

**МЕТОД ПОШУКУ ОПТИМАЛЬНИХ СТРАТЕГІЙ
МАРКОВСЬКОГО ПРОЦЕСУ ПРИЙНЯТТЯ
РІШЕННЯ НА ОСНОВІ ВЛАСТИВОСТЕЙ
ВЛАСНОГО ВЕКТОРА**

©2007 р. Андрій БОБИЛЯК

Львівський національний університет імені Івана Франка,
вул. Університетська 1, Львів 79000

Редакція отримала статтю 14 червня 2007 р.

Встановлено нестандартні властивості власних векторів, на основі яких побудовано алгоритм пошуку оптимальної стратегії для марковського процесу прийняття рішення для однокритеріального випадку, а також сформульовано узагальнення алгоритму для багатокритеріального випадку.

1. У даній статті розглядаються можливі модифікації та альтернативні алгоритми до відомого ітераційного алгоритму Ховарда. Відомо, що алгоритм Ховарда побудовано для знаходження оптимальних стратегій в сенсі одного критерію, а саме, максимізації середнього доходу. Однак у багатьох практичних задачах бажано знайти оптимальну стратегію в іншому сенсі, наприклад, стратегію, яка не тільки максимізує середній дохід, а й мінімізує дисперсію. Таким чином, практика вимагає пошуку оптимальних стратегій для багатокритеріальних випадків. Пошук оптимальних стратегій марковського процесу прийняття рішень є головною метою алгоритму, побудованого у даній статті.

2. Опишемо процедуру прийняття рішення. Нехай задано множину $S = \{1, 2, \dots, N\}$ станів середовища — це означає, що в кожен момент часу середовище може перебувати в одному з N станів. Кожному стану $i \in S$ ставимо у відповідність скінченну множину M_i рішень (альтернатив), елементи якої позначимо через $k = 1, 2, \dots, M_i$. Простором політик

M назвемо прямий добуток множин рішень, тобто $M = M_1 \times \dots \times M_N$. Розглядається задача про прийняття послідовних рішень, яка полягає у виборі рішень при настанні біжучих станів у моменти $n = 0, 1, 2, \dots$. Під стратегією π розуміємо послідовність політик $\{f_n \in M, n = 1, 2, \dots\}$, тобто $\pi = (f_1, f_2, \dots, f_n, \dots)$, де f_n — вектор, i -й елемент якого позначимо через $f_n(i)$; цей елемент є рішенням, що приймається в стані $i \in S$ в момент n . Стратегію (g, f_1, f_2, \dots) позначимо через (g, π) , де $g \in M$, а $\pi = (f_1, f_2, \dots)$. Стратегію (f, f, \dots, f, \dots) , де $f \in M$, позначимо через f^∞ і назвемо її *стаціонарною*. Нарешті, стратегію вигляду $(\underbrace{g, g, \dots, g}_n, f_1, f_2, \dots)$ позначимо через (g^n, π) , де $g \in M$, $\pi = (f_1, f_2, \dots)$.

Вважаємо, що якщо система знаходиться в стані $i \in S$ та приймається рішення $m \in M_i$, то стан системи в наступний момент часу визначається ймовірнісним законом p_{ij}^m ($j \in S$), де p_{ij}^m — ймовірність того, що система зі стану i при виборі рішення m перейде в стан j , тобто

$$\sum_{j \in S} p_{ij}^m = 1, \quad p_{ij}^m \geq 0, \quad m \in M_i, \quad i, j \in S.$$

Нехай задано початковий розподіл станів середовища $a = (a_1, a_2, \dots, a_N)$, $\sum_{i \in S} a_i = 1$, $a_i \geq 0$, $i \in S$. Зрозуміло, що незалежно від стратегії π даний процес в загальному являє собою неоднорідний ланцюг Маркова, для якого матриця ймовірностей переходу за n кроків має вигляд

$$P_n(\pi) = P(f_1)P(f_2) \dots P(f_n), \quad n = 1, 2, \dots,$$

де $P(f_n)$ — матриця переходу розміру $N \times N$, (i, j) -й елемент якої дорівнює p_{ij}^m , $m = f_n(i) \in M_i$. Нехай $P_0(\pi) = \mathbf{I}$, де \mathbf{I} — одинична матриця розміру $N \times N$.

