала, заданого на траєкторіях моделі на відрізку часу $[0, \tau_{\varepsilon}]$, де τ_{ε} — момент поглинання (попадання в стан поглинання), в той час, як у другому випадку розглядається неадитивний функціонал на траєкторіях моделі, яка функціонує на нескінченному інтервалі часу. Ясно, що для існування адитивного та неадитивного функціоналів доводиться вимагати, щоб елементи $E, A, P_{\varepsilon}^{f}$ та r^f, які характеризують відповідні моделі, задовольняли певні умови.

АДИТИВНИЙ КРИТЕРІЙ. СЛАБКОКЕРОВАНІ МОДЕЛІ З МАЛОЮ ЙМОВІРНІСТЮ ПОГЛИНАННЯ

Розглянемо випадок, коли в поданні матриці P_{ε}^{f} (1) матриця збурень P_1^f задовольняє умові $P_1^f 1 > 0$, де 1 — вектор з одиничними компонентами розмірності E, і матриця P_{ε}^{f} є напівстохастичною, тобто для СКМП з ймовірністю більшою нуля існує стан поглинання.

Слабкокеровані марковські послідовності з малою ймовірністю поглинання

Слабкокерована марковська послідовність (СКМП) $\{x_n^{\varepsilon}, a_n, n \geq 0\}$ із скінченним або зчисленним ФПС $E_0 := E \cup \{0\}$, де $E = \{1, 2, \ldots, N, \ldots\}$, $\{0\}$ — стан поглинання, та компактною множиною керувань A, задається МЙП (1), у якої $f(x_n^{\varepsilon}) = a_n \in A$.

Незбурена марковська послідовність $\{x_n^0,\ n \geq 0\}$ з матрицею $P_0 = \{p_{kr}^0;\ k, r \in E\}$ є ергодичною зі стаціонарним розподілом $\rho = \{\rho_k;\ k \in E\}$, $\rho_k > 0$, $k \in E$.

Елементи матриці збурень P_1^f залежать від керувань та задовольняють умови $\exists\, k \in E:\ p_{k_0}^{1, f(k)} > 0$; $p_{00}^0 - \varepsilon\, p_{00}^{1, f(0)} = 1$.

Введемо поняття доходу, пов'язаного з функціонуванням СКМП. Якщо в стані $k \in E$ прийнято рішення або керування $a \in A$, то очікуваний однокроковий доход задається функцією $r_k(a)$, обмеженою за k, a.

Критерій оптимальності стратегії f визначається функціоналом, заданим на траєкторіях СКМП, і має вигляд

$$u_\varepsilon^f(k) = M_\varepsilon^f\{L \mid x_0 = k\}, \quad L = \sum_{n=1}^{\tau_\varepsilon} r_{x_{n-1}}(a_n) + \psi(a_{\tau_\varepsilon}), \tag{9}$$

або у векторній формі $u_\varepsilon^f = M_\varepsilon^f L$, де u_ε^f — ціна стратегії f; τ_ε — момент попадання СКМП у стан поглинання $\{0\}$; ψ — обмежена функція на A, що характеризує штраф у момент поглинання.

Стратегія f^* — оптимальна, якщо

$$u_\varepsilon^{f^k(k)}(k) - u_\varepsilon^{f(k)} \geq 0, \quad \forall f \in F, \ \forall k \in E.$$

Відповідна стратегії f^* ціна $u_\varepsilon^{f^*}$ позначається u_ε^*.

Рівняння Беллмана для оптимальної середньої ціни має вигляд

$$u_\varepsilon^* = \sup_{f \in F} \left\{\omega^f + P_\varepsilon^f u_\varepsilon^f\right\}, \tag{10}$$

де

$$\omega^f = \left\{r_k\big(f(k)\big) + \varepsilon\, p_{k_0}^{1, f(k)} \psi\big(f(k)\big); \ k \in E_0\right\}. \tag{11}$$

Метод розв'язання рівняння оптимальності (10) передбачає одночасне виконання операцій обернення та оптимізації за стратегією f оператора $\left(I - P_\varepsilon^f\right)$. Операція обернення ґрунтується на методах теорії збурення звідно-оборотного оператора $\left(I - P_\varepsilon^f\right)$ на спектрі, в той час як в основі операції оптимізації лежать методи рекурентного виділення підкласів стратегій із F — класу всіх можливих стратегій f.

