Pb207.fm Page 46 Monday, March 12, 2007 3:43 PM

УДК 623.4.083

РАНГОВЫЕ АДАПТИВНЫЕ МЕХАНИЗМЫ

В. В. Цыганов[(1)], М. В. Аржаков[(2)], Р. А. Багамаев[(3)]

[(1)] *Институт проблем управления им. В. А. Трапезникова РАН, г. Москва;*
[(2)] *ЗАО НПК «Атомтехнопром», г. Москва;*
[(3)] *Коммерческий банк «Витас», г. Москва*

Рассмотрены задачи синтеза прогрессивных адаптивных механизмов функционирования двухуровневой активной системы, Центр которой настраивает параметр решающего правила классификации. По ее результатам производится ранжирование и стимулирование дальновидного элемента системы. Найдены достаточные условия прогрессивности ранговых адаптивных механизмов для дальновидного и недальновидного Центра.

ВВЕДЕНИЕ

Традиционно, основным объектом исследований в теории активных систем является двухуровневая система, на верхнем уровне которой находится Центр, а на нижнем — дальновидный элемент (ДЭ). Многие задачи управления такими системами сводятся к классификации, по результатам которой этому элементу присваивается ранг, определяющий его стимулирование. Такие механизмы были названы ранговыми [1]. В ранговом адаптивном механизме (РАМ) информация о состояниях ДЭ, получаемая в процессе управления, используется Центром для настройки параметров процедур классификации и стимулирования. Необходимость в применении РАМ возникает, когда Центр должен работать в условиях неопределенности, а имеющаяся априорная информация настолько мала, что невозможно заранее спроектировать эффективный детерминированный ранговый механизм. Теоретическое направление, связанное с построением РАМ, нашло свое отражение в монографии [2]. В РАМ часто используются эвристические процедуры настройки параметра решающего правила, подобные применяемому на практике «планированию от достигнутого». Дальновидный элемент, зная процедуры РАМ, может предсказывать будущее управление Центра, в зависимости от собственного выбора сегодня. Он выбирает состояние, при котором его целевая функция максимальна, даже если это противоречит интересам Центра. Например, ДЭ может занижать показатели своей эффективности по сравнению с потенциально возможными. Поэтому представляет интерес решение задачи синтеза прогрессивного РАМ, обеспечивающего максимальное раскрытие потенциала ДЭ.

1. ДАЛЬНОВИДНЫЙ ОБУЧАЮЩИЙСЯ ЦЕНТР

Предположим, что Центр дальновиден и обучается, настраивая параметр решающего правила при классификации так, чтобы минимизировать средние потери. Такая настройка основана на наблюдениях состояния управляемого объекта и процедурах обучения [1—3]. Как и в работе [3], обозначим через ξ случайную величину, характеризующую состояние объекта, $\xi \in \Delta \subset R^1$. Рассмотрим задачу обучения дихотомической классификации ситуаций путем отнесения их к одной из двух областей, составляющих множество Δ. Обозначим Δ_1, Δ_2 некоторое разбиение множества Δ на две области, $\bigcup\limits_{k=1}^{2} \Delta_k = \Delta$. Задача состоит в определении разбиения, минимизирующего средний риск, связанный с классификацией. Предположим, что $q(\xi)$ — известная стационарная плотность распределения случайной величины ξ. Введем для каждой, пока неизвестной области Δ_k, $k = \overline{1, 2}$ функции потерь $F_k(c, \xi)$, где c — неизвестный параметр. Минимизируется средний риск, оценивающий качество классификации

$$J(c) = \sum_{k=1}^{2} \int_{\Delta k} F_k(c, \xi)\, q(\xi)\, d\xi \underset{c}{\mapsto} \min. \qquad (1)$$

Принадлежность ситуации ξ множествам Δ_1 или Δ_2 определяется знаком решающего правила $\mu_{12}(c, \xi) = F_1(c, \xi) - F_2(c, \xi)$:

Black process 45.0• 175.0 LPI

Pb207.fm Page 47 Monday, March 12, 2007 3:43 PM

$$\xi \in \Delta_1, \text{ если } \mu_{12}(c, \xi) < 0 \text{ и } \xi \in \Delta_2, \text{ если } \mu_{12}(c, \xi) \geqslant 0. \quad (2)$$

Положим $F_1(c, \xi) = \xi - vc$, $F_2(c, \xi) = d(c - \xi)$, где v, d — параметры функции потерь, $0 < v < 1$, $d > 0$, $vc \leqslant \xi \leqslant c$. Подставляя эти выражения в решающее правило (2), получаем его в виде

$$\xi \in \Delta_1, \text{ если } \xi < \frac{d+v}{d+1}c \text{ и } \xi \in \Delta_2, \text{ если } \xi \geqslant \frac{d+v}{d+1}c. \quad (3)$$

где параметр решающего правила c определяется решением задачи (1).

