

УДК 517.97

УПРАВЛЕНИЕ СЛУЧАЙНОЙ СКАЧКООБРАЗНОЙ СТРУКТУРОЙ ОБЪЕКТА В УСЛОВИЯХ ПРОТИВОДЕЙСТВИЯ

В.А. Бухалёв

АО «Московский научно-исследовательский телевизионный институт»,
Россия, 105094, Москва, ул. Гольяновская, 7а, стр. 1
E-mail: vadim.bukhalev@yandex.ru

В.А. Болдинов

Московский авиационный институт (национальный исследовательский университет),
Россия, 125993, Москва, Волоколамское шоссе, д. 4
E-mail: viktorboldiniv@mail.ru

А.А. Скрынников

ФГУП «Государственный НИИ авиационных систем»,
Россия, 125319, Москва, ул. Викторенко, 7
E-mail: a1260@mail.ru

Ключевые слова: система со случайной скачкообразной структурой, информационно-управляющий алгоритм, индикатор структуры, марковский процесс, байесовский алгоритм.

Аннотация: 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. Оптимальные управления находятся в классе детерминированных зависимостей от результатов наблюдений, предшествующих текущему моменту.

1. Введение

В настоящей статье рассматривается стохастическая система со случайной скачкообразной структурой (ССС), имеющей конечное число возможных состояний. Переходы из одного состояния в другое происходят в случайные моменты времени и управляются двумя противоборствующими сторонами, преследующими строго противоположные интересы. При этом каждый из противников располагает конечным числом возможных стратегий (управлений) и руководствуется некоторым своим априорным представлением об управляемом объекте и информацией, получаемой от своего индикатора структуры, регистрирующего с ошибками текущее состояние структуры объекта.

Ставится задача построения алгоритмов управления противников («игроков»), состоящая в нахождении оптимальных управлений в каждый момент времени k в классе

детерминированных зависимостей от показаний индикаторов структуры – на отрезке времени от начального момента до текущего k .

2. Постановка задачи

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

Информация, которой располагают игроки о вероятностях переходов из состояния s_k в состояние s_{k+1} , неодинакова:

$$(1) \quad q_k^A(s_{k+1}|s_k, \theta_k, \vartheta_k) \text{ и } q_k^B(s_{k+1}|s_k, \theta_k, \vartheta_k),$$

где θ_k, ϑ_k – управления игроков A и B , имеющие конечное число возможных стратегий: $\theta_k = \overline{1, n^\theta}, \vartheta_k = \overline{1, n^\vartheta}$.

Состояние структуры регистрируется с ошибками индикаторами, которые описываются условно-марковскими цепями с конечным числом возможных состояний $r_k = \overline{1, n^r}$ и $\rho_k = \overline{1, n^\rho}$. Условно-марковские цепи заданы условными вероятностями переходов из r_k в r_{k+1} и из ρ_k в ρ_{k+1} при фиксированных $s_{k+1}, \theta_k, \vartheta_k$:

$$(2) \quad \pi_{k+1}^A(r_{k+1}|r_k, s_{k+1}, \theta_k, \vartheta_k) \text{ и } \pi_{k+1}^B(\rho_{k+1}|\rho_k, s_{k+1}, \theta_k, \vartheta_k).$$

Зависимость $\pi_{k+1}^A(\cdot)$ от θ_k означает, что игрок A может управлять как структурой объекта, так и характеристикой индикатора структуры. Зависимость $\pi_{k+1}^A(\cdot)$ от ϑ_k означает, что игрок B может управлять не только структурой объекта s_k , но и осуществлять информационное противодействие игроку A . Аналогичный смысл имеет зависимость $\pi_{k+1}^B(\cdot)$ от ϑ_k и θ_k .

Так как интересы игроков строго противоположны, то показатели качества (эффективности) игры для обоих аналогичны:

$$(3) \quad J^A = \sum_{k=1}^n M[W_k(s_k, \theta_{k-1}, \vartheta_{k-1})|r_{\overline{0, k-1}}],$$

$$(4) \quad J^B = \sum_{k=1}^n M[W_k(s_k, \theta_{k-1}, \vartheta_{k-1})|\rho_{\overline{0, k-1}}],$$

