DOI: 10.25728/mlsd.2022.0834

ОБУЧЕНИЕ С ПОДКРЕПЛЕНИЕМ И УПРАВЛЕНИЕ ИЕРАРХИЧЕСКИМ ВАГОНОРЕМОНТНЫМ ПРОИЗВОДСТВОМ

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

Аннотация: Рассмотрена модель корпорации, в которой региональный менеджер и подчиненный ему директор завода организуют производство. Их работу контролируют управляющий корпорации с консультантом. Найдены достаточные условия синтеза механизма обучения с подкреплением и управления, обеспечивающего использование потенциала производства, который иллюстрируется на примере вагоноремонтной корпорации.

Ключевые слова: корпорация, управление, обучение с подкреплением, активность, механизм функционирования.

Введение

Развитие производства базируется на концепции ИНДУСТРИИ 4.0 и искусственного интеллекта [1]. Основным направлением его использования в управлении производством является машинное обучение (МО) [2, 3]. Спектр приложений, где используется МО, постоянно расширяется. Одной из причин является универсальность, поскольку МО может быть использовано в контексте контролируемого, неконтролируемого и подкрепляемого обучения (ПО) [4]. В процессе ПО, агент учится, взаимодействуя со средой. Сигналы подкрепления — это реакция окружающей среды на принимаемые решения. Правила подкрепления основаны на использовании явных и неявных учителей. ПО выигрывает от своей способности распознавать структуры, изучать сложные элементы управления или имитировать поведение, без необходимости более глубокого понимания и математического моделирования [5, 6]. Однако применение ПО в моделях управления затруднено из-за отсутствия соответствующей теоретической базы [7].

Другая проблема управления производством в корпорации связана с человеческим фактором. В практике управления многоуровневыми компаниями, руководители более низкого уровня лучше знают возможности производства, чем руководители более высоких уровней. В таких случаях говорят об асимметричной осведомленности сторон [8]. Введение в литературу и специальный выпуск по проблеме неосведомленности был подготовлен Шиппером [9].

При асимметричной осведомленности, человеческий фактор часто проявляется в активности людей, направленной на использование информации для достижения собственных целей. Поэтому, при использовании ПО в многоуровневой корпорации, необходимо учитывать активность ее руководителей. Для достижения собственных целей, они могут манипулировать, чтобы влиять на результаты ПО, планирование и стимулирование в свою пользу. Чтобы избежать такого рода нежелательной активности, используется теория активных систем и теория организационного управления [10]. В последние годы на их основе были проведены исследования и разработки в сфере управления производством в условиях неопределенности. Одно из их направлений связано с адаптивными механизмами цифрового управления крупномасштабными промышленными системами [11] и корпоративным производством [12]. Неконтролируемое обучение использовалось при проектировании адаптивных информационных моделей управления производством [13] и для его обновления [14]. Еще одна область исследований и их применений в производстве связана с контролируемым обучением для принятия решений [15]. Различные комбинации этих моделей были использованы для построения сложных моделей управления производством, основанных на концепции системотехники (Model Based Systems Engineering) [16]. Для этого адаптивные информационные модели элементов, расположенных на разных уровнях корпорации, объединяются, как архетипы адаптивного управления [13]. Например, исследованы механизмы управления производством с контролируемым обучением в вертикальном концерне [17].

Ниже мы рассмотрим иерархическую модель крупномасштабной корпорации, в которой её региональный Менеджер и подчиненный ему Директор местного завода организуют выпуск продукции. Их контролирует представитель центрального аппарата корпорации (Управляющий), с целью увеличить выпуск. Чтобы научиться управлять Менеджером и Директором в условиях неопределенности, Управляющий пользуется услугами Консультанта. При этом Менеджер знает производство лучше, чем Управляющий и Консультант. Таким образом, Менеджер может

манипулировать выпуском, чтобы влиять на результаты обучения Управляющего и Консультанта в свою пользу. С другой стороны, и сам Менеджер не знает максимальных производственных возможностей (потенциала) местного завода. Таким образом, Директор тоже может манипулировать производством на заводе, чтобы влиять на результаты обучения Менеджера в свою пользу. Поэтому Менеджер также должен научиться контролировать Директора.