Предположим теперь, что плотность распределения $q(\xi)$ неизвестна, и непосредственное определение параметра c, как решения задачи оптимизации (1), невозможно. Возникает необходимость в настройке параметра решающего правила по наблюдениям ξ_t, где t — номер периода, $t = 1, 2...$, для минимизации среднего риска (1). Применяя метод стохастической аппроксимации для решения задачи (1), с учетом (3) можно показать, что процедура настройки оценки c_t параметра решающего правила имеет вид:

$$c_{t+1} = I^k(c_t, \xi_t) = \begin{cases} c_t + \gamma_t v \text{ при } \xi_t < \dfrac{d+v}{d+1}c_t, \\ c_t - \gamma_t d \text{ при } \xi_t \geqslant \dfrac{d+v}{d+1}c_t, \end{cases} \quad (4)$$

где γ_t — коэффициент усиления [2].

Будем предполагать, что ДЭ может выбирать свое состояние y_t, причем $y_t \leqslant \xi_t$, $t = 1, 2...$ Величину ξ_t назовем случайным потенциалом ДЭ. Дальновидный обучающийся Центр, наблюдая состояние y_t и используя процедуру обучения (4), формирует собственную оценку a_t параметра решающего правила, чтобы минимизировать средний риск. Классификация проводится путем сопоставления оценки a_t и состояния y_t. Затем определяется стимул ДЭ $\varphi_t = f(a_t, y_t)$. Тем самым дальновидный Центр формирует ранговый обучающийся механизм (РОМ) $\Sigma^k = (I^k, f)$, где I^k — процедура обучения (4), f — процедура стимулирования. Предполагается, что цель Центра состоит в синтезе прогрессивного механизма, обеспечивающего раскрытие потенциала ДЭ: $y_t = \xi_t$, $t = 1, 2...$

2. ДАЛЬНОВИДНЫЙ ЭЛЕМЕНТ

Целевая функция ДЭ в периоде t имеет вид

$$V_t = \sum_{\tau = t}^{t+T} \rho^{\tau - t}\varphi^{\tau}, \quad (5)$$

где ρ — коэффициент дисконтирования, используемый для приведения будущих стимулов к текущему моменту времени t, $0 < \rho \leqslant 1$; T — дальновидность ДЭ, исчисляемая в периодах времени. Поскольку цель ДЭ заключается в максимизации критерия (5), необходим прогноз потенциалов и состояний в будущем. Поскольку состояние y_τ (при заданном потенциале ξ_τ) зависит от самого ДЭ, в качестве прогнозных рассматриваются состояния, максимизирующие критерий (5). Введем оператор максимизации на множестве возможных состояний ДЭ в периоде τ: $M_\tau = \max\limits_{y_\tau \leqslant \xi_\tau}$. Введем также оператор E_τ устранения неопределенности относительно потенциала ДЭ в периоде τ, $\tau = \overline{t+1, t+T}$. Применение оператора E_τ к любой функции $g(\xi_\tau)$, непрерывной при $\xi_\tau \in \Delta$, определяет некоторое ее значение $g(\xi^0)$, $\xi^0 \in \Delta$: $E_\tau g(\xi_\tau) = g(\xi^0)$.

При выборе состояния y_t в периоде t, при заданном потенциале ξ_t, ДЭ должен решить задачу оптимизации критерия (5) с прогнозом потенциалов и состояний на периоды $t + 1, ..., t + T$ (кратко — задачу ОППС). Иными словами, ДЭ необходимо определить оптимальную позиционную стратегию в виде набора оптимальных состояний $(\tilde{y}_t, ..., \tilde{y}_{t+T})$ как функций его потенциала в текущем и будущих периодах: $y_\tau^* = \tilde{y}_\tau(\xi_\tau)$, $\tau = \overline{t, t+T}$. После того, как ДЭ становится известно значение потенциала ξ_t, он выбирает, в качестве оптимального, состояние $y_t^* = \tilde{y}_t(\xi_t)$. Рассмотрим процедуру выбора ДЭ оптимальной позиционной стратегии $(\tilde{y}_t, ..., \tilde{y}_{t+T})$, используя метод динамического программирования. Начнем с периода $t + T$. Потенциал ДЭ ξ_{t+T} и оценку a_{t+T}, зависящую от y_τ, $\tau = \overline{t, t+T-1}$, считаем заданными. С помощью оператора M_{t+T} проведем оптимизацию целевой функции (5) по y_{t+T}. Тем самым определяется состояние $y_{t+T}^* = \tilde{y}_{t+T}(\xi_{t+T})$. В периоде $t + T - 1$ потенциал ξ_{t+T} неизвестен. Поэтому, перед оптимизацией целевой функции (5) по y_{t+T-1}, необходимо устранить неопределенность в отношении потенциала ξ_{t+T}. Для этого к целевой функции (5), в которой положено $y_{t+T}^* = \tilde{y}_{t+T}(\xi_{t+T})$, применяется оператор E_{t+T} устранения неопределенности в отношении потенциала ξ_{t+T}. В результате получаем «однократно усеченную» целевую функцию, которая отличается от функции (5) тем, что в ней устранена неопределенность в отношении потенциала и состояния ДЭ в периоде $t + T$.

