МЕХАНИЗМЫ ОБУЧЕНИЯ В ВАГОНОРЕМОНТНОМ ПРОИЗВОДСТВЕ

Цыганов В.В.

Институт проблем управления им. В.А. Трапезникова РАН,
Россия, г. Москва, ул. Профсоюзная д.65
bbc@ipu.ru

Аннотация: Рассмотрены механизмы управления ремонтом вагонов в условиях неопределенности. При этом персонал может использовать неизвестный менеджменту производственный потенциал так, чтобы обеспечить себе лучшие стимулы сегодня и в будущем. Разработаны механизмы управления с помощью обучающих инструкций, в котором управляющие воздействия – стимулы формируются на основе оценок, получаемых посредством процедур стохастической аппроксимации. При этом обеспечивается полное использование персоналом неизвестного менеджменту потенциала вагоноремонтного производства.

Ключевые слова: транспорт, вагон, производство, неопределенность, управление, обучение, дальновидность.

Многие задачи управления вагоноремонтным производством в условиях неопределенности сводятся к классификации наблюдаемых ситуаций и событий [1]. Будем предполагать, что потенциал этого производства в периоде t есть случайная величина $\xi_t \in \Delta$, где Δ - компакт. При этом плотность ее распределения $q(\xi_t)$ ограничена: $q(\xi_t) \le q^*$. Величина ξ_t принадлежит, с условной плотностью распределения $q(\xi / k) = q(\xi)$ и априорной вероятностью, к одному из двух неизвестных заранее классов Δ_k^o, $k = \overline{1,2}$, $\Delta_1^o \cup \Delta_2^o = \Delta$. Рассмотрим вначале случай, когда ξ_t известно. Классификация, т.е. отнесение ситуации ξ к одному из двух классов Δ_k^o, $k = \overline{1,2}$, связана с риском. Проблема заключается в определении разбиения, минимизирующего средний риск, связанный с классификацией.

Обозначим через $\{\Delta_1, \Delta_2\}$ некоторое разбиение множества Δ на два подмножества $\Delta_1 \cup \Delta_2 = \Delta$, через ω_{km} - потери, возникающие при отнесении ситуации класса Δ_k^o к классу Δ_m^o (или, иначе, при попадании ситуации класса Δ_k^o в подмножество Δ_m). Предполагается, что $\omega_{11} < \omega_{12}$, $\omega_{22} < \omega_{21}$. Минимизируется средний риск, оценивающий качество классификации:

(1) $$\sum_{k=1}^{2}\sum_{m=1}^{2}\omega_{km}\int_{\Delta_m} Q_k q_k(\xi)d\xi \to min$$

Уравнение для определения точки ξ^*, разделяющей области Δ_1 и Δ_2, при условии минимизации среднего риска (1), имеет вид

(2) $$\mu_{12}(\xi^*) = \sum_{k=1}^{2}(\omega_{k1} - \omega_{k2})Q_k q_k(\xi^*) = 0.$$

Оптимальное решающее правило имеет вид: $\xi \in \Delta_1$, если $\mu_{12}(\xi) < 0$, в противном случае $\xi \in \Delta_2$.

Предположим теперь, что априорные вероятности Q_k, $k = \overline{1,2}$ неизвестны лицу, принимающему решения (кратко - Менеджеру). Рассмотрим задачу выделения указанных классов с помощью решающего правила на основе процедуры обучения Менеджера Тьютором [1]. Именно, предположим

наличие указаний Тьютора Менеджера, касающихся принадлежности любой ситуации ξ_t двум непересекающимся классам Δ_1^0 и Δ_2^0, $\Delta_1^0 \cup \Delta_2^0 = [0,b]$:

$$(3) \qquad S(\xi_t) = \begin{cases} 0, & \text{если} \quad \xi_t \in \Delta_1^0, \\ 1, & \text{если} \quad \xi_t \in \Delta_2^0. \end{cases}$$

Заметим, что (3) эквивалентно существованию ξ^* такого, что $\Delta_1^0 = [0,\xi^*)$ и $\Delta_2^0 = [\xi^*,b]$. Поэтому выражение (3) можно записать в виде

$$S(\xi_t) = \Theta(\xi_t - \xi^*) = \begin{cases} 1, & \text{если} \quad \xi_t \geq \xi^* \\ 0, & \text{если} \quad \xi_t < \xi^* \end{cases},$$