Будемо шукати представлення ціни u_ε^* у вигляді степеневого ряду за малим параметром ε

$$u_\varepsilon^* = \sum_{m=1}^{\infty} \varepsilon^m u_m^*. \tag{12}$$

Підставляючи цей ряд у рівняння (10) і прирівнюючи коефіцієнти при однакових степенях ε, отримаємо систему рівнянь

$$[I - P_0]u_{-1}^* = 0,$$

$$[I - P_0]u_0^* = \sup_{f \in F}\left\{\omega^f - P_1^f u_{-1}^*\right\}, \tag{13}$$

$$[I - P_0]u_m^* = -\inf_{f \in F_m} P_1^f u_{m-1}^*, \ m \geq 1,$$

де

$$F_1 = \left\{ f \in F_{m-1} : [I - P_0] u_0^* = \omega^f - P_1^f u_{-1}^* \right\},$$

$$F_m = \left\{ f \in F_{m-1} : [I - P_0] u_m^* = - P_1^f u_{m-1}^* \right\}, \quad m \geq 2 .$$

Рівняння системи (13) являються операторними рівняннями в E із виродженим матричним оператором $[I - P_0]$, $\dim [I - P_0] = 1$, правий і лівий власні вектори якого визначаються відповідно рівняннями

$$[I - P_0]\, 1 = 0 , \quad \rho [I - P_0] = 0 .$$

Важливу роль у визначенні умов розв'язності кожного з рівнянь (13) відіграє власний проектор оператора $[I - P_0]$, що визначається у вигляді тензорного добутку векторів 1 та ρ, а саме, $\Pi = [1 \otimes \rho]$, для якого виконуються рівності

$$\Pi [I - P_0] = [I - P_0] \Pi = 0 .$$

Оператор $[I - P_0]$ є звідно-оборотним у просторі V, який подається у вигляді

$$V = N [I - P_0] \oplus R\, [I - P_0],$$

де $N[I - P_0]$ — ядро оператора $[I - P_0]$; $R[I - P_0]$ — множина його значень. Для оператора $[I - P_0]$ на $R[I - P_0]$ існує узагальнений обернений $R_0 = (I - P_0 + \Pi)^{-1} - \Pi$ такий, що виконуються рівності

$$\Pi R_0 = R_0 \Pi = 0 .$$

Що стосується припущень щодо матричного оператора збурень P_1^f, то будемо вважати виконаною умову

$$\Pi P_1^f \Pi > 0, \quad \forall\, f \in F.$$

Застосовуючи до системи (13) модифікований алгоритм Вишика–Люстерника [7] отримаємо

$$u_{-1}^* = \omega^* / \hat{q}^*\, 1 , \tag{14}$$

$$u_m^* = R_0 \varphi_{0m}^* + \hat{\psi}_{m+1}^* \;\; / \hat{q}^*\, 1, \;\; m \geq 0 ,$$

де

$$\hat{\omega}^* / \hat{q}^* = \sup_{f \in F} \left\{ \hat{\omega}^f / \hat{q}^f \right\}, \; \hat{\omega}^f = \sum_{k \in E} \rho_k \omega_k^f , \;\; \hat{q}^f = \sum_{k \in E} \rho_k P_{k0}^{1, f(k)},$$

$$\varphi_m^* = - \inf_{f \in F} \left\{ \psi_m^f \hat{q}^* + q^f \psi_m^* \right\} / \hat{q}^* ,$$

причому

$$\psi^f_{m+1} = P^f_1 R_0 \varphi^*_m, \;\; \psi^f_{m+1} / \hat{q}^{\,*} = - \inf_{f \in F_m} \left\{ \hat{\psi}^f_{m+1} / \hat{q}^{\,f} \right\}.$$

Підставляючи вирази для u^*_m, $m \geq 0$ із (14) в (12), отримаємо

$$u^*_\varepsilon = \left(\hat{\omega}^* / \hat{q}^{\,f} \right) 1 \varepsilon^{-1} + \sum \left(R_0 \varphi^*_m + \hat{\psi}^f_{m+1} \middle| \hat{q}^* 1 \right) \varepsilon^m . \qquad (15)$$

Ряд у правій частині формули (15) абсолютно збігається в нормі V для достатньо малих ε.