Для устранения неопределенности в отношении потенциала и состояния ДЭ в периоде $t + T - 1$, проведем оптимизацию «усеченной» целевой функции по состоянию y_{t+T-1} с помощью оператора M_{t+T-1}. Тем самым, определяется состояние $y_{t+T-1}^* = \tilde{y}_{t+T-1}(\xi_{t+T-1})$. После этого к «усеченной» целевой функции ДЭ, в которой положено $y_{t+T-1}^* = \tilde{y}_{t+T-1}(\xi_{t+T-1})$, применяется оператор E_{t+T-1} устранения неопределенности в отношении потенциала ξ_{t+T-1}. В результате получаем «двукратно усеченную» целевую функцию, в которой, по сравнению с функцией (5), устранена неопределенность в отношении его потенциалов и состояний в периодах

Black process 45.0• 175.0 LPI

Pb207.fm Page 48 Monday, March 12, 2007 3:43 PM

$t + T$ и $t + T - 1$. Повторяя эту процедуру вплоть до периода $t + 1$ включительно, получаем «T-кратно усеченную» целевую функцию $\hat{V}_t(a_t, y_t)$. Она отличается от целевой функции (5) тем, что в ней устранена неопределенность в отношении потенциалов и состояний ДЭ в периодах $t + 1, ..., t + T$. Формально функция $\hat{V}_t(a_t, y_t)$ определяется путем последовательного применения к критерию (5) операторов $M_{t+T}, E_{t+T}, ..., M_{t+1}, E_{t+1}$, устраняющих неопределенность в отношении будущих состояний и потенциалов ДЭ в периодах $t + T, ..., t + 1$. Полагая $M_\nu^\mu = E_\nu M_\nu ... E_\mu M_\mu$, $E_\nu^\mu = E_\nu ... E_\mu$, имеем

$$\hat{V}_t(a_t, y_t) = M_{t+1}^{t+T} V_t = \varphi_t + \sum_{\tau = t+1}^{t+T} \rho^{\tau+t} M_{t+1}^{t+T} \varphi_\tau, \ \varphi_\tau =$$
$$= f(a_\tau, y_\tau), \ a_{\tau+1} = I^k(a_\tau, y_\tau), \ \tau = \overline{t, t+T}. \qquad (6)$$

Дальновидный элемент решает задачу ОППС путем выбора состояния y_t, максимизирующего ожидаемое значение критерия $\hat{V}_t(a_t, y_t)$. При выборе состояния y_t ДЭ известен потенциал ξ_t. Множество решений задачи ОППС в периоде t как множество состояний y_t^*, при которых достигается максимальное значение ожидаемого критерия (6), имеет вид:

$$R_t(\Sigma, \xi_t) = \operatorname*{Arg\,max}_{y_t \in Y(\xi_t)} \hat{V}_t(a_t, y_t).$$

Далее предполагается, что справедлива гипотеза благожелательности ДЭ по отношению к Центру: при одинаковых значениях целевой функции ДЭ выбирает состояние, наиболее благоприятное для Центра. Тогда, если $\xi_t \in R_t(\Sigma, \xi_t)$, то $y_t^* = \xi_t$.

3. ПРОГРЕССИВНЫЙ МЕХАНИЗМ ПРИ ДАЛЬНОВИДНОМ ЦЕНТРЕ

Рассмотрим РОМ $\Sigma^k = (I^k, f)$, в котором процедура обучения (4) используется для получения текущих оценок параметра решающего правила:

$$a_{t+1} = I^k(a_t, y_t). \qquad (7)$$

Будем говорить, что РОМ $\Sigma^k = (I^k, f)$ прогрессивен, если $y_t^* = \xi_t$, $t = 1, 2...$

Теорема 1. *Для прогрессивности РОМ* $\Sigma^k = (I^k, f)$ *с процедурой настройки параметра решающего правила* (8) *достаточно, чтобы*

$$f(a_t, y_t) = \Theta(y_t - a_t(d + v)/(d + 1)) =$$
$$= \begin{cases} 1 \text{ при } y_t \geq a_t(d + v)/(d + 1), \\ 0 \text{ при } y_t < a_t(d + v)/(d + 1). \end{cases} \qquad (8)$$

Доказательство этой и следующей теорем дано в Приложении.

Заметим, что, условия (7) и (8) аналогичны условиям прогрессивности самообучающегося механизма, приведенным в теореме 1 работы [3]. Однако достаточность условий (7) и (8) для прогрессивности РОМ $\Sigma^k = (I^k, f)$ доказана для случая ДЭ, решающего задачу оптимизации критерия (5) с прогнозом потенциалов и состояний на периоды $t + 1, ..., t + T$ (задачу ОППС). Это предположение приводит к необходимости разработки особого метода доказательства, основанного на динамическом программировании и математической индукции. Таким образом, адаптивный механизм, удовлетворяющий условиям теоремы 1 работы [3], оказывается прогрессивным и при гипотезе ОППС с оператором E_τ устранения неопределенности относительно потенциала ДЭ, что существенно расширяет сферу его применимости в организационных системах.