где ξ^* - параметр решающего правила Тьютора. Если бы были известны Q_k, $k = \overline{1,2}$, и путем решения (2) удалось найти ξ^*, то оптимальное решающее правило Менеджера имело бы вид $\mu_{12}(\xi) = \xi - \xi^*$. Однако это невозможно, поскольку Менеджеру неизвестны соответствующие априорные вероятности. В связи с этим, рассмотрим стохастическую аппроксимацию $\mu_{12}(\xi)$ в виде:

$$(4) \qquad \mu_{12}(c,\xi) = \xi - c.$$

Воспользуемся следующим решающим правилом: при $\xi_t < c$ ситуация относится к классу 1 $(\xi_t \in \Delta_1)$, в противном случае - к классу 2 $(\xi_t \in \Delta_2)$. Здесь c - параметр, настраиваемый таким образом, чтобы минимизировать критерий качества стохастической аппроксимации параметра ξ^* оптимального решающего правила $\mu_{12}(\xi)$:

$$(5) \qquad J_{\xi^*}(c) = \int_{\Delta} [\mu_{12}(\xi) - \mu_{12}(c,\xi)]^2 d\xi.$$

С учетом (1)-(4), условие минимума критерия (5) имеет вид

$$(6) \quad \frac{dJ_{\xi^*}(c)}{dc} = c\ell + E_\xi\{\omega_{11} - \omega_{12} + \tilde{\omega} S(\xi) - h\} = 0, \quad \ell = \int_{\Delta} d\xi, \; h = \int_{\Delta} \xi d\xi, \; \tilde{\omega} = \sum_{k,m=1}^{2} (-1)^{k+m+1} \omega_{km}$$

где E_ξ - символ математического ожидания. Для решения уравнения (6) можно использовать алгоритм стохастической аппроксимации:

$$(7) \quad c_{t+1} = I^S(c_t, \xi_t) = c_t - \gamma_t\{c_t + [\omega_{11} - \omega_{12} + \tilde{\omega} S(\xi_t) - h]/l\} \xrightarrow[t]{} a(\xi^*) = \arg\min_c J_{\xi^*}(c)$$

Здесь $a(\xi^*)$ - наилучшая аппроксимация параметра решающего правила Тьютора ξ^*.

На практике персонал часто более осведомлен о потенциале производства, чем менеджмент. В таких случаях говорят об асимметричной осведомленности сторон. Обычно руководство заинтересовано в использовании неизвестного ему потенциала. Тем не менее, персонал, зная истинный потенциал, может выбрать своё состояние так, чтобы обеспечить лучшие стимулы сегодня и в будущем.

Предположим, что потенциал ξ_t неизвестен как Менеджеру, так и Тьютору. Однако ξ_t становится известен персоналу в начале периода t, то есть до выбора его выхода y_t. Дальновидный персонал (ДП) выбирает в своих интересах y_t, который не обязательно равен потенциалу ξ_t ($y_t \neq \xi_t$). Менеджеру и Тьютору остается только наблюдать за выходом ДП y_t.

Предположим, что Тьютор может установить, к какому классу (Δ_1^0 или Δ_2^0) относится наблюдаемое состояние y_t. Однако величина потенциала ДП ξ_t ему неизвестна. Поэтому Тьютор не в состоянии выявить случаи неполного использования элементом своего потенциала (т.е. ситуации y_t, в которых $y_t < \xi_t$). Далее, предположим, что для поддержки своих решений Менеджер использует инструкции Тьютора:

(9) $$S(y_t) = \begin{cases} 0 & if \quad y_t \in D_1^0 \\ 1 & if \quad y_t \in D_2^0 \end{cases}$$

Кроме того, для управления в условиях неопределенности относительно ξ_t, Менеджер формирует собственные оценки с помощью процедуры стохастической аппроксимации (7):

(10) $$a_{t+1} = I^S(a_t, y_t),$$

где y_t - наблюдаемое Менеджером состояние ДП, не обязательно совпадающее с его потенциалом $(y_t \le \xi_t)$.

Рассмотрим механизм управления с помощью инструкций Тьютора (кратко – ТМ) $\Sigma^S = (I^S, f)$, в котором управляющие воздействия в периоде t – стимулы φt формируются на основе оценок, получаемых посредством процедуры стохастической аппроксимации (10). Менеджер должен установить такой ТМ $\Sigma^S = (I^S, f)$, который обеспечивает полное использование неизвестного ему потенциала ξ_t, т.е. выполнении равенства: $y_t = \xi_t$. Персонал же, зная потенциал ξ_t и ТМ $\Sigma^S = (I^S, f)$, должен выбрать своё состояние так, чтобы обеспечить себе лучшие стимулы сегодня и в будущем.

