

**Институт проблем управления
им. В.А. Трапезникова РАН**

УПРАВЛЕНИЕ БОЛЬШИМИ СИСТЕМАМИ

***Выпуск 77
Январь 2019***

**СБОРНИК
ТРУДОВ**

ISSN 1819-2467

Регистрационный номер Эл. №ФС77-44158 от 09 марта 2011 г.

Москва – 2018

РОССИЙСКАЯ АКАДЕМИЯ НАУК
Институт проблем управления
им. В.А. Трапезникова

**УПРАВЛЕНИЕ
БОЛЬШИМИ
СИСТЕМАМИ**

СБОРНИК ТРУДОВ

Выпуск 77

Москва – 2019

КООРДИНАЦИОННЫЙ СОВЕТ

Академики РАН: Васильев С.Н., Емельянов С.В., Куржанский А.Б., Федосов Е.А., Черноусько Ф.Л.; члены-корреспонденты РАН: Желтов С.Ю., Каляев И.А., Пархоменко П.П., Попков Ю.С.; д-ра техн. наук: Кузнецов О.П., Кульба В.В., Лотоцкий В.А., Павлов Б.В., Поляк Б.Т., Рутковский В.Ю.

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

Главный редактор: член-корр. РАН Новиков Д.А. **Зам. главного редактора:** д-р физ.-мат. наук Губко М.В. **Отв. секретарь:** канд. техн. наук Калимулина Э.Ю. **Редактор:** канд. техн. наук Квинто Я.И.

Д-ра техн. наук: проф. Алескеров Ф.Т. (ГУ ВШЭ), проф. Алчинов А.И. (ИПУ РАН), проф. Андриевский Б.Р. (ИПМ РАН), проф. Афанасьев В.Н. (МИЭМ), проф. Бахтадзе Н.Н. (ИПУ РАН), проф. Бурков В.Н. (ИПУ РАН), проф. Вишневский В.М. (ИПУ РАН), Галаев А.А. (ИПУ РАН), д-р физ.-мат. наук проф. Ерешко Ф.И. (ВЦ РАН), д-ра техн. наук Зоркальцев В.И. (ИСЭМ СО РАН), проф. Калашников А.О. (ИПУ РАН), проф. Калянов Г.Н. (ГУ ВШЭ), проф. Каравай М.Ф. (ИПУ РАН), д-р экон. наук, проф. Клочков В.В. (ИПУ РАН), д-ра техн. наук, Коргин Н.А. (ИПУ РАН), проф. Курдюков А.П. (ИПУ РАН), д-ра физ.-мат. наук, проф. Кушнер А.Г., проф. Лазарев А.А. (МФТИ), д-ра техн. наук: проф. Лебедев В.Г. (ИПУ РАН), проф. Мандель А.С. (ИПУ РАН), д-р биол. наук проф. Михальский А.И., д-р физ.-мат. наук, проф. Непейвода Н.Н. (ИПС РАН), д-р экон. наук, проф. Нижегородцев Р.М. (ИПУ РАН), д-ра техн. наук: проф. Орлов А.И. (МГТУ), д-ра физ.-мат. наук: проф. Рапопорт Л.Б. (ИПУ РАН), проф. Райгородский А.М. (МГУ), проф. Савватеев А.В. (РЭШ), д-ра техн. наук: проф. Самуйлов К.Е. (РУДН), проф. Сидельников Ю.В. (МАИ), Совлуков А.С. (ИПУ РАН) д-ра физ.-мат. наук: проф. Соловьев С.Ю. (МГУ), проф. Угольницкий Г.А. (ЮФУ), проф. Уткин В.А. (ИПУ РАН), проф. Хоботов Е.Н. (МГТУ), д-ра физ.-мат. наук: доцент Чеботарев П.Ю. (ИПУ РАН), проф. Чхартишвили А.Г. (ИПУ РАН), проф. Щербаков П.С. (ИПУ РАН).

РЕГИОНАЛЬНЫЕ РЕДАКЦИОННЫЕ СОВЕТЫ

Арзамас – д-р физ.-мат. наук проф. Пакшин П.В. **Волгоград** – д-ра физ.-мат. наук: проф. Воронин А.А., проф. Лосев А.Г. (ВолГУ); **Воронеж** – д-р техн. наук, проф. Баркалов С.А., д-р физ.-мат. наук, проф. Головинский П.А. (ВГАСУ), д-р техн. наук, проф. Подвальный С.Л. (ВГТУ); **Иркутск** – академик РАН Бычков И.В., д-р физ.-мат. наук, проф. Лакеев А.В. (ИДСТУ СО РАН); **Казань** – д-р физ.-мат. наук, проф. Маликов А.И., д-р техн. наук, проф. Сиразетдинов Р.Т. (КГТУ-КАИ); **Липецк** – д-ра техн. наук: проф. Погодаев А.К., Сараев П.В. (ЛГТУ); **Самара** – д-ра экон. наук: проф. Богатырев В.Д., проф. Гераськин М.И., д-р техн. наук, проф. Засканов В.Г. (СГАУ); **Петрозаводск** – д-р физ.-мат. наук, проф. Мазалов В.В., д-р техн. наук, доц. Печников А.А. (ИПМИ КарНЦ РАН); **Санкт-Петербург** – д-р физ.-мат. наук: проф. Петросян Л.А. (СПбГУ), д-р техн. наук проф. Фуртат И.Б. (ИПМ РАН); **Старый Оскол** – д-р техн. наук, проф. Еременко Ю.И. (СТИ).

Адрес редакции: 117997, г. Москва, ул. Профсоюзная, д. 65.

Адрес в интернете: ubs.mtas.ru.

СОДЕРЖАНИЕ

Системный анализ

Соболев В.Н.

О законе стационарной очереди для одной системы массового обслуживания с групповым поступлением требований 6

Горелов М.А.

Иерархическая игра с ограничениями на содержание и объем передаваемой информации 20

Математическая теория управления

Фургат И.Б., Гущин П.А.

Алгоритм управления в условии насыщения сигнала управления и его производных..... 47

Квинто Я.И., Хлебников М.В.

Верхние границы максимального отклонения траектории в линейных дискретных системах: робастная постановка..... 70

Управление в социально-экономических системах

Белов М.В.

Статистические модели процессов приёма и увольнения сотрудников..... 85

**Управление в медико-биологических
и экологических системах**

Подвальный С.Л., Васильев Е.М. <i>Многоальтернативность: эволюционная стратегия биологических систем</i>	125
--	-----

**Управление техническими системами
и технологическими процессами**

Белов М.В., Новиков Д.А. <i>Модели разработки и освоения технологии комплексной деятельности.....</i>	171
---	-----

Гужа Е.Д., Романенко В.А., Скороход М.А. <i>Нечёткая оценка эффективности трансферной системы авиаперевозок</i>	219
---	-----

О ЗАКОНЕ СТАЦИОНАРНОЙ ОЧЕРЕДИ ДЛЯ ОДНОЙ СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ С ГРУППОВЫМ ПОСТУПЛЕНИЕМ ТРЕБОВАНИЙ¹

Соболев В. Н.²
(Москва)

Рассматривается однолинейная система массового обслуживания с групповым поступлением требований, в которой моменты поступления групп требований образуют процесс восстановления, длительности обслуживания имеют показательное распределение, число заявок в группе ограничено, а число мест ожидания неограничено. Для данной системы массового обслуживания найдены условия выполнения основного закона стационарной очереди Хинчина. Показано, что в случае выполнения основного закона стационарной очереди Хинчина для описанной выше системы массового обслуживания стационарные вероятности числа заявок в системе по времени имеют один и тот же вид при любом входящем потоке, и совпадают с соответствующими вероятностями однолинейной системы массового обслуживания с групповым поступлением требований, в которой моменты поступления групп требований образуют простейший входной поток, длительности обслуживания имеют показательное распределение, число мест ожидания неограничено. Доказано одно новое представление для производящей функции стационарных вероятностей числа заявок в системе по времени. Для этого вводится производящая функция «хвостов» распределения числа требований во входящей группе заявок и производящая функция стационарных вероятностей числа заявок в системе по времени вложенной однородной цепи Маркова.

Ключевые слова: система массового обслуживания, групповое поступление, стационарное распределение, производящая функция вероятностей, вложенная цепь Маркова, процесс восстановления, основной закон стационарной очереди Хинчина.

1. Введение

В сборнике «Аналитические и вычислительные методы в теории вероятностей и её приложениях (АВМТВ-2017)» [3, с. 171–175] была опубликована статья А.Д. Соловьёва и В.Н. Со-

¹ Автор пользуется случаем и благодарит В.В. Козлова за постоянное внимание к его работе, плодотворное обсуждение и редакционные замечания.

² Виталий Николаевич Соболев, к.ф.-м.н., свободный исследователь (sobolev_vn@mail.ru).

болева «Одна система массового обслуживания с групповым поступлением требований», в которой рассматривалась система $GI^{\nu}|M_{\mu}|1|\infty$.

В системе массового обслуживания $GI^{\nu}|M_{\mu}|1|\infty$ моменты поступления групп требований $t_0 = 0, t_0 < t_1 < t_2 < \dots < t_n < \dots$ образуют процесс восстановления [2] с функцией распределения $P\{X_n < t\} = F(t)$, где $X_n = t_n - t_{n-1}, n \geq 1$.

В каждый момент t_n поступает группа из ν_n требований, причем величины ν_n независимы, одинаково распределены и ограничены.

В системе имеется один обслуживающий прибор, время обслуживания распределено по показательному закону с параметром μ , а число мест для ожидания неограниченно.

Пусть $\xi(t)$ – число требований в системе в момент t . В рамках данной системы в [3, с. 171–175] было найдено стационарное распределение процесса

$$P(z) = \lim_{t \rightarrow \infty} Mz^{\xi(t)} = \sum_{n=0}^{\infty} p_n z^n,$$

в котором искомое распределение определялось через стационарное распределение вложенной однородной цепи Маркова

$$\xi_n = \xi(t_n - 0), n = 1, 2, \dots, \xi_0 = 0$$

с производящей функцией

$$\pi(z) = \lim_{t \rightarrow \infty} Mz^{\xi_n} = \sum_{k=0}^{\infty} \pi_k z^k.$$

В данной работе исследуются условия, при выполнении которых в модели $GI^{\nu}|M_{\mu}|1|\infty$ выполнен основной закон стационарной очереди Хинчина [4, с. 76], т.е. для любого неотрицательного k выполняются соотношения

$$p_k = \pi_k.$$

Данные соотношения эквивалентны совпадению производящих функции распределений $P(z) = \pi(z)$.

В статье доказано, что для системы $M_\lambda^\nu | M_\mu | 1 | \infty$ выполняется основной закон стационарной очереди Хинчина; а если данный закон выполняется для произвольной модели $GI^\nu | M_\mu | 1 | \infty$, то в последней стационарные вероятности p_k и π_k совпадают с соответствующими вероятностями системы $M_\lambda^\nu | M_\mu | 1 | \infty$.

Также в статье найдено одно новое представление для производящей функции стационарных вероятностей $P(z)$, имеющее простой вероятностный смысл.

2. Предварительные сведения и обозначения

2.1. ВХОДЯЩИЙ ПОТОК

Как было указано выше, в модели $GI^\nu | M_\mu | 1 | \infty$ длины интервалов между поступлениями групп требований $X_n = t_n - t_{n-1}$, которые представляют собой время между приходами двух соседних групп требований, независимы и одинаково распределены с функцией распределения $F(x)$.

Определим T как среднее время между поступлениями групп заявок в систему:

$$T = MX_n = \int_0^\infty x dF(x).$$

В частности, для модели $M_\lambda^\nu | M_\mu | 1 | \infty$ имеем $F(x) = 1 - e^{-\lambda x}$ и среднее время между поступлениями групп заявок в систему определяется через параметр распределения по следующей формуле:

$$T = \int_0^\infty \lambda x e^{-\lambda x} dx = \frac{1}{\lambda}.$$

2.2. ВХОДЯЩАЯ ГРУППА ТРЕБОВАНИЙ

В каждый момент времени t_n поступает группа требований из ν_n независимых, одинаково распределённых требований с производящей функцией

$$\alpha(z) = Mz^{\nu_n} = \alpha_1 z + \alpha_2 z^2 + \dots + \alpha_m z^m, \quad \alpha_m \neq 0,$$

где $\alpha_k = P \{ \nu_n = k \}$, $k = 1, 2, \dots, m$. Величины ν_n независимы в совокупности от величин X_n .

С помощью производящей функции $\alpha(z)$ легко находится среднее число заявок в группе:

$$\nu = M\nu_n = \alpha'(z) \Big|_{z=1} = \sum_{k=1}^m k\alpha_k.$$

2.3. ПРИБОР

Пусть случайная величина Y_n обозначает время обслуживания n -й заявки. Величины Y_n независимы друг от друга и от величин X_n . По условию задачи величины Y_n имеют одинаковое распределение с функцией распределения $G(t) = 1 - e^{-\mu t}$.

В этом случае среднее время обслуживания n -й заявки τ конечно и равно

$$\tau = MY_n = \int_0^{\infty} x dG(x) = \int_0^{\infty} x \mu e^{-\mu x} dx,$$

где $\mu = 1/\tau$ – интенсивность обслуживания.

2.4. ОБСЛУЖИВАНИЕ

Пусть η_n – это число точек пуассоновского потока с параметром μ , приходящих на интервале (t_n, t_{n+1}) . Поскольку в системе имеется один обслуживающий прибор, а время обслуживания распределено по показательному закону с параметром μ , то величину η_n можно интерпретировать как количество заявок обслуженных за время $X_{n+1} = t_{n+1} - t_n$.

Определим случайные величины $S_n = Y_1 + Y_2 + \dots + Y_n$ и величину

$$\mu(t) = \max \{ n \geq 0 \mid S_n \leq t \}.$$

Тогда $\eta_n = \mu(X_{n+1})$ и

$$P \{ \mu(t) = n \} = \frac{(\mu t)^n}{n!} e^{-\mu t}.$$

Неотрицательные целочисленные случайные величины η_n независимы и поэтому можно вычислить для каждой из них производящую функцию

$$\begin{aligned}\omega(z) = Mz^{\eta_n} &= \sum_{s=0}^{\infty} \omega_s z^s = \sum_{s=0}^{\infty} z^s \int_0^{\infty} \frac{(\mu x)^s}{s!} e^{-\mu x} dF(x) = \\ &= \int_0^{\infty} e^{-\mu x(1-z)} dF(x) = \varphi(\mu - \mu z),\end{aligned}$$

где через $\varphi(z)$ обозначено преобразование Лапласа–Стильтьеса функции распределения $F(x)$:

$$\varphi(z) = \int_0^{\infty} e^{-zx} dF(x), \quad z \geq 0,$$

а через

$$\omega_s = \int_0^{\infty} \frac{(\mu x)^s}{s!} e^{-\mu x} dF(x), \quad s \geq 0,$$

– вероятности события $\{\eta_n = s\}$, т.е. ω_s – вероятности того, что за время между поступлениями двух соседних групп заявок будет обслужено ровно s заявок при условии, что после прихода первой из них общее число заявок в системе было больше s .

Как известно [1, с. 94], в случае $F(x) = 1 - e^{-\lambda x}$ преобразование Лапласа–Стильтьеса имеет следующий вид:

$$\varphi(z) = \lambda \int_0^{\infty} e^{-(z+\lambda)x} dx = \frac{\lambda}{z + \lambda}.$$

Откуда

$$(1) \quad \omega(z) = Mz^{\eta_n} = \varphi(\mu - \mu z) = \frac{\lambda}{\mu - \mu z + \lambda} = \frac{\rho_0}{1 + \rho_0 - z},$$

где $\rho_0 = \lambda/\mu$.

С помощью производящей функции $\omega(z)$ в общем случае можно найти среднее число обслуженных на интервале (t_n, t_{n+1}) требований:

$$M\eta_n = \varphi'(\mu - \mu z)|_{z=1} = \mu T = \frac{1}{\rho_0}.$$

3. Явный вид производящих функций $\pi(z)$ и $P(z)$ в модели $GI^\nu|M_\mu|1|_\infty$

3.1. СТАЦИОНАРНЫЕ РАСПРЕДЕЛЕНИЯ ПРОЦЕССА $\xi(t)$ И ВЛОЖЕННОЙ ЦЕПИ МАРКОВА

В указанной выше статье [3, с. 171–175] (подробнее см. [6, с. 97–108]) для стационарного распределения вложенной цепи Маркова системы $GI^\nu|M_\mu|1|_\infty$ доказана следующая теорема:

Теорема 1 [3, с. 173]. *Если выполнено условие $\nu < \mu T$, то стационарное распределение вложенной цепи Маркова существует и задается производящей функцией*

$$(2) \quad \pi(z) = \frac{(1 - \lambda_1)(1 - \lambda_2) \dots (1 - \lambda_m)}{(1 - \lambda_1 z)(1 - \lambda_2 z) \dots (1 - \lambda_m z)},$$

где числа $\lambda_1, \dots, \lambda_m$ являются корнями уравнения (считая каждый корень столько раз, какова его кратность)

$$(3) \quad \alpha \left(\frac{1}{z} \right) \varphi(\mu - \mu z) = 1$$

принадлежащими области $|z| < 1$.

Отметим, что в [3, с. 171–175] доказано, что у уравнения (3) при $|z| < 1$ имеет ровно m корней с учётом кратности, а уравнение

$$(4) \quad \alpha(z) \varphi \left(\mu - \mu \frac{1}{z} \right) = 1$$

имеет при $|z| > 1$ ровно m корней с учётом их кратности.

Обозначим корни уравнения (4) через z_1, z_2, \dots, z_m . Тогда корни уравнений (3) и (4) связаны равенством $\lambda_k = z_k^{-1}$ для $k = 1, \dots, m$.

Иногда технически оказывается удобнее использовать уравнение (4) и его корни. При этом после замены $\lambda_k = z_k^{-1}$, $k = 1, \dots, m$, производящая функция (2) принимает вид

$$(5) \quad \pi(z) = \frac{(1 - z_1)(1 - z_2) \dots (1 - z_m)}{(z - z_1)(z - z_2) \dots (z - z_m)}.$$

Вторая теорема, доказанная в [3, с. 171–175], позволяет находить для системы $GI^\nu|M_\mu|1|_\infty$ производящую функцию стационарного распределения процесса $\xi(t)$ через производящую функцию вложенной цепи Маркова $\pi(z)$:

Теорема 2 [3, с. 175]. Если выполнено условие $\nu < \mu T$, то стационарное распределение процесса $\xi(t)$ существует и задается производящей функцией

$$(6) \quad P(z) = \sum_{n=0}^{\infty} p_n z^n = \frac{1}{\mu T} z \pi(z) \frac{1 - \alpha(z)}{1 - z} + 1 - \frac{\nu}{\mu T},$$

где $\pi(z)$ определяется равенством (5).

3.2. НОВОЕ ПРЕДСТАВЛЕНИЕ ПРОИЗВОДЯЩЕЙ ФУНКЦИИ СТАЦИОНАРНОГО РАСПРЕДЕЛЕНИЯ ПРОЦЕССА $\xi(t)$ ДЛЯ СИСТЕМЫ $GI^{\nu} | M_{\mu} | 1 | \infty$

Из равенства $\rho_0 = \lambda/\mu = 1/\mu T$ следует, что выражение (6) можно записать в виде

$$(7) \quad P(z) = \rho_0 z \frac{1 - \alpha(z)}{1 - z} \pi(z) + 1 - \nu \rho_0,$$

а условие $\nu < \mu T$ как $\nu \rho_0 < 1$.