Стратегія f^*_ε, яка реалізує екстремум правої частини системи (13), є оптимальною. Відшукання f^*_ε пов'язано з алгоритмом, на k-му кроці якого обчислюється наближення $f^{(k)}_\varepsilon$ стратегії f^*_ε із нерівностей

$$\hat{\omega}^{f^{(k)}_\varepsilon} / \hat{q}^{\,f^{(k)}_\varepsilon} \geq \hat{\omega}^{f^{(k-1)}_\varepsilon} / \hat{q}^{\,f^{(k-1)}_\varepsilon} ,$$

$$R_0 \varphi^{f^{(k)}_\varepsilon}_m + \hat{\psi}^{f^{(k)}_\varepsilon}_{m+1} / \hat{q}^{\,f^{(k)}_\varepsilon} \geq R_0 \varphi^{f^{(k-1)}_\varepsilon}_m + \hat{\psi}^{f^{(k-1)}_\varepsilon}_{m+1} / \hat{q}^{\,f^{(k-1)}_\varepsilon} , \; m \geq 1 . \qquad (16)$$

Зупинка алгоритму відбувається за перетворення нерівностей (16) у рівності.

Алгоритм поліпшення стратегій (16) значно спрощується при достатньо малих $\varepsilon \geq 0$. За $\varepsilon \to 0$ $\lim \varepsilon u^*_\varepsilon = u^*_{-1}$. Тоді для відшукання асимптотично оптимальної стратегії f^*_0 розглядається алгоритм, на k-му кроці якого вибирається $f^{(k)}_0$-е наближення стратегії f^*_0 за умови

$$\hat{\omega}^{f^{(k)}_0} / \hat{q}^{\,f^{(k)}_0} \geq \hat{\omega}^{f^{(k-1)}_0} / \hat{q}^{\,f^{(k-1)}_0} . \qquad (17)$$

Зупинка алгоритму відбувається за перетворення нерівності (17) у рівність.

Слабкокеровані процеси марковського відновлення з малою ймовірністю поглинання

Слабкокерований процес марковського відновлення (СКПМВ) $\left\{ x^\varepsilon_n, a_n, \theta_n;\; n \geq 0 \right\}$ з скінченним або зчисленним ФПС $E_0 : E \bigcup \{0\}$ ($E = \{1, 2, \ldots, N, \ldots\}$, $\{0\}$ — стан поглинання) і компактною множиною A рішень або керувань задається напівмарковським ядром [7].

$$Q^{\varepsilon, f}_{kr} = P^{\varepsilon, f}_{kr}\, G_k(t), \;\; k, r \in E, \;\; 0 < \varepsilon < 1 , \qquad (18)$$

де f — марковська стаціонарна нерандомізована стратегія, яка ототожнюється з функцією f, що відображає ФПС E у простір керувань A; $G_k(t)$ — функція розподілу тривалості перебування СКПМВ у стані k, а саме

$$G_k(t) = P\left\{\theta_{n+1} \le t \middle| x_n^{\varepsilon} = k\right\} = P\{\theta_k \le t\}. \tag{19}$$

Ймовірності переходу слабкокерованого вкладеного ланцюга Маркова (СКВЛМ) залежать від стратегії f і задані у вигляді

$$p_{kr}^{\varepsilon,f} = P\left\{x_{n+1}^{\varepsilon} = r \middle| x_n^{\varepsilon} = k\right\} = \int_0^{\infty} p_{kr}^{\varepsilon,f}\, dG_k(t) = p_{kr}^0 - \varepsilon p_{kr}^1 ,$$

$$\exists\, k \in E:\, p_{k0}^{1,f} > 0, \quad p_{00}^{\varepsilon,f} = 1 \tag{20}$$

або в матричному вигляді (1).

