ИЗВЕСТИЯ РАН. ТЕОРИЯ И СИСТЕМЫ УПРАВЛЕНИЯ, 2022, № 5, с. 14–22

УПРАВЛЕНИЕ В СТОХАСТИЧЕСКИХ СИСТЕМАХ И В УСЛОВИЯХ НЕОПРЕДЕЛЕННОСТИ

УДК 517.977

ИГРОВОЕ УПРАВЛЕНИЕ СЛУЧАЙНОЙ СКАЧКООБРАЗНОЙ СТРУКТУРОЙ ОБЪЕКТА В СМЕШАННЫХ СТРАТЕГИЯХ[1]

В. А. Болдинов[a],*, В. А. Бухалев[a], А. А. Скрынников[a]

[a] *Московский научно-исследовательский телевизионный ин-т МАИ (национальный исследовательский ун-т), ГНЦ ФАУ "ГосНИИАС", Москва, Россия*

**e-mail: victorboldinov@mail.ru*

Поступила в редакцию 13.12.2021 г.
После доработки 23.01.2022 г.
Принята к публикации 28.03.2022 г.

Рассматривается задача оптимального управления случайной скачкообразной структурой объекта в условиях противодействия. Смена состояний структуры объекта наблюдается противоборствующими сторонами с помощью индикаторов, работающих с ошибками. Критерием оптимальности управлений является некоторый функционал состояния объекта, который один из противников стремится минимизировать, а другой – максимизировать. Игроки управляют структурой объекта в *смешанных* стратегиях, применяя случайным образом управление. Оптимальные управления находятся в классе детерминированных зависимостей вероятностей управлений от результатов наблюдений, предшествующих текущему моменту. Приводится пример решения задачи оптимизации управления структурой объекта с двумя состояниями методами теории систем со случайной скачкообразной структурой в игровой постановке.

DOI: 10.31857/S0002338822050043

Введение. В работе [1] рассматривалась задача игрового управления случайной скачкообразной структурой (ССС) [1–8] некоторого динамического объекта *в чистых стратегиях*.

Структура объекта имеет конечное множество состояний, переходы которых из одного в другое происходят в случайные моменты времени и управляются двумя противоборствующими сторонами, преследующими строго противоположные цели.

При этом каждый из противников ("игроков") располагает конечным числом возможных стратегий (управлений) и руководствуется некоторым своим априорным представлением об управляемом объекте и информацией, которую он получает от своего индикатора структуры, регистрирующего с ошибками текущее состояние структуры объекта.

Как показано в [1], задача представляет собой игру с ненулевой суммой и не имеет седловой точки. Основная причина этого – неодинаковая информированность игроков о состоянии структуры.

Согласно теореме фон Неймана–Моргенштерна [9, 10], переход от управления в чистых стратегиях к управлению *в смешанных стратегиях* приводит к появлению седловой точки игры в задачах с полной информацией о состоянии управляемого объекта.

В задачах с неполной информацией при различной информированности игроков о состоянии объекта седловой точки игры в общем случае не должно быть. Однако следует ожидать определенного сближения верхней и нижней цены игры при переходе от управления в чистых стратегиях к управлению в смешанных стратегиях.

В статье рассматривается задача построения информационно-управляющих алгоритмов противников ("игроков"), состоящая в нахождении оптимальных вероятностей случайных управлений в текущий момент времени k в классе детерминированных зависимостей от показаний индикаторов структуры на отрезке времени от начального момента до текущего k.

Для решения задачи используются теория систем ССС [1–4], теория игр [5, 7, 9–15], теория стохастического динамического программирования [16, 17] на основе метода динамического

[1] Работа выполнена при финансовой поддержке РФФИ (проекты № 19-08-00502а, 19-08-00487).

программирования Беллмана [18], байесовская обработка информации [1–4, 9, 16, 17] и марковские математические модели [1–4, 9, 16, 17, 19]. Применение этих методов позволяет построить алгоритмы, сочетающие точность решения с простотой реализации. Их достоинствами являются: обратная связь управляющих сигналов с состоянием объекта, комплексирование априорной и апостериорной информации о состоянии объекта и рекуррентная форма алгоритмов, не требующая запоминания всей совокупности наблюдений на отрезке времени, предшествующих текущему моменту. Это особенно важно для систем с ограниченной памятью цифровых вычислительных систем (ЦВС), в бортовых ЦВС систем навигации и наведения летательных аппаратов [4, 20].

1. Постановка задачи. Дано: рассматривается объект ССС, управляемый двумя игроками – A и B, преследующими строго противоположные цели. Структура s_k описывается управляемой марковской цепью с конечным числом возможных состояний $s_k = \overline{1, n^{(s)}}$, где k – текущий момент времени: $k = \overline{0, n}$.

Априорная и текущая информированности игроков о состоянии структуры неодинаковы:

$$q_k^A(s_{k+1} | s_k, \Theta_k, \vartheta_k); \quad q_k^B(s_{k+1} | s_k, \Theta_k, \vartheta_k); \quad p_0^A(s_0); \quad p_0^B(s_0), \tag{1.1}$$

$$\pi_{k+1}^A(r_{k+1} | r_k, s_{k+1}, \Theta_k, \vartheta_k); \quad \pi_{k+1}^B(\rho_{k+1} | \rho_k, s_{k+1}, \Theta_k, \vartheta_k), \tag{1.2}$$

$$s_k = \overline{1, n^{(s)}}; \quad \Theta_k = \overline{1, n^{(\Theta)}}; \quad \vartheta_k = \overline{1, n^{(\vartheta)}}; \quad r_k = \overline{1, n^{(r)}}; \quad \rho_k = \overline{1, n^{(\rho)}},$$

где $q_k^A(\cdot)$, $q_k^B(\cdot)$ – вероятности переходов структуры из состояния s_k в состояние s_{k+1} при фиксированных управлениях: Θ_k – игрока A, ϑ_k – игрока B; $p_0^A(s_0)$, $p_0^B(s_0)$ – начальные вероятности состояний структуры; $\pi_{k+1}^A(\cdot)$, $\pi_{k+1}^B(\cdot)$ – вероятности переходов индикаторов структуры, описывамых управляемыми условными марковскими цепями, из состояния r_k в состояние r_{k+1} – для игрока A и из ρ_k в ρ_{k+1} – для игрока B при фиксированных s_k, Θ_k, ϑ_k.