Введение понятия загрузки системы

$$(8) \quad \rho = \nu \rho_0.$$

показывает, что условие $\nu < \mu T$ является известным условием на загрузку системы массового обслуживания

$$(9) \quad \rho < 1.$$

Из (7) сразу следует, что вероятность того, что система свободна, равна

$$(10) \quad p_0 = P(0) = 1 - \nu \rho_0 = 1 - \rho.$$

Как хорошо известно (см, например, [5, т. 1, с. 271]), выражение

$$\frac{1 - \alpha(z)}{1 - z}$$

представляет собой производящую функцию «хвоста» распределения случайной величины ν_n , т.е. производящую функцию вероятностей $\alpha_k = P\{\nu_n = k\}$, $k = 1, 2, \dots, m$.

Однако в рассматриваемом случае, в силу того, что случайные величины ν_n неотрицательные и не принимают значение равное нулю, т.е.

$$P\{\nu_n > 0\} = P\{\nu_n \geq 1\},$$

удобнее взять в качестве вероятностей «хвоста»

$$A_k = P \{ \nu_n \geq k \} = P \{ \nu_n > k - 1 \} = \sum_{l=k}^m \alpha_l,$$

где $k = 1, \dots, m$. При таком определении индексы в обозначениях вероятностей сдвигаются на единицу относительно обозначений, принятых в [5, т. 1, с. 271].

В силу равенства

$$\sum_{k=1}^m A_k = \sum_{k=1}^m (\alpha_k + \dots + \alpha_m) = (\alpha_1 + 2\alpha_2 + \dots + m\alpha_m) = \nu$$

скалярные величины

$$q_k = \frac{A_k}{\nu}, \quad k = 1, \dots, m,$$

представляют собой распределение вероятностей с производящей функцией

$$(11) \quad A(z) = \frac{1}{\nu} \sum_{k=1}^m A_k z^k = \sum_{k=1}^m q_k z^k.$$

Для $A(z)$ справедлива следующая лемма:

Лемма 1. Производящая функция «нормированного хвоста» $A(z)$ и производящая функция числа заявок в группе $\alpha(z)$ связаны соотношением

$$(12) \quad A(z) = \frac{z}{\nu} \frac{1 - \alpha(z)}{1 - z}.$$

Доказательство. Следующая цепочка равенств

$$\begin{aligned} \frac{\nu}{z} A(z) &= \sum_{k=1}^m A_k z^{k-1} = \\ &= \alpha_1 + \alpha_2 (1 + z) + \dots + \alpha_m (1 + z + \dots + z^{m-1}) = \\ &= \frac{\alpha_1 (1 - z) + \alpha_2 (1 - z^2) + \dots + \alpha_m (1 - z^m)}{1 - z} = \end{aligned}$$

$$= \frac{1 - \alpha_1 z - \alpha_2 z^2 - \dots - \alpha_m z^m}{1 - z} = \frac{1 - \alpha(z)}{1 - z}$$

полностью доказывает лемму.

Прямым следствием леммы (1) и представления (7) является следующая теорема:

Теорема 3. *Если выполнено условие $\rho < 1$, то стационарное распределение процесса $\xi(t)$ существует и задается производящей функцией*

$$(13) \quad P(z) = (1 - p_0) \pi(z)A(z) + p_0,$$

где производящая функция $A(z)$ определяется равенством (11), а вероятность $p_0 = (1 - \rho)$ – равенством (10).

Замечание 1. Формула (13) показывает, что искомое распределение есть смесь вырожденного распределения и распределения, представляющего собой сумму распределений: стационарного вложенной цепи Маркова с распределением «хвоста» входящей группы требований.

Замечание 2. Представление (13) также может быть записано в виде

$$(14) \quad P(z) = \rho \pi(z)A(z) + 1 - \rho.$$

4. О математическом законе стационарной очереди

4.1. О ЗНАМЕНАТЕЛЕ $\pi(z)$ В МОДЕЛИ $M_\lambda^\nu | M_\mu | 1 | \infty$

Для системы $M_\lambda^\nu | M_\mu | 1 | \infty$ уравнение (4), определяющее корни z_1, z_2, \dots, z_m в представлении (5), может быть сведено к нахождению корней многочлена $R_m(z)$ степени m . Данный многочлен выражается через производящую функцию $A(z)$ «хвоста» распределения числа элементов в группе поступающих требований (11). Кроме того, знаменатель производящей функции $\pi(z)$ в представлении (5) может быть записан с помощью данного многочлена.

Лемма 2. *В системе $M_\lambda^\nu | M_\mu | 1 | \infty$ в качестве знаменателя производящей функции $\pi(z)$ в представлении (5) можно взять многочлен*

$$(15) \quad R_m(z) = 1 - \rho A(z).$$

Доказательство. Легко видеть, что из представления (1) производящей функции $\omega(z) = \varphi(\mu - \mu z)$ следует равенство

$$\varphi\left(\mu - \frac{\mu}{z}\right) = \frac{\rho_0}{1 - z^{-1} + \rho_0} = \frac{\rho_0 z}{z + \rho_0 z - 1},$$

используя которое, уравнение (4) можно записать в виде

$$\alpha(z) \frac{\rho_0 z}{z + \rho_0 z - 1} = 1,$$

а после элементарных преобразований свести к уравнению

$$\rho_0 z \frac{1 - \alpha(z)}{1 - z} = 1,$$

которое в свою очередь с учётом (12) преобразуется к виду

$$1 - \rho A(z) = 0.$$

Выражение $1 - \rho A(z)$ есть некоторый многочлен $R_m(z)$ степени m , который определяется через интенсивности λ , μ и вероятности $\alpha_2, \dots, \alpha_m$ формулой

$$R_m(z) = 1 - \rho_0 z (1 + (\alpha_2 + \dots + \alpha_m)z + \dots + \alpha_m z^{m-1}).$$

Следовательно, корни z_1, \dots, z_m уравнения (4) при $|z| > 1$ суть корни многочлена $R_m(z)$, т.е. удовлетворяют уравнению (16)

$$R_m(z) = 0.$$

Таким образом, лемма доказана.

4.2. ПРЕДСТАВЛЕНИЕ ПРОИЗВОДЯЩИХ ФУНКЦИЙ $\pi(z)$ И $P(z)$ В МОДЕЛИ $M_\lambda^\nu | M_\mu | 1 |_\infty$ ЧЕРЕЗ ФУНКЦИЮ $A(z)$

Представление (15) для многочлена $R_m(z)$ позволяет выписать для модели $M_\lambda^\nu | M_\mu | 1 |_\infty$ явный вид производящих функций $\pi(z)$ и $P(z)$ без нахождения корней z_1, z_2, \dots, z_m . При этом оказывается, что распределения $\pi(z)$ и $P(z)$ совпадают. Докажем следующую теорему:

Теорема 4. Если для системы $M_\lambda^\nu | M_\mu | 1 | \infty$ выполнено условие $\rho < 1$, то стационарное распределение процесса $\xi(t)$ существует и совпадает со стационарным распределением последовательности $\xi(t_n - 0)$, а для их производящих функцией справедливо представление

$$(17) \quad P(z) = \pi(z) = \frac{1 - \rho}{1 - \rho A(z)}.$$

Доказательство. Поскольку $A(z)$ – производящая функция вероятностей, то $A(1) = 1$ и

$$R_m(1) = 1 - \rho A(z)|_{z=1} = 1 - \rho.$$

Поэтому (5) с учётом (15) можно записать в виде

$$\pi(z) = \frac{R_m(1)}{R_m(z)} = \frac{1 - \rho}{1 - \rho A(z)}.$$

Из последнего представления справедливо равенство

$$1 - \rho = \pi(z) - \rho A(z) \pi(z),$$

с учётом которого из (14) следует совпадение стационарных распределений $\pi(z)$ и $P(z)$. Действительно,

$$\begin{aligned} P(z) &= \rho \pi(z) A(z) + 1 - \rho = \\ &= \rho \pi(z) A(z) + \pi(z) - \rho A(z) \pi(z) = \pi(z). \end{aligned}$$

Последнее равенство завершает доказательство теоремы.

4.3. О СОВПАДЕНИИ ПРОИЗВОДЯЩИХ ФУНКЦИЙ $\pi(z)$ И $P(z)$ СИСТЕМЫ $GI^\nu | M_\mu | 1 | \infty$

Рассмотрим случай, когда для системы $GI^\nu | M_\mu | 1 | \infty$ выполнен закон стационарной очереди. Тогда оказывается, что искомые стационарные распределения всегда совпадают с соответствующими стационарными распределениями (см. (17)) модели $M_\lambda^\nu | M_\mu | 1 | \infty$. Докажем данное утверждение.

Теорема 5. Если для системы $GI^\nu | M_\mu | 1 | \infty$ выполнен закон стационарной очереди, то для производящих функций $\pi(z)$ и $P(z)$ справедливо представление (17). Если для системы $GI^\nu | M_\mu | 1 | \infty$ выполнено условие $\rho < 1$, а при $|z| > 1$ корни уравнения $\alpha(z) \varphi(\mu - \mu z^{-1}) = 1$ совпадают с учётом их кратности с корнями уравнения $A(z) = \rho^{-1}$, то для системы выполняется основной закон стационарной очереди.

Доказательство. При выполнении закона стационарной очереди верно равенство $P(z) = \pi(z)$. Подставляя $\pi(z)$ вместо $P(z)$ в формулу (14), убеждаемся в справедливости равенства

$$\pi(z) = \rho \pi(z) A(z) + 1 - \rho,$$

из которого элементарными преобразованиями для $\pi(z)$ получается представление (17). Следовательно, корни уравнения (4) должны совпадать с корнями уравнения $A(z) = \rho^{-1}$.

Обратное очевидно. На этом доказательство теоремы является законченным.

Представление (5) показывает, что в системе $GI^\nu | M_\mu | 1 | \infty$ при выполнении закона стационарной очереди стационарное распределение $P(z)$ зависит только от интенсивностей μ , λ и распределения числа заявок в группе.

Как было показано выше, закон стационарной очереди выполнен для модели $M_\lambda^\nu | M_\mu | 1 | \infty$. Естественно, возникает вопрос о существовании (или отсутствии) других видов входящих потоков, при которых для системы массового обслуживания выполняется основной закон стационарной очереди Хинчина. Ответ на этот вопрос в данный момент нам не известен. Вполне может оказаться, что закон стационарной очереди верен только для $M_\lambda^\nu | M_\mu | 1 | \infty$.

5. Заключение

Выше было показано, что для системы $M_\lambda^\nu | M_\mu | 1 | \infty$ основной закон стационарной очереди Хинчина выполняется. При выполнении данного закона для системы $GI^\nu | M_\mu | 1 | \infty$ искомые стационарные вероятности совпадают с соответствующими стационарными вероятностями системы $M_\lambda^\nu | M_\mu | 1 | \infty$. Совпадение корней уравнения (4) с корнями уравнения (16) гарантирует выполнения закона стационарной очереди для системы $GI^\nu | M_\mu | 1 | \infty$.

Литература

1. БОЧАРОВ П.П., ПЕЧИНКИН А.В. *Теория массового обслуживания*. – М.: РУДН, 1995. – 529 с.
2. ГНЕДЕНКО Б.В., КОВАЛЕНКО И.Н. *Лекции по теории массового обслуживания*. – К.: МИР, 1963. – 315 с.
3. СОЛОВЬЁВ А.Д., СОБОЛЕВ В.Н. *Одна система массового обслуживания с групповым поступлением требований // Аналитические и вычислительные методы в теории вероятностей и её приложениях (АВМТВ-2017): материалы Международной научной конференции. Россия, Москва, 23-27 октября 2017. / МГУ им. М.В.Ломоносова, РУДН; [под общ. ред. А. В. Лебедева]. – Москва: Изд-во «РУДН», 2017. – 743 с.*
4. ХИНЧИН А.Я. *Математическая теория стационарной очереди // Матем. сб. – 1932. – Т. 39, №4. – С. 73–84.*
5. ФЕЛЛЕР В. *Введение в теорию вероятностей и её приложения*. – М.: МИР, 1964. – 498 с.
6. SOLOVIEV A.D., SOBOLEV V.N. *One Server Queue with Bulk Arrivals // In: Analytical and Computational Methods in Probability Theory. ACMPT 2017. Lecture Notes in Computer Science. Vol 10684. / Eds.: Rykov V., Singpurwalla N., Zubkov A. – Cham: Springer, 2017. – 540 p.*

KHINCHIN'S BASIC LAW OF A STATIONARY QUEUE FOR SINGLE-SERVER QUEUEING SYSTEMS WITH BATCH ARRIVALS

Vitaly Sobolev, Moscow, Cand.Sc., free researcher
(sobolev_vn@mail.ru).

Abstract: This paper deals with a queueing system with general renewal arrivals, exponential service times, single service channel and infinite number of waiting positions, customers are serviced in the order of their arrival. For this queueing system, a condition for the fulfilment of the Khinchin's basic law of a stationary queue is given. The article shows that, in the case of basic law of a stationary queue for our system, the stationary distribution of the number of the customers in the system always coincides with the corresponding probability distribution of the queueing system with exponential interarrival times. In stationary case a new form of the probability generating functions of the number of clients in the system is also derived. This new form is written in terms of the probability generating functions of the tail distribution function of the number of customers per group and of the probability generating functions of a embedded discrete time homogeneous Markov chain.

Keywords: queueing system, batch arrivals, stationary distribution, probability generating functions, embedded Markov chain, renewal process, Khinchin's basic law of a stationary queue.

УДК 519.872

ББК 22.171

DOI: <https://doi.org/10.25728/ubs.2019.77.1>

*Статья представлена к публикации
членом редакционной коллегии Э.Ю. Калимулиной.*

Поступила в редакцию 23.12.2018.

Дата опубликования 31.01.2019.

ИЕРАРХИЧЕСКАЯ ИГРА С ОГРАНИЧЕНИЯМИ НА СОДЕРЖАНИЕ И ОБЪЕМ ПЕРЕДАВАЕМОЙ ИНФОРМАЦИИ

Горелов М. А.¹

(Вычислительный центр им. А.А. Дородницына
ФИЦ ИУ РАН, Москва)

Рассматривается иерархическая игра двух лиц. Считается, что игрок верхнего уровня обладает правом первого хода. Предполагается, что он располагает некой информацией о выборе партнера, но имеется два типа ограничений на доступ к такой информации. С одной стороны, некоторые варианты такого выбора остаются для игрока верхнего уровня неразличимыми. Кроме того, учитывается наличие ограничения на объем используемой игроком верхнего уровня информации о выбранной партнером стратегии. Для измерения количества информации используется комбинаторный подход (в терминах А.Н. Колмогорова). «Смысл» информации о действиях партнера игрок верхнего уровня вправе определять сам (в пределах ограничения первого типа). Считается, что игрок верхнего уровня точно знает интересы и возможности партнера и рассчитывает на его рациональное поведение. В этих предположениях задача вычисления максимального гарантированного результата игрока верхнего уровня представляет собой задачу вычисления максимина на сложных функциональных пространствах. В статье эта задача сводится к вычислению кратных максиминов на «конечномерных» пространствах. Предложено два подхода к вычислению этого результата. Выясняется структура оптимальной стратегии и, в частности, оптимальная семантика используемой игроком верхнего уровня информации. Приведен иллюстративный пример, демонстрирующий возможность применения предложенной техники.

Ключевые слова: иерархические игры, максимальный гарантированный результат, информация.

1. Введение

В процессе принятия решений в условиях конфликта или неопределенности вопросы информированности занимают одно из центральных мест. На «бытовом» уровне это понимает каждый. Да и теоретики осознали это очень давно. Но здесь имеется две проблемы.

¹ Михаил Александрович Горелов, к.ф.-м.н. (griever@ccas.ru).

С одной стороны, информации обычно не хватает: всегда хотелось бы, принимая решение, знать выборы партнеров. А последним, чаще всего, выгодно скрывать свои действия, намерения и т.д. И во многих случаях существенная информация становится известной лишь тогда, когда решение уже принято и изменить ничего нельзя.

А с другой стороны, информации слишком много. Действительно, знать все детали поведения контрагентов, не влияющие или мало влияющие на результат своей деятельности, лицу, принимающему решения, обычно не нужно и даже вредно, если учесть затраты времени и ресурсов, необходимые для обработки «лишней» информации. В эпоху цифровизации экономики это становится особенно заметно, поскольку объемы доступной информации растут очень быстро. Вопрос в том, как отличить «немногие существенные вещи от многих несущественных вещей»? Ответ на этот вопрос, разумеется, должно искать лицо, принимающее решение. А вот ограничение на объем информации, которую можно эффективно обработать в нужные сроки, во многих случаях можно считать объективно заданным.

В данной работе рассматривается простейшая модель, учитывающая наличие этих двух проблем.

Модели, учитывающие наличие первой из отмеченных проблем, изучаются весьма активно. Впервые модели такого рода появились в ранних работах Дж. Фон Неймана. В настоящее время в исследовании таких моделей основное внимание сосредоточено на разработке методов поиска оптимальных решений типа равновесия по Нэшу в играх со сложной динамикой принятия решений [11, 13]. Ниже рассматривается игра с достаточно простой динамикой, но исследуется иной принцип оптимальности. В его определении появляется максимин со связанными переменными, что приводит к появлению дополнительных сложностей.

Ближе всего к теме данной статьи модель из работы [1]. Эта модель обобщалась и модифицировалась в разных направлениях в работах тех же авторов. Кроме того, различные варианты этой модели изучались в рамках теории активных систем [2–3] и теории контрактов [10, 12].

В известном смысле все подобные модели сложнее модели из [1]. Они учитывают наличие большего числа игроков, динамики, внешней неопределенности и т.п. Но во многих случаях задачи оказываются слишком сложными, чтобы их можно было решить в общем виде. Поэтому на параметры модели накладываются дополнительные ограничения. Число рассмотренных вариантов постановок задач и этих дополнительных ограничений слишком велико, чтобы на них можно было остановиться в рамках данной статьи.

Изучение моделей, учитывающих наличие второй проблемы, было начато позже [5]. Модель из этой статьи также варьировалась в различных направлениях в других работах автора. Но к теме данной статьи эти вариации прямого отношения не имеют. Поэтому останавливаться на них не станем.

Учет наличия обеих проблем, естественно, усложняет задачу. Но с возникающими трудностями все-таки удастся справиться, по крайней мере в случае, рассмотренном далее.

В данной статье исследуется статическая игра двух лиц без неопределенных факторов. Предполагается, что один из игроков обладает правом первого хода и ориентируется на свой гарантированный результат в предположении о рациональном поведении партнера.

Статья построена следующим образом. В разделе 2 описываются возможности игроков, их интересы и, главное, информированность. В результате получается некоторая игра в нормальной форме, множества стратегий в которой наделены сложной структурой. В следующем разделе дается два определения максимального гарантированного результата: одно традиционное, другое более удобное для использования. Доказывается их эквивалентность. В разделе 3 структура множеств стратегий игроков в рассматриваемой игре используется для того, чтобы охарактеризовать максимальный гарантированный результат в более простых терминах. Это позволяет, в частности, выяснить структуру оптимальной стратегии игрока, обладающего правом первого хода. В разделе 4 задача вычисления максимального гарантированного результата сводится к задаче вычисления максиминов.

2. Интересы, возможности, информированность

Рассмотрим игру двух лиц в нормальной форме $\Gamma = \langle U, V, g, h \rangle$. Здесь U и V – множества управлений первого и второго игроков соответственно, а $g: U \times V \rightarrow \mathbf{R}$ и $h: U \times V \rightarrow \mathbf{R}$ – их функции выигрыша (буква \mathbf{R} обозначает множество действительных чисел).

Дабы избежать в дальнейшем ненужных технических сложностей, будем считать, что множества U и V наделены топологиями и компактны, а функции g и h непрерывны в топологии декартова произведения $U \times V$.

В рассматриваемой модели игра Γ задает «технологические» возможности и интересы игроков. Опишем их информированность.

Как отмечалось во введении, в данной статье рассматривается случай, когда первый игрок, принимая решение, может использовать информацию о выборе партнера, но при этом имеются и некие ограничения. Это означает, что для некоторых пар управлений $v \in V$ и $v' \in V$ доступная первому игроку информация не позволяет отличить выбор v от выбора v' .