Рассмотрим игру ДП и Менеджера, возникающую вследствие их асимметричной информированности о потенциале ξ_t. Исходя из того, что выход y_t не может превышать потенциал ξ_t, ДП выбирает y_t, $y_t \le \xi_t$, так, чтобы максимизировать свою целевую функцию:

(11) $$V(a_t, y_t) = \sum_{\tau=t}^{t+T} \rho^{\tau-t} \varphi_\tau, \quad 0 < \rho < 1,$$

где ρ - коэффициент дисконтирования, T - дальновидность персонала.

Заметим, что при $y_t \ne \xi_t$ из (6) и (7), вообще говоря, следует $a_t \ne c_t$, так что a_t не сходится к оптимальному параметру $a(\xi^*)$. Поэтому оценки, получаемые с помощью ТМ $\Sigma^S = (I^S, f)$, могут быть далеки от оптимальных. Следовательно, другая важная задача менеджмента - проектирование корректного ТМ $\Sigma^S = (I^S, f)$, при котором $a_t \underset{t}{\to} a(\xi^*)$. Будем говорить, что ТМ $\Sigma^S = (I^S, f)$ корректен, если оценки параметра, получаемые на основе стохастической процедуры (10), сходятся к $a(\xi^*)$- наилучшей аппроксимации параметра решающего правила ξ^* (7).

Будем предполагать, что справедлива гипотеза благожелательности персонала по отношению к менеджменту: если множество выходов y_t, на которых достигается максимум $V(a_t, y_t)$ (11), включает точку $y_t^* = \xi_t$, то ДП выбирает выход y_t^*, равный потенциалу: $y_t^* = \xi_t$ (т.е. персонал использует весь потенциал производства).

Теорема. Для использования потенциала производства ($y_t^* = \xi_t$) и корректности ТМ $\Sigma^S = (I^S, f)$ достаточно:

(12) $$f(a_t, y_t) = \Theta(y_t - a_t) = \begin{cases} 1 & при \quad y_t \ge a_t \\ 0 & при \quad y_t < a_t \end{cases}.$$

Доказательство. Согласно (11), целевая функция ДП V_t зависит как от текущих, так и от будущих стимулов φt,…, φt+T. По условию (12), текущий стимул $\varphi t = f(a_t, y_t)$ возрастает (не убывает) с ростом показателя y_t. Далее, в ТМ $\Sigma^S = (I^S, f)$ используется процедура обучения с Тьютором I^S (10), определяемая с учетом (7). Согласно (9), величина $S(y_t)$ возрастает (не убывает) с ростом y_t. Кроме того, согласно (7) и (10), нормы a_τ, с ростом $S(y_t)$, убывают (не возрастают), $\tau = \overline{t+1, t+T}$. С другой стороны, согласно (12), с убыванием нормы a_τ, будущий стимул в периоде τ - $\varphi\tau = f(a_\tau, y_\tau)$ возрастает (не убывает) при любом τ, $\tau = \overline{t+1, t+T}$. Следовательно, с ростом показателя y_t, будущие стимулы со стороны Центра $\varphi\tau = f(a_\tau, y_\tau)$ возрастают (не убывают) при $\tau = \overline{t+1, t+T}$.

Но целевая функция ДП V_t - монотонно возрастающая функция φ_τ, $\tau = \overline{t,t+T}$. Следовательно, с ростом показателя y_t, возрастает (не убывает) и величина $V(a_t, y_t)$. Поскольку $y_t \le \xi_t$, то максимум $V(a_t, y_t)$ достигается при $y_t = \xi_t$. Следовательно, согласно гипотезе благожелательности ДП по отношению к Центру, $y_t^* = \xi_t$. При этом выполняется (7), и оценки a_t сходятся к $a(\xi^*)$, ч.т.д.

Заметим, что согласно (7), (10), чем выше показатели ДП (y_t), там ниже норма его оценки на следующий период (a_{t+1}). Это соответствует дополнительным стимулам для развития элемента – при повышении показателя y_t, ДП получает не только более высокое поощрение, но и «планка оценки» для него в будущем (a_{t+1}) понижается. Далее использование ТМ $\Sigma^S = (I^S, f)$ иллюстрируется на примере вагоноремонтного производства ОАО «РЖД» [2].

Литература

1. *Tsyganov V.* Tutoring mechanisms of business management / 21st IEEE International Conference on Business Informatics, July 15-17, 2019, NRU HSE, Moscow: IEEE, 2019.
2. *Tsyganov V.* Tutoring mechanisms of wagon-repairing / Management of Large-Scale System Development, October 1-3, 2019, Moscow: IEEE, 2019.