Незбурений вкладений ланцюг Маркова (ВЛМ) $\left\{x_n^0;\ n \ge 0\right\}$ з матрицею $P_0 = \left\{p_{kr}^0;\ k, r \in E\right\}$ є ергодичним зі стаціонарним розподілом $\rho = \{\rho_k;\ k \in E\}$.

Введемо поняття доходу, пов'язаного з функціонуванням СКПМВ. Якщо в стані $k \in E$ прийнято рішення $a \in A$ і тривалість часу, проведеного у стані k, дорівнює t, то очікуваний доход за час t дорівнює $\varphi_k(t,a)$. Функція $\varphi_k(t,a)$ припускається вимірною за t та обмеженою за k, a.

Позначимо

$$r_k(a) = M\varphi_k(\theta_k, a) = \int_0^{\infty} \varphi_k(t,a) G_k(dt) . \tag{21}$$

В якості критерію оптимальності стратегії f розглядається функціонал вигляду (9). Зокрема, функціонал L за нульової штрафної функції ψ можна інтерпретувати як деяку наробку напівмарківської системи за час τ_{ε}.

Таким чином, за допомогою подання динаміки СКПМВ, вартісної структури та критерію оптимізації у вигляді (20), (21) та (9), задача оптимального керування СКПМВ з поглинанням зведена до задачі оптимального керування СКМП при наявності стану поглинання з дослівним повторенням усього викладеного вище у формулах (10) – (17).

НЕАДИТИВНИЙ КРИТЕРІЙ. СЛАБКОКЕРОВАНІ МОДЕЛІ БЕЗ ПОГЛИНАННЯ

Припускається, що в слабкокерованій моделі відсутній стан поглинання, тобто $P_1^f 1 = 0$, причому модель є рівномірно-зворотною за кожної стратегії $f \in F$, і для матриці P_{ε}^f існує неперервний на E проектор Π_{ε} такий, при якому

$$\Pi_\varepsilon^f = \lim_{n\to\infty} n^{-1}\sum_{k=0}^{n-1}\left(P_\varepsilon^f\right)^k, \quad \Pi_\varepsilon^f P_\varepsilon^f = P_\varepsilon^f \Pi_\varepsilon = \Pi_\varepsilon^f .$$

Критерій оптимальності стратегій f визначається функціоналом вигляду

$$g_\varepsilon^f = \lim_{n\to\infty} n^{-1}\sum_{m=0}^{n-1} M_\varepsilon^f r\left(X_m^\varepsilon\right)^2, \tag{22}$$

де r^f — обмежена на E та неперервна на F функція однокрокового доходу.

Рівняння оптимальності для неадитивного критерію (22) являють собою систему нелінійних рівнянь типу Беллмана вигляду

$$\begin{aligned} & g_\varepsilon^* = \sup_{f\in F} P_\varepsilon^f g_\varepsilon^f, \quad \upsilon_\varepsilon^* = \sup_{f\in F'}\left[r^f - g_\varepsilon^f + P_\varepsilon^f \upsilon\right], \\ & \Pi_\varepsilon^* \upsilon_\varepsilon^* = 0, \end{aligned} \tag{23}$$

де $F' = \left\{f : g_\varepsilon^* = P_\varepsilon^f g_\varepsilon^f\right\}$; g_ε^f та υ_ε^f — середній асимптотичний доход за один крок та відповідно деяка вагова функція, які відповідають стратегії f; g_ε^*, υ_ε^* — відповідають оптимальній стратегії f^*.