Отношение «неотличимости» естественно считать отношением эквивалентности, т.е. предполагать выполненными следующие условия:

- 1) выбор v не отличим от самого себя;
- 2) если выбор v неотличим от выбора v' , то и выбор v' не отличим от выбора v .
- 3) если выбор v не отличим от выбора v' , а выбор v' не отличим от выбора v'' , то выбор v не отличим от выбора v'' .

В пояснении нуждается, пожалуй, лишь последний пункт. Будем рассуждать от противного. Пусть выбор v можно отличить от выбора v'' . Тогда существует какое-то доступное первому игроку «измерение», которое для выборов v и v'' дает разные результаты. Тогда то же измерение, примененное к выбору v' , даст результат, отличный либо от результата измерения v , либо от результата измерения v'' , т.е. позволит отличить либо v и v' , либо v' и v'' .

Если условия 1–3 выполнены, то множество V разбивается на непересекающиеся классы «неотличимых» элементов. Обозначим множество таких классов буквой Ω . Существует каноническое отображение $T: V \rightarrow \Omega$, ставящее в соответствие элементу $v \in V$ содержащий его класс эквивалентности $\omega \in \Omega$. Пожалуй, самый простой способ задания отношения эквивалентности состоит в указании фактормножества Ω и канонического отображения T . При этом можно считать, что первому игроку доступна любая информация об элементе $\omega = T(v)$, соответствующем выбранному вторым игроком управлению $v \in V$.

Сказанное в трех предыдущих абзацах является лишь изложением неких неформальных соображений. В сухом остатке будет лишь следующее.

Будем считать, что задано множество Ω и отображение $T: V \rightarrow \Omega$.

Можно считать, что множество Ω наделено фактортопологией, т.е. самой тонкой топологией, в которой отображение T непрерывно (см. [9]). В этой топологии пространство Ω компактно как образ компактного множества V при непрерывном отображении T .

В разделе 5 понадобится еще одно предположение относительно отношения «неотличимости». Сформулировано оно может быть следующим образом. Будем считать, что множество Ω , наделенное фактортопологией, является T_1 -пространством. Содержательно это означает, что если первый игрок может уверенно отличить управление v от управления w , то он может также отличить любое управление v' , достаточно близкое к v , от любого управления w' , достаточно близкого к w (близость понимается в смысле топологии на множестве V). Примеры задач, для которых это дополнительное ограничение было бы обременительным, автору не известны.

Вдобавок к описанным ограничениям на информированность первого игрока будем предполагать, что объем информации, которую он способен получить и своевременно обработать, тоже ограничен. Сказанное формализуем способом, впервые предложенным в [5].

Будем предполагать, что вся информация о $\omega = T(v)$, которую первый игрок получает и обрабатывает в процессе игры, может быть закодирована двоичным словом $r = (r_1, \dots, r_n)$ из множества $N = \{0,1\}^n$. Таким образом, объем обрабатываемой информации не превосходит n бит. Способ кодировки $P: \Omega \rightarrow N$ выбирает первый игрок.

Кроме того, для каждого сообщения $r \in N$ он вправе выбрать свое управление $u \in U$. То есть фактически он выбирает функцию $u_*: N \rightarrow U$.

Сказанное приводит к следующим формальным конструкциям.

Будем обозначать через $\Phi(X, Y)$ класс всех функций, отображающих множество X в множество Y .

Положим $U_* = \Phi(N, U) \times \Phi(\Omega, N)$, $V_* = V$. Определим функции $g_*: U_* \times V_* \rightarrow \mathbf{R}$ и $h_*: U_* \times V_* \rightarrow \mathbf{R}$ условиями $g_*(u_*, P, v) = g(u_*(P(T(v))), v)$ и $h_*(u_*, P, v) = h(u_*(P(T(v))), v)$. Получим новую игру в нормальной форме $\Gamma_* = \langle U_*, V, g_*, h_* \rangle$, которая будет основным объектом исследования в данной работе.

Будем считать, что все параметры рассматриваемой модели, а именно, четверка $\Gamma = \langle U, V, g, h \rangle$, а также множество Ω , отображение T и число n точно известны обоим игрокам.

Приведем простейший пример, позволяющий интерпретировать веденные конструкции в содержательных терминах (разумеется, с большой долей условности).

Пусть второй игрок выбирает объемы v^{ij} ресурсов типа j , выделяемых на производство продукции вида i , $i = 1, 2, \dots, k$, $j = 1, 2, \dots, l$, из некоторого множества V , принадлежащего пространству размерности $k \times l$. Если такой выбор сделан, то объем произведенной продукции вида i составит $\pi^i(v^{i1}, v^{i2}, \dots, v^{il})$. Пусть первому игроку доступна лишь информация о прибыли

$$T(v) = \sum_{i=1}^k p^i \pi^i(v^{i1}, v^{i2}, \dots, v^{il}) - \sum_{j=1}^l c^j \sum_{i=1}^k v^{ij},$$

где p^i и c^j – цены на продукцию и ресурсы, а v – набор всех v^{ij} .

Представим себе теперь, что первый игрок работает одновременно со многими агентами, поэтому обрабатывать всю эту

информацию для него накладно. Он хочет выделять стимулирующие выплаты агенту в зависимости от того, будет ли полученная им прибыль высокой, средней, низкой или отрицательной, т.е. он готов обработать $n = 2$ бит информации. Естественно встает вопрос о том, как определить понятие «средняя прибыль» и какую надбавку в этом случае выплатить, чтобы второй игрок сделал выбор, наиболее выгодный для первого?

Для этого, разумеется, нужно еще описать отношение первого игрока к неопределенности.

3. Отношение к неопределенности

Начнем с классического определения максимального гарантированного результата первого игрока в игре $\Gamma = \langle U, V, g, h \rangle$ (см. [4]).

Предположим, что игрок номер один обладает правом первого хода, т.е. он первым выбирает свою стратегию $u \in U$ и этот выбор становится известным второму игроку до того, как он зафиксировал свою стратегию $v \in V$.

В таком случае второй игрок принимает свое решение «в условиях полной определенности», соответственно его поведение становится предсказуемым. Поскольку первый игрок знает интересы партнера, он вправе ожидать, что при фиксированной стратегии u тот выберет управление из множества

$$(1) \quad BR_{\kappa}(u) = \left\{ v \in V : h(u, v) = \max_{w \in V} h(u, w) \right\}.$$

Проблема возникает тогда, когда стратегия u такова, что максимум в предыдущей формуле не достигается. Будем считать, что в таком случае первый игрок предполагает, что возможные выборы второго игрока содержатся в множестве

$$(2) \quad BR_{\kappa}(u) = \left\{ v \in V : h(u, v) \geq \sup_{w \in V} h(u, w) - \kappa \right\}$$

при некотором значении $\kappa > 0$.

Предполагая, что по отношению к оставшейся неопределенности первый игрок осторожен, придем к следующему определению.

Определение 1. Максимальный гарантированный результат первого игрока в игре Γ равен

$$R_{\kappa}(\Gamma) = \sup_{u \in U} \inf_{v \in BR_{\kappa}(u)} g(u, v).$$

Можно исходить из иной логики рассуждений второго игрока. Естественно предположить, что при известной стратегии u все выборы второго игрока разделятся на «рациональные» и нерациональные. Вполне логично полагать, что это разделение происходит по пороговому принципу: существует такое число λ , что управления, для которых выполняется неравенство $h(u, w) \geq \lambda$, рациональны, а все прочие – нет. Вновь считая, что первый игрок осторожен по отношению к рациональным выборам партнера, придем к следующему определению.

Определение 2. Число γ называется гарантированным результатом первого игрока в игре Γ , если существуют такие стратегия $u \in U$ и число λ , что выполняются условия

1°. Существует стратегия $w \in V$, для которой $h(u, w) \geq \lambda$.

2°. Для любой стратегии $v \in V$ либо $g(u, v) \geq \gamma$; либо $h(u, v) < \lambda$.

Точная верхняя грань $R(\Gamma)$ гарантированных результатов называется максимальным гарантированным результатом первого игрока в игре Γ .

Рассуждения, приведшие к этим двум определениям, вполне разумны. Поэтому не удивительно, что «для большинства» игр эти определения эквивалентны в том смысле, что $R(\Gamma) = R_{\kappa}(\Gamma)$. Чтобы сделать это утверждение точным, в [7] было предложено следующее определение.

Определение 3. Игру Γ назовем хорошей, если для любой стратегии $\omega \in U$ найдется такая стратегия $u \in U$, что

$$\inf_{v \in BR_{\kappa}(u)} g(u, v) \geq \inf_{v \in BR_{\kappa}(\omega)} g(\omega, v)$$

и максимум $\max_{v \in V} h(u, v)$ достигается.

В [7] было доказано следующее утверждение.

Лемма 1. Для любой хорошей игры Γ и любого $\kappa > 0$ справедливо равенство $R(\Gamma) = R_{\kappa}(\Gamma)$.

Поскольку в данной статье основной интерес представляет игра $\Gamma_* = \langle U_*, V, g_*, h_* \rangle$, установим следующий факт.

Лемма 2. При сделанных топологических предположениях относительно параметров игры $\Gamma = \langle U, V, g, h \rangle$ соответствующая игра $\Gamma_* = \langle U_*, V, g_*, h_* \rangle$ является хорошей.

Доказательство. Фиксируем произвольную стратегию $(w_*, Q) \in U_*$. Обозначим

$$l(w_*, Q) = \sup_{v \in V} h(w_*(Q(T(v))), v).$$

Если верхняя грань в этой формуле достигается, то все очевидно. Поэтому можно, не ограничивая общности, считать, что это не так. А в таком случае для всех $v \in V$ выполняется неравенство

$$(3) \quad \sup_{v \in V} h(w_*(Q(T(v))), v) < l(w_*, Q).$$

Выберем последовательность v_1, v_2, \dots так, что

$$\lim_{n \rightarrow \infty} h(w_*(Q(T(v_n))), v_n) = \sup_{v \in V} h(w_*(Q(T(v))), v).$$

Для достаточно больших значений n выполняется неравенство $h(w_*(Q(T(v_n))), v_n) > l(w_*, Q) - \kappa$, значит, $v_n \in BR_\kappa(w_*, Q)$. Поэтому без ограничения общности можно считать, что условие $v_n \in BR_\kappa(w_*, Q)$ выполняется для всех n .

Значения $Q(T(v_n))$ принадлежат конечному множеству N , поэтому, не уменьшая общности, можно считать, что они одинаковы (в противном случае можно перейти к подпоследовательности). Обозначим общее значение $Q(T(v_n))$ буквой r и положим $w_*(Q(T(v_n))) = w_*(r) = u_0$. Поскольку множество V компактно, можно считать, что последовательность v_1, v_2, \dots сходится к некоторому элементу v_0 (иначе можно вновь перейти к подпоследовательности).

Тогда в силу непрерывности функции h

$$\begin{aligned} h(u_0, v_0) &= \lim_{n \rightarrow \infty} h(u_0, v_n) = \lim_{n \rightarrow \infty} h(w_*(Q(T(v_n))), v_n) = \\ &= \sup_{v \in V} h(w_*(Q(T(v))), v) = l(w_*, Q). \end{aligned}$$

Положим $u_* = w_*$ и определим отображение P условием

$$P(v) = \begin{cases} r, & \text{если } v = v_0, \\ Q(v), & \text{если } v \neq v_0. \end{cases}$$

Покажем, что верхняя грань $\sup_{v \in V} h(u_*(P(T(v))), v)$ достигается

в единственной точке v_0 . В самом деле, по построению

$$h(u_*(P(T(v_0))), v_0) = h(u_0, v_0) = l(w_*, Q),$$

а для $v \neq v_0$ в силу условия (3) выполняется неравенство

$$h(u_*(P(T(v))), v) = h(w_*(Q(T(v))), v) < l(w_*, Q).$$

Таким образом, $BR_\kappa(u_*, P) = \{v_0\}$ и

$$(4) \quad \inf_{v \in BR_\kappa(u_*, P)} g(u_*(P(T(v))), v) = g(u_*(P(T(v_0))), v_0) = g(u_0, v_0).$$

В силу непрерывности функции g имеем

$$(5) \quad g(u_0, v_0) = \lim_{n \rightarrow \infty} g(u_0, v_n) = \lim_{n \rightarrow \infty} g(w_*(Q(T(v_n))), v_n).$$

А так как для всех n выполняется условие $v_n \in BR_\kappa(w_*, Q)$, справедливы неравенства

$$\inf_{v \in BR_\kappa(w_*, Q)} g(w_*(Q(T(v))), v) \leq g(w_*(Q(T(v_n))), v_n), \quad n = 1, 2, \dots,$$

а, значит,

$$(6) \quad \inf_{v \in BR_\kappa(w_*, Q)} g(w_*(Q(T(v))), v) \leq \lim_{n \rightarrow \infty} g(w_*(Q(T(v_n))), v_n).$$

Сопоставляя условия (4), (5) и (6), получим

$$\inf_{v \in BR_\kappa(u_*, P)} g(u_*(P(T(v))), v) \geq \inf_{v \in BR_\kappa(w_*, Q)} g(w_*(Q(T(v))), v).$$

Лемма доказана.

Поскольку определения 1 и 2 оказываются эквивалентными для игры Γ_* , можно пользоваться тем из них, которое удобнее. Таковым является определение 2.

4. Структура оптимальной стратегии

Вряд ли может вызвать сомнение утверждение о том, что игра Γ проще, чем игра Γ_* . Связь между этими играми задает дополнительную структуру игры Γ_* . Благодаря наличию такой структуры, свойство числа γ быть гарантированным результатом в игре Γ_* может быть выражено в терминах игры Γ . Этому и будет посвящен данный раздел. Как следствие, удастся найти

структуру стратегий, позволяющих гарантированно получить результат γ , если таковая существует.

Соответствующие рассуждения относятся к области чистой логики. Они не сложны, но достаточно длинны. Поэтому удобно перейти с русского языка на язык исчисления предикатов.

Согласно определению 2 число γ является гарантированным результатом в игре Γ^* тогда и только тогда, когда выполняется условие

$$\begin{aligned} & \exists(u_*, P) \in U \exists \lambda \in \mathbf{R} : [\exists w_* \in V_* : h_*((u_*, P), w_*) \geq \lambda] \& \\ & \& [\forall v_* \in V_* g_*((u_*, P), v_*) \geq \gamma \vee h_*((u_*, P), v_*) < \lambda]. \end{aligned}$$

Используя структуру игры Γ^* , эту формулу можно переписать в виде

$$\begin{aligned} & \exists(u_*, P) \in \Phi(N, U) \times \Phi(\Omega, N) \exists \lambda \in \mathbf{R} : \\ & [\exists w \in V : h(u_*(P(T(w))), w) \geq \lambda] \& \\ & \& [\forall v \in V g(u_*(P(T(v))), v) \geq \gamma \vee h(u_*(P(T(v))), v) < \lambda], \end{aligned}$$

или

$$\begin{aligned} & \exists u_* \in \Phi(N, U) \exists P \in \Phi(\Omega, N) \exists \lambda \in \mathbf{R} : \\ & [\exists w \in V : h(u_*(P(T(w))), w) \geq \lambda] \& \\ & \& [\forall v \in V g(u_*(P(T(v))), v) \geq \gamma \vee h(u_*(P(T(v))), v) < \lambda]. \end{aligned}$$

Поменяем порядок кванторов существования

$$\exists \lambda \in \mathbf{R} \exists u_* \in \Phi(N, U) \exists P \in \Phi(\Omega, N) :$$

$$(7) \quad \begin{aligned} & [\exists w \in V : h(u_*(P(T(w))), w) \geq \lambda] \& \\ & \& [\forall v \in V g(u_*(P(T(v))), v) \geq \gamma \vee h(u_*(P(T(v))), v) < \lambda]. \end{aligned}$$

Согласно общей тактике исследования подобных задач (см., например, [7]), на следующем этапе нужно «расцепить» две части условия (7), соответствующие пунктам 1° и 2° определения 2 (в формуле (7) они разделены знаком конъюнкции). В данной задаче это требует некоторых усилий.

Обозначим

$$W(\omega) = \{v \in V : T(v) = \omega\}.$$

Тогда условие (7) можно переписать в виде

$$\begin{aligned} & \exists \lambda \in \mathbf{R} \exists u_* \in \Phi(N, U) \exists P \in \Phi(\Omega, N) : \\ (8) \quad & [\exists \varpi \in \Omega \exists w \in W(\varpi) : h(u_*(P(\varpi)), w) \geq \lambda] \& \\ & \& [\forall \omega \in \Omega \forall v \in W(\omega) g(u_*(P(\omega)), v) \geq \gamma \vee h(u_*(P(\omega)), v) < \lambda]. \end{aligned}$$

Для агрегата $\varpi \in \Omega$, существование которого предусмотрено первой частью условия (8), должна выполняться вторая часть этого условия. Следовательно, условие (8) эквивалентно следующему:

$$\begin{aligned} & \exists \lambda \in \mathbf{R} \exists u_* \in \Phi(N, U) \exists P \in \Phi(\Omega, N) : \\ (9) \quad & \{ \exists \varpi \in \Omega : [\exists w \in W(\varpi) : h(u_*(P(\varpi)), w) \geq \lambda] \& \\ & \& [\forall v \in W(\varpi) g(u_*(P(\varpi)), v) \geq \gamma \vee h(u_*(P(\varpi)), v) < \lambda] \} \& \\ & \& [\forall \omega \in \Omega \forall v \in W(\omega) g(u_*(P(\omega)), v) \geq \gamma \vee h(u_*(P(\omega)), v) < \lambda]. \end{aligned}$$

Теперь заметим, что условие (9) эквивалентно условию $\exists \lambda \in \mathbf{R} \exists u_* \in \Phi(N, U) \exists P \in \Phi(\Omega, N) \exists Q \in \Phi(\Omega, N) :$

$$\begin{aligned} (10) \quad & \{ \exists \varpi \in \Omega : [\exists w \in W(\varpi) : h(u_*(Q(\varpi)), w) \geq \lambda] \& \\ & \& [\forall v \in W(\varpi) g(u_*(Q(\varpi)), v) \geq \gamma \vee h(u_*(Q(\varpi)), v) < \lambda] \} \& \\ & \& [\forall \omega \in \Omega \forall v \in W(\omega) g(u_*(P(\omega)), v) \geq \gamma \vee h(u_*(P(\omega)), v) < \lambda]. \end{aligned}$$

Необходимость условия (10) для выполнения условия (9) очевидна. Докажем достаточность.

Пусть условие (10) выполнено. Фиксируем $\lambda \in \mathbf{R}$, $u_* \in \Phi(N, U)$, $P \in \Phi(\Omega, N)$, $Q \in \Phi(\Omega, N)$, $\varpi \in \Omega$ и $w \in W(\varpi)$ так, что

$$\begin{aligned} & h(u_*(Q(\varpi)), w) \geq \lambda \& \\ (11) \quad & \& [\forall v \in W(\varpi) g(u_*(Q(\varpi)), v) \geq \gamma \vee h(u_*(Q(\varpi)), v) < \lambda] \& \\ & \& [\forall \omega \in \Omega \forall v \in W(\omega) g(u_*(P(\omega)), v) \geq \gamma \vee h(u_*(P(\omega)), v) < \lambda]. \end{aligned}$$

С помощью условия (11) непосредственно проверяется, что для функции

$$P^0(\omega) = \begin{cases} Q(\omega), & \text{если } \omega = \varpi, \\ P(\omega), & \text{если } \omega \neq \varpi, \end{cases}$$

выполняется условие

$$\begin{aligned}
 & h(u_*(P^0(\varpi)), w) \geq \lambda \ \& \\
 & \& \left[\forall v \in W(\varpi) \ g(u_*(P^0(\varpi)), v) \geq \gamma \vee h(u_*(P^0(\varpi)), v) < \lambda \right] \& \\
 & \& \left[\forall \omega \in \Omega \forall v \in W(\omega) \ g(u_*(P^0(\omega)), v) \geq \gamma \vee h(u_*(P^0(\omega)), v) < \lambda \right].
 \end{aligned}$$

Тем более справедливо условие (9).