Другими словами, индикаторы структуры регистрируют с ошибками состояния структуры s_{k+1}, обладают некоторой инерционностью (зависимость от r_k, ρ_k) и зависят от управлений Θ_k ϑ_k. Зависимость $\pi_{k+1}^A(\cdot)$ от ϑ_k означает, что игрок B может управлять характеристикой индикатора своего противника, осуществляя, таким образом, информационное противодействие игроку A. Аналогичный смысл имеет зависимость $\pi_{k+1}^B(\cdot)$ от Θ_k.

Игроки управляют структурой объекта в *смешанных стратегиях*, применяя случайным образом управление $\Theta_k = i$, $i = \overline{1, n^{\Theta}}$, с вероятностями $\alpha_k(\Theta_k)$ и $\vartheta_k = j$, $j = \overline{1, n^{\vartheta}}$, с вероятностями $\beta_k(\vartheta_k)$.

Показатели качества игры J^A и J^B аналогичны для обеих игроков. Разница между ними состоит в использовании различной информации, которую каждый из противников получает от своего индикатора структуры: $r_{\overline{0,n}}$ для игрока A и $\rho_{\overline{0,n}}$ для игрока B:

$$J^A(\alpha_{\overline{0,n-1}}, \beta_{\overline{0,n-1}}, r_{\overline{0,n-1}}) \triangleq \sum_{k=1}^{n} \mathrm{M}[W_k(s_k, \theta_{k-1}, \vartheta_{k-1}) | r_{\overline{0,k-1}}], \tag{1.3}$$

$$J^B(\alpha_{\overline{0,n-1}}, \beta_{\overline{0,n-1}}, \rho_{\overline{0,n-1}}) \triangleq \sum_{k=1}^{n} \mathrm{M}[W_k(s_k, \theta_{k-1}, \vartheta_{k-1}) | \rho_{\overline{0,k-1}}], \tag{1.4}$$

где $W_k(\cdot)$ – заданная текущая функция потерь. При этом

$$\mathrm{M}[W_k(s_k, \theta_{k-1}, \vartheta_{k-1}) | r_{\overline{0,k-1}}] = \sum_{s_k} \sum_{\theta_{k-1}} \sum_{\vartheta_{k-1}} W_k(s_k, \theta_{k-1}, \vartheta_{k-1}) p_k^A(s_k | \theta_{k-1}, \vartheta_{k-1}) \alpha_{k-1}(\theta_{k-1}) \beta_{k-1}(\vartheta_{k-1}), \tag{1.5}$$

$$\mathrm{M}[W_k(s_k, \theta_{k-1}, \vartheta_{k-1}) | \rho_{\overline{0,k-1}}] = \sum_{s_k} \sum_{\theta_{k-1}} \sum_{\vartheta_{k-1}} W_k(s_k, \theta_{k-1}, \vartheta_{k-1}) p_k^B(s_k | \theta_{k-1}, \vartheta_{k-1}) \alpha_{k-1}(\theta_{k-1}) \beta_{k-1}(\vartheta_{k-1}). \tag{1.6}$$

Здесь функции $p_k^A(\cdot) \triangleq \mathrm{P}[s_k\,|\,\theta_{k-1},\vartheta_{k-1},r_{\overline{0,k-1}}]$, $p_k^B(\cdot) \triangleq \mathrm{P}[s_k\,|\,\theta_{k-1},\vartheta_{k-1},\rho_{\overline{0,k-1}}]$ являются вероятностями состояний структур для игроков A и B; $\mathrm{M}[\cdot]$, $\mathrm{P}[\cdot]$ и $\triangleq$ – символы соответственно математического ожидания, вероятности и равенства по определению.

Критерии оптимальности управлений игроков отражают диаметральную противоположность их интересов и описывается формулами

$$J^{A*}(r_{\overline{0,n-1}}) = \min_{\alpha_{\overline{0,n-1}}}\max_{\beta_{\overline{0,n-1}}} J^A(\alpha_{\overline{0,n-1}},\beta_{\overline{0,n-1}},r_{\overline{0,n-1}}), \tag{1.7}$$

$$J^{B*}(\rho_{\overline{0,n-1}}) = \max_{\beta_{\overline{0,n-1}}}\min_{\alpha_{\overline{0,n-1}}} J^B(\alpha_{\overline{0,n-1}},\beta_{\overline{0,n-1}},\rho_{\overline{0,n-1}}), \tag{1.8}$$

т.е. игрок A выбирает оптимальную стратегию на отрезке $[0,n-1]$, добиваясь *минимума* показателя качества $J^A(\cdot)$ и предполагая, что его противник будет придерживаться стратегии, *максимизирующей* этот показатель. Противоположным образом действует игрок B, который *максимизирует* показатель качества $J^B(\cdot)$ в расчете на стратегию игрока A, *минимизирующего* этот показатель.

Требуется найти: алгоритмы оптимального управления структурой для обоих игроков в виде детерминированных зависимостей вероятностей α_k, β_k от показателей индикаторов структуры $r_{\overline{0,k}}$, $\rho_{\overline{0,k}}$.

2. Алгоритм игрока A. 2.1. Регулятор структуры. Найдем уравнения регулятора структуры (блока управления), связывающие оптимальные вероятности управлений с вероятностью состояния структуры.

С учетом специфики поставленной задачи применим метод динамического программирования Р. Беллмана, обобщенный и модифицированный для синтеза оптимальных управлений с обратной связью по состоянию объекта в стохастических системах [1, 2, 4, 16, 17].