Розв’язок системи (23) має вигляд

$$g_\varepsilon^* = \sup_{f\in F} \Pi_\varepsilon^f r^f, \quad \upsilon_\varepsilon^* = \sum_{m=0}^{\infty} \upsilon_m^* \varepsilon^m , \tag{24}$$

де проектор Π_ε^f подається у вигляді степеневого ряду за малим параметром ε, а саме

$$\Pi_\varepsilon^f = \sum_{m=0}^{\infty} \Pi_m^f \varepsilon^m , \tag{25}$$

де $\Pi_0 = [\mathbf{1}\otimes\rho]$ — власний проектор оператора $[I - P_0]$; $\otimes$ — знак тензорного множення;

$$\Pi_1^f = \Pi_0 P_1^f R_0 + R_0 P_1^f \Pi_0 ,$$

$$\begin{aligned} \Pi_2^f = {} & \Pi_0 P_1^f R_0 + R_0 P_1^f \Pi_0 P_1^f R_0 + R_0 P_1^f R_0 P_1^f \Pi_0 - \Pi_0 P_1^f \Pi_0 P_1^f (R_0)^2 - \\ & - (R_0)^2 P_1^f \Pi_0 P_1^f (R_0)^2 - (R_0)^2 P_1^f \Pi_0 P_1^f \Pi_0 - \Pi_0 P_1^f (R_0)^2 P_1^f \Pi_0 . \end{aligned}$$

Загальний член Π_m^f, $m \ge 3$ розкладу Π_ε^f $(\varepsilon \ne 1)$ можна за аналогією із [8] записати у вигляді

$$\Pi_m^f = \sum_{l=1}^{m} \sum_{i_1+\ldots+i_l=m} \sum_{j_0+\ldots+j_l} (-1)^{l_j} H^{(j_0)} P_{1,i_1}^f H^{(j_2)} P_{1,i_2}^f \ldots P_{1,i_l}^f H^{(j_l)} ,$$

де

$$J_m^+ = -\min\left(J_{m,0}\right), \quad \text{а} \quad l_j = \sum_{m=0}^{l} J_m^+ ; \quad P_{1,i_s}^f = P_{1,i_t}^f = P_1^f , \quad s,t = \overline{1,l} ;$$

$$H^{(j)} = \begin{cases} \Pi, & j = -1, \\ V_0^{j+1}, & j \geq 0, \end{cases} \qquad V_0 = -R_0 .$$

Коефіцієнти υ_m^* розкладу степеневого ряду за ε для оптимальної вагової функції υ_ε^* мають вигляд

$$\upsilon_m^* = R_0 \varphi_m^*, \quad \varphi_0 = \operatorname*{Sup}_{f \in F'} r^f ,$$

$$\varphi_m^* = \operatorname*{Sup}_{f \in F_m'} \left[P_1^f \upsilon_{m-1}^f - \Pi_m^f r^f\right], \quad m \geq 1 ,$$

де R_0 — узагальнений обернений оператор до оператора $[1 - P_0]$.

Класи стратегій F_m' визначаються співвідношеннями

$$F_m' = \left\{f : [I - P_0]\upsilon_m^* = P_1^f \upsilon_{m-2}^f - \Pi_{m-1}^f r^f \right\}, \quad m \geq 1 .$$

Для доведення співвідношень (23) скористаємось допоміжним твердженням.

Лема. Розв'язок системи рівнянь (лінійного аналогу рівнянь оптимальності (23))

$$g_\varepsilon^f = P_\varepsilon^f g_\varepsilon^f , \quad \upsilon_\varepsilon^f = r^f - g_\varepsilon^f + P_\varepsilon^f \upsilon_\varepsilon^f , \quad \Pi_\varepsilon^f \upsilon_\varepsilon^f = 0 \qquad (26)$$

може бути поданий у вигляді

$$g_\varepsilon^f = \Pi_\varepsilon^f r^f , \quad \upsilon_\varepsilon^f = \sum_{m=0}^{\infty} \upsilon_\varepsilon^f \varepsilon^m . \qquad (27)$$

Перше співвідношення в (26) випливає із (22). Справді,

$$g_\varepsilon^f (i) = \lim_{n \to \infty} n^{-1} \sum_{m=0}^{\infty} \left(P_{ij}^{\varepsilon, f(i)}\right)^n r^f (j) = \sum_{i \in E} \Pi_\varepsilon^f (i,j) r^f (i), \quad x_0 = i ,$$

що у векторному вигляді означає подання g_ε^f першою формулою (27), яке задовольняє, очевидно, перше рівняння (26).