Формулу (10) можно переписать в эквивалентном виде $\exists \lambda \in \mathbf{R} \exists u_* \in \Phi(N, U)$:

$$\begin{aligned}
 & \left\{ \exists Q \in \Phi(\Omega, N) \exists \varpi \in \Omega : \left[\exists w \in W(\varpi) : h(u_*(Q(\varpi)), w) \geq \lambda \right] \& \right. \\
 (12) \ & \left. \& \left[\forall v \in W(\varpi) \ g(u_*(Q(\varpi)), v) \geq \gamma \vee h(u_*(Q(\varpi)), v) < \lambda \right] \right\} \& \\
 & \& \exists P \in \Phi(\Omega, N) \left[\forall \omega \in \Omega \forall v \in W(\omega) \ g(u_*(P(\omega)), v) \geq \gamma \vee \right. \\
 & \left. \vee h(u_*(P(\omega)), v) < \lambda \right].
 \end{aligned}$$

Условие

$$\begin{aligned}
 & \exists Q \in \Phi(\Omega, N) \exists \varpi \in \Omega : \left[\exists w \in W(\varpi) : h(u_*(Q(\varpi)), w) \geq \lambda \right] \& \\
 & \& \left[\forall v \in W(\varpi) \ g(u_*(Q(\varpi)), v) \geq \gamma \vee h(u_*(Q(\varpi)), v) < \lambda \right],
 \end{aligned}$$

очевидно, равносильно условию

$$\begin{aligned}
 & \exists b \in N \exists \varpi \in \Omega : \left[\exists w \in W(\varpi) : h(u_*(b), w) \geq \lambda \right] \& \\
 & \& \left[\forall v \in W(\varpi) \ g(u_*(b), v) \geq \gamma \vee h(u_*(b), v) < \lambda \right]
 \end{aligned}$$

Преобразуем формулу

$$\begin{aligned}
 & \exists P \in \Phi(\Omega, N) : \forall \omega \in \Omega \forall v \in W(\omega) \left[g(u_*(P(T(v))), v) \geq \gamma \vee \right. \\
 & \left. \vee h(u_*(P(T(v))), v) < \lambda \right].
 \end{aligned}$$

Можно поменять местами кванторы существования и общности

$$\forall \omega \in \Omega \exists r \in N : \forall v \in W(\omega) \left[g(u_*(r), v) \geq \gamma \vee h(u_*(r), v) < \lambda \right].$$

Таким образом, условие (12) равносильно следующему:

$$\begin{aligned}
 & \exists \lambda \in \mathbf{R} \exists u_* \in \Phi(N, U) : \\
 & \exists b \in N \exists \varpi \in \Omega : \left[\exists w \in W(\varpi) : h(u_*(b), w) \geq \lambda \right] \& \\
 & \& \left[\forall v \in W(\varpi) \ g(u_*(b), v) \geq \gamma \vee h(u_*(b), v) < \lambda \right] \& \\
 & \& \forall \omega \in \Omega \exists r \in N : \forall v \in W(\omega) \left[g(u_*(r), v) \geq \gamma \vee h(u_*(r), v) < \lambda \right].
 \end{aligned}$$

В этой формуле осталось одно функциональное пространство $\Phi(N, U)$. Чтобы избавиться от него, воспользуемся тем, что

множество N содержит $m = 2^n$ элементов. Поэтому последнее условие равносильно условию

$$(13) \quad \begin{aligned} & \exists \lambda \in \mathbf{R} \exists (u_0, \dots, u_{m-1}) \in U^m : \\ & \exists b \in N \exists \varpi \in \Omega : [\exists w \in W(\varpi) : h(u_b, w) \geq \lambda] \& \\ & \& [\forall v \in W(\varpi) g(u_b, v) \geq \gamma \vee h(u_b, v) < \lambda] \& \\ & \& \forall \omega \in \Omega \exists r \in N : \forall v \in W(\omega) [g(u_r, v) \geq \gamma \vee h(u_r, v) < \lambda], \end{aligned}$$

где, как обычно, U^m означает декартову степень множества U .

Таким образом, установлен следующий факт.

Теорема 1. Для того чтобы число γ было гарантированным результатом первого игрока в игре Γ_* , необходимо и достаточно выполнение условия (13).

Если число γ является гарантированным результатом, то формула (13) позволяет построить стратегию (u_*, P) , гарантирующую получение такого результата.

Фиксируем число λ и набор управлений $(u_1, \dots, u_m) \in U^m$ так, что

$$\begin{aligned} & \exists b \in N \exists \varpi \in \Omega : [\exists w \in W(\varpi) : h(u_b, w) \geq \lambda] \& \\ & \& [\forall v \in W(\varpi) g(u_b, v) \geq \gamma \vee h(u_b, v) < \lambda] \& \\ & \& \forall \omega \in \Omega \exists r \in N : \forall v \in W(\omega) [g(u_r, v) \geq \gamma \vee h(u_r, v) < \lambda], \end{aligned}$$

Прежде всего заметим, что в силу симметрии задачи можно считать, что $u_*(r) = u_r$ для всех $r \in N$ (подробнее об этом написано в [6]).

Далее позаботимся о том, чтобы выполнялся первый пункт определения 2. Для этого фиксируем сообщение $b \in N$ и агрегат $\varpi \in \Omega$ так, чтобы выполнялось условие

$$\begin{aligned} & [\exists w \in W(\varpi) : h(u_b, w) \geq \lambda] \& \\ & \& [\forall v \in W(\varpi) g(u_b, v) \geq \gamma \vee h(u_b, v) < \lambda], \end{aligned}$$

и положим $P(\varpi) = b$.

Остается выполнить условие 2° определения 2. Для ω , отличного от уже фиксированного агрегата ϖ , выберем произвольное $r \in N$, для которого выполняется условие

$$\forall v \in W(\omega) [g(u_r, v) \geq \gamma \vee h(u_r, v) < \lambda],$$

и положим $P(\omega) = r$.

Тем самым функция P будет определена для всех $\omega \in \Omega$. Непосредственно проверяется, что так построенная стратегия (u_*, P) гарантирует первому игроку получение выигрыша γ .

Скажем несколько слов об интерпретации построенной стратегии. Наличие в приведенных выше конструкциях неравенств вида $h(u_r, v) < \lambda$ позволяет говорить об использовании первым игроком стратегии наказания партнера. Впрочем, в данном случае говорить об этом можно с известной натяжкой, что демонстрирует следующий пример.

Пример. Пусть выигрыши игроков задаются матрицами

$$\begin{pmatrix} (2,2) & (0,0) & (1,1) & (1,1) \\ (1,1) & (1,1) & (0,0) & (2,2) \end{pmatrix}$$

(как обычно, первый игрок выбирает строку, второй – столбец, первое число в соответствующей паре означает выигрыш первого игрока, а второе – выигрыш его партнера). Положим $\Omega = \{0, 1\}$, $T(1) = T(2) = 0$, $T(3) = T(4) = 1$. Будем считать, что первый игрок способен обработать один бит информации.

В этих условиях максимальный гарантированный результат первого игрока равен 2. В самом деле, положим $P(\omega) = \omega$, $u_*(r) = r$. В этом случае можно взять $\lambda = 2$, и условия определения 2 будут выполнены.

При таком выборе стратегии для $v = 2$ и $v = 3$ соответствующее управление $u_*(P(v))$ можно рассматривать как наказание второго игрока. Но при $v = 1$ и $v = 4$ то же управление стоит трактовать скорее как «пряник».

В данном случае, чтобы не увеличивать размеры матрицы, множество Ω было выбрано состоящим из двух элементов, поэтому ограничение на объем передаваемой информации оказалось несущественным. Но понятно, что добавление такого рода ограничения может лишь усугубить обсуждаемый эффект.

5. Вычисления

Кванторы « \exists » в формуле (13) гарантируют существование элементов, из которых строится оптимальная стратегия, но не

дают инструмента для их поиска. Эти элементы могут быть найдены как решения некоторых оптимизационных задач. Сведение условия (13) к такого рода задачам осуществляется в целом так же, как и в других аналогичных случаях. Но в задаче с ограничением на содержание используемой информации появляются дополнительные трудности как содержательного, так и чисто технического характера. Поэтому проведем соответствующие рассуждения достаточно подробно.

Выше было сделано предположение о том, что множество Ω , наделенное фактортопологией, является T_1 -пространством. Следовательно, при каждом $\omega \in \Omega$ одноточечное множество $\{\omega\}$ замкнуто (см. [9]). Значит, замкнут, а, следовательно, компактен его прообраз $W(\omega)$ при непрерывном отображении T .

Поэтому при любых $u \in U$ и $\omega \in \Omega$ достигается максимум

$$\bar{h}(u, \omega) = \max_{v \in W(\omega)} h(u, v).$$

Тогда множество

$$\Sigma(u, \omega) = \{v \in V : h(u, v) = \bar{h}(u, \omega)\}$$

не пусто и замкнуто (как прообраз одного числа при непрерывном отображении), а потому компактно. Значит, достигается минимум

$$\bar{g}(u, \omega) = \min_{v \in \Sigma(u, \omega)} g(u, v).$$

Перепишем условие (13) в эквивалентном виде

$$(14) \quad \begin{aligned} & \exists(u_0, \dots, u_{m-1}) \in U^m \exists \lambda \in \mathbf{R} : \\ & \exists b \in N \exists \varpi \in \Omega : [\exists w \in W(\varpi) : h(u_b, w) \geq \lambda] \& \\ & \& [\forall v \in W(\varpi) g(u_b, v) \geq \gamma \vee h(u_b, v) < \lambda] \& \\ & \& [\forall \omega \in \Omega \exists r \in N : \forall v \in W(\omega) [g(u_r, v) \geq \gamma \vee h(u_r, v) < \lambda]]. \end{aligned}$$

Для начала зафиксируем произвольный набор $(u_0, \dots, u_{m-1}) \in U^m$.

Обратим внимание на последнюю часть условия (14). Для достаточно больших значений λ выполняется уже условие

$$\forall \omega \in \Omega \exists r \in N : \forall v \in W(\omega) h(u_r, v) < \lambda.$$

Достаточно взять λ большим, чем

$$L(u_0, \dots, u_{m-1}) = \sup_{\omega \in \Omega} \min_{r \in N} \max_{v \in W(\omega)} h(u_r, v) = \sup_{\omega \in \Omega} \min_{r \in N} \bar{h}(u_r, \omega).$$

При таких значениях λ ограничение на возможное значение гарантированного результата γ задает только первая часть условия (14):

$$\exists b \in N \exists \varpi \in \Omega: [\exists w \in W(\varpi): h(u_b, w) \geq \lambda] \& \\ \& [\forall v \in W(\varpi) g(u_b, v) \geq \gamma \vee h(u_b, v) < \lambda].$$

Его можно переписать в виде

$$\exists b \in N \exists \varpi \in \Omega: \max_{w \in W(\varpi)} h(u_b, w) \geq \lambda \& \\ \& [\forall v \in W(\varpi) g(u_b, v) \geq \gamma \vee h(u_b, v) < \lambda],$$

или

$$(15) \quad \exists b \in N \exists \varpi \in \Omega: \bar{h}(u_b, \varpi) \geq \lambda \& \\ \& [\forall v \in W(\varpi) g(u_b, v) \geq \gamma \vee h(u_b, v) < \lambda].$$

А теперь заметим, что если выбрать значение λ так, что $\lambda = \bar{h}(u_r, \varpi)$ (увеличивать значение λ можно без опаски), то первая часть условия (15) будет выполнена, а вторую часть можно будет записать в более простом виде:

$$\forall v \in \Sigma(u_b, \varpi) g(u_b, v) \geq \gamma.$$

Это условие можно переписать в виде

$$\min_{v \in \Sigma(u_b, \varpi)} g(u_b, v) \geq \gamma,$$

или $\bar{g}(u_b, \varpi) \geq \gamma$.

Напомним, что все эти рассуждения справедливы при $\lambda > L(u_0, \dots, u_{m-1})$, т.е. когда b и ϖ выбраны из множества

$$D(u_0, \dots, u_{m-1}) = \{(b, \varpi) \in N \times \Omega: \bar{h}(u_b, \varpi) > L(u_0, \dots, u_{m-1})\}.$$

Таким образом, при $\lambda > L(u_0, \dots, u_{m-1})$ величина γ не может превосходить значения

$$K(u_0, \dots, u_{m-1}) = \sup_{(b, \varpi) \in D(u_0, \dots, u_{m-1})} \bar{g}(u_b, \varpi),$$

а при всяком $\gamma < K(u_0, \dots, u_{m-1})$ условие (14) может быть выполнено.

Рассмотрим теперь случай $\lambda = L(u_0, \dots, u_{m-1})$. При таком выборе λ для любого ω из множества

$$E(u_0, \dots, u_{m-1}) = \left\{ \omega \in \Omega : \min_{r \in N} \bar{h}(u_r, \omega) = \sup_{\omega \in \Omega} \min_{r \in N} \bar{h}(u_r, \omega) \right\},$$

любого $r \in N$ и любого $v \in \Sigma(u_r, \omega)$ выполняется неравенство $h(u_r, v)$. Поэтому для выполнения условия (14) необходима справедливость условия

$$\forall \omega \in E(u_0, \dots, u_{m-1}) \exists r \in N \forall v \in \Sigma(u_r, \omega) : g(u_r, v) \geq \gamma.$$

Поэтому γ не может превосходить значения

$$\begin{aligned} M(u_0, \dots, u_{m-1}) &= \inf_{\omega \in E(u_0, \dots, u_{m-1})} \max_{r \in N} \min_{v \in \Sigma(u_r, \omega)} g(u_r, v) = \\ &= \inf_{\omega \in E(u_0, \dots, u_{m-1})} \max_{r \in N} \bar{g}(u_r, \omega). \end{aligned}$$

Обратно, если $M(u_0, \dots, u_{m-1}) > K(u_0, \dots, u_{m-1})$, то для любого $\gamma < M(u_0, \dots, u_{m-1})$ условие (14) может быть выполнено.

В самом деле, первая часть этого условия выполняется при $\omega \in E(u_0, \dots, u_{m-1})$, любом $b \in N$ и $v \in \Sigma(u_b, \omega)$.

Если $\omega \notin E(u_0, \dots, u_{m-1})$, то можно выбрать $r \in N$ так, что при любом $v \in \Sigma(u_r, \omega)$ будет справедливо неравенство $h(u_r, v) < \lambda$, и тем более будет выполняться вторая часть условия (14).

Остается разобраться со случаем $\omega \in E(u_0, \dots, u_{m-1})$. В таком случае для любого r , удовлетворяющего условию $\bar{g}(u_r, \omega) = \max_{b \in N} \bar{g}(u_b, \omega)$, выполняется неравенство

$$\bar{h}(u_r, \omega) \leq \lambda = L(u_0, \dots, u_{m-1}).$$

В самом деле, в противном случае точка (u_r, ω) принадлежала бы множеству $D(u_0, \dots, u_{m-1})$ и, как следствие, было бы $\bar{g}(u_r, \omega) \leq K(u_0, \dots, u_{m-1})$. Но с другой стороны в силу выбора r имеем

$$\begin{aligned} \bar{g}(u_r, \omega) &= \max_{b \in N} \bar{g}(u_b, \omega) \geq \inf_{\omega \in E(u_0, \dots, u_{m-1})} \max_{b \in N} \bar{g}(u_b, \omega) = \\ &= M(u_0, \dots, u_{m-1}) > K(u_0, \dots, u_{m-1}). \end{aligned}$$

Получено противоречие.

А тогда становится ясно, что при выборе r , удовлетворяющего условию $\bar{g}(u_r, \omega) = \max_{b \in N} \bar{g}(u_b, \omega)$ для $v \in \Sigma(u_r, u)$ имеет место неравенство $g(u_r, v) \geq M(u_0, \dots, u_{m-1}) > \gamma$, а для $v \notin \Sigma(u_r, u)$ справедливо неравенство $h(u_r, v) < L(u_0, \dots, u_{m-1}) = \lambda$, т.е. вторая часть условия (14) выполнена.

Рассмотрение случая $\lambda < L(u_0, \dots, u_{m-1})$ не дает ничего нового. Действительно, при таком выборе λ при любом $\omega \in E(u_0, \dots, u_{m-1})$, любом $r \in N$ и $v \in \Sigma(u_r, \omega)$ справедливо неравенство $h(u_r, v) > \lambda$. Поэтому для выполнения условия (14), уж во всяком случае, необходимо, чтобы

$$\gamma \leq \max\{K(u_0, \dots, u_{m-1}), M(u_0, \dots, u_{m-1})\}.$$

Таким образом, при фиксированном наборе $(u_1, \dots, u_m) \in U^m$ для выполнения условия (14) необходимо, чтобы

$$\gamma \leq \max\{K(u_0, \dots, u_{m-1}), M(u_0, \dots, u_{m-1})\},$$

и достаточно, чтобы

$$\gamma < \max\{K(u_0, \dots, u_{m-1}), M(u_0, \dots, u_{m-1})\}.$$

Значит, для того чтобы число γ было гарантированным результатом, необходимо, чтобы

$$\gamma \leq \sup_{(u_0, \dots, u_{m-1}) \in U^m} \max\{K(u_0, \dots, u_{m-1}), M(u_0, \dots, u_{m-1})\},$$

и достаточно, чтобы

$$\gamma < \sup_{(u_0, \dots, u_{m-1}) \in U^m} \max\{K(u_0, \dots, u_{m-1}), M(u_0, \dots, u_{m-1})\}.$$

Следовательно, максимальный гарантированный результат равен $\sup_{(u_0, \dots, u_{m-1}) \in U^m} \max\{K(u_0, \dots, u_{m-1}), M(u_0, \dots, u_{m-1})\}$.

Замечание. Дабы не отвлекаться от главного, в рассуждениях данного раздела были опущены доказательства некоторых технических фактов. Все они доказываются стандартными топологическими рассуждениями и потому не представляют особого интереса. Пожалуй, наиболее существенным и наименее очевидным из них является непустота множества $E(u_0, \dots, u_{m-1})$. Доказательство этого факта дословно повторяет доказательство леммы 1 из [1].

6. Модельный пример

Продemonстрируем, как работают полученные результат при исследовании простой линейной модели.

Пусть $U = [0, u^+]$, $V = [0, 1] \times [0, 1]$, $g(u, v) = pv - u$, $h(u, v) = qv + u$, $T(v) = v^1 + v^2$, где $v = (v^1, v^2)$, $p = (p^1, p^2)$, $q = (q^1, q^2)$, а $pv = p^1v^1 + p^2v^2$ и $qv = q^1v^1 + q^2v^2$ – скалярные произведения. Предположим, что первому игроку доступен один бит информации, т.е. $N = \{0, 1\}$.

С известной долей условности пример может быть интерпретирован следующим образом. Второй игрок производит продукт и продает его первому. Производство осуществляется на двух технологических линиях и ограничено их мощностями. Объемы производства на этих линиях – v^1 и v^2 соответственно. Продукция, произведенная на разных линиях, различается себестоимостью и качеством, скажем, сроком службы. На момент расчетов первый игрок не может различать сорта продукции и покупает ее по единой фиксированной цене. Но он может осуществлять некие дополнительные выплаты u . Полезность купленной продукции для первого игрока линейно зависит от объемов v^1 и v^2 и дополнительных выплат u . Цель второго игрока – максимизация прибыли, а q^1 и q^2 – разности между ценой продукции и себестоимости ее производства.

Будем считать, что $q^1 > 0$, $q^2 < 0$, $p^1 > p^2$.