Введем понятие “функция оставшихся потерь” J_k^A, которую определим формулами

$$J_k^A(\alpha_{\overline{k-1,n-1}},\beta_{\overline{k-1,n-1}},r_{\overline{0,k-1}}) \triangleq \sum_{i=k}^{n} W_i^A(\alpha_{i-1},\beta_{i-1},r_{\overline{0,k-1}}), \tag{2.1}$$

$$W_i^A(\alpha_{i-1},\beta_{i-1},r_{\overline{0,k-1}}) \triangleq M[W_i(s_i,\theta_{i-1},\vartheta_{i-1})\,|\,r_{\overline{0,k-1}}]. \tag{2.2}$$

Представив $J_k^A(\cdot)$ в виде суммы двух слагаемых $W_i^A(\cdot)$ и оставшейся части суммы из (2.1), (2.2), получаем

$$\begin{aligned} J_k^A(\alpha_{\overline{k-1,n-1}},\beta_{\overline{k-1,n-1}},r_{\overline{0,k-1}}) &= W_k^A(\alpha_{i-1},\beta_{i-1},r_{\overline{0,k-1}}) + \sum_{i=k+1}^{n} W_i^A(\alpha_{i-1},\beta_{i-1},r_{\overline{0,k-1}}) = \\ &= W_k^A(\alpha_{i-1},\beta_{i-1},r_{\overline{0,k-1}}) + \sum_{r_k} J_{k+1}^A(\alpha_{\overline{k,n-1}},\beta_{\overline{k,n-1}},r_k,r_{\overline{0,k-1}})\gamma_k^A(r_k), \end{aligned} \tag{2.3}$$

где

$$\gamma_k^A(r_k) \triangleq P[r_k\,|\,r_{\overline{0,k-1}}] = \sum_{s_k}\sum_{\theta_{k-1}}\sum_{\vartheta_{k-1}} \pi_k^A(r_k\,|\,r_{k-1},s_k,\theta_{k-1},\vartheta_{k-1})p_k^A(s_k\,|\,\theta_{k-1},\vartheta_{k-1})\alpha_{k-1}(\theta_{k-1})\beta_{k-1}(\vartheta_{k-1}), \tag{2.4}$$

$$p_k^A(s_k\,|\,\theta_{k-1},\vartheta_{k-1}) = \sum_{s_{k-1}} q_{k-1}^A(s_k\,|\,s_{k-1},\theta_{k-1},\vartheta_{k-1})\hat{p}_{k-1}^A(s_{k-1}\,|\,\theta_{k-1},\vartheta_{k-1}). \tag{2.5}$$

Здесь $\hat{p}_{k-1}^A(s_{k-1}\,|\,\theta_{k-1},\vartheta_{k-1}) \triangleq P[s_{k-1}\,|\,\theta_{k-1},\vartheta_{k-1},r_{\overline{0,k-1}}]$ – апостериорная вероятность состояния структуры.

Так как в силу марковских свойств вероятность $\hat{p}_{k-1}(s_{k-1}\,|\,\theta_{k-1},\vartheta_{k-1})$ не зависит от θ_{k-1}, ϑ_{k-1}, то формула (2.5) принимает вид

$$p_k^A(s_k\,|\,\theta_{k-1},\vartheta_{k-1}) = \sum_{s_{k-1}} q_{k-1}^A(s_k\,|\,s_{k-1},\theta_{k-1},\vartheta_{k-1})\hat{p}_{k-1}^A(s_{k-1}). \tag{2.6}$$

Оптимизируем функцию оставшихся потерь $J_k^A(\cdot)$ по минимаксному критерию:

$$J_k^{A*}(r_{\overline{0,k-1}}) = \min_{\alpha_{\overline{k-1,n-1}}} \max_{\beta_{\overline{k-1,n-1}}} J_k^A(\alpha_{\overline{k-1,n-1}}, \beta_{\overline{k-1,n-1}}, r_{\overline{0,k-1}}) =$$

$$= \min_{\alpha_{\overline{k-1,n-1}}} \max_{\beta_{\overline{k-1,n-1}}} \left[W_k^A(\alpha_{k-1}, \beta_{k-1}, r_{\overline{0,n-1}}) + \sum_{r_k} J_{k+1}^A(\alpha_{\overline{k,n-1}}, \beta_{\overline{k,n-1}}, r_{\overline{0,k}}) \gamma_k^A(r_k) \right] = \tag{2.7}$$

$$= \min_{\alpha_{k-1}} \max_{\beta_{k-1}} \min_{\alpha_{\overline{k,n-1}}} \max_{\beta_{\overline{k,n-1}}} \left[W_k^A(\alpha_{k-1}, \beta_{k-1}, r_{\overline{0,k-1}}) + \sum_{r_k} J_{k+1}^A(\alpha_{\overline{k,n-1}}, \beta_{\overline{k,n-1}}, r_{\overline{0,k}}) \gamma_k^A(r_k) \right].$$

Так как $W_k^A(\cdot)$ в (2.7) от α_{k-1}, β_{k-1} не зависит, а

$$\min_{\alpha_{\overline{k,n-1}}} \max_{\beta_{\overline{k,n-1}}} \sum_{r_k} J_{k+1}^A(\alpha_{\overline{k,n-1}}, \beta_{\overline{k,n-1}}, r_{\overline{0,k}}) \gamma_k^A(r_k) = \sum_{r_k} J_{k+1}^{A*}(r_{\overline{0,k}}) \gamma_k^A(r_k),,$$

то из (2.7) следует рекуррентное уравнение для $J_k^{A*}(\cdot)$:

$$J_k^{A*}(r_{\overline{0,k-1}}) = \min_{\alpha_{k-1}} \max_{\beta_{k-1}} \left[W_k^A(\alpha_{k-1}, \beta_{k-1}, r_{\overline{0,k-1}}) + \sum_{r_k} J_{k+1}^{A*}(\alpha_{k-1}, \beta_{k-1}, r_{\overline{0,k}}) \gamma_k^A(r_k) \right],$$
$$k = n, n-1, \ldots, 1; \quad J_{n+1}^{A*} \equiv 0. \tag{2.8}$$

Пара минимаксных смешанных стратегий, согласно (2.8), определяется формулой

$$(\alpha_{k-1}^*, \beta_{k-1}^A) = \arg \min_{\alpha_{k-1}} \max_{\beta_{k-1}} \left[W_k^A(\hat{p}_{k-1}(s_{k-1}), \alpha_{k-1}, \beta_{k-1}) + \sum_{r_k} J_{k+1}^{A*}(\hat{p}_{k-1}(s_{k-1}), \alpha_{k-1}, \beta_{k-1}) \gamma_k^A(r_k) \right], \tag{2.9}$$

где α_{k-1}^* – оптимальная стратегия игрока A, а β_{k-1}^A – предполагаемая игроком A оптимальная стратегия игрока B, основанные на показаниях индикатора структуры $r_{\overline{0,k-1}}$, принадлежащего игроку A.