Для учета особенностей задачи, обратимся к существующей практике управления производством в многоуровневой транспортной корпорации [18]. Во-первых, производственники стимулируются вышестоящими руководителями, если фактический выпуск не ниже установленного плана (норматива). Поэтому, чем выше план, тем сложнее получить стимулы. Во-вторых, план на следующий период часто увеличивается на определенный процент от текущего выпуска [12]. Поэтому планы на будущее будут тем выше, чем выше выпуск сегодня. В результате, производственники могут быть не заинтересованы в превышении планов (потому что чем выше планы на будущее, тем труднее получить стимулы в будущем). Применительно к рассматриваемой иерархической модели корпорации, возникает проблема незаинтересованности Менеджера и Директора в увеличении выпуска. Для её решения, Управляющему необходимо выбрать процедуры обучения и управления так, чтобы мотивировать Менеджера и Директора к раскрытию возможностей увеличения выпуска.

1 Обучение с подкреплением

1.1 Производство и выпуск продукции

Менеджер несет ответственность за выпуск продукции. Обозначим t период времени, t=0,1,... Объем выпуска c_t в периоде t равен сумме выпуска завод m_t и вспомогательного выпуска a_t: $c_t = a_t + m_t$. Менеджер может выбрать a_t, $a_t \in A_t = [\alpha, A_t]$, где A_t — стохастическая переменная, которая определяет максимально возможный вспомогательный выпуск, $A_t \in [\varepsilon, \gamma], \varepsilon \geq \alpha$. Менеджер знает A_t. Кроме того, Менеджер может выбрать $m_t, m_t \in M_t = [\mu, M_t]$. Здесь M_t — выпуск продукции заводом, который определяется его Директором, $M_t \in \varphi = [\chi, \delta], \chi \geq \mu$. Таким образом, максимально возможный выпуск $C_t = A_t + M_t$. Поэтому $c_t \in X_t = [\alpha + \mu, C_t]$, $X_t \subset X = [\alpha + \mu, \delta + \gamma]$. Для простоты предполагается, что $\alpha + \mu = 0, \delta + \gamma = 1$. На первый взгляд, это выглядит как ограничение. Но поскольку всегда возможно масштабирование, проводимые ниже рассуждения нетрудно обобщить и для произвольных границ $\alpha + \mu$ и $\delta + \gamma$.

1.2 Контролируемое обучение Управляющего

В соответствии с гипотезой асимметричной осведомленности [8], предполагается, что и Управляющий, и Консультант не знают A_t, M_t, и C_t. Но A_t и M_t становятся известными Менеджеру до выбора a_t, m_t и c_t. Поэтому Управляющий должен контролировать Менеджера, чтобы увеличить выпуск c_t до максимума C_t.

Предположим, что Управляющий оценивает работу Менеджера по увеличению выпуска c_t как удовлетворительную (оценка 1) или неудовлетворительную (0). Неправильная оценка приводит к потерям Управляющего. Обозначим: l_{10} — потери от несправедливой оценки 0 (хотя справедливая оценка Менеджера равна 1); l_{10} — потери от несправедливой оценки 1 (хотя справедливая оценка Менеджера равна 0). Для формирования справедливых оценок, Управляющий использует контролируемое обучение. Формально Управляющий учится с помощью консультаций. Именно, если Консультант считает, что выпуск в периоде t занижен, то $b_t = 1$. В противном случае $b_t = 0$.

Предположим, что c_t — стационарная стохастическая величина. Тогда оценка Менеджера должна определяться так, чтобы минимизировать ожидаемые средние потери Управляющего. Для этого можно использовать алгоритм контролируемого обучения [15]. В нем параметр правила принятия решения (сокращенно — норматив s_τ, τ=1,2,...,) определяется с помощью алгоритма стохастической аппроксимации:

$$s_{t+1} = S(s_t, c_t) = s_t - \gamma_t[s_t - 0{,}5l_{01} + (l_{01} + l_{10})b_t], \; s_o = s^0, \; m = 0{,}1{,}..., \tag{1}$$

где $S(s_t, c_t)$ — нормативная процедура, $\gamma_t > 0$, $\sum_{t=0}^{\infty} \gamma_t < \infty$. Используя (1), Управляющий определяет оценку Менеджера

$$h_t = H(s_t, c_t) = \begin{cases} 1, & \text{если } c_t \geq s_t \\ 0, & \text{если } c_t < s_t \end{cases}. \tag{2}$$

1.3 Обучение Консультанта с подкреплением

Управляющий консультируется по поводу занижения Менеджером выпуска. Формально, если выпуск c_t занижен, то консультация $b_t = 0$, в противном случае $b_t = 1$. Ошибочная консультация связана со штрафом для Консультанта. Чтобы избежать штрафа, Консультант использует ПО.