Щоб запровадити поняття оптимальної стратегії, нам знадобиться поняття *вектора сумарної середньої корисності стратегії* та *вектора дисперсії стратегії*. Для кожної політики $f \in M$ визначимо такий N -вимірний вектор-стовпець корисності $I'(f)$, i -й елемент якого дорівнює $I_i^m = I(m)$, де $m = f(i) \in M_i$. Надалі для зручності запису замість $I'(f)$ будемо писати $I(f)$. У цих позначеннях вектор-стовпці розмірності N , i -й елемент яких відповідає початковому стану процесу $i \in S$, будемо називати вектором сумарної середньої корисності стратегії π . Цей вектор та вектор дисперсії стратегії π мають вигляд

$$V_\beta(\pi) = \sum_{n=0}^{\top} \beta^n P_n(\pi) I(f_{n+1}),$$

$$\sigma_{\beta}^2(\pi) = \sum_{n=0}^{\top} \beta^n P_n(\pi) \text{sqr}(I(f_{n+1}) - \max_{\pi} V_{\beta}(\pi)),$$

де $\text{sqr}(\cdot)$ — функція покомпонентного піднесення до квадрату елементів вектора, β — коефіцієнт переоцінки, $0 \leq \beta < 1$. Описану модель називають *моделлю з переоцінкою*. Зауважимо, що $\sigma_{\beta}^2(\pi)$ можна розглядати як $V_{\beta}(\pi)$, де $\widehat{I}(f) = \text{sqr}(I(f) - V_{\beta}(\pi))$. Якщо $\beta = 1$, то модель називають *моделлю без переоцінки* і розглядають наступний аналог в перерахунку за одиницю часу, тобто

$$\Gamma(\pi) \equiv \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} P_i(\pi) I_k(f_{i+1}).$$

У даній статті розглядаємо модель без переоцінки з одним ергодичним класом. Відомо [3], що для такої моделі існує оптимальна стаціонарна стратегія $\pi = f^{\infty}$, тому

$$\Gamma(\pi) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} P^i(f) \cdot I(f) = P^*(f) \cdot I(f),$$

де P^* — гранична матриця, яка складається з однакових рядків $\pi(f) = [\pi_i(f)]$. Якщо $P(f)$ — нерозкладна, то з матричного аналізу відомо, що $\pi(f)$ є власним вектором для $P^{\top}(f)$. Оскільки рядки однакові, то вектор $\Gamma(\pi)$ складається з однакових елементів і можна розглядати відповідний скаляр $\tau(\pi)$. Тоді сумарний вектор середніх доходів $V_{\beta}(\pi)$ замінимо відповідним скаляром $\tau(\pi)$, що характеризує середній дохід від стратегії π за одиницю часу, а дисперсію замінимо на

$$D\tau(\pi) = \sum_{j \in S} \pi_j(f) \cdot [\text{sqr}(I(f) - \tau(\pi) \cdot 1)]^j.$$

Доведемо властивість розкладу власного вектора, на основі якої побудуємо новий алгоритм пошуку оптимальної стратегії. Нагадаємо, що марковською (рядково стохастичною) матрицею називається невід’ємна квадратна матриця, сума елементів довільного рядка якої дорівнює одиниці (кожен рядок є ймовірнісним вектором). З узагальнення теорему Перрона для випадку невід’ємної, нерозкладної матриці [1, 4] випливає, що довільна нерозкладна марковська матриця має одновимірний власний підпростір, що відповідає власному значенню 1, причому існує єдиний додатний власний вектор, сума координат якого дорівнює 1 (тобто ймовірнісний вектор), який називають перроновим вектором.