Для відшукання коефіцієнтів розкладу степеневого ряду υ_ε^f підставимо його у друге рівняння (26) і, прирівнюючи коефіцієнти при однакових степенях ε, прийдемо до системи рівнянь

$$\left[I - P_0\right]\upsilon_0^f = r^f - \Pi_0 r^f , \qquad (28)$$

$$\left[I - P_0\right]\upsilon_m^f = \varphi_m^f, \quad m \geq 1 ,$$

де

$$\varphi_1^f = -P_1^f R_0 r^f - \Pi_1^f ,$$

$$\varphi_m^f = -P_1^f R_0 \varphi_{m-1}^f - \Pi_m^f, \quad m \geq 2 .$$

Частковий розв'язок системи (27) має вигляд

$$\upsilon_0^f = R_0 r^f ,$$

$$\upsilon_m^f = R_0 \varphi_m^f, \quad m \geq 1 . \qquad (29)$$

Степеневий ряд υ_ε^f у (26) абсолютно збігається за достатньо малих ε. ■

Для розв'язання системи рівнянь оптимальності (23) використовується метод послідовної оптимізації розв'язків (27) системи рівнянь (26) з урахуванням представлень (25), (29).

АГРЕГОВАНЕ КЕРУВАННЯ УКРУПНЕНОЮ МАРКОВСЬКОЮ МОДЕЛЛЮ

Розглянемо слабкокеровану марковську модель, у якої ФПС Е має вигляд $E = \bigcup\limits_{i=1}^{\infty} E_i \cup E'$; $E_k \cup E_j = \varnothing$; $k \neq r$; E_i — замкнені стани класів; E' — клас перехідних станів; A — компактна множина рішень; P_ε^f — МЙП для якої МЙП незбуреного ергодичного ланцюга Маркова має блочно-діагональний вигляд зі стаціонарним розподілом $\rho = (\rho_1, ..., \rho_N, ...)$; P_1^f — матриця збурень, залежних від керувань згідно стратегії f; r^f — однокроковий доход, який описується дійснозначною обмеженою функцією, що відповідає нерандомізованій стаціонарній стратегії f .

Описана модель являє собою зчисленну сім'ю слабкозалежних та слабкокерованих моделей. В залежності від властивостей цих моделей можна провадити оптимізацію за адитивним (за наявності поглинання) або за неадитивним (за відсутності поглинання та можливості укрупнення або агрегування) критеріями оптимальності.

Розглянемо агреговане керування, яке ґрунтується на розв'язку рівняння оптимальності, що виникає в результаті оптимізації за неадитивним критерієм укрупненої моделі, пов'язаної з вихідною моделлю таким чином.

Укрупнена модель задається набором $\left(\hat{E}, \hat{A}, \hat{P}^{\hat{f}}, \hat{r}^{\hat{t}}\right)$, де $\hat{E}$ — зчисленний ФПС, який одержується із E заміною класів $E_1, E_2, ..., E_N, ...$ ергодичних станів їх індексами $1,2,...,N,...$; $\hat{A}$ — множина значень укрупнених (агрегованих) рішень

$$\hat{f} = \left\{ \hat{f}(1), \hat{f}(2), ..., \hat{f}(N), ... \right\},$$

де

$$\hat{f}(i) = \int\limits_{E_i} \rho_i(dx) f(x) . \tag{30}$$

МЙП укрупненої моделі має вигляд

$$\hat{P}^{\hat{f}} = \left\{ \hat{p}_{ij}^{\hat{f}(i)} = \int\limits_{E_i} \rho_i(dx) P_1^{f(x)}(x, E_j) \Big/ \int\limits_{E_i} \rho_i(dx) P_1^{f(x)}(x, E_i), \right.$$

$$\left. x \in E_i, \;\; \overline{E}_i = E/E_i, \;\; i \neq j \right\} . \tag{31}$$