4. НЕДАЛЬНОВИДНЫЙ ЦЕНТР

До сих пор предполагалась дальновидность Центра, понимаемая как его заинтересованность в обучении, с целью минимизации среднего риска при классификации. В РОМ будущая оценка параметра решающего правила понижается при росте состояния y_t^* ДЭ. Предположим теперь, что Центр использует эвристическую процедуру настройки, повышая оценку параметра решающего правила при увеличении состояния y_t^*. Такое «планирование от достигнутого» не способствует заинтересованности ДЭ в раскрытии своего потенциала и связано с «проклятием координации» [2]. Кроме того, такая процедура настройки параметра решающего правила не обеспечивает обучение Центра и минимизацию среднего риска при классификации. Назовем использующий ее Центр недальновидным. Задача состоит в том, чтобы найти условия прогрессивности РАМ при недальновидном Центре, использующем при настройке параметра решающего правила e_t процедуру «планирования от достигнутого»:

$$e_{t+1} = E(e_t, y_t) = \begin{cases} e_t - \beta \text{ при } y_t < e_t, \\ e_t + \beta d \text{ при } y_t \geq e_t, \end{cases} \qquad (9)$$

где β, d — неотрицательные величины. Согласно процедуре (9), если фактический выход y_t больше параметра e_t, то она увеличивается (и наоборот). Если $y_t < e_t$, то ДЭ относится к классу 1, в противном случае — к классу 2. По результатам классификации осуществляется ранжирование и стимулирование. Ранговый адаптивный механизм с процедурой настройки (9) будем обозначать $\Sigma^e = (E, f)$.

Рассмотрим вероятностный подход к построению критерия (6), основанный на усреднении целевой функции (5) по возможным значениям потенциала на период дальновидности ДЭ. В качестве оператора устранения

Black process 45.0• 175.0 LPI

неопределенности в отношении потенциала будем пользоваться оператором математического ожидания: $E_\tau = \int_\Delta q(\xi_\tau)d\xi_\tau$, где $q(\xi_\tau)$ — плотность распределения случайной величины ξ_τ. Предполагается, что

$$q(\xi_\tau) \leqslant q^*,\ \xi_\tau \in \Delta,\ \tau = \overline{t, t+T}. \qquad (10)$$

Будем говорить, что РАМ $\Sigma^e = (E, f)$ прогрессивен, если $y_t^* = \xi_\tau$, $t = 1, 2...$

Теорема 2. *Для прогрессивности РАМ* $\Sigma^e = (E, f)$ *с процедурой настройки* (9) *достаточно, чтобы процедура стимулирования была ранговой*

$$f(e, y) = \Theta(y - e) = \begin{cases} 1 \text{ при } y \geqslant e, \\ 0 \text{ при } y < e, \end{cases} \qquad (11)$$

и выполнялось неравенство

$$\beta q^*(1 + d)T \leqslant 1. \qquad (12)$$

Условие (12) есть условие прогрессивности РАМ при недальновидном Центре. Оно определяет ограничения на характеристики случайного процесса, дальновидность элемента и процедуру настройки оценки параметра решающего правила, при которых ДЭ не занижает свои показатели, несмотря на недальновидность Центра. Если потенциал ξ имеет гауссово распределение ($q(\xi) = N(0, \sigma^2)$), то $q^* = 1/\sqrt{2\pi}\sigma$ и условие прогрессивности (12) имеет вид $\beta \leqslant \sqrt{2\pi}\sigma/T(1 + d)$. В случае, если потенциал распределен с одинаковой плотностью на отрезке длиной l, условие (12) имеет вид $\beta \leqslant l^2/T(1 + d)$. Для стохастического потенциала с плотностью распределения Лапласа ($q(\xi) = \exp[-|\xi|/r]/2r$) имеем $q^* = 1/2r$, и условие (12) имеет вид $\beta(1 + d)T \leqslant 2r$.

ЗАКЛЮЧЕНИЕ

Поставлены и решены задачи синтеза прогрессивных ранговых адаптивных механизмов функционирования двухуровневой активной системы, в которых Центр осуществляет настройку параметра решающего правила классификации, используя алгоритмы обучения, а также ранжирование и стимулирование. Найдены достаточные условия прогрессивности для дальновидного и недальновидного Центра. Перспективы дальнейших исследований ранговых адаптивных механизмов связаны с поиском необходимых условий их прогрессивности, а также поиском более общих постановок задач и методов их решения, в том числе на основе новых адаптивных алгоритмов.