Теорема. *Нехай*

$$A = \begin{pmatrix} \lambda_1 & a_{12} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots \\ \lambda_r & a_{r2} & \dots & a_{rn} \\ \dots & \dots & \dots & \dots \\ \lambda_n & a_{n2} & \dots & a_{nn} \end{pmatrix}, \quad \forall i, j \quad \lambda_i, a_{ij} \in [0; 1], -$$

довільна нерозкладна стовпцево стохастична матриця, а

$$A_r = \begin{pmatrix} 0 & a_{12} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots \\ 1 & a_{r2} & \dots & a_{rn} \\ \dots & \dots & \dots & \dots \\ 0 & a_{n2} & \dots & a_{nn} \end{pmatrix}, \quad 1 \leq r \leq n, -$$

стовпцево стохастична матриця, перший стовпець якої складається з однієї одиниці на r -му місці та нулів на всіх решті місцях. Нехай y, y_r — ймовірнісні власні вектори матриць A та A_r , $1 \leq r \leq n$, відповідно, які відповідають власному значенню 1. Тоді вектор y є опуклою комбінацією y_1, \dots, y_n , а саме, $y = k_1 y_1 + \dots + k_n y_n$. Вектор $k = (k_1, \dots, k_n)^\top$ визначається рівністю

$$k = \frac{1}{\lambda^\top \tilde{y}} \lambda \circ \tilde{y},$$

де $\lambda = (\lambda_1, \dots, \lambda_n)^\top$, $\tilde{y} \equiv \left(\frac{1}{[y_1]^1}, \dots, \frac{1}{[y_n]^1} \right)^\top$ — додатний вектор, а символ \circ означає покомпонентний добуток векторів (добуток Адамара).

Доведення. Неважко переконатись, що марковська матриця A_r має єдиний ймовірнісний власний вектор з першою додатною компонентою, тому всі ймовірнісні власні вектори y, y_1, \dots, y_n , що відповідають власному значенню 1 визначаються однозначно, а тому формулювання теореми є коректним. Щоб уникнути зайвих позначень, вважаємо, що всі власні вектори y, y_1, \dots, y_n мають однакове значення норми l_1 , тобто сума їхніх компонент є однаковою (але не обов'язково дорівнює 1).

Розв'яжемо векторне рівняння $y = k_1 y_1 + \dots + k_n y_n$. З означення власного вектора випливає, що $Ay = y$. Оскільки $A = \sum_{i=1}^n \lambda_i A_i$, то

$$\sum_{i=1}^n \lambda_i A_i \sum_{i=1}^n k_i y_i = \sum_{i=1}^n k_i y_i,$$

тобто

$$\sum_{\substack{i,j=1, \\ i \neq j}}^n \lambda_i k_j A_i y_j = \sum_{i=1}^n (1 - \lambda_i) k_i y_i. \quad (1)$$

З'ясуємо, який вигляд може набувати вектор $A_i y_j$, де $i, j \in \{1, 2, \dots, n\}$, $i \neq j$. Якщо $i > j$, то

$$A_i y_j = y_j + \left(\underbrace{0, \dots, 0, -[y_j]^1}_{j}, \underbrace{0, \dots, 0, [y_j]^1}_{i-j}, 0, \dots, 0 \right)^T,$$

якщо ж $i < j$, то

$$A_i y_j = y_j + \left(\underbrace{0, \dots, 0, [y_j]^1}_i, \underbrace{0, \dots, 0, -[y_j]^1}_{j-i}, 0, \dots, 0 \right)^T.$$

Запровадимо таке позначення: $\Delta y_{ij} \equiv A_i y_j - y_j$. Тоді за допомогою еквівалентних перетворень рівняння (1) отримуємо, що

$$\sum_{j=1}^n \sum_{\substack{i=1, \\ i \neq j}}^n \lambda_i k_j (y_j + \Delta y_{ij}) = \sum_{i=1}^n (1 - \lambda_i) k_i y_i.$$

Оскільки $\sum_{\substack{i=1, \\ i \neq j}}^n \lambda_i k_j y_j = (1 - \lambda_j) k_j y_j$, то після скорочення дістаємо векторне рівняння

$$\sum_{j=1}^n \sum_{\substack{i=1, \\ i \neq j}}^n \lambda_i k_j \Delta y_{ij} = \vec{0},$$