Введем следующие обозначения: $F_0(c_t, e) = c_t \text{-} \eta e$ — штраф Консультанту за ошибочную консультацию $b_t = 0$ (тогда как правильная консультация — $b_t = 1$); $F_1(c_t, e) = \iota(e\text{-}c_t)$ — штраф Консультанту за ошибочную консультацию $b_t = 1$ (тогда как правильная консультация — $b_t = 0$). Здесь e — параметр, настраиваемый для максимизации среднего штрафа, $0 < \eta < 1, \iota > 0$. Для настройки параметра e, Консультант может использовать алгоритм самообучения [14]. В нем Консультант формирует оценку e_{t+1} параметра e:

$$e_{t+1} = E(c_t, e_t) = \begin{cases} e_t + \eta\kappa_t, & \text{если } c_t < g_t \\ e_t - \iota\kappa_t, & \text{если } c_t \geq g_t \end{cases}, \; e_0 = e^0, \; m = 0,1,..., \tag{3}$$

где E — процедура самообучения, $g_t = e_t(\eta + \iota)/(\iota + 1)$, $0 < \kappa_{t+1} < \kappa_t$, $\sum_{\tau=1}^{\infty} \kappa_\tau < \infty$. С помощью (3), Консультант определяет консультацию:

$$b_t = B(c_t, e_t) = \begin{cases} 1, & \text{если } c_t \geq g_t \\ 0, & \text{если } c_t < g_t \end{cases}, \qquad m = 0,1,..., \tag{4}$$

где $B(\bullet)$ — процедура консультации. Набор нормативной (S), оценочной (H), самообучающейся (E) и консультационной (B) процедур, которые определяются в соответствии с (1), (2), (3) и (4), называется механизмом обучения и обозначается $K_L = (S, H, E, B)$.

1.4 Выбор Менеджера

Фактически, норматив s_t — это порог выпуска c_t, приемлемый для Управляющего, при котором Менеджер получает оценку 1 ($h_t = 1$). Если выпуск c_t меньше $s_t(c_t < s_t)$, то $h_t = 0$, и Управляющий может наказать Менеджера. Таким образом, полезность Менеджера увеличивается вместе с его оценкой. Формально, полезность дальновидного Менеджера растет вместе с текущими и будущими оценками:

$$V_t = V(h_t, h_{t+1}, \dots h_{t+\Theta}), \; V_\tau \uparrow h_\tau, \; \tau = \overline{t, t+\Theta}, \tag{5}$$

где Θ — число периодов, учитываемых дальновидным Менеджером. Для увеличения полезности (5), Менеджер выбирает выпуск c_t в периоде t, зная C_t, но не зная будущих выпусков $\mathrm{C}_\tau, \tau = \overline{t+1, t+\Theta}$. Предположим, что Менеджер руководствуется принципом максимально гарантированного результата [10], предполагая, что $C_\tau \in X$ и $c_\tau \in X_\tau$, $\tau = \overline{t+1, t+\Theta}$. Тогда целевая функция Менеджера $D_t(c_t)$ — это максимальное гарантированное значение (5):

$$D_t(c_t) = \min_{\tau=\overline{t+1,t+\Theta}} \; \min_{C_\tau \in X} \; \min_{c_\tau \in X_\tau} \; V_t, \; t = 0,1,\dots \tag{6}$$

Обозначим c_t^* значение c_t, которое максимизирует целевую функцию Менеджера (6). Тогда множество оптимальных c_t^* выглядит следующим образом:

$$\Psi_t(\mathrm{C}_t, \Phi_C) = \{\mathrm{c}_t^* | D_t(c_t^*) \geq D_t(c_t)\}. \tag{7}$$

Ниже предполагается, что Менеджер благожелателен к Управляющему: если $C_t \in \Psi_t(C_t, \Phi_C)$, то $c_t^* = C_t$, t=0,1,... Это означает, что Менеджер скрывает резервы производства только в том случае, если это увеличивает его целевую функцию (6).

1.5 Механизм обучения

Необходимо разработать механизм, который мотивировал бы Менеджера увеличивать выпуск продукции в каждом периоде: $c_t^* = C_t$, t=0,1,....