По сути, здесь предполагается, что более высокому качеству соответствует более высокая себестоимость.

Воспользуемся теоремой 1 для вычисления максимального гарантированного результата в соответствующей игре Γ^* .

Из второй и третьей строк формулы (13) видно, что λ – максимальное значение функции h на множестве

$$T^{-1}(\varpi) = \{(v^1, v^2) \in V: v^1 + v^2 = \varpi\},$$

а w – точка максимума функции h на этом множестве. Поэтому $v^1 = \varpi$, $v^2 = 0$, если $\varpi \leq 1$, и $v^1 = 1$, $v^2 = \varpi - 1$ в противном случае. Соответственно, $\lambda = q^1\varpi + u_b$ при $\varpi \leq 1$, и $\lambda = q^1 + q^2(\varpi - 1) + u_b$ при $\varpi > 1$. Из тех же двух строк формулы (13) видно, что в найденной точке максимума значение функции g не может быть

меньше γ . Значит, $\gamma \geq p^1 \varpi - u_b$ при $\varpi \leq 1$, и $\gamma \geq p^1 + p^2(\varpi - 1) - u_b$ при $\varpi > 1$.

Положим $h^+(u, \omega) = q^1 \omega + u$ при $\omega \leq 1$, а если $\omega > 1$, то $h^+(u, \omega) = q^1 + q^2(\omega - 1) + u$. Пусть $g^+(u, \omega) = p^1 \omega - u$ при $\omega \leq 1$, и $g^+(u, \omega) = p^1 + p^2(\omega - 1) - u$ при $\omega > 1$. Соображения из предыдущего абзаца применимы и при анализе условия из последней строки формулы (13), поэтому эта формула может быть переписана в виде

$$\begin{aligned} & \exists \lambda \in \mathbf{R} \exists (u_0, \dots, u_{m-1}) \in U^m : \\ & \exists b \in N \exists \varpi \in \Omega : h^+(u_b, \varpi) \geq \lambda \ \& \ g^+(u_b, \varpi) \geq \gamma \ \& \\ & \ \& \ \forall \omega \in \Omega \exists r \in N : [g^+(u_r, \omega) \geq \gamma \vee h^+(u_r, \omega) < \lambda]. \end{aligned}$$

Теперь можно воспользоваться одной идеей общего характера. Так как условие из последней строки этой формулы должно выполняться, в частности, при $\omega = \varpi$, то она эквивалентна условию

$$(16) \quad \begin{aligned} & \exists \lambda \in \mathbf{R} \exists (u_0, \dots, u_{m-1}) \in U^m : \exists b \in N \exists \varpi \in \Omega : h^+(u_b, \varpi) \geq \lambda \ \& \\ & \ \& \ \forall \omega \in \Omega \exists r \in N : [g^+(u_r, \omega) \geq \gamma \vee h^+(u_r, \omega) < \lambda]. \end{aligned}$$

Далее заметим, что при фиксированном u функции $h^+(u, \omega)$ и $g^+(u, \omega)$ монотонно возрастают по ω , когда $\omega \in [0, 1]$. Поэтому если условие $\exists r \in N : [g^+(u_r, \omega) \geq \gamma \vee h^+(u_r, \omega) < \lambda]$ выполняется при $\omega = 1$, то оно выполняется и при всех $\omega < 1$. Поэтому в условии (16) можно, не ограничивая общности считать, что $\omega \in [1, 2]$. По аналогичной причине можно считать, что $\varpi \in [1, 2]$. А тогда формула (16) переписывается в виде

$$\begin{aligned} & \exists \lambda \exists (u_0, u_1) \in U^2 : \exists b \in N \exists \varpi \in [1, 2] : q^1 + q^2(\varpi - 1) + u_b \geq \lambda \ \& \\ & \ \& \ \forall \omega \in [1, 2] \exists r \in N : \\ & \quad [p^1 + p^2(\omega - 1) - u_r \geq \gamma \vee q^1 + q^2(\omega - 1) + u_r < \lambda]. \end{aligned}$$

С учетом того, что функции g^+ и h^+ зависят от u монотонно, это условие переписывается в виде

$$\begin{aligned} & \exists \lambda \exists (u_0, u_1) \in U^2 : \exists \varpi \in [1, 2] : q^1 + q^2(\varpi - 1) + \max\{u_0, u_1\} \geq \lambda \ \& \\ & \ \& \forall \omega \in [1, 2] [p^1 + p^2(\omega - 1) - \min\{u_0, u_1\} \geq \gamma \vee \\ & \ \vee q^1 + q^2(\omega - 1) + \min\{u_0, u_1\} < \lambda]. \end{aligned}$$

В первую часть этого условия меньшее из чисел u_0 и u_1 не входит, а вторую выполнить тем легче, чем меньше это число. Поэтому можно, не ограничивая общности, считать, что меньшее из чисел u_0 и u_1 равно нулю. В силу симметрии задачи можно считать, что $u_0 = 0$ и тогда получим условие

$$\begin{aligned} & \exists \lambda \exists u_1 \in U : \exists \varpi \in [1, 2] : q^1 + q^2(\varpi - 1) + u_1 \geq \lambda \ \& \\ & \ \& \forall \omega \in [1, 2] [p^1 + p^2(\omega - 1) \geq \gamma \vee q^1 + q^2(\omega - 1) < \lambda]. \end{aligned}$$

В последней строке этой формулы, по сути, написано, что отрезок $[1, 2]$ покрывается двумя множествами:

$$\left\{ \omega \in [1, 2] : p^1 + p^2(\omega - 1) \geq \gamma \right\} = \left\{ \omega \in [1, 2] : \omega \geq \frac{\gamma - p^1}{p^2} + 1 \right\}$$

и

$$\left\{ \omega \in [1, 2] : q^1 + q^2(\omega - 1) < \lambda \right\} = \left\{ \omega \in [1, 2] : \omega > \frac{\lambda - q^1}{q^2} + 1 \right\},$$

(напомним, что $q^2 < 0$). Это равносильно тому, что точка $\omega = 1$ принадлежит одному из этих множеств, т.е. выполняется одно из условий $\gamma \leq p^1$ или $\lambda > q^1$. Как станет ясно из следующего абзаца, первый случай можно не рассматривать. Во втором случае для определения числа γ имеем условия $h^+(u_b, \varpi) > q^1$ и $g(u_b, \varpi) \geq \gamma$, или $q^2(\varpi - 1) + u_b > 0$ и $p^1 + p^2(\varpi - 1) - u_b \geq \gamma$. Отсюда $p^1 + (p^2 + q^2)(\varpi - 1) \geq \gamma$. Далее придется разобрать несколько случаев.

Если $p^2 + q^2 \leq 0$ (т.е. себестоимость второго вида продукции слишком высока), то выгодно взять $\varpi = 1$. В этом случае информация о действиях партнера первому игроку не нужна, и его оптимальная стратегия состоит в отсутствии дополнительных выплат, т.е. $u_*(\omega) \equiv 0$ (а отображение P может быть любым). Максимальный гарантированный результат первого игрока в этом случае достигается и равен p^1 . Оптимальным ответом вто-

рого игрока на такую стратегию первого является выбор $v = (1, 0)$.

Если $p^2 + q^2 > 0$, то чем больше ϖ , тем больше может быть сделано значение γ . Но само значение ϖ не может быть сделано очень большим, поскольку оно стеснено двумя ограничениями: $q^2(\varpi - 1) + u^+ > 0$ и $\varpi \leq 2$.

Пусть более сильным является первое ограничение. Тогда максимальный гарантированный результат первого игрока равен $p^1 - \left(\frac{p^2}{q^2} + 1 \right) u^+$. Он не достигается, но результат, сколь угодно близкий к такому, может быть получен применением следующей стратегии. Выберем ϖ достаточно близким к $1 - \frac{u^+}{q^2}$.

Положим $u_*(0) = 0$, $u_*(1) = u^+$. Пусть $P(\omega) = 1$ при $\omega \geq \varpi$ и $P(\omega) = 0$ в остальных случаях. Рациональным ответом второго игрока на такую стратегию будет выбор $v^1 = 1$ и $v^2 = \varpi - 1$, и результат первого игрока будет искомым.

Если же критическим является ограничение $\varpi \leq 2$, то первый игрок гарантированно может получить результат сколь угодно близкий к $p^1 + p^2 + q^2$. Оптимальная стратегия выглядит примерно так же, как описано в предыдущем абзаце.

Большая часть шагов в проведенном рассуждении может быть интерпретирована в содержательных терминах. Более того, если использовать содержательные соображения, выкладки в данном случае можно существенно сократить. Это не сделано исключительно из-за желания продемонстрировать возможности формальной техники.

Разумеется, этот пример может быть исследован и с помощью результатов раздела 5. В данном случае это сделать даже немного проще. Но эта техника более стандартна, поэтому здесь на ней не останавливаться не будем.

7. Заключение

Полученные в данной статье результаты можно рассматривать как начало исследований в двух направлениях.

Первое связано с введением внешней неопределенности. Допустим, имеется некий фактор, который не контролируется рассматриваемыми в модели игроками, но может существенно влиять на их результаты. Может оказаться, что первому игроку не доступна информация об этом неопределенном факторе и действиях партнера, но он может получить информацию о некой «смеси». Скажем, производство продукции может зависеть от напряженности действий второго игрока и погодных условий. Первый игрок не может наблюдать ни того, ни другого, но объем произведенной продукции он может узнать. Здесь имеется большой спектр задач, отличающихся характером зависимости выигрышей игроков от неопределенного фактора (в приведенном примере: зависит ли выигрыш первого игрока только от объема произведенной продукции или по отдельности от действий партнера и погоды), отношения игроков к неопределенности и т.п.

Без учета ограничений на объем передаваемой информации многие такие модели исследовались в рамках теории активных систем и теории контрактов. Но использованный выше метод открывает и некие новые перспективы. Практика показывает, что во многих случаях этот метод позволяет избавиться от «лишних» ограничений и получить результат в более общем виде. Кроме того, в ряде случаев метод позволяет понять, что рассматриваемая задача в общем виде в принципе не поддается решению, поскольку какие-то кванторы просто нельзя «переставить». Это позволяет надеяться, что удастся выяснить вид дополнительных предположений, при которых эта задача все-таки допускает решение.

Другое направление связано с динамикой. Можно считать, что в рассмотренной выше модели решения принимались в следующем порядке. Сначала второй игрок выбирает информационный агрегат $\omega \in \Omega$, затем первый игрок выбирает свое управление $u \in U$, и, наконец, второй игрок выбирает свое управле-

ние ν из множества $W(\omega)$. Таким образом, по сути, имеется «полугорашаговая» игра. Естественно желание рассмотреть ее многошаговые варианты. Здесь тоже имеется целый спектр разумных постановок (см. по этому поводу [8]).

Но на сегодняшний день более существенным представляется другое. Обсуждавшийся двумя абзацами выше метод хорошо зарекомендовал себя при исследовании статических моделей. Но пока есть проблемы с тем, чтобы применить его для решения динамических задач. Тот факт, что с исследованием рассмотренной в данной статье модели удалось справиться, внушает определенный оптимизм.

Литература

1. АЛИЕВ В.С., КОНОНЕНКО А.Ф. *Некоторые вопросы принятия решений в играх двух лиц при агрегированной информации* // Ж. вычисл. матем. и матем. физ. – 1997. – Т. 37. – №10. – С. 1163–1173.
2. БУРКОВ В.Н., НОВИКОВ Д.А. *Теория активных систем: состояние и перспективы*. – М.: Синтег, 1999. – 128 с.
3. ВОРОНИН А.А., ГУБКО М.В., МИШИН С.П., НОВИКОВ Д.А. *Математические модели организаций*. – М.: ЛЕНАНД, 2008. – 360 с.
4. ГЕРМЕЙЕР Ю.Б. *Игры с противоположными интересами*. – М.: Наука, 1976. – 327 с.
5. ГОРЕЛОВ М.А. *Максимальный гарантированный результат при ограниченном объеме передаваемой информации* // Автоматика и телемеханика. – 2011. – №3. – С. 124–144.
6. ГОРЕЛОВ М.А. *Игры с обменом недостоверной информацией* // Управление большими системами. – 2013. – Вып. 41. – С. 5–27.
7. ГОРЕЛОВ М.А. *Максимальный гарантированный результат в иерархических играх* // Управление большими системами. – 2017. – Вып. 67. – С. 4–31.
8. ГОРЕЛОВ М.А., КОНОНЕНКО А.Ф. *Динамические модели конфликтов. III. Иерархические игры* // Автоматика и телемеханика. – 2015. – №2. – С. 89–106.

9. ЭНГЕЛЬКИНГ Р. *Общая топология*. – М.: Наука, 1986. – 752 с.
10. BOLTON P., DEWATRIPONT M. *Contract Theory*. – Mass.: MIT Press, 2005. – 740 p.
11. GILPIN A., SANDHOLM T. *Lossless abstraction of imperfect information games* // J. of the ACM. – 2007. – Vol. 54, No. 5, 9. – P. 1–32.
12. LAFFONT J.-J., MARTIMORT D. *The Theory of Incentives: The Principal–Agent Model*. – Princeton: Princeton University Press, 2002. – 440 p.
13. WULF DE M., DOYEN L., RASKIN JF *A Lattice Theory for Solving Games of Imperfect Information*. // Hespanha J.P., Tiwari A. (eds) *Hybrid Systems: Computation and Control*. – HSCC 2006. – Lecture Notes in Computer Science. – Vol. 3927. – Berlin, Heidelberg: Springer, 2006. – P. 153–168.

HIERARCHICAL GAMES WITH RESTRICTIONS ON CONTENT AND VOLUME OF INFORMATION TRANSFERED

Mikhail Gorelov, Dorodnicyn Computing Centre, FRC CSC RAS, Moscow, Cand.Sc. (griefer@ccas.ru).

Abstract: Two players hierarchical game is investigated. The top-level player is supposed to have right to first move. It is supposed that he has access to some information about his partner's choice. But two types of restrictions on such information are taken into consideration. From one hand there are such pairs of bottom-level player choices that elements of pair are not distinguished one from another from the top level-player's point of view. From other hand the volume of information on the bottom-level player's choice which the top-level player can handle is restricted. The combinatorial approach (in terms of A.N. Kolmogorov) is used for measuring of the amount of information. Top level player is supposed to have the right of choice of the "sense" of information obtained (in the framework of restrictions of the first type). It is assumed that the top-level player knows the opportunity and goals of his partner and he can expect to rational behavior of his partner. In such assumptions the problem of calculating of the top level player's maximal guaranteed result is a problem of calculating a maxima on complex

functional spaces. In the article the problem is reduced to calculation of multiple maximin on “finite-dimensional” spaces. Two approaches to computing of this result are proposed. A structure of top level player’s optimal strategy is estimated. In particular the optimal semantics of information which top layer player handle is estimated. An illustrative example is provided which demonstrates the possibilities of use of methods proposed.

Keywords: hierarchical games, maximal guaranteed result, information.

УДК 519.865 + 519.95

ББК 22.165

DOI: <https://doi.org/10.25728/ubs.2019.77.2>

*Статья представлена к публикации
членом редакционной коллегии О.П. Кузнецовым.*

Поступила в редакцию 18.05.2018.

Опубликована 31.01.2019.

АЛГОРИТМ УПРАВЛЕНИЯ В УСЛОВИИ НАСЫЩЕНИЯ СИГНАЛА УПРАВЛЕНИЯ И ЕГО ПРОИЗВОДНЫХ¹

Фуртат И. Б.², Гущин П. А.³

(Институт проблем машиноведения РАН, Университет
ИТМО, Санкт-Петербург)

Приведен синтез системы управления объектами в условиях параметрической неопределенности, внешних ограниченных возмущений, насыщения регулируемой переменной и ее производных по времени. Такие задачи типичны при управлении различного рода техническими системами. Игнорирование неопределенностей и ограничений может приводить к невыполнению поставленной цели или потери устойчивости системы управления. Например, катастрофические последствия могут возникать в системе управления полетом летательного аппарата из-за насыщения регулируемой переменной и ее производной. Другими примерами являются управление компрессорами реактивных двигателей и исследование поведения объекта управления с медленным приводом. Для решения задачи в условиях неопределенностей и ограничений на регулирующий сигнал предложен новый способ формирования закона управления, который позволяет обеспечить нахождение регулирующего сигнала и его производных в заданных множествах. Использован принцип компенсации возмущений, который позволил получить информацию о параметрической неопределенности и внешних возмущениях в виде некоторой функции. Для компенсации данных неопределенностей формировался сигнал управления, значение которого противоположно значению возмущений. Таким образом для регулирования в условиях ограничений на сигнал управления требуется наложить соответствующие ограничения на функцию, содержащую информацию о данных неопределенностях. Получены условия на параметры объекта, внешнее возмущение, эталонную модель и регулятор при выполнении которых система управления будет работоспособной. Приведены результаты моделирования, иллюстрирующие работоспособность приведенной схемы.

Ключевые слова: управление в условии неопределенности, насыщение сигнала управления и его производных, компенсация возмущений.

¹ Результаты раздела 3 получены при поддержке гранта РФФИ №17-08-01266.

² Игорь Борисович Фуртат, д.т.н., доцент (cainenash@mail.ru).

³ Павел Александрович Гущин, к.т.н. (guschin.p@mail.ru).

1. Введение

Насыщение в регулирующих сигналах – одна из типичных проблем управления техническими системами. В ряде задач дополнительно насыщению (*magnitude saturation*) в качестве регулирующей переменной рассматривается насыщение ее производной по времени (*rate saturation*). Например, в [10, 22, 23] изучались катастрофические последствия, возникающие в системе управления полетом летательного аппарата из-за насыщения регулируемой переменной и ее производной. Другими примерами являются управление компрессорами реактивных двигателей [15, 18, 26] и исследование поведения объекта управления с медленным приводом [11].

В настоящее время для решения задачи в условии насыщения регулируемой переменной и ее первой производной по времени предложено достаточно большое количество методов и алгоритмов. В [12, 20, 25, 27] построение системы управления базируется на современных методах исследования нелинейных систем. В [6, 7, 8, 16, 17] синтез алгоритмов управления основан на использовании выпуклых вычислительных методов, таких как линейных матричных неравенств. Достаточно много решений, например, [6, 20, 21, 25], предложено на базе метода «*anti-windup*». Проблема заключается в том, что из-за насыщения управляющего сигнала переменные интегрирующего звена в ПИ- и ПИД-регуляторах могут быть неограниченными (*windup*), что приводит к невыполнению поставленной цели или потери устойчивости замкнутой системы. В таких случаях работоспособность системы управления с ПИ- и ПИД-регулятором может достигаться введением контура, предотвращающего рост параметров в регуляторе (*anti-windup*). В работах [9, 24] рассмотрено решение задачи с использованием статической и динамической обратных связей и нахождением областей устойчивости с привлечением аппарата линейных матричных неравенств. В [24] дополнительно решена задача для несимметричного насыщения регулируемой переменной в дискретных и непрерывных системах.

Вышеописанные работы посвящены решению задачи управления в условиях насыщения регулируемой переменной и ее производной для объекта с известными параметрами. К тому же для реализации закона управления использовался принцип подавления возмущений. В отличие от работ [6–8, 12, 16, 17, 20–22, 25, 26], данная статья посвящена синтезу алгоритма слежения выхода объекта за эталонным сигналом в условиях параметрической неопределенности, внешних неконтролируемых возмущений, насыщения сигнала управления и его производных. Для синтеза закона управления будет использоваться принцип компенсации возмущений, согласно которому значение сигнала управления должно быть противоположно значению возмущений. Таким образом, при наличии ограничений на сигнал управления и его производные соответствующие ограничения должны быть наложены на возмущения и его производные.