Рекуррентные уравнения (2.8), (2.9) описывают алгоритм регулятора структуры игрока A. Выходными сигналами регулятора являются стратегии $\alpha_k^*(\theta_k)$, $\beta_k^A(\vartheta_k)$ – вероятности управлений игроков θ_k, ϑ_k, а входным сигналом – апостериорная вероятность состояния структуры $\hat{p}_k^A(s_k)$, которая определяется классификатором структуры.

2.2. Классификатор структуры. Апостериорная вероятность состояния структуры $\hat{p}_k^A(s_k)$, согласно формуле Байеса, обобщенной на класс систем ССС [1–4], и формуле полной вероятности, определяется рекуррентными уравнениями

$$\hat{p}_{k+1}^A(s_{k+1}) = [\gamma_{k+1}^A(r_{k+1})]^{-1} \times \sum_{\Theta_k} \sum_{\vartheta_k} \pi_{k+1}^A(r_{k+1} \,|\, r_k, s_{k+1}, \theta_k, \vartheta_k) p_{k+1}^A(s_{k+1} \,|\, \theta_k, \vartheta_k) \alpha_k^*(\theta_k) \beta_k^A(\vartheta_k),$$

$$p_{k+1}^A(s_{k+1} \,|\, \theta_k, \vartheta_k) = \sum_{s_k} q_k^A(s_{k+1} \,|\, s_k, \theta_k, \vartheta_k) \hat{p}_k^A(s_k),$$

$$\gamma_{k+1}^A(r_{k+1}) = \sum_{s_{k+1}} \sum_{\theta_k} \sum_{\vartheta_k} \pi_{k+1}^A(r_{k+1} \,|\, r_k, s_{k+1}, \theta_k, \vartheta_k) p_{k+1}^A(s_{k+1} \,|\, \theta_k, \vartheta_k) \alpha_k^*(\theta_k) \beta_k^A(\vartheta_k), \tag{2.10}$$

$$k = \overline{0, n-1}; \quad s_k = \overline{1, n^{(s)}}; \quad \hat{p}_0^A(s_0) = p_0^A(s_0).$$

В целом, оптимальный *минимаксный* информационно-управляющий алгоритм игрока A описывается замкнутой системой рекуррентных уравнений (2.8)–(2.10), в которой уравнения регулятора (2.8), (2.9) решаются в “обратном” времени ($k = n, n-1, \ldots, 1$) при начальных условиях $J_{n+1}^{A*} \equiv 0$, а уравнения классификатора (2.10) – в прямом времени ($k = 0, 1, \ldots, n-1$) при начальных условиях $\hat{p}_0(s_0) = p_0^A(s_0)$ (двухточечная краевая задача (ДКЗ)).

При этом оптимальные стратегии определяются как функции апостериорных вероятностей состояния структуры: $\alpha_k^*(\hat{p}_k^A(s_k))$, $\beta_k^A(\hat{p}_k^A(s_k))$ и запоминаются в ЦВМ. Таким образом, мы имеем закон управления с обратной связью по состоянию структуры. Для того, чтобы воспользоваться им, необходим второй этап – нахождение апостериорной вероятности $\hat{p}_k^A(s_k)$. Эта вероятность находится в результате решения уравнений (2.10) в процессе управления структурой s_k. При этом в качестве управляющих воздействий используются запомненные функции $\alpha_k^*(\hat{p}_k^A(s_k))$, $\beta_k^A(s_k)$.

3. Алгоритм игрока *B*. Аналогичный информационно-управляющий *максиминный* алгоритм игрока B описывается уравнениями (2.8)–(2.10), в которых производятся следующие замены: $\min\limits_{\alpha_{k-1}} \max\limits_{\beta_{k-1}} [\cdot]^A \to \max\limits_{\beta_{k-1}} \min\limits_{\alpha_{k-1}} [\cdot]^B$; индекс $A \to$ индекс B; $r_k \to \rho_k$; $\alpha_k^*(\theta_k) \to \alpha_k^B(\Theta_k)$, $\beta_k^A(\vartheta_k) \to \beta_k^*(\vartheta_k)$, где $\beta_k^*(\vartheta_k)$ – оптимальная стратегия B, а $\alpha_k^B(\Theta_k)$ – предполагаемая игроком B оптимальная стратегия его противника A, основанная на показаниях индикатора структуры $\rho_{\overline{0,k}}$, принадлежащего игроку B.

Алгоритм имеет следующий вид.