перша компонента якого має вигляд

$$-\sum_{\substack{i=1, \\ i \neq 1}}^n \lambda_i [y_1]^1 k_1 + \lambda_1 [y_2]^1 k_2 + \dots + \lambda_1 [y_n]^1 k_n = 0.$$

Аналогічно знаходимо решту рівнянь і тоді рівняння $y = k_1 y_1 + \dots + k_n y_n$ можна записати в еквівалентній матричній формі:

$$\begin{pmatrix} (\lambda_1 - 1) [y_1]^1 & \dots & \lambda_1 [y_n]^1 \\ \dots & \dots & \dots \\ \lambda_n [y_1]^1 & \dots & (\lambda_n - 1) [y_n]^1 \end{pmatrix} \begin{pmatrix} k_1 \\ \dots \\ k_n \end{pmatrix} = \vec{0}.$$

Розв'язок рівняння $y = k_1 y_1 + \dots + k_n y_n$, якщо він існує, визначається з точністю до множника. Покладемо $k_1 = 1$ (випадок, коли $k_1 = 0$, розглянемо пізніше). Тоді для визначення k_2, \dots, k_n отримаємо систему

$$\begin{pmatrix} (\lambda_2 - 1) [y_2]^1 & \dots & \lambda_2 [y_n]^1 \\ \dots & \dots & \dots \\ \lambda_n [y_2]^1 & \dots & (\lambda_n - 1) [y_n]^1 \end{pmatrix} \begin{pmatrix} k_2 \\ \dots \\ k_n \end{pmatrix} = - \begin{pmatrix} \lambda_2 \\ \dots \\ \lambda_n \end{pmatrix}.$$

Позначимо $C \equiv ([y_2]^1, \dots, [y_n]^1)(k_2, \dots, k_n)^\top$. Тоді отримаємо:

$$([y_2]^1 k_2, \dots, [y_n]^1 k_n)^\top = (1 + C)(\lambda_2, \dots, \lambda_n)^\top.$$

Якщо $\lambda_1 = 1$, то це означає, що $\lambda_2 = \dots = \lambda_n = 0$, тобто $A = A_1$, $y = y_1$, а, отже, $k_1 = 1$, $k_2 = \dots = k_n = 0$. Тому розглянемо випадок $\lambda_1 \neq 1$ та $\lambda_1 \neq 0$ (варіант $\lambda_1 = 0$ розглянемо пізніше). Тоді якщо підсумувати компоненти векторів (або взяти норму від обох частин) останньої рівності отримаємо, що $C = (1 + C)(1 - \lambda_1)$, тобто $\lambda_1(1 + C) = 1$. Нагадаємо, що $\forall i [y_i]^1 \neq 0$, тому шукані коефіцієнти можна знайти зі співвідношень

$$(k_2, \dots, k_n) = \left(\frac{\lambda_2}{\lambda_1 [y_2]^1}, \dots, \frac{\lambda_n}{\lambda_1 [y_n]^1} \right).$$

З точністю до множника можна записати розв'язок початкового рівняння $y = k_1 y_1 + \dots + k_n y_n$ у формі:

$$(k_1, \dots, k_n) = \left(\frac{\lambda_1}{[y_1]^1}, \dots, \frac{\lambda_n}{[y_n]^1} \right),$$

тобто $k_i = \frac{\lambda_i}{[y_i]^1}$ для всіх $i = 1, 2, \dots, n$. Зауважимо, що формула залишається правильною при $\lambda_1 = 1$, оскільки в цьому випадку $k_1 = 1$, а $k_2 = \dots = k_n = 0$. Запишемо вектор k як імовірнісний вектор так, щоб сума його компонент дорівнювала 1, тоді отримаємо вираз

$$k = \frac{1}{\lambda^\top \tilde{y}} \lambda \circ \tilde{y},$$

де $\lambda = (\lambda_1, \dots, \lambda_n)^\top$, $\tilde{y} = \left(\frac{1}{[y_1]^1}, \dots, \frac{1}{[y_n]^1} \right)^\top$. Зазначимо, що означений вектор k є розв'язком навіть при $\lambda_1 = 0$ ($k_1 = 0$). У цьому легко переконатися, безпосередньо підставивши вираз для k у вихідну систему.