Лемма 1. Механизма обучения $K_L = (S, H, E, B)$ достаточно для $c_t^* = C_t$, t=0,1,...

Доказательство. Согласно (5), целевая функция Менеджера (6) не уменьшается с ростом h_τ, $\tau = \overline{t, t+\Theta}$. Механизм $K_L = (S, H, E, B)$ включает в себя процедуры (1)-(4). Согласно (2), h_t не уменьшается с ростом c_t. Рассмотрим зависимости $h_\tau, \tau = \overline{t+1, t+\Theta}$, от c_t. Из (3) легко видеть, что e_τ (а значит и g_τ) не увеличивается с ростом $c_t, \tau = \overline{t+1, t+\Theta}$. Поэтому, согласно (4), b_τ не уменьшается с ростом c_t. Поэтому, согласно (1), s_τ не увеличивается с ростом c_t. Кроме того, согласно (2), b_τ не увеличивается с ростом c_t. Следовательно, h_τ не уменьшается с ростом c_t, $\tau = \overline{t+1, t+\Theta}$. Таким образом, из (5) и (6) следует, что целевая функция Менеджера $D_t(c_t)$ не уменьшается с ростом c_t. Поскольку $c_t \leq C_t$, то $D_t(c_t) \leq D_t(C_t)$. Следовательно, согласно (7), $C_t \in \Psi_t(\mathrm{C}_t, \Phi_C)$. Из того, что Менеджер благожелателен к Управляющему, получаем $c_t^* = C_t$, $t = 0,1,\dots$, ч.т.д.

2 Адаптация к изменениям

2.1 Заинтересованность Менеджера в росте выпуска продукции заводом

Согласно Лемме 1, если Управляющий использует механизм $K_L = (S, H, E, B)$, то Менеджер заинтересован в увеличении выпуска: $c_t^* = C_t = A_t + M_t$. Из доказательства Леммы 1 следует, что целевая функция $D_t(c_t)$ не уменьшается с ростом c_t. Поэтому целевая функция $D_t(c_t)$ не уменьшается с ростом M_t. Поэтому Менеджер заинтересован в увеличении выпуска завода M_t.

Однако M_t определяет Директор завода. Предположим, что M_t ограничено максимально возможным объемом выпуска завода (кратко — потенциалом) P_t: $M_t \le P_t$, где P_t — стационарная случайная величина, $P_t \in \varphi = [\chi, \delta]$. Тогда $M_t \in \Pi_t = [\chi, P_t]$. В духе гипотезы асимметричной осведомленности [8], будем предполагать, что Директор узнает реализацию P_t в начале периода t (т.е. до выбора M_t). Но реализация P_t не известна Менеджеру. Таким образом, Директор может манипулировать выпуском завода в свою пользу, выбирая M_t, при $M_t < P_t$. Это характерно, например, для вагоноремонтного производства [15]. Таким образом, Менеджер должен мотивировать Директора использовать потенциал завода: $M_t = P_t$, t=0,1,...

2.2 Планирование выпуска

Если Менеджеру известна статистика P_t, t=0,1,..., то он может использовать аналитику больших данных для прогнозирования производственного потенциала. В противном случае можно использовать адаптивные методы прогнозирования. Например, в [19] предложено прогнозирование на основе адаптивной модели Брауна. Обст с соавторами [20] разработали адаптивные методы краткосрочного прогнозирования. Эти методы прогнозирования могут быть оптимизированы с помощью методов идентификации с теоретико-информационными критериями [21].

Обозначим f_t прогноз потенциала P_t в периоде t, $t = 0,1,\ldots$ Задача состоит в том, чтобы минимизировать средние потери прогнозирования $V_P\{L(\zeta_t)\}$. Здесь функция потерь $L(\zeta_t)$ является выпуклой дважды дифференцируемой функцией невязки $\zeta_t = P_t - f_t$, $L(0) = 0$, V_P — оператор усреднения по всем реализациям P_t. Обозначим через $R_t(f, P^t) = (\sum_{\tau=1}^{t} L(\zeta_\tau)|_{f_\tau = f})/t$ — эмпирические средние потери, характеризующие качество прогноза f, $P^t = (P_0, \ldots, P_t)'$. Тогда оптимальный выборочный прогноз f_{t+1} в периоде $t+1$ рассчитывается по рекуррентной формуле:

$$f_{t+1} = arg\min_f R_t(f, P^t) = f_t - \gamma_t L_f'(\zeta_t), f_0 = f^0, \tag{8}$$

$$\gamma_t \in \Gamma = \{\gamma_t > 0 | \gamma_t \ge \gamma_{t+1}, \textstyle\sum_{t=1}^{\infty} \gamma_t < \infty\}, \quad t = 0,1,\ldots \tag{9}$$