ПРИЛОЖЕНИЕ

Доказательство теоремы 1 проводится с помощью методов динамического программирования и математической индукции. Для выбора ДЭ оптимального состояния (действия) y_t^* в периоде t необходимо решить задачу ОППС, определив оптимальную позиционную стратегию $(\tilde y_t, ..., \tilde y_{t+T})$ как функцию потенциала ДЭ в текущем и будущих периодах: $y_\tau^* = \tilde y_\tau(\xi_\tau)$, $\tau = \overline{t, t+T}$. После того, как ДЭ становится известно значение потенциала ξ_t, он выбирает действие $y_t^* = \tilde y_t(\xi_t)$. Определим оптимальную позиционную стратегию ДЭ $(\tilde y_t, ..., \tilde y_{t+T})$. Целевую функцию ДЭ (5), определяемую с точностью до членов, не содержащих выбираемую в периоде τ переменную y_τ, будем для краткости называть полезностью ДЭ в периоде τ.

Определим вначале состояние $y_{t+T}^* = \tilde y_{t+T}(\xi_{t+T})$ в периоде $t + T$, когда потенциал ДЭ ξ_{t+T} и оценка a_{t+T} известны. Полезность ДЭ в периоде $t + T$ определяется стимулом (8): $f(a_{t+T}, y_{t+T}) = \Theta(y_{t+T} - a_{t+T})$. Ее оптимизация по y_{t+T} проводится с помощью оператора M_{t+T}. Согласно условиям (8), $\Theta(y_{t+T} - a_{t+T})$ — неубывающая функция y_{t+T} при любых ξ_{t+T}, $y_{t+T} \leqslant \xi_{t+T}$. Следовательно, множество $R_{t+T}(\Sigma, \xi_{t+T}) = \underset{y_{t+T} \in Y(\xi_{t+T})}{\mathrm{Arg\,max}} f(a_{t+T}, y_{t+T})$ действий ДЭ y_{t+T}^* в периоде $t + T$, максимизирующих полезность ДЭ в периоде $t + T$, включает в себя потенциал ξ_{t+T}. Но тогда, в соответствии с гипотезой благожелательности ДЭ по отношению к Центру, $y_{t+T}^* = \xi_{t+T}$. Значение стимула ДЭ (значение полезности) в периоде $t + T$

$$v_{t+T}^* = f(a_{t+T},\ y_{t+T}^*) = \Theta(y_{t+T}^* - a_{t+T}) = \\ = \Theta(\xi_{t+\mathrm{T}} - a_{t+T}) \qquad (\mathrm{T}1.1)$$

зависит от a_{t+T} и, следовательно, от состояния ДЭ y_{t+T-1} в периоде $t + T - 1$. Определим теперь действие ДЭ y_{t+T-1}^*. Проведем оптимизацию полезности ДЭ в периоде $t + T — 1$ по состоянию y_{t+T-1} с помощью оператора M_{t+T-1}. Для этого необходимо знать зависимость полезности ДЭ от состояния y_{t+T-1}. Согласно функции (5), эта полезность включает в себя стимул ДЭ в периоде $t + T — 1$, определяемый согласно условиям (8), и ожидаемый стимул ДЭ (Т1.1) в периоде $t + T$, зависящий от случайного потенциала ξ_{t+T}, значение которого в периоде $t + T — 1$ неизвестно. Поэтому в периоде $t + T - 1$ неизвестна и зависимость стимула ДЭ v_{t+T}^* (Т1.1) от состояния y_{t+T-1}.

Для определения ожидаемого значения стимула (Т1.1) в периоде $t + T$ необходимо устранить неопределенность в отношении ξ_{t+T} с помощью оператора E_{t+T}. В результате, ожидаемое значение стимула в периоде $t + T$

$$E_{t+T}v_{t+T}^* = E_{t+T}\Theta(\xi_{t+\mathrm{T}} - a_{t+T}). \qquad (\mathrm{T}1.2)$$

Полезность ДЭ в периоде $t + T - 1$, как функция y_{t+T-1}, согласно целевой функции (5) имеет вид суммы стимула в текущем периоде (см. выражение (8)) и дисконтированного ожидаемого стимула (Т1.2) — в будущем:

$$v_{t+T-1} = \Theta(y_{t+T-1} - a_{t+T-1}) + \\ + \rho E_{t+T}\Theta(\xi_{t+\mathrm{T}} - a_{t+T}). \qquad (\mathrm{T}1.3)$$

Для определения множества $R_{t+T-1}(\Sigma, \xi_{t+T-1})$ оптимальных действий ДЭ в периоде $t + T - 1$ при потенциале ξ_{t+T-1} проведем оптимизацию полезности (Т1.3) по y_{t+T-1} с помощью оператора M_{t+T-1}. По условиям (8), с ростом показателя