где $W_k(\cdot)$ – текущая функция потерь; $M[\cdot]$ – символ математического ожидания, а критерии оптимальности управления разные:

$$(5) \quad J^{A*} = \min_{\theta_{\overline{0, n-1}}} \max_{\vartheta_{\overline{0, n-1}}} J^A,$$

$$(6) \quad J^{B*} = \max_{\vartheta_{\overline{0, n-1}}} \min_{\theta_{\overline{0, n-1}}} J^B,$$

$$\theta_{\overline{0, n-1}} = (\theta_0, \theta_1, \dots, \theta_{n-1}), \vartheta_{\overline{0, n-1}} = (\vartheta_0, \vartheta_1, \dots, \vartheta_{n-1}),$$

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

Априорные сведения о начальных значениях вероятностей состояний структуры, которыми располагают игроки, различны: $p_0^A(s_0)$ и $p_0^B(s_0)$.

Требуется найти: оптимальные управления $\theta_k^*, \vartheta_k^*$ в классе детерминированных зависимостей от наблюдений $r_{\overline{0, k}}, \rho_{\overline{0, k}}$.

3. Решение

3.1. Алгоритм игрока A

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

Найдем уравнения регулятора структуры (блока управления), связывающие оптимальное управление с вероятностью состояния структуры.

С учетом специфики поставленной задачи применим подход, разработанный Р. Беллманом и известный как метод динамического программирования [1, 2].

Обозначим

$$(7) \quad J_k^A \triangleq \sum_{i=k}^n M[W_i(s_i, \theta_{i-1}, \vartheta_{i-1}) | r_{\overline{0,k-1}}],$$

$$(8) \quad W_k^A \triangleq M[W_k(s_k, \theta_{k-1}, \vartheta_{k-1}) | r_{\overline{0,k-1}}],$$

где J_k^A – функционал оставшихся потерь на отрезке $[k, n]$; W_k^A – оценка текущей функции потерь W_k , прогнозируемая на один дискретный шаг вперед при известных наблюдениях на отрезке $[0, k-1]$; \triangleq – символ равенства по определению.

Из (7), (8) следует

$$(9) \quad \begin{aligned} J_k^A &= W_k^A + \sum_{i=k+1}^n M[W_i(s_i, \theta_{i-1}, \vartheta_{i-1}) | r_{\overline{0,k-1}}] = \\ &= W_k^A + \sum_{r_k} J_{k+1}^A(r_k) \sigma_k^A(r_k) = W_k^A + \tilde{J}_{k+1}^A, \end{aligned}$$

где $\sigma_k^A(r_k) \triangleq P[r_k | r_{\overline{0,k-1}}]$, $\tilde{J}_{k+1}^A \triangleq \sum_{r_k} J_{k+1}^A(r_k) \sigma_k^A(r_k)$, P – символ вероятности.

Из (9) следует

$$(10) \quad \begin{aligned} J_k^{A*} &= \min_{\theta_{\overline{k-1,n-1}}} \max_{\vartheta_{\overline{k-1,n-1}}} [W_k^A + \sum_{r_k} J_{k+1}^A(r_k) \sigma_k^A(r_k)] = \\ &= \min_{\theta_{k-1}} \max_{\vartheta_{k-1}} \min_{\theta_{\overline{k,n-1}}} \max_{\vartheta_{\overline{k,n-1}}} [W_k^A + \sum_{r_k} J_{k+1}^A(r_k) \sigma_k^A(r_k)]. \end{aligned}$$

Так как W_k^A от $\theta_{\overline{k,n-1}}$ не зависит, а $\min_{\theta_{\overline{k,n-1}}} \max_{\vartheta_{\overline{k,n-1}}} \tilde{J}_{k+1}^A = \tilde{J}_{k+1}^{A*}$, то из (10) следует рекуррентное уравнение для J_k^{A*} :

$$(11) \quad \begin{aligned} J_k^{A*} &= \min_{\theta_{k-1}} \max_{\vartheta_{k-1}} [W_k^A + \sum_{r_k} J_{k+1}^{A*}(r_k) \sigma_k^A(r_k)] = \min_{\theta_{k-1}} \max_{\vartheta_{k-1}} [W_k^A + \tilde{J}_{k+1}^{A*}], \\ &k = n, n-1, \dots, 1; \tilde{J}_{n+1}^{A*} \equiv 0, \end{aligned}$$

где

$$(12) \quad \tilde{J}_{k+1}^{A*} \triangleq \sum_{r_k} J_{k+1}^{A*}(r_k) \sigma_k^A(r_k).$$