Вектор однокрокового агрегованого доходу подається як

$$\hat{r}^{\hat{f}} = \Pi_0 r^f . \tag{32}$$

Для критерію середнього асимптотичного однокрокового доходу рівняння оптимальності агрегованого керування укрупненою керованою моделлю являє собою систему нелінійних рівнянь типу Беллмана

$$\hat{g}^* = \operatorname*{Sup}_{\hat{f} \in \hat{F}} \hat{P}^{\hat{f}} \hat{g}^{\hat{f}}, \;\; \hat{\upsilon}^* = \operatorname*{Sup}_{\hat{f} \in \hat{F}'} \left[\hat{r}^{\hat{f}} - \hat{g}^{\hat{f}} + \hat{P}^{\hat{f}} \hat{\upsilon}^{\hat{f}} \right], \tag{33}$$

де $\hat{F}$ — множина агрегованих стратегій, відповідне рішення $\hat{a} \in \hat{A}$ кожної з яких приймається в станах укрупненої моделі; $\hat{F}' = \left\{ \hat{f} : \hat{g}^* = \operatorname*{Sup}_{\hat{f} \in \hat{F}} \hat{P}^{\hat{f}} \hat{g}^{\hat{f}} \right\}$; $\hat{\Pi}^{\hat{f}}$ — власний проектор оператора $\left[\hat{I} - \hat{P}^{\hat{f}} \right]$, що задовольняє співвідношення

$$\hat{\Pi}^{\hat{f}} \hat{P}^{\hat{f}} = \hat{\Pi}^{\hat{f}} .$$

Для зчисленного ФПС $\hat{E}$ Гордайком одержано розв'язок системи рівнянь (33) у вигляді [9]

$$\begin{aligned} \hat{g}^* &= \operatorname*{Sup}_{\hat{f} \in \hat{F}} \sum_{m=0}^{\infty} \left(\hat{P}^{\hat{f}} \right)^m \hat{r}^{\hat{f}}(1) \Big/ \sum_{m=0}^{\infty} \left(\hat{P}^{\hat{f}} \right)^m 1(1), \\ \hat{\upsilon}^* &= \operatorname*{Sup}_{\hat{f} \in \hat{F}'} \hat{R}_0 \left(\hat{r}^{\hat{f}} - \hat{g}^* \right). \end{aligned} \tag{34}$$

Для отримання подання $\hat{g}^*$, що виражається першою формулою в (33), розглядається спочатку лінійне рівняння в i-му стані

$$\hat{w}^{\hat{f}}(i) + \hat{g}^{\hat{f}}(i) = \hat{r}^{\hat{f}} 1(i) + \hat{P}^{\hat{f}} \hat{w}^{\hat{f}}(i) .$$

З цього співвідношення формально отримується

$$\left(\hat{I}-\hat{P}^{\hat{f}}\right)^{-1}\hat{g}^{\hat{f}}(i)=\left(\hat{I}-\hat{P}^{\hat{f}}\right)^{-1}\hat{r}^{\hat{f}}-\hat{w}^{\hat{f}}(i),\quad \forall\, i\,. \tag{35}$$

За умови Говарда [10] $\hat{w}^{\hat{f}}(1)=0$ із (35) випливає, що

$$\hat{g}^{\hat{f}}=\left(\hat{I}-\hat{P}^{\hat{f}}\right)^{-1}\hat{r}^{\hat{f}}(1)\Big/\left(\hat{I}-\hat{P}^{\hat{f}}\right)^{-1}1\,(1)\,. \tag{36}$$

За умови Ляпунова

$$\hat{r}^{\hat{f}}+1+\hat{P}^{\hat{f}}\hat{w}^{\hat{f}}\le\hat{w}^{\hat{f}}$$

має місце нерівність

$$\left\|\hat{P}^{\hat{f}}\right\|<1\,. \tag{37}$$

При узятті супремума в (36) за $\hat{f}\in\hat{F}$ з урахуванням (35), отримується вираз для $\hat{g}^{*}$ в (34).