y_{t+T-1}, стимул $\Theta(y_{t+T-1} - a_{t+T-1})$ не убывает. Далее, Центр использует процедуру обучения I^k (7), при которой оценка a_{t+T} не возрастает с увеличением показателя y_{t+T-1}. Следовательно, согласно условиям (8), дисконтированный ожидаемый стимул (Т1.2) не убывает с ростом показателя y_{t+T-1}, полезность (Т1.3) — неубывающая функция y_{t+T-1}, множество $R_{t+T-1}(\Sigma, \xi_{t+T-1})$ оптимальных действий ДЭ y^*_{t+T-1} в периоде $t+T-1$, максимизирующих полезность ДЭ в периоде $t+T-1$, включает в себя потенциал ξ_{t+T-1}. Согласно гипотезе благожелательности ДЭ по отношению к Центру, если $\xi_{t+T-1} \in R_{t+T-1}(\Sigma, \xi_{t+T-1})$, то $y^*_{t+T-1} = \xi_{t+T-1}$. Тогда полезность ДЭ в периоде $t+T-1$, согласно выражению (Т1.3), имеет вид: $v^*_{t+T-1} = \Theta(\xi_{t+T-1} - a_{t+T-1}) + \rho E_{t+T}\Theta(\xi_{t+T} - a^*_{t+T})$, $a^*_{t+T} = I^k(a_{t+T-1}\xi_{t+T-1})$.

Далее доказательство проводится по индукции. Предположим, что для некоторого s, $t+1 \leqslant s \leqslant t+T-1$, выполняется условие $y^*_\tau = \xi_\tau$, $s \leqslant \tau \leqslant t+T-1$, и полезность ДЭ в периоде s

$$v^*_s = \Theta(\xi_s - a_s) + \sum_{\tau=s+1}^{t+T} \rho^{\tau-s} E_{s+1} \dots E_\tau \Theta(\xi_\tau - a^*_\tau),$$

$$a^*_\tau = I^k(a_{\tau-1}, \xi_{\tau-1}),\ \tau = \overline{s+1, t+T}. \qquad (\text{Т1.4})$$

Докажем, что $y^*_{s-1} = \xi_{s-1}$. Для устранения неопределенности в отношении состояния y_{s-1} проведем оптимизацию полезности ДЭ в периоде $s-1$ с помощью оператора M_{s-1}. Эта полезность включает в себя стимул ДЭ в периоде $s-1$, определяемый согласно условиям (9), а также полезность ДЭ в периоде s (Т1.4), которая зависит от состояния ДЭ y_{s-1} в периоде $s-1$. Но полезность (Т1.4) зависит от случайного потенциала ξ_s, значение которого в периоде $s-1$ неизвестно. Поэтому в периоде $s-1$ неизвестна и зависимость v^*_s от y_{s-1}. Устраняя неопределенность в отношении ξ_s с помощью оператора математического ожидания E_s, получаем ожидаемое значение полезности ДЭ в периоде s:

$$E_s v^*_s = \sum_{\tau=s}^{t+N} \rho^{\tau-s} E_s \dots E_\tau \Theta(\xi_\tau - a^*_\tau). \qquad (\text{Т1.5})$$

Полезность ДЭ в периоде $s-1$ как функция y_{s-1}, согласно целевой функции (5), имеет вид суммы текущего стимула (см. выражение (9)) и дисконтированного ожидаемого значения полезности ДЭ в будущем (Т1.5):

$$v_{s-1} = \Theta(y_{s-1} - a_{s-1}) + \\ + \sum_{\tau=s}^{t+T} \rho^{\tau-s+1} E_s \dots E_\tau \Theta(\xi_\tau - a^*_\tau). \qquad (\text{Т1.6})$$

Для определения множества $R_{s-1}(\Sigma, \xi_{s-1})$ оптимальных действий ДЭ в периоде $s-1$ при потенциале ξ_{s-1} проведем оптимизацию полезности (Т1.6) по y_{s-1} с помощью оператора M_{s-1}. По условиям (8), с ростом показателя y_{s-1}, стимул $\Theta(y_{s-1} - a_{s-1})$ не убывает. Далее, Центр использует процедуру обучения I^k (7), при которой оценка a_s не возрастает с увеличением показателя y_{s-1}. Следовательно, согласно условиям (8), дисконтированный ожидаемый стимул (Т1.2) не убывает с ростом показателя y_{s-1}. Таким образом, полезность ДЭ (Т1.6) — неубывающая функция y_{s-1}. Но, согласно гипотезе благожелательности ДЭ по отношению к Центру, если $\xi_{s-1} \in R_{s-1}(\Sigma, \xi_{s-1})$, то $y^*_{s-1} = \xi_{s-1}$. Поскольку $s \geqslant t+1$, то, полагая $s = t+1$, получаем: $y^*_s = \xi_t$. Таким образом, РОМ $\Sigma^k = (I^k, f)$ является прогрессивным.