где $L_f'(\zeta_t)$ — производная функции $L(\zeta_t)$ по f. Но Менеджер не знает P_t. Ему известно только M_t. Подставляя в (8) M_t вместо P_t, Менеджер может получить рекуррентную формулу для определения оценки r_{t+1} оптимального выборочного прогноза f_{t+1} в виде:

$$r_{t+1} = arg\min_r R_t(r, M^t) = r_t - \gamma_t L_r'(M_t - r_t), \quad r_0 = f^0. \tag{10}$$

Менеджер хочет заинтересовать Директора в раскрытии потенциала: $M_t = P_t, t = 0,1,\ldots$ Предположим, что для этого Менеджер в периоде t устанавливает заводу производственный план p_t на основе оценки r_t: $p_t = r_t, t = 0,1,\ldots$Тогда, согласно (10), рекуррентное уравнение для плана p_{t+1} имеет вид:

$$p_{t+1} = p_t - \gamma_t L_p'(M_t - p_t) \equiv R_t(p_t, M_t), \quad t = 0,1,\ldots, \tag{11}$$

$$R = \{R_t(p_t, M_t), t = 0,1,\ldots\}, \ \gamma_t L_p^{"}(M_t - p_t) < 1, \ p_0 = f^0, \tag{12}$$

где $R = \{R_t(p_t, M_t), t = 0,1,\ldots\}$ — процедура планирования, $L_p^{"}(M_t - p_t)$ — вторая производная $L_p(M_t - p_t)$ по p_t.

2.3 Стимулирование выпуска

Стимул Директору за выполнение плана p_t в периоде t:

$$i_t = I(p_t, M_t), I(p_t, M_t) \downarrow p_t, I(p_t, M_t) \uparrow M_t, \tag{13}$$

где $I(p_t, M_t)$ — дифференцируемая функция. Назовем $I(p_t, M_t)$ процедурой стимулирования. Директор суммирует стимулы (13) на θ будущих периодов, взвешивая их с помощью коэффициента дисконтирования ρ:

$$U_t = \textstyle\sum_{\tau=t}^{t+\theta} \rho^{\tau - t} i_\tau, \quad 0 < \rho < 1, \tag{14}$$

Предположим, что Директор знает только множество φ будущих потенциалов $P_\tau, \tau = \overline{t+1, t+\theta}$. Тогда Директор выбирает M_t, чтобы максимизировать гарантированное значение (14):

$$W_t(p_t, M_t, P_t) = \min_{P_\tau \in \varphi,\ \gamma_\tau \in \Gamma,\ \tau = \overline{t+1,t+\theta}} \min_{M_\tau \in \Pi_\tau,\ \tau = \overline{t+1,t+\theta}} U_t. \quad (15)$$

При этом множество оптимальных выборов M_t^* Директора

$$Y_t(P_t) = \{M_t^* \in \Pi_t \mid W_t(p_t, M_t^*, P_t) \geq W_t(p_t, M_t, P_t), M_t \in \Pi_t\} \quad (16)$$

Предполагается доброжелательность Директора, по отношению к Менеджеру: если $P_t \in Y_t(P_t)$, то $M_t^* = P_t$. Другими словами, Директор не занижает выпуск, если это ему не выгодно.