Теорему доведено.

3. Побудуємо ітераційний алгоритм пошуку оптимальної стратегії для моделі марковського процесу прийняття рішення у випадку одного ергодичного класу, як альтернативний варіант відомого алгоритму Ховарда [3].

Нехай модель прийняття рішень характеризується такими множинами станів та політик:

$$\{S = \{1, \dots, N\}, M = M_1 \times \dots \times M_N\}.$$

Зміну станів системи вважаємо ланцюгом Маркова з одним ергодичним класом та матрицею переходу за n кроків $P_n(\pi)$. Вважаємо, що суб'єкт прийняття рішення шукає таку стратегію $\pi = \{f_1, \dots, f_n, \dots\}$, яка задовольняла б такі умови:

$$\tau(\pi) \rightarrow \max,$$

де $\tau(\pi)$ — середній дохід стратегії π за одиницю часу.

3.1. Процедура визначення ваг ключової матриці. Вибираємо довільну початкову політику $f \in M$ (наприклад, за принципом максимуму $I(f)$), де $A \equiv P^\top(f) = (a_{ij})_{i,j=1}^N$ — відповідна марковська матриця. Знаходимо для неї $\tilde{A}1_1^{-1} \equiv (\tilde{A}_1 - I)^{-1}$, де \tilde{A}_1 — матриця, яка утворюється з матриці A шляхом викреслювання 1-го рядка та 1-го стовпця.

3.2. Процедура поліпшення рішення. Для кожного індексу s , де $s = 1, \dots, N$, знаходимо оптимальне або близьке до оптимального рішення в стані s . Позначимо:

$$A_s(\lambda) \equiv \begin{pmatrix} a_{11} & \dots & a_{1,s-1} & \lambda_1 & a_{1,s+1} & \dots & a_{1,n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ a_{s-1,1} & \dots & a_{s-1,s-1} & \lambda_{s-1} & a_{s-1,s+1} & \dots & a_{s-1,n} \\ a_{s,1} & \dots & a_{s,s-1} & \lambda_s & a_{s,s+1} & \dots & a_{s,n} \\ a_{s+1,1} & \dots & a_{s+1,s-1} & \lambda_{s+1} & a_{s+1,s+1} & \dots & a_{s+1,n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ a_{n,1} & \dots & a_{n,s-1} & \lambda_n & a_{n,s+1} & \dots & a_{n,n} \end{pmatrix}.$$

3.2. а) Розглядаючи можливі рішення $m \in M_s$ (розподіли λ), знаходимо ймовірнісні власні вектори $\hat{y}_s^*(\lambda)$ матриць $A_s^*(\lambda)$ ($A_s^*(\lambda)$ відрізняється від $A_s(\lambda)$ тим, що $1, \dots, s-1$ рядки відповідають розподілам знайдених оптимальних рішень в станах $1, \dots, s-1$), тобто

$$\hat{y}_s^* = \frac{1}{\|-(\tilde{A}1_s^*)^{-1} \cdot \tilde{\lambda}\| + 1} \left([\tilde{y}_s^*]^1, \dots, [\tilde{y}_s^*]^{N-1} \right),$$

де $\tilde{y}_s^* = -(\widetilde{A1}_s^*)^{-1} \cdot (\lambda_1, \dots, \lambda_N)^\top$. Далі знаходимо відповідні середні доходи політик

$$\widehat{f}_s^* = (f_1^*, \dots, f_{s-1}^*, m, f(s+1), \dots, f(N))^\top,$$

враховуючи, що $\tau(\widehat{f}_s^*) = \widehat{y}_s^\top \cdot \widehat{r}_s^*$, де $\widehat{r}_s^* = I(\widehat{f}_s^*)$. Серед них обираємо максимальну f_s^* , таку, що $\tau(f_s^*) = \max_{\widehat{f}_s^*} \tau(\widehat{f}_s^*)$.