Доказательство теоремы 2 проводится методами динамического программирования и математической индукции, подобно доказательству теоремы 1. Определим вначале состояние $y^*_{t+T} = \tilde{y}_{t+T}(\xi_{t+T})$ в периоде $t+T$, когда ДЭ известны потенциал ξ_{t+T} и оценка e_{t+T}. Полезность ДЭ в периоде $t+T$ определяется стимулом (11): $f(e_{t+T}, y_{t+T}) = \Theta(y_{t+T} - e_{t+T})$. Ее оптимизация по y_{t+T} проводится с помощью оператора M_{t+T}. Согласно выражению (11), $\Theta(y_{t+T} - e_{t+T})$ — неубывающая функция y_{t+T} при любых ξ_{t+T}, $y_{t+T} \leqslant \xi_{t+T}$. Но тогда, в соответствии с гипотезой благожелательности ДЭ по отношению к Центру, $y^*_{t+T} = \xi_{t+T}$. Значение стимула ДЭ (значение полезности) в периоде $t+T$

$$v^*_{t+T} = f(e_{t+T}, y^*_{t+T}) = \Theta(y^*_{t+T} - e_{t+T}) = \\ = \Theta(\xi_{t+T} - e_{t+T}) \qquad (\text{Т2.1})$$

зависит от оценки e_{t+T} и, следовательно, от состояния ДЭ y_{t+T-1} в периоде $t+T-1$.

Определим теперь состояние y^*_{t+T-1}, оптимизируя полезность ДЭ в периоде $t+T-1$ по состоянию y_{t+T-1} с помощью оператора M_{t+T-1}. Согласно функции (5), эта полезность включает в себя стимул ДЭ в периоде $t+T-1$, определяемый согласно выражению (11), и ожидаемый стимул (Т2.1) в периоде $t+T$, зависящий от случайного потенциала ξ_{t+T}, значение которого в периоде $t+T-1$ неизвестно. Поэтому в периоде $t+T-1$ неизвестна и зависимость стимула (Т2.1) от y_{t+T-1}. Чтобы определить ожидаемое значение стимула (Т2.1) в периоде $t+T$, необходимо устранить неопределенность в отношении потенциала ξ_{t+T} с помощью оператора математического ожидания E_{t+T}. В результате получаем математическое ожидание полезности ДЭ $\langle v^*_{t+T} \rangle$ в периоде $t+T$:

$$\langle v^*_{t+T} \rangle = E_{t+T} v^*_{t+T} = \int_\Delta \Theta(\xi_{t+T} - e_{t+T})\, q(\xi_{t+T}) d\xi_{t+T}. \qquad (\text{Т2.2})$$

Полезность ДЭ в периоде $t+T-1$, как функция y_{t+T-1}, согласно целевой функции (5) имеет вид суммы стимула в текущем периоде (11) и дисконтированной будущей полезности (Т2.2):

$$v_{t+T-1} = \Theta(y_{t+T-1} - e_{t+T-1}) + \\ + \rho \int_\Delta \Theta(\xi_{t+T} - e_{t+T}) q(\xi) d\xi. \qquad (\text{Т2.3})$$

Для определения действия ДЭ y^*_{t+T-1} в периоде $t+T-1$ проведем оптимизацию полезности (Т2.3) по y_{t+T-1} с помощью оператора M_{t+T-1}. Нетрудно показать, что $\frac{\partial v_{t+T-1}(e_{t+T-1}, y_{t+T-1})}{\partial y_{t+T-1}} = k_{t+T-1}\delta(y_{t+T-1}, e_{t+T-1})$, где $\delta(y - e) = \begin{cases} 0, \text{ если } y \ne e, \\ \infty, \text{ если } y = e, \end{cases}$ причем $\int_0^y \delta(z-e) q(z) dz = q(e)\Theta(y-e)$, и k_{t+T-1} удовлетворяет неравенству

$$k_{t+T-1} \geqslant 1 - \beta q^* \rho(1+d) \geqslant 1 - \beta q^*(1+d) \geqslant \\ \geqslant 1 - \beta q^*(1+d)T \geqslant 0.$$

Первое неравенство имеет место согласно предположению (10), второе — в силу неравенств $0 \leqslant \rho \leqslant 1$, третье — в силу неравенства $T \geqslant 1$ и четвертое — по условию (12) теоремы. Следовательно, $v_{t+T-1}(e_{t+T-1}, y_{t+T-1}) = k_{t+T-1}\Theta(y_{t+T-1}, e_{t+T-1})$ — неубывающая функция y_{t+T-1} при любых $y_{t+T-1} \leqslant \xi_{t+T-1}$. Но тогда, в соответствии с гипотезой благожелательности ДЭ по отношению к Центру, $y^*_{t+T-1} = \xi_{t+T-1}$. Полезность ДЭ в периоде $t + T - 1$, согласно формуле (Т2.3), имеет вид:

$$v^*_{t+T-1} = \Theta(\xi_{t+T-1} - e_{t+T-1}) + \rho\int_\Delta \Theta(\xi_{t+T} - e_{t+T})q(\xi_{t+T})d\xi_{t+T}.$$

Далее доказательство проводится по индукции. Предположим, что для некоторого s, $t + 1 \leqslant s \leqslant t + T - 1$, выполняется условие $y^*_\tau = \xi_\tau$, $s \leqslant \tau \leqslant t + T - 1$ и полезность ДЭ в периоде s

$$v^*_s = \Theta(\xi_s - e_s) + \int_{\tau=s+1}^{t+T} \rho^{\tau-s}\int_\Delta d\xi_{s+1}q(\xi_{s+1})... \\ ... \int_\Delta d\xi_\tau q(\xi_\tau)\Theta(\xi_\tau - e_\tau). \quad (\text{Т2.4})$$