В данной статье решена задача управления линейными объектами в условии насыщения регулируемой переменной и ее производных по времени. Рассматривается модель объекта с неизвестными параметрами, которая подвержена действию внешних ограниченных возмущений. Синтез алгоритма управления условно разбивается на два этапа. На первом этапе формируется закон управления, учитывающий ограничения на входной сигнал и его производные; на втором этапе осуществляется синтез алгоритма компенсации параметрических и внешних возмущений. Получены условия на параметры объекта, эталонной модели и регулятора при выполнении которых система управления обеспечивает слежение выхода объекта за эталонным сигналом с заданной точностью. Приведены результаты моделирования, иллюстрирующие работоспособность приведенной схемы.

2. Постановка задачи

Рассмотрим объект управления, математическая модель которого описывается уравнением

$$(1) \quad \dot{x}(t) = Ax(t) + Bg_0(u_0) + Df(t), \quad y(t) = Lx(t), \quad x(0) = x_0,$$

где $x(t) \in R^n$ – вектор состояния; $u_0(t) \in R$ и $y(t) \in R$ – входной и выходной сигналы соответственно; $f(t) \in R$ – гладкое неконтролируемое внешнее возмущение, причем $|f^{(i)}(t)| \leq \bar{f}_i$, \bar{f}_i – известные величины, $i = 0, 1, \dots, k$; элементы матрицы $A \in R^{n \times n}$ и коэффициенты векторов $B \in R^n$, $D \in R^n$ – неизвестные числа; $L = [1, 0, \dots, 0]$ – матрица соответствующей размерности, x_0 – неизвестные начальные условия; функция $g_0(u_0) \in R$ определена выражением

$$(2) \quad g_0(u_0) = \bar{u}_0 \operatorname{sat} \left(\frac{u_0(t)}{\bar{u}_0} \right) = \begin{cases} u_0(t), & |u_0(t)| \leq \bar{u}_0, \\ \bar{u}_0 \operatorname{sgn}(u_0(t)), & |u_0(t)| > \bar{u}_0, \end{cases}$$

где $\bar{u}_0 > 0$ – величина насыщения. Кроме того, на входной сигнал объекта (1) наложены ограничения следующего вида:

$$u_0(t) = \int_0^t g_1(u_1(z)) dz,$$

$$g_1(u_1) = \bar{u}_1 \operatorname{sat} \left(\frac{u_1(t)}{\bar{u}_1} \right) = \begin{cases} u_1(t), & |u_1(t)| \leq \bar{u}_1, \\ \bar{u}_1 \operatorname{sgn}(u_1(t)), & |u_1(t)| > \bar{u}_1, \end{cases}$$

$$u_1(t) = \int_0^t g_2(u_2(z)) dz,$$

$$(3) \quad g_2(u_2) = \bar{u}_2 \operatorname{sat} \left(\frac{u_2(t)}{\bar{u}_2} \right) = \begin{cases} u_2(t), & |u_2(t)| \leq \bar{u}_2, \\ \bar{u}_2 \operatorname{sgn}(u_2(t)), & |u_2(t)| > \bar{u}_2, \end{cases}$$

⋮

$$u_{k-1}(t) = \int_0^t g_k(u_k(z)) dz,$$

$$g_k(u_k) = \bar{u}_k \operatorname{sat} \left(\frac{u_k(t)}{\bar{u}_k} \right) = \begin{cases} u_k(t), & |u_k(t)| \leq \bar{u}_k, \\ \bar{u}_k \operatorname{sgn}(u_k(t)), & |u_k(t)| > \bar{u}_k. \end{cases}$$

Зададим уравнение эталонной модели в виде

$$(4) \quad \dot{x}_m(t) = A_m x_m(t) + B_m r(t), \quad y_m(t) = L x_m(t), \quad x_m(0) = x_{m0},$$

где $x_m(t) \in R^n$ – вектор состояния; $r(t) \in R$ – гладкое задающее воздействие, причем $|r^{(i)}(t)| \leq \bar{r}_i$, $i = 0, 1, \dots, k$; $y_m(t) \in R$ – выход

эталонной модели; матрица $A_m \in R^{n \times n}$ и вектор $B_m \in R^n$ – известны, причем матрица A_m – гурвицева; x_{m0} – известные начальные условия.

Требуется разработать алгоритм, обеспечивающий выполнение целевого условия

$$(5) \quad |y(t) - y_m(t)| < \delta \quad \text{при } t > T,$$

где величина $\delta > 0$ характеризует точность слежения, $T > 0$ – время переходного процесса. Также необходимо, чтобы все переменные в замкнутой системе были ограниченными.

Предположения.

1. Неизвестные элементы матрицы A и коэффициенты векторов B и D принадлежат известному ограниченному множеству Ξ . Пара (A, B) управляема, пара (A, C) наблюдаема.

2. Выполнены условия: $A = A_m + B_m c_{01}^T$, $B = B_m + B_m c_{02}$, $D = B_m c_{03}$, где $c_{01} \in R^n$, $c_{02} \in R$, $c_{03} \in R$ – неизвестные вектор и числа.

3. Числитель передаточной функции $L(\lambda I - A)^{-1}B$ – гурвицев, где λ – комплексная переменная, I – единичная матрица соответствующего порядка.

Прокомментируем необходимость выполнения предположений 2 и 3. Идея компенсации возмущений состоит в поиске сигнала управления, значение которого было бы противоположно значению возмущений. Тогда при наличии ограничений на сигнал управления необходимо потребовать соответствующие ограничения и на возмущения. Предположение 2 требует согласованность возмущений, т.е. возможность напрямую повлиять сигналом управления на параметрические и внешние возмущения с учетом ограничений на регулируемый сигнал. Предположение 3 необходимо для синтеза устойчивого регулятора. К сожалению, предложенный алгоритм не обоснован, а моделирование показывает, что зачастую и не работоспособен для несогласованных возмущений и передаточных функций объекта с неустойчивым числителем.

3. Структура основного закона управления

Принимая во внимание предположение 2, перепишем уравнение (1) в виде

$$(6) \quad \dot{x}(t) = A_m x(t) + B_m \hat{u}_0(t) + B_m \psi(t), \quad y(t) = Lx(t),$$

где $\hat{u}_0(t)$ – новый сигнал управления, $\psi(t) = c_{01}^T x(t) + (1 + c_{02})g_0(u_0) - \hat{u}_0(t) + c_{03}f(t)$. В условиях ограничений (2) и (3) зададим функцию $\hat{u}_0(t)$ в виде

$$(7) \quad \begin{aligned} \hat{u}_0(t) &= \frac{1}{1 + \sigma_0} \left(g_1(t) + \sigma_0 \tilde{u}_0 \text{sat} \left(\frac{g_1(t)}{\tilde{u}_0} \right) \right), \\ \dot{g}_1(t) &= \hat{u}_1(t) - \mu g_1(t), \\ \hat{u}_1(t) &= \frac{1}{1 + \sigma_1} \left(g_2(t) + \sigma_1 \tilde{u}_1 \text{sat} \left(\frac{g_2(t)}{\tilde{u}_1} \right) \right), \\ \dot{g}_2(t) &= \hat{u}_2(t) - \mu g_2(t), \\ &\vdots \\ \dot{g}_{k-1}(t) &= \hat{u}_{k-1}(t) - \mu g_{k-1}(t), \\ \hat{u}_{k-1}(t) &= \frac{1}{1 + \sigma_{k-1}} \left(g_k(t) + \sigma_{k-1} \tilde{u}_{k-1} \text{sat} \left(\frac{g_k(t)}{\tilde{u}_{k-1}} \right) \right), \\ \dot{g}_k(t) &= \hat{u}_k(t) - \mu g_k(t), \\ \hat{u}_k(t) &= \frac{1}{1 + \sigma_k} \left(u_c(t) + \sigma_k \tilde{u}_k \text{sat} \left(\frac{u_c(t)}{\tilde{u}_k} \right) \right), \end{aligned}$$

где $\sigma_i > 0$ – коэффициенты, выбираемые разработчиком из условия нахождения сигналов $\hat{u}_i(t)$ в соответствующих множествах $[-\bar{u}_i, \bar{u}_i]$, $0 < \tilde{u}_i < \bar{u}_i$, $i = 0, \dots, k$; $\mu > 0$ – достаточно малое число; $u_c(t)$ – функция, необходимая для компенсации параметрической неопределенности и внешнего возмущения в (1). В (7) слагаемые $\mu g_i(t)$ необходимы для предотвращения возможного неограниченного роста функции $g_i(t)$ при насыщении сигналов $\hat{u}_i(t)$.

Из (7) видно, что при $\sigma_i < \infty$ функция $|\hat{u}_i(t)|$ может принимать значения больше, чем \tilde{u}_i . Однако при $\sigma_i \rightarrow \infty$ следует, что $\hat{u}_k(t) \rightarrow \tilde{u}_k \text{sat}\left(\frac{u_c(t)}{\tilde{u}_k}\right)$ и $\hat{u}_{i-1}(t) \rightarrow \tilde{u}_{i-1} \text{sat}\left(\frac{g_i(t)}{\tilde{u}_{i-1}}\right)$. Значит, при достаточно больших σ_i величину \tilde{u}_i можно выбирать достаточно близкой к \bar{u}_i . Так, при $\tilde{u}_i = \bar{u}_i$ и $\sigma_i \rightarrow \infty$ следует, что $\hat{u}_k(t) \rightarrow \bar{u}_k \text{sat}\left(\frac{u_c(t)}{\bar{u}_k}\right)$ и $\hat{u}_{i-1}(t) \rightarrow \bar{u}_{i-1} \text{sat}\left(\frac{g_i(t)}{\bar{u}_{i-1}}\right)$. В дальнейшем будут получены условия выбора величин σ_i для обеспечения $|\hat{u}_i(t)| \leq \bar{u}_i$.

Дополнительно отметим, что в (7) коэффициенты σ_i можно выбирать из условия $\sigma_i \in (-\infty; -1) \cup (0; +\infty)$. Однако ради простоты синтеза системы управления положим в (7) $\sigma_i > 0$.

4. Алгоритм компенсации возмущений

Перепишем (7) в виде

$$\begin{aligned}
 \hat{u}_0(t) &= g_1(t) + \sigma_0 \left[\tilde{u}_0 \text{sat}\left(\frac{g_1(t)}{\tilde{u}_0}\right) - \hat{u}_0(t) \right], \\
 \dot{g}_1(t) &= \hat{u}_1(t) - \mu g_1(t), \\
 \hat{u}_1(t) &= g_2(t) + \sigma_1 \left[\tilde{u}_1 \text{sat}\left(\frac{g_2(t)}{\tilde{u}_1}\right) - \hat{u}_1(t) \right], \\
 (8) \quad \dot{g}_2(t) &= \hat{u}_2(t) - \mu g_2(t), \\
 &\vdots \\
 \dot{g}_{k-1}(t) &= \hat{u}_{k-1}(t) - \mu g_{k-1}(t), \\
 \hat{u}_{k-1}(t) &= g_k(t) + \sigma_{k-1} \left[\tilde{u}_{k-1} \text{sat}\left(\frac{g_k(t)}{\tilde{u}_{k-1}}\right) - \hat{u}_{k-1}(t) \right], \\
 \dot{g}_k(t) &= \hat{u}_k(t) - \mu g_k(t), \\
 \hat{u}_k(t) &= u_c(t) + \sigma_k \left[\tilde{u}_k \text{sat}\left(\frac{u_c(t)}{\tilde{u}_k}\right) - \hat{u}_k(t) \right].
 \end{aligned}$$

Выразим в (8) функцию $\hat{u}_0(t)$ через $u_c(t)$ в виде

$$(9) \quad \hat{u}_0(t) = v(t) + w(t),$$

где

$$(10) \quad \begin{aligned} v(t) &= \int_0^t \dots \int_0^t u_c(z_1) dz_1 \dots dz_k, \\ w(t) &= \sigma_0 \left(\tilde{u}_0 \operatorname{sat} \left(\frac{g_1(t)}{\tilde{u}_0} \right) - \hat{u}_0(t) \right) + \\ &+ \sigma_k \int_0^t \dots \int_0^t \left(\tilde{u}_k \operatorname{sat} \left(\frac{u_c(z_1)}{\tilde{u}_k} \right) - \hat{u}_k(z_1) \right) dz_1 \dots dz_k + \\ &+ \sum_{i=1}^{k-1} \sigma_i \int_0^t \dots \int_0^t \left(\tilde{u}_i \operatorname{sat} \left(\frac{g_{i+1}(z_1)}{\tilde{u}_i} \right) - \hat{u}_i(z_1) \right) dz_1 \dots dz_i - \\ &- \mu \sum_{i=1}^{k-1} \int_0^t \dots \int_0^t g_i(z_1) dz_1 \dots dz_i. \end{aligned}$$

Принимая во внимание (4), (6), (9) и (10), составим уравнение для ошибки $\varepsilon(t) = x(t) - x_m(t)$ в виде

$$(11) \quad \dot{\varepsilon}(t) = A_m \varepsilon(t) + B_m v(t) + B_m \varphi(t), \quad e(t) = L \varepsilon(t),$$

где $\varphi(t) = \psi(t) - r(t) + w(t)$. Для компенсации возмущений воспользуемся подходом [3, 4]. Согласно [3, 4], введем вспомогательный контур

$$(12) \quad \dot{\varepsilon}_a(t) = A_m \varepsilon_a(t) + B_m v(t), \quad e_a(t) = L \varepsilon_a(t), \quad \varepsilon_a(0) = 0,$$

где $\varepsilon_a(t) \in R^n$. С учетом (11) и (12) составим уравнение для рассогласования $\zeta(t) = \varepsilon(t) - \varepsilon_a(t)$ в виде

$$(13) \quad \dot{\zeta}(t) = A_m \zeta(t) + B_m \varphi(t), \quad z(t) = L \zeta(t).$$

Преобразуем уравнение (13) к форме «вход–выход»:

$$(14) \quad Q_m(p)z(t) = R_m(p)\varphi(t),$$

где $Q_m(p)$, $R_m(p)$ – линейные дифференциальные операторы, полученные при переходе от (13) к (14), $p = d/dt$. Выразив в (14) функцию $\varphi(t)$ и подставив ее в (11), получим

$$(15) \quad \dot{\varepsilon}(t) = A_m \varepsilon(t) + B_m \left[v(t) + \frac{Q_m(p)}{R_m(p)} z(t) \right].$$

Если бы $(\gamma + k)$ производных сигнала $y(t)$ были бы доступны измерению, то, принимая во внимание (15) и первое выражение из (10), закон компенсации возмущений формировался бы

в виде $u_c(t) = -\frac{p^k Q_m(p)}{R_m(p)} z(t)$. Однако из постановки задачи

производные сигнала $y(t)$ недоступны измерению. Перепишем $\lambda^k Q_m(\lambda) / R_m(\lambda)$ в виде $\lambda^k Q_m(\lambda) / R_m(\lambda) = F(\lambda) + \tilde{Q}_m(\lambda) / R_m(\lambda)$, где $F(\lambda)$ и $\tilde{Q}_m(\lambda)$ получены при выделении целой части в дробно-рациональной функции $\lambda^k Q_m(\lambda) / R_m(\lambda)$, $\deg F(\lambda) = \gamma + k$, $\deg \tilde{Q}_m(\lambda) \leq m - 1$. Тогда, принимая во внимание первое выражение из (10), закон компенсации возмущений $u_c(t)$ зададим в виде

$$(16) \quad u_c(t) = -q^T \xi(t) - \frac{\tilde{Q}_m(p)}{R_m(p)} z(t),$$

где q – вектор, составленный из коэффициентов оператора $F(p)$, которые записаны в обратном порядке, $\xi(t) = [\bar{z}(t), \dot{\bar{z}}(t), \dots, \bar{z}^{(\gamma+k)}(t)]^T$, $\bar{z}^{(i)}(t)$ – оценка i -й производной сигнала $z(t)$. Вектор $\xi(t)$ получен с помощью наблюдателя [5], который записан в форме

$$(17) \quad \dot{\xi}(t) = G_0 \xi(t) + D_0 (\bar{z}(t) - z(t)), \quad \bar{z}(t) = L \xi(t), \quad \xi(0) = 0,$$

где $\xi(t) \in R^{\gamma+k+1}$, $G_0 = \begin{bmatrix} 0 & I_{\gamma+k} \\ 0 & 0 \end{bmatrix}$, $I_{\gamma+k}$ – единичная матрица

порядка $\gamma + k$, $D_0 = -[d_1 \mu^{-1}, d_2 \mu^{-2}, \dots, d_{\gamma+k+1} \mu^{-\gamma-1}]^T$, $d_1, \dots, d_{\gamma+k+1}$ выбираются из условия гурвицевости матрицы $G = G_0 - [d_1, \dots, d_{\gamma+k+1}]^T L$.

Перед формулировкой утверждения введем следующие обозначения:

$$\alpha_1 = \max_{\Xi} (|c_{01}|) \max_{\Xi} (\|A^k\|),$$

$$\alpha_2 = \max_{\Xi} (|c_{01}|) \max_{\Xi} \left| \sum_{i=0}^{k-1} A^i B \right| + \max_{\Xi} (c_{01}),$$

$$\alpha_3 = \max_{\Xi} (|c_{01}|) \max_{\Xi} \left| \sum_{i=0}^{k-1} A^i D \right| + \max_{\Xi} (c_{03}),$$

$$\bar{x}_m = \sup_t (|x_m(t)|), \quad \tilde{f} = [\bar{f}_0, \bar{f}_1, \dots, \bar{f}_k].$$

Утверждение. Пусть выполнены условия предположений 1-3. Тогда существуют $\mu > 0$ и $\sigma_i > 0, i = 0, \dots, k$, такие, что при $\mu \leq \mu_0$ и

$$(18) \alpha_1 \bar{x}_m + \alpha_3 \tilde{f} + \bar{r} < (1 - \alpha_2 k) \tilde{u} \text{ и } \alpha_2 k > 1,$$

$$(19) |\varepsilon(0)| \leq \alpha_1^{-1} \left((1 - \alpha_2 k) \tilde{u} - \alpha_1 \bar{x}_m - \alpha_3 \tilde{f} - \bar{r}_k \right),$$

система управления (7), (12), (16), (17) обеспечит выполнение целевого условия (5) и ограниченность всех сигналов в замкнутой системе.

Доказательство утверждения приведено в Приложении.

Замечание. Из доказательства утверждения следует, что система управления (7), (12), (16), (17) обеспечит выполнение условия (5) в момент времени T с точностью

$$(20) \delta = \sqrt{\lambda_{\min}^{-1}(P) \left[\left(\varepsilon^T(0) P \varepsilon(0) - \mu_0 \alpha^{-1} \rho \right) e^{-\alpha T} + \mu_0 \alpha^{-1} \rho \right]},$$

где $P = P^T > 0$ – решение уравнения $A_m^T P + P A_m = -Q_1$,

$$Q_1 = Q_1^T > 0, \alpha = \min \left\{ \frac{\lambda_{\min}(R_1)}{\lambda_{\max}(P)}, \frac{\lambda_{\min}(R_2)}{\mu_0 \lambda_{\max}(H)} \right\}, \lambda_{\max}(\cdot) \text{ } (\lambda_{\min}(\cdot))$$

наибольшее (наименьшее) собственное число соответствующей матрицы, $H = H^T > 0$ – решение уравнения $G^T H + H G = -Q_2$,

$$Q_2 = Q_2^T > 0, \quad R_1 = Q_1 - 2\mu_0 P B_m q^T T \left(P B_m q^T T \right)^T,$$

$$T = \text{diag} \{ \mu^\gamma, \dots, \mu, 1 \}, \quad R_2 = Q_2 - 2H b b^T H, \quad \rho = 2 \sup_t \left\{ \zeta^{(\gamma+k+1)}(t) \right\}.$$

Оценки для коэффициентов σ_i рассчитываются с помощью следующих условий:

$$(21) \sigma_i \geq \frac{(2 + \mu_0) \tilde{u}}{\mu_0 (\bar{u} - \tilde{u})}, \quad i = 0, 1, \dots, k-1,$$

$$\sigma_k \geq \frac{\bar{u} + |q| \pi + \kappa \bar{c}_{01} \sqrt{\lambda_{\min}^{-1}(P) \left[\left(\varepsilon^T(0) P \varepsilon(0) + \frac{\mu_0 \rho}{\alpha} \right) + \frac{\mu_0 \rho}{\alpha} \right]} + \kappa (\bar{c}_{01} \bar{x}_m + (2 + \bar{c}_{02}) \tilde{u} + \bar{c}_{03} \tilde{f} + \bar{r})}{\bar{u} - \tilde{u}}.$$

где $\pi = \sup_t (|\xi(t)|)$, $\kappa = \sup_{\omega} \left| \frac{\tilde{Q}_m(j\omega)}{Q_m(j\omega)} \right|$, j – мнимая единица.