2.4 Адаптивный механизм

Введем оператор для устранения неопределенности относительно потенциалов и выпусков, предшествующих выбору M_t в периоде *t*:

$$o_t x = \min_{P_t \in \varphi} \min_{P_{t-1} \in \varphi, \gamma_{t-1} \in \Gamma} \min_{M_{t-1} \in \Pi_{t-1}} \dots \min_{P_0 \in \varphi, \gamma_0 \in \Gamma} \min_{M_0 \in \Pi_0} \frac{\partial}{\partial x}. \quad (17)$$

Введем также оператор устранения неопределенности в отношении потенциалов и выпусков после выбора M_t:

$$O_t x = \min_{\tau = \overline{t+1,t+\theta}} \min_{P_{t+1} \in \varphi, \gamma_{t+1} \in \Gamma} \min_{M_{t+1} \in \Pi_{t+1}} \dots \min_{P_\tau \in \varphi, \gamma_\tau \in \Gamma} \min_{M_\tau \in \Pi_\tau} \frac{\partial}{\partial x}. \quad (18)$$

Предполагая дифференцируемость функций $R_t(p_t, M_t), t = 0,1, \dots$, обозначим $L_p'(M_\omega - p_\omega) \equiv L_p'(\omega), \omega = 0,1, \dots$, и

$$G_t = O_t p_\tau i_\tau, J_t = oM_t[-L_p'(t)], N_t = O_t p_\tau L_p'(\tau). \quad (19)$$

Лемма 2. Для максимизации выпуска завода в каждом периоде: $A_t^* = P_t$, *t*=0,1,..., достаточно использовать адаптивный механизм $K_A = (R, I)$ с процедурами, определенными в соответствии с (11) – (13), параметры которых удовлетворяют неравенствам:

$$o_t M_t s_t \geq \rho \gamma_t G_t J_t [1 - \rho^\theta (1 - \gamma_t N_t)^\theta] / [1 - \rho(1 - \gamma_t N_t)], t = 0,1, \dots \quad (20)$$

Доказательство. Дифференцируя U_t, из (14) получаем:

$$\frac{\partial U_t}{\partial M_t} = \frac{\partial i_t}{\partial M_t} + \sum_{\tau=t+1}^{t+\theta} \rho^{\tau - t} \frac{\partial i_\tau}{\partial p_\tau} \frac{\partial p_\tau}{\partial M_t}. \quad (21)$$

Подставив выражение для p_t из (11) в (18), и используя (21) в качестве рекуррентного уравнения, нетрудно получить:

$$\frac{\partial U_t}{\partial M_t} = \frac{\partial i_t}{\partial M_t} - \gamma_t \frac{\partial L_p'(t)}{\partial M_t} \sum_{\tau=t+1}^{t+T} \rho^{\tau - t} \frac{\partial i_\tau}{\partial p_\tau} \prod_{\xi=t+1}^{\tau-1} [1 - \gamma_\xi L_p''(\xi)]. \quad (22)$$

Согласно (13), $\frac{\partial i_t}{\partial M_t} \geq 0$. Тогда из (17) следует:

$$\partial i_t / \partial M_t \geq o_t M_t i_t. \quad (23)$$

Рассмотрим вычитаемое в формуле (22). Заметим, что $L(M_t - p_t)$ является выпуклой дважды дифференцируемой функцией, которая растет с увеличением остатка $(M_t - p_t)$. Поэтому $L_p'(M_t - p_t)$ монотонно уменьшается с ростом M_t, так что $\frac{\partial L_p'(t)}{\partial M_t} \leq 0$. В силу (13), $\partial i_t / \partial p_t \leq 0$. Из (9) и неравенства в (12) следует, что все факторы под знаком произведения положительны. Принимая во внимание знаки всех факторов вычитаемого в (22), получаем, что в целом вычитаемое в (22) положительно. Используя определения (17) – (19) и условие $\gamma_t \geq \gamma_{t+1}$ из (8), легко показать, что

$$\gamma_t \frac{\partial L_p'(t)}{\partial M_t} \sum_{\tau=t+1}^{t+T} \rho^{\tau - t} \frac{\partial i_\tau}{\partial p_\tau} \prod_{\xi=t+1}^{\tau-1} [1 - \gamma_\xi L_p'(\xi)] \leq \rho \gamma_t G_t J_t [1 - \rho^\theta (1 - \gamma_t N_t)^\theta] / [1 - \rho(1 - \gamma_t N_t)]. \quad (24)$$

Вычитая (24) из (23), с учетом (19), получаем: $\frac{\partial U_t}{\partial M_t} \geq o_t M_t - \rho \gamma_t G_t J_t [1 - \rho^\theta (1 - \gamma_t N_t)^\theta] / [1 - -\rho(1 - \gamma_t N_t)]$. Следовательно, $\partial U_t / \partial M_t \geq 0$ по условию Леммы 2 (20). Поэтому, согласно (15), максимум $W_t(p_t, M_t, P_t)$ по M_t достигается при $A_t = P_t$. Тогда из (16) $P_t \in Y_t(P_t)$, и, вследствие доброжелательности Директора к Менеджеру, $A_t^* = P_t$, ч.т.д.