Докажем, что $y^*_{s-1} = \xi_{s-1}$. Для устранения неопределенности в отношении состояния y_{s-1} проведем оптимизацию полезности ДЭ в периоде $s - 1$ с помощью оператора M_{s-1}. Эта полезность включает в себя стимул ДЭ в периоде $s - 1$, определяемый согласно выражению (11), а также полезность (Т2.4), которая зависит от состояния y_{s-1} ДЭ в периоде $s - 1$. Но полезность (Т2.4) зависит от случайного потенциала ξ_s, значение которого в периоде $s - 1$ неизвестно. Устраняя неопределенность в отношении ξ_s с помощью оператора математического ожидания E_s, получаем ожидаемое значение полезности ДЭ в периоде s:

$$\langle v^*_s\rangle = E_s v^*_s = \sum_{\tau=s}^{t+T}\rho^{\tau-s}\int_\Delta d\xi_s q(\xi_s)... \int_\Delta d\xi_\tau q(\xi_\tau)\Theta(\xi_\tau - e_\tau). \quad (\text{Т2.5})$$

Полезность ДЭ в периоде s — 1, как функция y_{s-1}, согласно целевой функции (5) имеет вид суммы текущего стимула (11) и дисконтированной ожидаемой полезности (Т2.5):

$$v_{s-1} = \Theta(y_{s-1} - e_{s-1}) + \sum_{\tau=s}^{t+T}\rho^{\tau-s+1}\int_\Delta d\xi_s q(\xi_s)... \\ ... \int_\Delta d\xi_\tau q(\xi_\tau)\Theta(\xi_\tau - e_\tau). \quad (\text{Т2.6})$$

Определим действие y^*_{s-1} в периоде $s - 1$, проводя оптимизацию полезности (Т2.6) по y_{s-1} с помощью оператора M_{s-1}. Нетрудно показать, что $\dfrac{\partial v_{s-1}(e_{s-1}, y_{s-1})}{\partial y_{s-1}} = k_{s-1}\delta(y_{s-1} - e_{s-1})$ причем k_{s-1} удовлетворяет неравенствам:

$$k_{s-1} \geqslant 1 - \beta q^*\rho(1+d)\left[1 + \sum_{\tau=s+1}^{t+T}\{\rho[1 - \beta q^*(1+d)]\}^{\tau-s}\right] \geqslant$$

$$\geqslant 1 - \beta q^*\rho(1+d)\left[1 + \sum_{\tau=1}^{t+T-s}\rho^\tau\right] \geqslant 1 - \beta q^*(t + T - s)(1+d) \geqslant$$

$$\geqslant 1 - \beta q^*(1+d)T \geqslant 0.$$

Второе и пятое неравенства имеют место в силу условия (12) теоремы, третье — в силу неравенств $0 \leqslant \rho \leqslant 1$, четвертое — в силу неравенства $t \leqslant s$. Следовательно, $v_{s-1}(e_{s-1}, y_{s-1}) = k_{s-1}\Theta(y_{s-1} - e_{s-1})$ — неубывающая функция y_{s-1} при любых $y_{s-1} \leqslant \xi_{s-1}$. Но тогда, в соответствии с гипотезой благожелательности ДЭ по отношению к Центру, $y^*_{s-1} = \xi_{s-1}$. Полагая $s = t + 1$, получаем: $y^*_t = \xi_t$. Следовательно, РАМ $\Sigma^e = (E, f)$ прогрессивен.

ЛИТЕРАТУРА

1. *Цыганов В .В.* Адаптивные механизмы в отраслевом управлении. М.: Наука, 1991. — 166 с.
2. *Цыганов В. В., Бородин В. А., Шишкин Г. Б.* Интеллектуальное предприятие: механизмы овладения капиталом и властью. — М.: Университетская книга, 2004. — 776 с.
3. *Агеев И. А., Гурлев И. В., Цыганов В. В.* Механизмы манипулирования корпорацией // Проблемы управления. — 2004. — № 3. — С. 34—38.

☎ *(495) 334-91-91*

e-mail: bbc@ipu.ru

Статья представлена к публикации членом редколлегии В. В. Кульбой. ❑

Новая книга

Новиков А.М., Новиков Д.А. Методология. — М.: СИНТЕГ. — 668 с.

С позиций системного анализа в логике современного проектно-технологического типа организационной культуры изложены основы методологии как учения об организации деятельности (научной, практической, художественной, учебной и игровой).

Предназначена для научных и практических работников, а также студентов, аспирантов и докторантов. В первую очередь — для преподавателей вузов и институтов повышения квалификации в целях использования при подготовке курсов лекций по теории систем, системному анализу, методологии научных исследований, инновационной деятельности, проектированию систем, управлению проектами и др.

Сайты авторов в Интернете — электронная библиотека (бесплатный доступ) www.anovikov@.ru ; www.mtas.ru