3.1. Регулятор структуры.

$$J_k^{B*} = \max_{\alpha_{k-1}} \min_{\beta_{k-1}} [W_k^B(\cdot) + \tilde{J}_{k+1}^{B*}(\cdot)], \tag{3.1}$$

$$\tilde{J}_{k+1}^{B*}(\cdot) \triangleq \sum_{\rho_k} J_{k+1}^{B*}(\rho_k)\gamma_k^B(\rho_k), \tag{3.2}$$

$$\gamma_k^B(\rho_k) = \sum_{s_k}\sum_{\Theta_{k-1}}\sum_{\vartheta_{k-1}} \pi_k^B(\rho_k \,|\, \rho_{k-1}, s_k, \Theta_{k-1}, \vartheta_{k-1}) \times p_k^B(s_k \,|\, \Theta_{k-1}, \vartheta_{k-1})\alpha_{k-1}(\Theta_{k-1})\beta_{k-1}(\vartheta_{k-1}), \tag{3.3}$$

$$W_k^B(\cdot) = \sum_{s_k}\sum_{\Theta_{k-1}}\sum_{\vartheta_{k-1}} W_k^B(s_k \,|\, \Theta_{k-1}, \vartheta_{k-1}) p_k^B(s_k \,|\, \Theta_{k-1}, \vartheta_{k-1})\alpha_{k-1}(\Theta_{k-1})\beta_{k-1}(\vartheta_{k-1}), \tag{3.4}$$

$$p_k^B(s_k \,|\, \Theta_{k-1}, \vartheta_{k-1}) = \sum_{s_{k-1}} q_{k-1}^B(s_k \,|\, s_{k-1}, \Theta_{k-1}, \vartheta_{k-1})\hat{p}_{k-1}^B(s_{k-1}), \tag{3.5}$$

$$(\beta_{k-1}^*, \alpha_{k-1}^B) = \arg\max_{\beta_{k-1}} \min_{\alpha_{k-1}} [W_k^B(\cdot) + \tilde{J}_{k+1}^{B*}(\cdot)], \tag{3.6}$$

$$k = n, n-1, \ldots, 1; \quad J_{n+1}^{B*} \equiv 0,$$

где

$$W_k^B(\cdot) \triangleq W_k^B(\hat{p}_{k-1}^B(s_{k-1}), \alpha_{k-1}, \beta_{k-1}),$$

$$\tilde{J}_{k+1}^{B*}(\cdot) \triangleq \tilde{J}_{k+1}^{B*}(\hat{p}_{k-1}^B(s_{k-1}), \alpha_{k-1}, \beta_{k-1}).$$

3.2. Классификатор структуры.

$$\hat{p}_{k+1}^B(s_{k+1}) = [\gamma_{k+1}^B(\rho_{k+1})]^{-1} \times \sum_{\Theta_k}\sum_{\vartheta_k} \pi_{k+1}^B(\rho_{k+1} \,|\, \rho_k, s_{k+1}, \Theta_k, \vartheta_k) p_{k+1}^B(s_{k+1} \,|\, \Theta_k, \vartheta_k)\alpha_k^B(\Theta_k)\beta_k^*(\vartheta_k), \tag{3.7}$$

$$\gamma_{k+1}^B(\rho_{k+1}) = \sum_{s_{k+1}}\sum_{\Theta_k}\sum_{\vartheta_k} \pi_{k+1}^B(\rho_{k+1} \,|\, \rho_k, s_{k+1}, \Theta_k, \vartheta_k) p_{k+1}^B(s_{k+1} \,|\, \Theta_k, \vartheta_k)\alpha_k^B(\Theta_k)\beta_k^*(\vartheta_k), \tag{3.8}$$

$$p_{k+1}^B(s_{k+1} \,|\, \Theta_k, \vartheta_k) = \sum_{s_k} q_k^B(s_{k+1} \,|\, s_k, \Theta_k, \vartheta_k)\hat{p}_k^B(s_k), \tag{3.9}$$

$$k = \overline{0, n-1}; \quad s_k = \overline{1, n^{(s)}}; \quad \hat{p}_0^B = p_0^B(s_0).$$

4. Пример. Рассмотрим задачу оптимизации управления ССС объекта с двумя состояниями в смешанных стратегиях как частный случай общей постановки задачи из разд. 1.

Дано.

1. Матрицы вероятностей переходов: Q_k – из состояния $s_k = 1$ в состояние $s_k = 2$ и G_k – из состояния $s_k = 2$ в состояние $s_k = 1$:

$$Q_k = \begin{bmatrix} q_k^{11} & q_k^{12} \\ q_k^{21} & q_k^{22} \end{bmatrix}; \quad G_k = \begin{bmatrix} g_k^{11} & g_k^{12} \\ g_k^{21} & g_k^{22} \end{bmatrix}, \tag{4.1}$$

где

$$q_k^{ij} \triangleq \mathrm{P}[s_{k+1} = 2 | s_k = 1, \Theta_k = i, \vartheta_k = j],$$

$$g_k^{ij} \triangleq \mathrm{P}[s_{k+1} = 1 | s_k = 2, \Theta_k = i, \vartheta_k = j], \quad i, j = 1, 2.$$

Индекс k у квадратных скобок у (4.1) означает, что все элементы матриц Q_k и G_k зависят от k.

2. Игрок A случайным образом с вероятностью α_k^i применяет управление $\theta_k = i$, а игрок B с вероятностью β_k^j применяет управление $\vartheta_k = j$:

$$\sum_{i=1}^{2} \alpha_k^i = \sum_{i=1}^{2} \beta_k^j = 1.$$

3. Стратегии игроков основываются на показаниях индикаторов структуры соответственно $r_{\overline{0,k}}$ и $\rho_{\overline{0,k}}$, которые описываются условно-марковскими цепями, заданными вероятностями переходов $\pi_{k+1}^A(r_{k+1} | r_k, s_{k+1}; \Theta_k, \vartheta_k)$, $\pi_{k+1}^B(\rho_{k+1} | \rho_k, s_{k+1}; \Theta_k, \vartheta_k)$, $r_k = 1,2$; $\rho_k = 1,2$; $\Theta_k = 1,2$; $\vartheta_k = 1,2$; $s_k = 1, 2$.