3 Корпоративный механизм

Рассмотрим, как создать механизм, обеспечивающий максимальный выпуск: $c_t^* = A_t + P_t$, *t*=0,1,...

Теорема. Механизм $K = \{K_L, K_A\}$, включающий обучающий механизм $K_L = (S, H, E, B)$ и адаптивный механизм $K_A = (R, I)$, параметры которых удовлетворяют (20), достаточен для максимизации выпуска корпорации: $c_t^* = A_t + P_t$, t=0,1,...

Доказательство. Если c_t — стационарная стохастическая величина, то согласно Лемме 1, $K_L = (S, H, E, B)$ достаточен для $\mathrm{c}_t^* = \mathrm{C}_t = A_t + M_t$. Далее, согласно Лемме 2, для $M_t^* = P_t$ достаточно механизма $K_A = (R, I)$, если его параметры удовлетворяют (20). Так как P_t — стационарная стохастическая величина, то и M_t^* — стационарная стохастическая величина. Учитывая, что A_t — также стационарная стохастическая величина, получаем, что $\mathrm{c}_t = A_t + M_t$ — тоже стационарная стохастическая величина. Но тогда, согласно лемме 1 и по условиям Теоремы, $c_t^* = A_t + P_t$, t=0,1,..., ч.т.д.

4 Пример: вагоноремонтное производство

Проиллюстрируем применение разработанного корпоративного механизма на примере управления ремонтом грузовых вагонов в вагоноремонтной корпорации ВРК-3, имеющей региональное представительство на Урале [18]. Прообразом Управляющего в исследуемой модели является один из руководителей ВРК-3, использующий обучающий механизм для управления производством в этом регионе. Уральский представитель ВРК-3 выступает в роли Менеджера. Подчиненный ему директор вагоноремонтного завода играет роль Директора.

Менеджер следит за ежемесячным количеством отремонтированных грузовых вагонов на этом заводе [15]. В условиях неопределенности, он использует адаптивный механизм $K_A = (R, I)$, чтобы мотивировать Директора увеличить число отремонтированных грузовых вагонов c_t. Для простоты, Менеджер использует линейный адаптивный механизм $K_A^l = (R^l, I^l)$ с линейными процедурами планирования и стимулирования:

$$p_{t+1} = R^l(p_t, M_t) = (1 - \gamma_t)p_t + \gamma_t M_t, \qquad p_0 = f^0, t = 0,1\ldots, \tag{25}$$

$$I^l(p_t, M_t) = v(M_t - p_t) + const, v > 0,\ t = 0,1.\ldots \tag{26}$$

Подставляя (25) и (26) в (20), получаем неравенство (20) в виде:

$$\rho\gamma_t[1 - \rho^\theta(1 - \gamma_t)^\theta] \le [1 - \rho(1 - \gamma_t)], t = 1,2\ldots. \tag{27}$$

В практике работы ВРК-3, γ_t уменьшается очень плавно: $\gamma_t << 1$. При этом неравенство (27) заведомо выполняется. Соответственно, неравенство (20) также выполняется, так что теорема применима. Таким образом, корпоративного механизма $K = \{K_L, K_A^l\}$ достаточно, чтобы максимально увеличить число отремонтированных грузовых вагонов. Этот пример иллюстрирует простоту, прозрачность и эффективность корпоративного механизма $K = \{K_L, K_A^l\}$, а также применимость условий теоремы.

Заключение

Рассмотрена модель транспортной корпорации, в которой региональный менеджер и подчиненный ему директор завода организуют местное производство продукции. Их работу контролируют управляющий корпорации с консультантом. Менеджер знает возможности выпуска только на региональном уровне. Директор знает производственный потенциал завода. Ни управляющий, ни его консультант не знают возможностей менеджера и директора. Таким образом, менеджер может манипулировать собственным результатом, чтобы влиять на решения консультанта и управляющего, чтобы увеличить собственное вознаграждение. Но и сам менеджер не знает потенциала завода. Это также может быть использовано его директором в свою пользу. Поэтому менеджеру необходимо научиться контролировать директора.