Вероятность $\sigma_k^A(r_k)$ находится по формуле полной вероятности:

$$(13) \quad \sigma_k^A(r_k) = \sum_{s_k} \pi_k^A(r_k | r_{k-1}, s_k, \theta_{k-1}, \vartheta_{k-1}) \tilde{p}_k^A(s_k),$$

а W_k^A , как следует из (8), – по формуле

$$(14) \quad W_k^A = \sum_{s_k} W_k(s_k, \theta_{k-1}, \vartheta_{k-1}) \tilde{p}_k^A(s_k),$$

где $\tilde{p}_k^A(s_k) \triangleq P[s_k | r_{\overline{0,k-1}}]$ – вероятность состояния структуры s_k , прогнозируемая на один шаг дискретности вперед, и определяемая по формуле полной вероятности:

$$(15) \quad \tilde{p}_k^A(s_k) = \sum_{s_{k-1}} q_{k-1}^A(s_k | s_{k-1}, \theta_{k-1}, \vartheta_{k-1}) \hat{p}_{k-1}^A(s_{k-1}),$$

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

Пара минимаксных управлений согласно (11) определяется формулой

$$(16) \quad (\theta_{k-1}^*, \vartheta_{k-1}^*) = \arg \min_{\theta_{k-1}} \max_{\vartheta_{k-1}} [W_k^A(\hat{p}_{k-1}^A(s_{k-1}), \theta_{k-1}, \vartheta_{k-1}) + \tilde{J}_{k+1}^{A*}(\hat{p}_{k-1}^A(s_{k-1}), \theta_{k-1}, \vartheta_{k-1})],$$

где θ_{k-1}^* – оптимальное управление игрока A , а ϑ_{k-1}^* – предполагаемое игроком A оптимальное управление игрока B , основанные на показаниях индикатора структуры $r_{\overline{0,k-1}}$, принадлежащего игроку A .

Рекуррентные уравнения (11)-(16) описывают алгоритм регулятора структуры игрока A . Выходными сигналами регулятора являются управления θ_k^* , ϑ_k^* , входным сиг-

налом – апостериорная вероятность $\hat{p}_k^A(s_k)$, которая определяется алгоритмом классификатора структуры (в блоке обработки информации).

Апостериорная вероятность состояния структуры $\hat{p}_k^A(s_k)$, согласно формуле Байеса, обобщенной на класс систем ССС [3] и формуле полной вероятности, определяется рекуррентными уравнениями

$$(17) \quad \hat{p}_{k=1}^A(s_{k+1}) = \frac{\pi_{k+1}^A(r_{k+1}|r_k, s_{k+1}, \theta_k^*, \vartheta_k^*) \tilde{p}_{k+1}^A(s_{k+1})}{\sum_{s_{k+1}} \pi_{k+1}^A(r_{k+1}|r_k, s_{k+1}, \theta_k^*, \vartheta_k^*) \tilde{p}_{k+1}^A(s_{k+1})}$$

$$(18) \quad \tilde{p}_{k+1}^A(s_{k+1}) = \sum_{s_k} q_k^A(s_{k+1}|s_k, \theta_k^*, \vartheta_k^*) \tilde{p}_k^A(s_k),$$

$$k = 0, 1, \dots, n-1; \tilde{p}_0^A(s_0) = p_0^A(s_0).$$

В целом, оптимальный *минимаксный* информационно-управляющий алгоритм игрока A описывается замкнутой системой рекуррентных уравнений (11)-(18), в которой уравнения регулятора (11)-(16) решаются в «обратном времени» ($k = 0, 1, \dots, n-1$) при начальных условиях $\tilde{p}_0^A = p_0^A(s_0)$.

3.2. Алгоритм игрока B

Аналогичный информационно-управляющий *максиминный* алгоритм игрока B описывается уравнениями (11)-(18), в которых производятся следующие замены:

$$\min_{\theta_{k-1}} \max_{\vartheta_{k-1}} [\cdot]^A \rightarrow \max_{\vartheta_{k-1}} \min_{\theta_{k-1}} [\cdot]^B;$$

индекс « A » \rightarrow индекс « B »; $r_k \rightarrow \rho_k$; $\theta_k^* \rightarrow \theta_k^B$; $\vartheta_k^A \rightarrow \vartheta_k^*$; где ϑ_k^* и θ_k^B – максиминные управления: ϑ_k^* – оптимальное управление игрока B , а θ_k^B – предполагаемое игроком B оптимальное управление его противника A , основанные на показателях индикатора структуры $\rho_{0,k}$, принадлежащего игроку B .