Стоит отметить, что оценки (18)–(21) достаточно грубы из-за использования грубых оценок в доказательстве. Для иллюстрации полученных результатов рассмотрим пример.

5. Пример

Рассмотрим объект управления, математическая модель которого имеет вид

$$(22) \quad \dot{x}(t) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ a_0 & a_1 & a_2 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 0 \\ b \end{bmatrix} g_0(u_0) + \begin{bmatrix} 0 \\ 0 \\ d \end{bmatrix} f(t),$$

$$y(t) = [1 \ 0 \ 0]x(t), \quad x(0) = x_0.$$

Множество возможных значений Ξ задано следующими неравенствами: $-1 \leq a_0 \leq 0,5$; $-2,5 \leq a_1 \leq -2$; $-0,5 \leq a_2 \leq 1$; $1 \leq b \leq 1,9$; $-1 \leq d \leq 1$. На возмущение и его первую производную наложены следующие ограничения: $\bar{f}_0 = 1$, $\bar{f}_1 = 1$. Дополнительно на входной сигнал объекта (22) наложены ограничения:

$$g_0(u_0) = 1,1 \text{sat} \left(\frac{u_0(t)}{1,1} \right) = \begin{cases} u_0(t), & |u_0(t)| \leq 1,1, \\ 1,1 \text{sgn}(u_0(t)), & |u_0(t)| > 1,1, \end{cases}$$

$$u_1(t) = \int_0^t g_2(u_2(z)) dz,$$

$$g_1(u_1) = 1,2 \text{sat} \left(\frac{u_1(t)}{1,21} \right) = \begin{cases} u_1(t), & |u_1(t)| \leq 1,21, \\ 1,21 \text{sgn}(u_1(t)), & |u_1(t)| > 1,21. \end{cases}$$

Выберем в (3) $A_m = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -3 & -3 \end{bmatrix}$, $B_m = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, $x_{m0} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$,

$\bar{r}_0 = 1$ и $\bar{r}_1 = 1$. Цель управления состоит в синтезе алгоритма, обеспечивающего выполнение целевого условия (5).

Зададим $\sigma_0 = \sigma_1 = 10^{10}$, $\tilde{u}_0 = 1$, $\tilde{u}_1 = 1,2$, $\mu = 0,01$ и сформируем закон управления (7) в виде

$$\hat{u}_0(t) = (1 + 10^{10})^{-1} (g_1(t) + 10^{10} \text{sat}(g_1(t))),$$

$$(23) \quad \dot{g}_1(t) = \hat{u}_1(t) - 0,01g_1(t),$$

$$\hat{u}_1(t) = (1 + 10^{10})^{-1} (u_c(t) + 10^{10} \cdot 1,2 \text{sat}(5u_c(t)/6)).$$

Согласно (12), вспомогательный контур определим уравнением

$$(24) \quad \dot{\varepsilon}_a(t) = A_m \varepsilon_a(t) + B_m \int_0^t u_c(z) dz, \quad \varepsilon_a(t) = [1 \quad 0 \quad 0] \varepsilon_a(t),$$

$$\varepsilon_a(0) = [0 \quad 0 \quad 0]^T.$$

Выберем $d_1 = 3$, $d_2 = 3$, $d_3 = 1$ и сформируем уравнения наблюдателя (17) в виде

$$\dot{\xi}_1(t) = \xi_2(t) - 4 \cdot 100 (\xi_1(t) - z(t)),$$

$$\dot{\xi}_2(t) = \xi_3(t) - 6 \cdot 100^2 (\xi_1(t) - z(t)),$$

$$(25) \quad \dot{\xi}_3(t) = \xi_4(t) - 6 \cdot 100^2 (\xi_1(t) - z(t)),$$

$$\dot{\xi}_4(t) = -100^3 (\xi_1(t) - z(t)),$$

$$\xi(0) = [0 \quad 0 \quad 0 \quad 0]^T.$$

Принимая во внимание уравнения (25) и структуры A_m , B_m и L , сформируем закон компенсации возмущений $u_c(t)$ в виде

$$(26) \quad u_c(t) = -\dot{\xi}_4(t) - 3\xi_4(t) - 3\xi_3(t) - \xi_2(t).$$

Теперь воспользуемся условиями (18) и (19) для задания остальных параметров в замкнутой системе. Принимая во внимание множество Ξ и параметры эталонной модели (3), перепишем условия (14) и (15) в виде $\tilde{f} + \bar{r} + 2\sqrt{6}\bar{x}_m < 0,1$ и $|x(0)| \leq 0,5 \cdot 6^{-0,5} (0,1 - \tilde{f} - \bar{r} - 2\sqrt{6}\bar{x}_m)$. Отметим, что для выбранных значений σ_0 и σ_1 оценки (21) выполнены. Как отмечалось, полученные оценки достаточно грубые. Результаты моделирования показали, что система управления будет работоспособной, например, при $\bar{r} = 1,2$, $\bar{x}_m = 0,6$, $\tilde{f} = 0,4$, $|x(0)| \leq 0,07\sqrt{3}$. Покажем работоспособность системы

управления (23)–(26) для следующих параметров объекта (22) и эталонной модели (4): $a_0 = -1$, $a_1 = -3$, $a_2 = 1$, $b = 1$, $d = 1$, $f(t) = 0,2 + 0,2\sin 0,7t$, $x(0) = 0,07 \cdot [1 \ 1 \ 1]^T$ и $r(t) = 0,2 + \sin t$. На рис. 1 представлены результаты моделирования по ошибке $e(t)$ и сигналам управления $\hat{u}_1(t)$ и $\hat{u}_0(t)$.

Анализ результатов моделирования показал, что замкнутая система робастна по отношению к внешним возмущениям и параметрической неопределенности из заданного класса Ξ . Так, из рис. 1 следует, что в системе управления динамическая ошибка не превышает значения 0,65 начиная с момента времени 5 с. Из рис. 1 видно, что сигналы $\hat{u}_1(t)$ и $\hat{u}_0(t)$ находятся в отрезках $[-1,2; 1,2]$ и $[-1; 1]$ соответственно. Без использования алгоритма (23) (т.е. как и в [4]: при $\hat{u}_0(t) = \int_0^t u_c(s)ds$) сигналы $\hat{u}_0(t)$ и $u_c(t)$ в начальный момент времени достигают значений $-1,5 \cdot 10^4$ и $-8 \cdot 10^6$ соответственно, а начиная с 10 с функции $\hat{u}_0(t)$ и $u_c(t)$ находятся в отрезке $[-1,3; 1,3]$ соответственно, что недопустимо по условию задачи.

Отметим, что данный пример был проверен для алгоритма из [19]. Однако в [19] для синтеза закона управления используется не принцип компенсации, а принцип подавления возмущений, что зачастую требует больше ресурсов управляющего устройства. В результате для алгоритма [19] не удалось подобрать параметры, чтобы обеспечить устойчивость замкнутой системы.

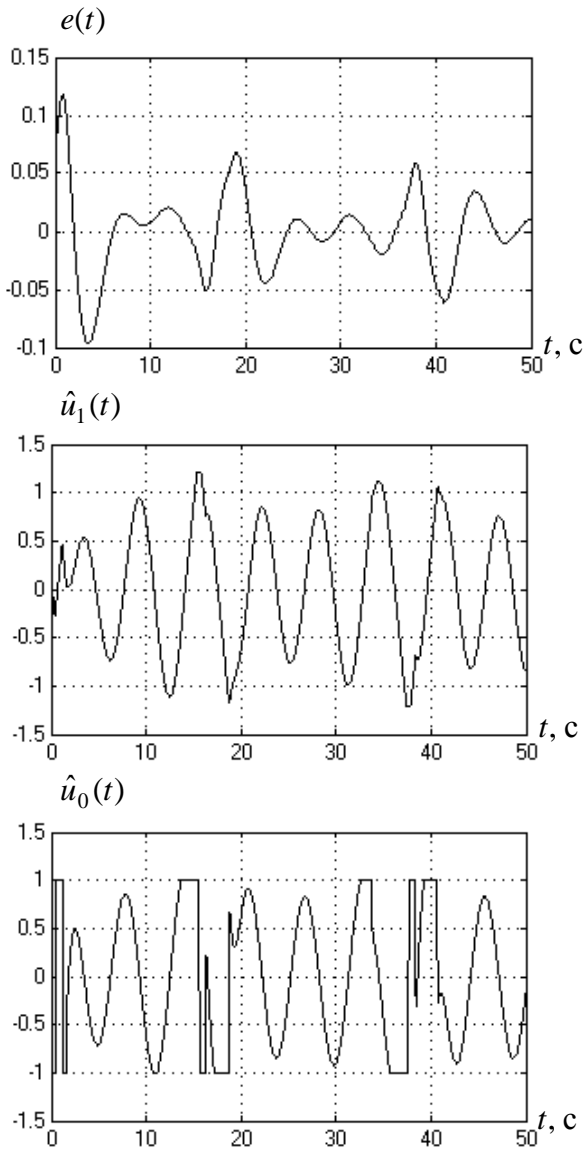


Рис. 1. Переходные процессы по $e(t)$, $\hat{u}_1(t)$ и $\hat{u}_0(t)$

6. Заключение

В статье приведен синтез системы управления для линейного объекта в условиях параметрической неопределенности, внешних ограниченных возмущений, насыщения регулируемой переменной и ее производных. Предложен способ формирования сигнала управления, обеспечивающего нахождение функции управления и ее производных в заданных множествах. Получены условия на параметры объекта управления, внешнего возмущения, эталонной модели и регулятора при выполнении которых система управления будет работоспособной. Отметим, что алгоритм, учитывающий насыщение сигнала управления и его производных, является независимым, т.е. им можно дополнить любой из существующих алгоритмов.

ПРИЛОЖЕНИЕ

Доказательство утверждения. Введем вектор $\theta(t) = [z(t), \dot{z}(t), \dots, z^{(\gamma+k)}(t)]^T$. Как и в [4], принимая во внимание (17), составим уравнение для ошибки оценки производных $\eta(t) = T^{-1}(\xi(t) - \theta(t))$ в виде

$$(П.1) \quad \dot{\eta}(t) = \mu^{-1}G\eta(t) + bz^{(\gamma+k+1)}(t), \quad \bar{z}(t) - z(t) = \mu^{\gamma+k}L\eta(t),$$

где $b = [0, \dots, 0, 1]^T$. С учетом (16), преобразуем (15) к виду

$$(П.2) \quad \dot{\varepsilon}(t) = A_m\varepsilon(t) - B_mq^T T\eta(t).$$

Перепишем уравнения (8), (П.1) и (П.2) в виде системы

$$\dot{\varepsilon}(t) = A_m\varepsilon(t) - B_mq^T T\eta(t),$$

$$\mu_1\dot{\eta}(t) = G\eta(t) + \mu_2bz^{(\gamma+k+1)}(t),$$

$$(П.3) \quad \hat{u}_0(t) = g_1(t) + \sigma_0 \left(\tilde{u}_0 \text{sat} \left(\frac{g_1(t)}{\tilde{u}_0} \right) - \hat{u}_0(t) \right),$$

$$\dot{g}_1(t) = \hat{u}_1(t) - \mu_2g_1(t),$$

⋮

$$\begin{aligned} \dot{g}_{k-1}(t) &= \hat{u}_{k-1}(t) - \mu_2 g_{k-1}(t), \\ \hat{u}_{k-1}(t) &= g_k(z) + \sigma_{k-1} \left(\tilde{u}_{k-1} \operatorname{sat} \left(\frac{g_k(t)}{\tilde{u}_{k-1}} \right) - \hat{u}_{k-1}(t) \right), \\ \dot{g}_k(t) &= \hat{u}_k(t) - \mu_2 g_k(t), \\ \hat{u}_k(t) &= u_c(t) + \sigma_k \left(\tilde{u}_k \operatorname{sat} \left(\frac{u_c(t)}{\tilde{u}_k} \right) - \hat{u}_k(t) \right), \end{aligned}$$

где $\mu_1 = \mu_2 = \mu$. Воспользуемся вспомогательной леммой [2, 13, 14].

Лемма. Если динамическая система описывается уравнением

$$(П.4) \quad \dot{x}(t) = f(x(t), t, \mu_1, \mu_2), \quad x(t) \in R^n, \quad \mu_1 > 0, \quad \mu_2 > 0,$$

где $f(x, t, \mu_1, \mu_2)$ – непрерывная функция, липшицева по x , и при $\mu_2 = 0$ имеет ограниченную замкнутую область диссипативности

$$(П.5) \quad \Omega_x = \{x : P(x) \leq C\},$$

где $P(x)$ – непрерывная, кусочно-гладкая, положительно определенная функция в R^n , такая, что при некоторых $C_1 > 0$ и $\mu_0 > 0$ выполнено условие

$$\sup_{\mu_1, \mu_2 \leq \mu_0} \left\langle \left[\frac{\partial P(x)}{\partial x} \right]^T, f(x, t, \mu_1, 0) \right\rangle \leq -C_1, \quad \text{при } P(x) = C,$$

тогда для всех достаточно малых $\mu_1 \leq \mu_0$ и $\mu_2 \leq \mu_0$ множество (П.5) остается областью диссипативности системы (П.4).

Лемма является обобщение первой леммы В.А. Брусина [1] для неавтономных систем. Доказательство леммы приведено в [13]. Проверим условия леммы для системы (П.3). Для этого рассмотрим (П.3) при $\mu_2 = 0$:

$$\begin{aligned}
 \dot{\varepsilon}(t) &= A_m \varepsilon(t) - B_m q^T T \eta(t), \\
 \dot{\eta}(t) &= \mu_1^{-1} G \eta(t), \\
 \hat{u}_0(t) &= g_1(t) + \sigma_0 \left(\tilde{u}_0 \operatorname{sat} \left(\frac{g_1(t)}{\tilde{u}_0} \right) - \hat{u}_0(t) \right), \\
 \dot{g}_1(t) &= \hat{u}_1(t), \\
 \text{(П.6)} \quad &\vdots \\
 \dot{g}_{k-1}(t) &= \hat{u}_{k-1}(t), \\
 \hat{u}_{k-1}(t) &= g_k(z) + \sigma_{k-1} \left(\tilde{u}_{k-1} \operatorname{sat} \left(\frac{g_k(t)}{\tilde{u}_{k-1}} \right) - \hat{u}_{k-1}(t) \right), \\
 \dot{g}_k(t) &= \hat{u}_k(t), \\
 \hat{u}_k(t) &= u_c(t) + \sigma_k \left(\tilde{u}_k \operatorname{sat} \left(\frac{u_c(t)}{\tilde{u}_k} \right) - \hat{u}_k(t) \right).
 \end{aligned}$$

Так как компенсация неопределенностей осуществляется в последнем уравнении (П.6), то рассмотрим случай, когда $|p^k \varphi(t)| \leq \tilde{u}$ и $|u_c(t)| \leq \tilde{u}$, где $\tilde{u} = \min\{\tilde{u}_0, \tilde{u}_1, \dots, \tilde{u}_k\}$. В этом случае насыщения в сигналах управления $\hat{u}_i(t)$ не будет, следовательно, $w(t) = 0$. Тогда из (П.6) имеем
$$\hat{u}_0(t) = \int_0^t \dots \int_0^t u_c(z_1) dz_1 \dots dz_k.$$
 Первые два уравнения (П.6)

асимптотически устойчивы в силу гурвицевости матриц A_m и G . Из (4) и ограниченности $\varepsilon(t)$ следует ограниченность $|x(t)|$, а из ограниченности $\eta(t)$ и уравнения (17) следует ограниченность $\xi(t)$. Тогда из (16) следует ограниченность функции $u_c(t)$, а из (15) следует ограниченность $v(t)$, а, следовательно, и ограниченность $\hat{u}_0(t)$. Тогда из (П.6) следует ограниченность функций $\hat{u}_i(t)$, $i = 1, 2, \dots, k$.

Найдем теперь множество притяжения для системы (П.6). Принимая во внимание структуру функции $\varphi(t)$ и уравнение (1), найдем $p^k \varphi(t)$ в виде

$$(П.7) \quad p^k \varphi(t) = c_{01}^T \left[A^k x(t) + \sum_{i=0}^{k-1} \left(A^i B \hat{u}_0^{(k-1-i)}(t) + A^i D f^{(k-1-i)}(t) \right) \right] + \\ + c_{02} \hat{u}_0^{(k)}(t) + c_{03} f^{(k)}(t) - r^{(k)}(t).$$

Подставив (П.7) в условие $|p^k \varphi(t)| \leq \tilde{u}$, получим

$$(П.8) \quad \left| c_{01}^T \left[A^k x(t) + \sum_{i=0}^{k-1} \left(A^i B \hat{u}_0^{(k-1-i)}(t) + A^i D f^{(k-1-i)}(t) \right) \right] + \right. \\ \left. c_{02} \hat{u}_0^{(k)}(t) + c_{03} f^{(k)}(t) - r^{(k)}(t) \right| \leq \tilde{u}.$$

Найдем верхнюю оценку для левой части (П.8) в виде

$$(П.9) \quad \left| c_{01}^T \left[A^k x(t) + \sum_{i=0}^{k-1} \left(A^i B \hat{u}_0^{(k-1-i)}(t) + A^i D f^{(k-1-i)}(t) \right) \right] + c_{02} \hat{u}_0^{(k)}(t) + \right. \\ \left. + c_{03} f^{(k)}(t) - r^{(k)}(t) \right| \leq \alpha_1 |\varepsilon(t)| + \alpha_1 \bar{x}_m + \alpha_2 k \tilde{u} + \alpha_3 \tilde{f} + \bar{r}_k.$$

Подставив оценку (П.9) в левую часть (П.8), потребуем выполнение условий (18) и (19). Тогда при выполнении (18) и (19) следует, что $\dot{V}(t) \leq 0$ и цель управления (5) будет достигнута и условия леммы будут выполнены. Однако из асимптотической устойчивости (П.6) при $\mu_2 = 0$ не следует асимптотическая устойчивость (П.3) при $\mu_2 \neq 0$. Поэтому рассмотрим случай, когда $\mu_1 = \mu_2 = \mu_0$. Для системы (П.3) рассмотрим функцию Ляпунова $V(t) = V(\varepsilon(t), \eta(t))$ в виде

$$(П.10) \quad V(t) = \varepsilon^T(t) P \varepsilon(t) + \eta^T(t) H \eta(t).$$

Возьмем производную от (П.10) вдоль траекторий первых двух уравнений (П.3):

$$(П.11) \quad \dot{V}(t) = -\varepsilon^T(t) Q_1 \varepsilon(t) - 2\varepsilon^T(t) P B_m q^T T \eta(t) - \\ - \mu_0^{-1} \eta^T(t) Q_2 \eta(t) + 2\eta^T(t) H b z^{(\gamma+k+1)}.$$

Воспользуемся оценками:

$$(П.12) \quad 2\varepsilon^T(t) P B_m q^T T \eta(t) \leq \\ \leq 2\mu_0 \varepsilon^T(t) P B_m q^T T (P B_m q^T T)^T \varepsilon(t) + \mu_0^{-1} \eta(t), \\ 2\eta^T(t) H b z^{(\gamma+k+1)}(t) \leq 2\mu_0^{-1} \eta^T(t) H b b^T H \eta(t) + 2\mu_0 \rho.$$

С учетом (П.12), оценим (П.11) в виде

$$(П.13) \quad \dot{V}(t) = -\varepsilon^T(t)R_1\varepsilon(t) - \mu_0^{-1}\eta^T(t)R_2\eta(t) + \mu_0\rho.$$

Перепишав (П.13) в виде $\dot{V}(t) \leq -\alpha V(t) + \mu_0\rho$ и решив его относительно $V(t)$, получим

$$(П.14) \quad V(t) \leq \left(V(0) - \mu_0\alpha^{-1}\rho\right)e^{-\alpha t} + \mu_0\alpha^{-1}\rho.$$

Тогда

$$(П.15) \quad \begin{aligned} |e(t)| &\leq |\varepsilon(t)| \leq \\ &\leq \sqrt{\lambda_{\min}^{-1}(P) \left[\left(\varepsilon^T(0)P\varepsilon(0) - \mu_0\alpha^{-1}\rho \right) e^{-\alpha t} + \mu_0\alpha^{-1}\rho \right]}. \end{aligned}$$

Из (5) и (П.15) следует (20).