4. Критерии оптимальности управлений игроков:

$$J^{A*} = \min_{\alpha_{\overline{0,n-1}}} \max_{\beta_{\overline{0,n-1}}} J^A; \quad J^{B*} = \max_{\beta_{\overline{0,n-1}}} \min_{\alpha_{\overline{0,n-1}}} J^B, \tag{4.2}$$

где $\alpha_k \triangleq (\alpha_k^1, \alpha_k^2)$, $\beta_k \triangleq (\beta_k^1, \beta_k^2)$;

$$J^A = \sum_{k=1}^{n} \mathrm{M}[W_k(s_k) | r_{\overline{0,k-1}}], \quad J^B = \sum_{k=1}^{n} \mathrm{M}[W_k(s_k) | \rho_{\overline{0,k-1}}], \tag{4.3}$$

$$W_k(s_k) = \delta(s_k, 1). \tag{4.4}$$

Здесь J^A, J^B – показатели качества; $W_k(s_k)$ – текущая функция потерь; $\delta(s_k, 1)$ – символ Кронекера:

$$\delta(s_k, 1) = \begin{cases} 1 & \text{при} \quad s_k = 1, \\ 0 & \text{при} \quad s_k = 2. \end{cases}$$

Требуется найти: оптимальные алгоритмы управления противников в виде детерминированных зависимостей вероятностей α_k^*, β_k^* от показателей индикаторов структуры $r_{\overline{0,k}}$, $\rho_{\overline{0,k}}$.

Решение

Регулятор структуры игрока A. Из (2.2), (2.8), (2.10), (4.3), (4.4) следует

$$W_k^A(\hat{p}_{k-1}^A(s_{k-1}, \alpha_{k-1}, \beta_{k-1})) = \sum_{s_k=1}^{2} \sum_{i,j=1}^{2} \delta(s_k, 1) p_k^A(s_k | i, j) \alpha_{k-1}^i \beta_{k-1}^j =$$

$$= \sum_{i,j=1}^{2} p_k^A(1 | i, j) \alpha_{k-1}^i \beta_{k-1}^j = \tilde{p}_k^A(1) = \sum_{i,j=1}^{2} [(1 - q_{k-1}^{ij}) \hat{p}_{k-1}^A(1) + g_{k-1}^{ij} \hat{p}_{k-1}^A(2)] \alpha_{k-1}^i \beta_{k-1}^j = \tag{4.5}$$

$$= [1 - q_{k-1}(\alpha_{k-1}, \beta_{k-1})] \hat{p}_{k-1}^A(1) + g_{k-1}(\alpha_{k-1}, \beta_{k-1}) \hat{p}_{k-1}^A(2),$$

где

$$q_{k-1}(\alpha_{k-1}, \beta_{k-1}) \triangleq \sum_{i,j=1}^{2} q_{k-1}^{ij} \alpha_{k-1}^i \beta_{k-1}^j, \tag{4.6}$$

$$g_{k-1}(\alpha_{k-1}, \beta_{k-1}) \triangleq \sum_{i,j=1}^{2} g_{k-1}^{ij} \alpha_{k-1}^i \beta_{k-1}^j, \tag{4.7}$$

$$\tilde{p}_k(1) \triangleq \mathrm{P}\,[s_k = 1 | r_{\overline{0,k-1}}], \tag{4.8}$$

$$J_{k+1}^{A*} = \psi_k^A \hat{p}_k^A(1) + m_k^A, \tag{4.9}$$

здесь ψ_k^A, m_k^A – неопределенные коэффициенты.

Из (2.5), (2.7), (4.5), (4.9) вытекает

$$\tilde{J}_{k+1}^{A*} = \psi_k^A \tilde{p}_k^A(1) + m_k^A = \psi_k^A\{[1 - q_{k-1}(\alpha_{k-1},\beta_{k-1})]\hat{p}_{k-1}^A(1) + g_{k-1}(\alpha_{k-1},\beta_{k-1})\hat{p}_{k-1}^A(2)\} + m_k^A. \quad (4.10)$$

Пара минимаксных смешанных стратегий $(\alpha_{k-1}^*,\beta_{k-1}^A)$ определяется формулой (2.9). Подставив $(\alpha_{k-1}^*,\beta_{k-1}^A)$ в (2.6), (4.5), (4.10) получаем

$$\psi_{k-1}\hat{p}_{k-1}^A(1) + m_{k-1}^A = [1 - q_{k-1}(\alpha_{k-1}^*,\beta_{k-1}^A)]\hat{p}_{k-1}^A(1) + g_{k-1}(\alpha_{k-1}^*,\beta_{k-1}^A)\hat{p}_{k-1}^A(2) +$$
$$+ \psi_k\{[1 - q_{k-1}(\alpha_{k-1}^*,\beta_{k-1}^A)]\hat{p}_{k-1}^A(1) + g_{k-1}(\alpha_{k-1}^*,\beta_{k-1}^A)\hat{p}_{k-1}^A(2)\} + m_k^A,$$

откуда, учитывая, что $\hat{p}_{k-1}^A(2) = 1 - \hat{p}_{k-1}^A(1)$, следует

$$\begin{gathered}\psi_{k-1}\hat{p}_{k-1}^A(1) + m_{k-1}^A = (1+\psi_k)\{[1 - q_{k-1}(\alpha_{k-1}^*,\beta_{k-1}^A) - \\ - g_{k-1}(\alpha_{k-1}^*,\beta_{k-1}^A)]\hat{p}_{k-1}^A(1) + g_{k-1}(\alpha_{k-1}^*,\beta_{k-1}^A)\} + m_k^A.\end{gathered} \quad (4.11)$$

Приравняв коэффициенты при одинаковых степенях $\hat{p}_{k-1}^A(1)$ в левой и правой частях уравнения (4.11) и обозначив

$$h_{k-1}^A(\alpha_{k-1}^*,\beta_{k-1}^A) \triangleq 1 - q_{k-1}(\alpha_{k-1}^*,\beta_{k-1}^A) - g_{k-1}(\alpha_{k-1}^*,\beta_{k-1}^A), \quad (4.12)$$

получаем рекуррентные уравнения для ψ_k^A, m_k^A:

$$\begin{gathered}\psi_{k-1}^A = h_{k-1}^A(\alpha_{k-1}^*,\beta_{k-1}^A)(1+\psi_k^A), \quad m_{k-1}^A = m_k^A + g_{k-1}(\alpha_{k-1}^*,\beta_{k-1}^A), \\ k = n, n-1,\ldots,1; \quad \psi_n^A \equiv 0.\end{gathered} \quad (4.13)$$