Вираз для $\hat{\upsilon}^{*}$ в (34) одержано за допомогою операції обернення оператора $\left\lfloor\hat{I}-\hat{P}_{0}\right\rfloor$ для лінійного аналога другого рівняння в (33) з подальшою оптимізацією результату цього обернення за стратегіями $\hat{f}\in\hat{F}'$.

ВИСНОВКИ

Окреслимо можливий напрямок застосувань розглянутої в цій статті моделі. При дослідженні різних сфер людської діяльності слід зазначити, що, незважаючи на весь різновид цих сфер, вони взаємнозалежні. Прогрес суспільного розвитку можна охарактеризувати деяким глобальним критерієм. Виникає проблема відшукання таких зв’язків між сферами діяльності, які б забезпечили оптимальні показники глобального критерію.

В нашій постановці сфери людської діяльності моделюються сім’єю слабкозалежних та слабкокерованих марковських та напівмарковських процесів у дискретному часі. Термін «слабкозалежні» — адекватно відповідає природі людських взаємин у розумінні самостійної їх активності в конкретній сфері діяльності. Термін «слабкокеровані» адекватно відповідає ролі держави у встановленні саме таких керованих зв’язків між сферами, за яких діяльність в усіх сферах характеризується оптимальними показниками суспільного розвитку.

Перехід до керованої агрегованої моделі ілюструє ідею системного підходу до політики суспільного розвитку. Агрегована стаціонарна нерандомізована стратегія керування встановлює за своєю суттю усталені «слабкі зв’язки» (умовні переходи) між станами (точками), які характеризують в агрегованому вигляді відповідні сфери суспільного розвитку. Критерій оптимальності шуканої стратегії характеризує середні в асимптотичному плані максимальні показники динаміки суспільного зростання.

ЛІТЕРАТУРА

1. *Bellman R.* Markovian decision processes // J.Math. Mech. — 1957. — № 6. — P. 679–684.
2. *Андрєєв М.В*. Синтез оптимальних стратегій контролю та керування в задачах неполадки // Теорія еволюційних рівнянь. Міжнар. конф. «П'яті Боголюбовські читання». — Кам.-Под. пед. університет. — 2002. — С. 20–21.
3. *Андреев Н.В*. Управление риском обмена валют // Україна: поступ у майбутнє. Міжнар. наук. конф., присвячена 290-річчю прийняття Конституції Пилипа Орлика. — Вісник АПСВ. — 2000. — С. 88–92.
4. *Андрєєв М.В*. Деякі аспекти оптимізації страхової діяльності // Україна шляхами віків. Міжнар. наук. конф., присвячена 175-річчю з дня народження Георгія Андрузького. — Вісник АПСВ. — 2002. — С. 79–81.
5. *Дынкин Е.Б., Юшкевич А.А*. Управляемые марковские процессы и их приложения. — М.: —Наука, 1975. — 338 с.
6. *Андрєєв М.В.* Прикладний статистичний аналіз марковських гауссових процесів в дискретному часі // Системні дослідження та інформаційні технології. — 2003. — № 1. — С. 112–120.
7. *Королюк В.С., Андреев Н.В*. Управляемые процессы марковского восстановления с малой вероятностью обрыва // Кибернетика. — 1986. — № 6. — С. 112–114.
8. *Korolyuk V.S., Turbin A.F.* Mathematical Foundations of the State Lumping of Large Systems, Kluwer. — 1993. — P. 264–278.
9. *Андреев Н.В.* Оптимальное управление слабоуправляемыми марковскими и полумарковскими моделями // Сучасні інформаційні технології — шлях до інформаційного суспільства. Ювілейний збірник наукових праць, присвячений 10-річчю кафедри математичних методів системного аналізу. — К.:ІПСА, 1998. — С. 91–98.
10. *Говард Р.* Динамическое программирование и марковские процессы. — М.: Сов. Радио, 1964. — 136 с.

Надійшла 25.12.2002