Интересы менеджера и директора отражаются в модели путем введения их целевых функций. На основе данной модели, найдены достаточные условия для синтеза механизма обучения с подкреплением и управления, обеспечивающего использование случайных возможностей увеличения выпуска. При этом управляющий использует советы самообучающегося консультанта. Посредством этих консультаций управляющий может нормировать результаты работы менеджера. Обучающий механизм побуждает менеджера, во-первых, максимизировать вспомогательный выпуск и, во-вторых, внедрить адаптивный механизм увеличения выпуска завода. Такой адаптивный механизм включает в себя процедуры планирования и стимулирования, которые побуждают директора максимизировать выпуск завода. Такого рода механизмы могут быть использованы для управления выпуском продукции и в других крупномасштабных корпорациях.

Литература

1. *Blanchet M., Rinn T., Thaden G., and Thieulloy G.* Industrie 4.0 – the new industrial revolution. In How Europe Will Succeed. München: Roland Berger Strategies, 2014.
2. *Bristow D., Tharayil M., and Alleyne A.* A survey of iterative learning control // IEEE Control Systems Magazine. Vol. 26. 2006, № 4. – P.96-114.
3. *Fradkov A.* Early history of machine learning. Berlin: Proc. of the 21st IFAC World Congress, 2020 – P. 3439.
4. *Sutton R. and Barto A.* Reinforcement learning: an introduction. Massachusetts: The MIT Press, 2nd edition, 2018.
5. *Recht B.* A tour of reinforcement learning: the view from continuous control. arXiv:1806.09460v2 [math.OC] 10 November 2018.
6. *Bertsekas D.* Reinforcement learning and optimal control. Athena: Athena Scientific, 2019.
7. *Recht B.* Reflections on learning-to-control renaissance. Berlin: Proc. of the 21st IFAC World Congress, 2020. – P. 4707.
8. *Auster S.* Asymmetric awareness and moral hazard // Games and Economic Behavior. Vol. 82. 2013. – P.503-521.
9. *Schipper B.* Unawareness – a gentle introduction to the literature // Mathematical social sciences. Vol. 70. 2014. – P.1-9.
10. *Burkov V., Gubko M., Kondratiev V., Korgin N., and Novikov D.* Mechanism design and management. New York: NOVA Publishers, 2013.
11. *Tsyganov V.* Learning mechanisms in digital control of large-scale industrial systems. In Global Smart Industry Conference. Chelyabinsk: IEEE, 2018. – P.1-9.
12. *Tsyganov V.* Corporative productivity adaptive mechanisms. In Control Systems, Mathematical Modeling, Automation and Energy Efficiency. Lipetsk: IEEE, 2020. – P.172-178.
13. *Tsyganov V.* Designing adaptive information models for production management // Procedia CIRP. Vol. 84. 2019. – P.1088-1093.
14. *Tsyganov V.* Self-tuning dichotomy and bonuses for renovation. In: Software Engineering Perspectives in Intelligent Systems. CoMeSySo 2020. Advances in Intelligent Systems and Computing. Vol. 1295. 2020. – P.644-656.
15. *Tsyganov V.* Decision making and learning in wagon-repairing. In Management of Large-Scale System Development. Moscow: IEEE, 2019. – P.261-267.
16. *Kossiakoff A., Sweet W., Seymour S., and Biemer S.* Systems Engineering. Principles and Practice. New York: John Wiley, 2011.
17. *Tsyganov V.* Mechanisms for learning and production management of a vertical concern. In Informatics and Cybernetics in Intelligent Systems. Vol. 228. 2021. – P.466-475.
18. Annual report of Carriage Repair Company – 3 (2017). https//raex-a.ru/annual_reports/reports/2017_vrk_3.pdf.
19. *Dibrivniy O., Onyshchenko V., and Grebenyuk V.* Forecasting based on the trend model and adaptive Brown's model. In Proc. of 14th Conf. on Advanced Trends in Computer Engineering. Lviv: IEEE. 2018. – P.944-947.
20. *Obst D., de Vilmarest J., Goude Y.* Adaptive methods for short-term electricity load forecasting during COVID-19 lockdown in France // IEEE Transactions on Power Systems. 2021. DOI: 10.1109/TPWRS.2021. 3067551.
21. *Stoorvogel A., van Schuppen J.* System identification with information theoretic criteria. In Bittanti S. and Picci G., (eds), Identification, Adaptation, Learning. London: Springer, 1996. – P. 289-338.