Согласно (2.11), (4.5), (4.9), (4.12), пара $(\alpha_{k-1}^*,\beta_{k-1}^A)$ определяется формулой

$$\begin{gathered}(\alpha_{k-1}^*,\beta_{k-1}^A) = \arg\min_{\alpha_{k-1}}\max_{\beta_{k-1}}[W_k^A(\cdot) + \tilde{J}_{k+1}^{A*}(\cdot)] = \\ = \arg\min_{\alpha_{k-1}}\max_{\beta_{k-1}}\{(1+\psi_k)[h_{k-1}^A(\alpha_{k-1},\beta_{k-1})\hat{p}_{k-1}^A(1) + g_{k-1}(\alpha_{k-1},\beta_{k-1})] + m_k^A\}.\end{gathered} \quad (4.14)$$

Так как ψ_k^A, m_k^A и $\hat{p}_{k-1}^A(1)$ не зависят от α_{k-1}, β_{k-1}, то из (4.12), (4.14) следует

$$\begin{gathered}(\alpha_{k-1}^*,\beta_{k-1}^A) = \arg\min_{\alpha_{k-1}}\max_{\beta_{k-1}}[-q_{k-1}(\alpha_{k-1},\beta_{k-1})\hat{p}_{k-1}^A(1) + g_{k-1}(\alpha_{k-1},\beta_{k-1})\hat{p}_{k-1}^A(2)] = \\ = \arg\min_{\alpha_{k-1}}\max_{\beta_{k-1}} l_{k-1}(\alpha_{k-1},\beta_{k-1}), \quad k = \overline{1,n},\end{gathered} \quad (4.15)$$

где

$$l_{k-1}(\cdot) \triangleq -q_{k-1}(\cdot)\hat{p}_{k-1}^A(1) + g_{k-1}(\cdot)\hat{p}_{k-1}^A(2). \quad (4.16)$$

Из теории игр известно, что для парной матричной игры 2 × 2 в смешанных стратегиях оптимальная пара $(\alpha_{k-1}^*,\beta_{k-1}^A)$ определяется формулами [15]

$$\begin{gathered}\alpha_{k-1}^{1*} = \frac{p_k^A(1|2,2) - p_k^A(1|2,1)}{p_k^{A^\#}}, \quad \beta_{k-1}^{1^A} = \frac{p_k^A(1|2,2) - p_k^A(1|1,2)}{p_k^{A^\#}}, \\ \alpha_{k-1}^{2*} = 1 - \alpha_{k-1}^{1*}; \quad \beta_{k-1}^{2^A} = 1 - \beta_{k-1}^{1^A}; \quad k = \overline{1,n},\end{gathered} \quad (4.17)$$

где

$$\begin{gathered}p_k^{A^\#} \triangleq p_k^A(1|1,1) + p_k^A(1|2,2) - p_k^A(1|1,2) - p_k^A(1|2,1), \\ p_k^A(1|i,j) = (1 - q_{k-1}^{ij})\hat{p}_{k-1}^A(1) + g_{k-1}^{ij}\hat{p}_{k-1}^A(2), \quad i,j = 1,2.\end{gathered} \quad (4.18)$$

Цена игры $\tilde{p}_k^{A*}(1)$ и критерий оптимальности находятся по формулам

$$\tilde{p}_k^{A*}(1) = \frac{p_k^A(1|1,1)p_k^A(1|2,2) - p_k^A(1|1,2)p_k^A(1|2,1)}{p_k^{A^\#}},$$
$$J^{A*} = \sum_{k=0}^{n} \tilde{p}_k^{A*}(1). \tag{4.19}$$

Классификатор структуры игрока А. Согласно (2.8), (2.10), (2.12), рекуррентные уравнения классификатора структуры принимают вид

$$\hat{p}_{k+1}^A(1) = \frac{\sum_{i,j=1}^{2} \pi_{k+1}^A(r_{k+1}|r_k,1,i,j)p_{k+1}^A(1|i,j)\alpha_k^{i*}\beta_k^{j^A}}{\sum_{s_{k+1}=1}^{2}\sum_{i,j=1}^{2} \pi_{k+1}^A(r_{k+1}|r_k,s_{k+1},i,j)p_{k+1}^A(s_{k+1}|i,j)\alpha_k^{i*}\beta_k^{j^A}},$$
$$\hat{p}_{k+1}^A(2) = 1 - \hat{p}_{k+1}^A(1), \tag{4.20}$$
$$p_{k+1}^A(1|i,j) = (1 - q_k^{ij})\hat{p}_k^A(1) + g_k^{ij}\hat{p}_k^A(2),$$
$$p_{k+1}^A(2|i,j) = 1 - p_{k+1}^A(1|i,j); \quad k = \overline{0, n-1}.$$

В целом, информационный-управляющий алгоритм игрока A в смешанных стратегиях описывается замкнутой системой рекуррентных уравнений (4.17)–(4.20) относительно (α_k^*, β_k^A) в "прямом времени" ($k = 0,1,\ldots,n$), где α_k^* – оптимальная минимаксная стратегия игрока A, а β_k^A – предполагаемая игроком A оптимальная максиминная стратегия игрока B, основанные на показаниях индикатора структуры, принадлежащего игроку А.