Алгоритм имеет следующий вид:

$$(19) \quad J_k^{B*} = \max_{\vartheta_{k-1}} \min_{\theta_{k-1}} [W_k^B(\cdot) + \tilde{J}_{k+1}^{B*}(\cdot)],$$

$$(20) \quad \tilde{J}_{k+1}^{B*} = \sum_{\rho_k} J_{k+1}^{B*}(\rho_k) \sigma_k^B(\rho_k),$$

$$(21) \quad \sigma_k^B(\rho_k) = \sum_{s_k} \pi_k^B(\rho_k | \rho_{k-1}, s_k, \theta_{k-1}, \vartheta_{k-1}) \tilde{p}_k^B(s_k),$$

$$(22) \quad W_k^B = \sum_{s_k} W_k(s_k, \theta_{k-1}, \vartheta_{k-1}) \tilde{p}_k^B(s_k),$$

$$(23) \quad \tilde{p}_k^B(s_k) = \sum_{s_{k-1}} q_{k-1}^B(s_k | s_{k-1}, \theta_{k-1}, \vartheta_{k-1}) \hat{p}_{k-1}^B(s_{k-1}),$$

$$(24) \quad (\vartheta_{k-1}^*, \theta_{k-1}^B) = \arg \max_{\vartheta_{k-1}} \min_{\theta_{k-1}} [W_k^B(\cdot) + \tilde{J}_{k+1}^{B*}(\cdot)],$$

$$k = n, n-1, \dots, 1; \tilde{J}_{n+1}^{B*} \equiv 0,$$

где $W_k^B(\cdot) \triangleq W_k^B(\hat{p}_{k-1}^B(s_{k-1}), \theta_{k-1}, \vartheta_{k-1})$, $\tilde{J}_{k+1}^{B*}(\cdot) \triangleq \tilde{J}_{k+1}^{B*}(\hat{p}_{k-1}^B(s_{k-1}), \theta_{k-1}, \vartheta_{k-1})$.

$$(25) \quad \hat{p}_{k=1}^B(s_{k+1}) = \frac{\pi_{k+1}^B(\rho_{k+1} | \rho_k, s_{k+1}, \theta_k^B, \vartheta_k^*) \tilde{p}_{k+1}^B(s_{k+1})}{\sum_{s_{k+1}} \pi_{k+1}^B(\rho_{k+1} | \rho_k, s_{k+1}, \theta_k^B, \vartheta_k^*) \tilde{p}_{k+1}^B(s_{k+1})},$$

$$(26) \quad \tilde{p}_{k+1}^B(s_{k+1}) = \sum_{s_k} q_k^B(s_{k+1} | s_k, \theta_k^B, \vartheta_k^*) \tilde{p}_k^B(s_k),$$

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

4. Заключение

Рассмотренная задача представляет собой игру с неполной информацией и ненулевой суммой и не имеет седловой точки вследствие различной информативности игроков о результатах игры. Прежде всего это объясняется различными в общем случае показателями индикаторов структуры ($r_k \neq \rho_k$), откуда следует $\hat{p}_k^A(s_k) \neq \hat{p}_k^B(s_k)$, $\theta_k^* \neq \theta_k^B$, $\vartheta_k^* \neq \vartheta_k^A$. Поэтому седловая точка игры отсутствует даже в том случае, когда противники располагают одинаковой априорной информацией ($q^A(\cdot) = q^B(\cdot)$, $\tilde{p}_0^A(s_0) = \tilde{p}_0^B(s_0)$), а текущая функция потерь $W_k(\cdot)$ сепарабельна относительно управлений игроков q_k, ϑ_k .

Работа выполнена при поддержке Российского фонда фундаментальных исследований (19-08-00502а).

Список литературы

1. Бухалёв В.А. Распознавание, оценивание и управление в системах со случайной скачкообразной структурой. М.: Наука, 1996. 287 с.
2. Бухалёв В.А., Скрынников А.А., Болдинов В.А. Алгоритмическая помехозащита беспилотных летательных аппаратов. М.: ФИЗМАТЛИТ, 2018, 192 с.
3. Болдинов В.А., Бухалёв В.А., Прядкин С.П., Скрынников А.А. Управление вероятностным распределением состоянием системы по сигналам индикаторов ее структуры // Изв. РАН. Теория и системы управления. 2016, № 3. С. 3-10.