3.2. б) Якщо $s < N$, знаходимо $(\widetilde{A1}_{s+1}^*)^{-1}$, а саме,

$$(\widetilde{A1}_{s+1}^*)^{-1} = B^{-1} - \frac{1}{1 + y_2^\top \cdot B^{-1} \cdot x_2} \cdot B^{-1} \cdot x_2 \cdot y_2^\top \cdot B^{-1},$$

$$B^{-1} = (\widetilde{A1}_s^*)^{-1} - \frac{1}{1 + y_1^\top \cdot (\widetilde{A1}_s^*)^{-1} \cdot x_1} \cdot (\widetilde{A1}_s^*)^{-1} \cdot x_1 \cdot y_1^\top \cdot (\widetilde{A1}_s^*)^{-1},$$

де $B = \widetilde{A1}_s^* + x_1 \cdot y_1^\top$, $y_1^\top = [\widetilde{A1}_{s+1}^*]^{s,\bullet} - [\widetilde{A1}_s^*]^{s,\bullet}$, а x_2 відрізняється від стовпця $[\widetilde{A1}_{s+1}^*]^{\bullet,s} - [\widetilde{A1}_s^*]^{\bullet,s}$ тільки s -ою компонентою, яка дорівнює нулю; $x_1 = y_2$ має всі нульові компоненти, крім однієї, яка знаходиться на s -му місці і дорівнює 1 (тут вжито позначення $[A]^{s,\bullet}$, $[A]^{\bullet,s}$ — s -ий рядок та s -ий стовпець матриці A відповідно).

Якщо $s < N$, переозначуємо $s + 1$ як s і переходимо до процедури 3.2. а).

3.2. с) Якщо $s = N$ і $f^* \neq f$, де $f^* \equiv (f_1^*, \dots, f_N^*)^\top$, то повертаємось до процедури 3.2, розглядаючи в ролі f політику f^* .

Якщо $s = N$ і $f^* = f$, то f^* — оптимальна стратегія.

Зауваження 1. Порівняно з алгоритмом Ховарда, у даному алгоритмі врахована властивість, яка дозволяє при переході з одного стану до іншого відразу враховувати заміну на краще рішення в попередніх станах без повторної процедури перерахунку ваг. Ця перевага в загальному випадку дозволяє заощадити на знаходженні N розв'язків N лінійних систем розміру $N \times N$.

Зауваження 2. На відміну від алгоритму Ховарда, описаний вище алгоритм без особливих змін можна перенести на складніші випадки початкової моделі з додатковими обмеженнями та критеріями. Для такого узагальнення слід розглянути оптимізацію за $\tau(\pi)$ або (та) $D\tau(\pi)$.

- [1] *Беллман Р.* Введение в теорию матриц. – М.: Наука, 1969. – 367 с.
- [2] *Бобилляк А.М.* Модель прийняття рішень з багатьма обмеженнями в умовах ризику // Трансформація економічної системи в Україні. – Наук. збірник / За ред. З.Г.Ватаманюка. – Львів: Інтереко, 1999. – (Формування ринкової економіки України. – Вип. 5). – С. 524–528.
- [3] *Майн Х., Осаки С.* Марковские процессы принятия решений. – М.: Наука, 1977. – 176 с.
- [4] *Хорн Р., Джонсон Ч.* Матричный анализ. – М.: Мир, 1989. – 655 с.
- [5] *Ховард Р.* Динамическое программирование и марковские процессы. – М.: «Советское радио», 1964. – 192 с.
- [6] *Blackwell D.* Discrete dynamic programming // Ann. Math. Statist., 1962. – **33**. – P. 719–726.
- [7] *Derman C.* On sequential decisions and Markov chains // Management Sci., 1962. – **9**. – P. 16–24.

**METHOD FOR FINDING OPTIMAL STRATEGY OF
MARKOV DECISION PROCESS BASED ON PROPERTIES
EIGEN VECTOR**

Andriy BOBYLIAK

Ivan Franko Lviv National University,
1 Universytetska Str., Lviv 79000, Ukraine

The specific properties of eigenvectors are proved. The algorithms for finding optimal strategy of Markov decision process created for single criteria based on that properties, also approach formulated for generalization of algorithm for multicriterial cases.