Регулятор структуры и классификатор структуры игрока В. Согласно (3.1)–(3.9) и по аналогии с (4.17)–(4.20) уравнения регулятора и классификатора для игрока B принимают вид

$$\alpha_{k-1}^{1^B} = \frac{p_k^B(1|2,2) - p_k^B(1|2,1)}{p_k^{B^\#}}, \quad \beta_{k-1}^{1*} = \frac{p_k^B(1|2,2) - p_k^B(1|1,2)}{p_k^{B^\#}},$$
$$\alpha_{k-1}^{2^B} = 1 - \alpha_{k-1}^{1^B}; \quad \beta_{k-1}^{2*} = 1 - \beta_{k-1}^{1*}; \quad k = \overline{1, n},$$

где

$$p_k^{B^\#} \triangleq p_k^B(1|1,1) + p_k^B(1|2,2) - p_k^B(1|1,2) - p_k^B(1|2,1),$$
$$p_k^B(1|i,j) = (1 - q_{k-1}^{ij})\hat{p}_{k-1}^B(1) + g_{k-1}^{ij}\hat{p}_{k-1}^B(2), \quad i,j = 1,2,$$
$$\tilde{p}_k^{B*}(1) = \frac{p_k^B(1|1,1)p_k^B(1|2,2) - p_k^B(1|1,2)p_k^B(1|2,1)}{p_k^{B^\#}}, \tag{4.21}$$
$$J^{B*} = \sum_{k=0}^{n} \tilde{p}_k^{B*}(1);$$

$$\hat{p}_{k+1}^B(1) = \frac{\sum_{i,j=1}^{2} \pi_{k+1}^B(\rho_{k+1}|\rho_k,1,i,j)p_{k+1}^B(1|i,j)\alpha_k^{i^B}\beta_k^{j*}}{\sum_{s_{k+1}=1}^{2}\sum_{i,j=1}^{2} \pi_{k+1}^B(\rho_{k+1}|\rho_k,s_{k+1},i,j)p_{k+1}^B(s_{k+1}|i,j)\alpha_k^{i^B}\beta_k^{j*}},$$
$$\hat{p}_{k+1}^B(2) = 1 - \hat{p}_{k+1}^B(1), \tag{4.22}$$
$$p_{k+1}^B(1|i,j) = (1 - q_k^{ij})\hat{p}_k^B(1) + g_k^{ij}\hat{p}_k^B(2),$$

$$p^B_{k+1}(2|i,j) = 1 - p^B_{k+1}(1|i,j); \quad k = \overline{0, n-1}.$$

Информационно-управляющий алгоритм игрока B в смешанных стратегиях описывается замкнутой системой рекуррентных уравнений (4.21), (4.22) относительно (α^B_k, β^*_k) в “прямом времени” ($k = 0,1,\ldots,n$), где β^*_k – оптимальная максиминная стратегия игрока B, а α^A_k – предполагаемая игроком B оптимальная минимаксная стратегия игрока A, основанные на показаниях $\rho_{\overline{0,k}}$ индикатора структуры, принадлежащего игроку B.

Заключение. Каждый из игровых информационно-управляющих алгоритмов противоборствующих сторон состоит из двух взаимосвязанных блоков: регулятора структуры и классификатора структуры.

Несмотря на то, что задача решается в смешанных стратегиях, она является, в отличие от классического варианта [9, 11–15], игрой с ненулевой суммой и не имеет седловой точки вследствие различной информированности игроков о результатах игры, так как каждый из игроков основывает свою оптимальную стратегию на данных своего индикатора структуры, показания которых в общем случае не совпадают ($r_k \neq \rho_k$).

СПИСОК ЛИТЕРАТУРЫ

1. *Бухалёв В.А., Болдинов В.А., Скрынников А.А.* Игровое управление случайной скачкообразной структурой объекта в чистых стратегиях // Изв. РАН. ТиСУ. 2020. № 4. С. 39–51.
2. *Бухалёв В.А.* Распознавание, оценивание и управление в системах со случайной скачкообразной структурой. М.: Наука, 1996. 287 с.
3. *Бухалёв В.А.* Оптимальное сглаживание в системах со случайной скачкообразной структурой. М.: Физматлит, 2013. 188 с.
4. *Бухалёв В.А., Скрынников А.А., Болдинов В.А.* Алгоритмическая помехозащита беспилотных летательных аппаратов. М.: Физматлит, 2018. 192 с.
5. *Piers B.D., Sworder D.D.* Bayes and Minimax Controllers for a Linear Systems for Stochastic Jump Parameters // IEEE Trans. AC-16. 1971. № 4. P. 677–685.
6. *Zhang C., Zhu H., Zhou H., Bin N.* Deterministic and Stochastic Differential Games // Non-cooperative Stochastic Differential Game Theory of Generalized Markov Jump Linear Systems. Studies in Systems, Decision and Control. V. 67. Cham: Springer, 2017. 187 p.
7. *Бухалёв А.И.* Игровая задача управления в системе со случайной скачкообразной структурой // Изв. АН. Тех. кибернетика. 1993. № 2. С. 122–132.
8. *Moon J.* A Sufficient Condition for Linear-Quadratic Stochastic Zero-Sum Differential Games for Markov Jump Systems // IEEE Transactions on Automatic Control. 2019. V. 64. № 4. P. 1619–1626.
9. *Брайсон А.Е., Хо Ю Ши.* Прикладная теория оптимального управления. М.: Мир, 1972. 544 с.
10. *Фон Нейманн Дж., Моргенштерн О.* Теория игр и экономическое поведение. М.: Наука, 1970. 707 с.
11. *Айзекс Р.* Дифференциальные игры. М.: Мир, 1967. 480 с.
12. *Гермейер Ю.Б.* Введение в теорию исследований операций. М.: Наука, 1971. 384 с.
13. *Вентцель Е.С.* Исследование операций: задачи, принципы, методология. М.: Кнорус, 2013. 192 с.
14. *Оуэн Г.* Теория игр. М.: Вузовская книга, 2007. 216 с.
15. *Абчук В.А., Матвейчук Ф.А., Томашевский Л.П.* Справочник по исследованию операций / Под ред. Ф.А. Матвейчука. М.: Воениздат, 1979. 368 с.
16. *Фельдбаум А.А.* Основы теории оптимальных автоматических систем. М.: Наука, 1966. 623 с.
17. *Аоки М.* Оптимизация стохастических систем. М.: Наука, 1971. 424 с.
18. *Беллман Р.* Динамическое программирование. М.: Изд-во иностр. лит., 1960. 400 с.
19. *Ховард Р.А.* Динамическое программирование и марковские процессы. М.: Сов. радио, 1964.
20. *Себряков Г.Г., Красильщиков М.Н.* Управление и наведение беспилотных маневренных летательных аппаратов на основе современных информационных технологий. М.: Физматлит, 2003. 280 с.