Теперь определим оценку для коэффициентов σ_i , при которых $|\hat{u}_i(t)| \leq \bar{u}$, где $\bar{u} = \min\{\bar{u}_i\}$, $i = 0, 1, \dots, k$. Оценим последнее выражение (7) в виде неравенства $|u_c(t)| + \sigma_k \tilde{u} \leq (1 + \sigma_k)\bar{u}$ и разрешим его относительно σ_k :

$$\sigma_k \geq \frac{\bar{u} - |u_c(t)|}{\bar{u} - \tilde{u}}.$$

Оценим сверху числитель правой части

последнего неравенства в виде

$$(П.16) \quad \begin{aligned} \bar{u} - |u_c(t)| &\leq \bar{u} + |u_c(t)| \leq \bar{u} + |q|\pi + \kappa|\varphi(t)| \leq \\ &\leq \bar{u} + |q|\pi + \\ &+ \kappa \left(c_{01}^T \varepsilon(t) + c_{01}^T x_m(t) + (1 + c_{02})g_0(u_0) - \hat{u}_0(t) + c_{03}f(t) + |r(t)| \right) \leq \\ &\leq \bar{u} + |q|\pi + \kappa \bar{c}_{01} |\varepsilon(t)| + \kappa \left(\bar{c}_{01} \bar{x}_m + (2 + \bar{c}_{02})\bar{u} + \bar{c}_{03} \bar{f} + \bar{r} \right). \end{aligned}$$

Принимая во внимание (П.15) и (П.16), получим вторую оценку (21). С учетом того, что $|\hat{u}_i(t)| \leq \bar{u}$, оценим σ_i в (7) в виде

$$\sigma_i \geq \frac{g_{i+1}(t) - \bar{u}}{\bar{u} - \tilde{u}}, \quad i = 0, 1, \dots, k.$$

Найдем оценки решений

неравенств $\dot{g}_{i+1}(t) \leq \bar{u} - \mu_0 g_{i+1}(t)$, полученных из (7), в виде $g_{i+1}(t) \leq 2\bar{u} / \mu_0$, $i = 0, 1, \dots, k-1$. Объединив полученные результаты получим первую оценку в (21).

Очевидно, что оценки (18)–(21) достаточно грубые, но из них видно, что существуют определенные значения параметров объекта, эталонной модели и регулятора, при которых в условиях ограничений можно обеспечить выполнение условия (5).

Литература

1. БРУСИН В.А. *Об одном классе сингулярно возмущенных адаптивных систем. I* // Автоматика и телемеханика. – 1995. – №4. – С. 119–127.
2. ФУРТАТ И.Б. *Робастное управление определенным классом неминимально-фазовых динамических сетей* // Известия РАН. Теория и системы управления. – 2014. – №1. – С. 35–48.
3. ФУРТАТ И.Б. *Динамическая компенсация возмущений в условии насыщения сигнала управления* // Управление большими системами. – 2017. – Вып. 65. – С. 24–40.
4. ЦЫКУНОВ А.М. *Алгоритмы робастного управления с компенсацией ограниченных возмущений* // Автоматика и телемеханика. – 2007. – №7. – С. 103–115.
5. ATASSI A.N., KHALIL H.K. *A separation principle for the stabilization of class of nonlinear systems* // IEEE Trans. on Automatic Control. – 1999. – Vol. 44, No. 9. – P. 1672–1687.
6. BARBU C., GALEANI S., TEEL A., ZACCARIAN L. *Nonlinear anti-windup for manual flight control* // Int. J. of Control. – 2005. – Vol. 78, No. 14. – P. 1111–1129.
7. BATEMAN A., LIN Z. *An analysis and design method for discrete-time linear systems under nested saturation* // IEEE Trans. Automat. Control. – 2002. – Vol. 47. – P. 1305–1310.
8. BATEMAN A., LIN Z. *An analysis and design method for linear systems under nested saturation* // Systems Control Lett. – 2003. – Vol. 48. – P. 41–52.
9. BENZAOUIA A., MESQUINE F., BENHAYOUN M. *Saturated Control of Linear Systems*. – Springer International Publishing, 2018. – 226 p.
10. BERG J., HAMMETT K., SCHWARTZ C., BANDA S. *An analysis of the destabilizing effect of daisy chained rate-limited actuators* // IEEE Trans. Control Systems Technology. – 1996. – Vol. 4. – P. 171–176.
11. CHEN P., SHAMMA J.S. *Gain-scheduled H_1 -optimal control for boiler-turbine dynamics with actuator saturation* // J. Process Control. – 2004. – Vol. 14. – P. 263–277.

12. FREEMAN R., PRALY L. *Integrator backstepping for bounded controls and control rates* // IEEE Trans. Automat. Control. – 1998. – Vol. 43, No. 2. – P. 258–262.
13. FURTAT I., FRADKOV A., TSYKUNOV A. *Robust synchronization of linear dynamical systems with compensation of disturbances* // Int. J. Robust and Nonlinear Control. – 2014. – Vol. 24, No. 17. – P. 2774–2784.
14. FURTAT I.B. *Robust Control for a Specific Class of Non-Minimum Phase Dynamical Networks* // Int. J. of Computer and Systems Sciences. – 2014. – Vol. 53, No. 1. – P. 33–46.
15. HARDT M., HELTON W., KREUTZ-DELGADO W. *Numerical solution of nonlinear H_2 and H_∞ control problems with application to jet engine compressors* // IEEE Trans. Automat. Control. – 2000. – Vol. 8, No. 1. – P. 98–111.
16. JABBARI F., KOSË I. *Rate and magnitude-bounded actuators: scheduled output feedback design* // Int. J. of Robust and Nonlinear Control. – 2004. – Vol. 14. – P. 1169–1184.
17. KAPILA V., PAN H., DE QUEIROZ M. *LMI-based control of linear systems with actuator amplitude and rate nonlinearities* // Proc. 38th IEEE Conf. on Decision and Control. –1999. – P. 1413–1418.
18. KRSTIC M., KRUPADANAM A., JACOBSON C. *Self-tuning control of a nonlinear model of combustion instabilities* // IEEE Trans. Automat. Control. – 1999. – Vol. 7. – P. 424–436.
19. LEONESSA A., HADDAD W.M., HAYAKAWA T., MOREL Y. *Adaptive control for nonlinear uncertain systems with actuator amplitude and rate saturation constraints* // Int. J. of Adaptive Control and Signal Processing. – 2009. – Vol. 23. – P. 73–96.
20. MEGRETSKI A. *New IQC for quasi-concave nonlinearities* // Proc. of American Control Conf. – 1999. – P. 2380–2385.
21. MILLER R., PACHTER M. *Manual flight control with saturating actuators* // IEEE Control Systems. – 1998. – Vol. 18, No. 1. – P. 10–19.
22. SABERI A., STOORVOGEL A., SANNUTI P. *Control of Linear Systems with Regulation and Input Constraints*. – London: Springer, 2000.

23. SHIFRIN C.A. *Sweden seeks cause of gripen crash* // Aviation Week and Space Tech. – 1993. – Vol. 139. – P. 78–79.
24. TARBOURIECH S., GARCIA G., GOMES DA SILVA JR. J.M., QUEINNEC I. *Stability and Stabilization of Linear Systems with Saturating Actuators*. – Springer London, Dordrecht, Heidelberg, New York, 2011. – 425 p.
25. TEEL A., BUFFINGTON J. *Anti-windup for an F-16's daisy chain control allocator* // Proc. AIAA GNC Conf., New Orleans, LA, USA, 1997. – P. 748–754.
26. TYAN F., BERNSTEIN D. *Dynamics output feedback compensation for linear systems with independent amplitude and rate saturations* // Int. J. of Control. – 1997. – Vol. 67. – P. 89–116.
27. WANG Y., YEUNG S., MURRAY R. *Bifurcation control of rotating stall with actuator magnitude and rate limits: part II – Control synthesis and comparison with experiments* // Automatica. – 2002. – Vol. 38. – P. 611–625.

CONTROL ALGORITHM UNDER SATURATION OF CONTROL SIGNAL AND ITS DERIVATIVES

Igor Furtat, Institute of Problems of Mechanical Engineering, Russian Academy of Sciences, ITMO University, St. Petersburg, Doctor of Sciences (cainenash@mail.ru).

Pavel Gushchin, Institute of Problems of Mechanical Engineering, Russian Academy of Sciences, ITMO University, St. Petersburg, Cand.Sc. (gushchin.p@mail.ru).

Abstract: The synthesis of the control system under conditions of parametric uncertainties, bounded external disturbances, saturation of the control signal and its time derivatives is given. Such tasks are typical in the control of various kinds of technical systems. Ignoring uncertainties and constraints can lead to objective failure or loss of control system stability. For example, catastrophic consequences may occur in an aircraft's flight control system due to the saturation of the controlled signal and its derivative. Other examples are the control of jet engine compressors and the study of the behavior of the control plant with a slow drive. To solve the problem in the conditions of uncertainties and restrictions on the regulating signal a new method of design the control law is proposed, which allows to ensure the finding of the regulating signal and its derivatives in given sets. The principle of compensation of disturbances is used, which allowed to obtain

information about parametric uncertainty and external disturbances in the form of some function. To compensate these uncertainties a control signal is designed, the value of which is opposite to the value of perturbations. Thus, for control in the conditions of constraints on the control signal it is required to impose the corresponding constraints on the function containing information on these uncertainties. The conditions on the parameters of the plant, external disturbance, the reference model and the controller when performing the control system is feasible are obtained. The simulation results illustrate the performance of the proposed control scheme.

Keywords: control under uncertainty, saturation of the control signal and its derivatives, compensation of disturbances.

УДК 519.7

ББК Ж 50

DOI: <https://doi.org/10.25728/ubs.2019.77.3>

*Статья представлена к публикации
членом редакционной коллегии П.В. Пакиным.*

Поступила в редакцию 16.08.2018.

Опубликована 31.01.2019.

ВЕРХНИЕ ГРАНИЦЫ МАКСИМАЛЬНОГО ОТКЛОНЕНИЯ ТРАЕКТОРИИ В ЛИНЕЙНЫХ ДИСКРЕТНЫХ СИСТЕМАХ: РОБАСТНАЯ ПОСТАНОВКА¹

Квинто Я. И.², Хлебников М. В.³
(ФГБУН Институт проблем управления
им. В.А. Трапезникова РАН, Москва)

Исследуется практически важный эффект максимального отклонения траектории в линейных динамических системах при ненулевых начальных условиях. Исследование переходного процесса является актуальным и практически значимым направлением в изучении линейных систем. В качестве основного способа получения оценок в настоящей работе используется построение общей квадратичной функции Ляпунова для семейства систем с неопределенностями, а также метод инвариантных эллипсоидов. Все полученные результаты остаются справедливыми также для случая нестационарной неопределенности, поскольку единственное требование к ней – это ее ограниченность в спектральной норме. Поставлены и решены задачи анализа и синтеза, а также получены верхние оценки отклонений для линейных дискретных систем, содержащих структурированную матричную неопределенность. Полученные результаты сформулированы в виде задач полуопределенного программирования, легко решаемых численным образом с помощью стандартных программных пакетов. Применение техники линейных матричных неравенств позволило минимизировать величину отклонений при стабилизации системы с помощью статической линейной обратной связи по состоянию. Результаты численного моделирования демонстрируют низкую степень консерватизма полученных оценок и обладают большим потенциалом для обобщений.

Ключевые слова: линейная дискретная система, максимальное отклонение, структурированная матричная неопределенность, линейные матричные неравенства, функция Ляпунова.

1. Введение

Изучение переходного процесса, т.е. поведения всей траектории x_k устойчивой дискретной линейной системы

$$(1) \quad x_{k+1} = Ax_k, \quad x_k \in \mathbb{R}^n, \quad A \in \mathbb{R}^{n \times n},$$

¹ Исследование выполнено при частичной финансовой поддержке РФФИ (проект №18-08-00140).

² Яна Игоревна Квинто, к.т.н. (yanakvinto@mail.ru).

³ Михаил Владимирович Хлебников, д.ф.-м.н. (khlebnik@ipu.ru).

с ненулевым начальным условием x_0 , является актуальным и практически значимым направлением. При этом одной из важнейших характеристик переходного процесса является величина максимального отклонения траектории системы от нуля:

$$\xi(x_0) = \max_{k=1,2,\dots} \frac{|x_k|}{|x_0|},$$

где $|\cdot|$ — некоторая векторная норма.

В работах [3, 6] исследовался эффект больших отклонений траекторий при ненулевых начальных условиях для непрерывных систем, в том числе с неопределенностями или при наличии внешних возмущений. В частности, было показано, что техника линейных матричных неравенств [7, 9] позволяет синтезировать законы управления, минимизирующие максимальное отклонение траектории в системе, а также дает простые, но достаточно точные оценки верхней границы максимального отклонения. Задачи синтеза для дискретных систем с различного вида неопределенностями рассматривались, например, в [2, 4].

Настоящая статья является естественным развитием идеи, предложенной в [3] для непрерывных динамических систем. А именно, в разделе 2 будут получены верхние оценки отклонений в дискретной системе, содержащей структурированную матричную неопределенность (задача анализа). В разделе 3 будет рассмотрена задача минимизации отклонений в дискретной системе управления при помощи статической линейной обратной связи по состоянию (задача синтеза). В разделе 4 рассматриваются результаты численного моделирования.

В качестве основного способа получения оценок используются построение общей квадратичной функции Ляпунова для семейства систем с неопределенностями, а также метод инвариантных эллипсоидов [5]. В дальнейшем изложении все матричные неравенства понимаются в смысле знакоопределенности матриц.

2. Задача анализа

Рассмотрим линейную динамическую систему в дискретном времени

$$(2) \quad x_{k+1} = (A + F\Delta H)x_k,$$

с заданными матрицами $A \in \mathbb{R}^{n \times n}$, $F \in \mathbb{R}^{n \times p}$, $H \in \mathbb{R}^{q \times n}$, фазовым состоянием $x_k \in \mathbb{R}^n$, ненулевым начальным условием x_0 и матричной неопределенностью

$$\Delta \in \mathbb{R}^{p \times q}: \quad \|\Delta\|_2 \leq \gamma.$$

Единственное требование к матричной неопределенности Δ — ее ограниченность по норме. Таким образом, все полученные ниже результаты остаются справедливыми и для случая нестационарной неопределенности $\Delta(t)$.

Матрица A системы (2) предполагается шуровской, т.е. все ее собственные значения лежат внутри единичного круга.

Для дискретной системы максимальное отклонение траектории представляется выражением

$$\begin{aligned} \widehat{\xi} &= \max_{\|\Delta\|_2 \leq \gamma} \xi(\Delta) = \max_{\|\Delta\|_2 \leq \gamma} \max_{k=1,2,\dots} \max_{|x_0|_2=1} |x_k|_2 = \\ &= \max_{\|\Delta\|_2 \leq \gamma} \max_{k=1,2,\dots} \|A^k\|_2. \end{aligned}$$

Получение оценок величины $\widehat{\xi}$ представляет собой весьма сложную задачу [6]. Ниже мы получим простые и пригодные для практического применения верхние оценки отклонения путем построения общей квадратичной функции Ляпунова с применением техники линейных матричных неравенств.

Как хорошо известно, достаточное условие робастной квадратичной устойчивости семейства (2) состоит в наличии общей квадратичной функции Ляпунова. А именно, выполнение неравенства Ляпунова

$$(3) \quad (A + F\Delta H)P(A + F\Delta H)^T - P \prec 0$$

с некоторой матрицей $P \succ 0$ при всех допустимых значениях неопределенности Δ означает, что у семейства (2) есть общая квадратичная функция Ляпунова

$$V(x) = x^T P^{-1} x.$$

Последовательно применяя лемму Шура и лемму Питерсена [11], представим матричное неравенство (3) в виде эквивалентного линейного матричного неравенства относительно скалярной переменной ε и матричной переменной $P \succ 0$:

$$(4) \quad \begin{pmatrix} P - \varepsilon\gamma^2 FF^\top & AP & 0 \\ PA^\top & P & PH^\top \\ 0 & HP & \varepsilon I \end{pmatrix} \succ 0.$$

Таким образом, разрешимость неравенства (4) является достаточным условием робастной квадратичной устойчивости семейства (2), см. подробнее [7].

Далее, рассмотрим эллипсоид

$$\mathcal{E} = \{x \in \mathbb{R}^n: \quad x^\top P^{-1}x \leq 1\}$$

с матрицей P , удовлетворяющей условию (4). Эллипсоид \mathcal{E} является инвариантным, т.е. траектория системы (2), начинаясь в нем ($x(0) \in \mathcal{E}$), будет оставаться в этот эллипсоиде во всех моменты времени t . Следовательно, если эллипсоид \mathcal{E} содержит единичный шар

$$\mathcal{B} = \{x \in \mathbb{R}^n: \quad |x|_2 \leq 1\},$$

то для любого начального условия из шара \mathcal{B} траектория системы не покинет эллипсоид \mathcal{E} , и в каждый момент времени для ее 2-нормы верна оценка

$$|x(t)|_2 \leq \lambda_{\max}(P) = \sqrt{\|P\|_2}.$$

Условие $\mathcal{B} \subseteq \mathcal{E}$ эквивалентно требованию $P \succcurlyeq I$, следовательно, получаем задачу

$$\min \|P\|_2$$

при ограничениях

$$\begin{pmatrix} P - \varepsilon\gamma^2 FF^\top & AP & 0 \\ PA^\top & P & PH^\top \\ 0 & HP & \varepsilon I \end{pmatrix} \succ 0, \quad P \succcurlyeq I,$$

где переменными являются матрица $P = P^\top \in \mathbb{R}^{n \times n}$ и скаляр ε .

Итак, установлен следующий результат.

Теорема 1. Пусть \hat{P} – решение задачи выпуклой оптимизации

$$\min \|P\|_2$$

при ограничениях

$$\begin{pmatrix} P - \varepsilon\gamma^2 FF^\top & AP & 0 \\ PA^\top & P & PH^\top \\ 0 & HP & \varepsilon I \end{pmatrix} \succ 0, \quad P \succcurlyeq I,$$

относительно матричной переменной $P = P^\top \in \mathbb{R}^{n \times n}$ и скалярной переменной ε .

Тогда для решений системы (2) при всех допустимых неопределенностях Δ справедлива оценка отклонения

$$\hat{\xi} \leq \sqrt{\|\hat{P}\|_2}.$$

Задача, сформулированная в теореме 1, представляет собой задачу полуопределенного программирования, легко решаемую численно.

В рамках рассматриваемого подхода нетрудно вычислить радиус квадратичной устойчивости семейства (2), т.е. максимальный размах γ_{\max} неопределенности Δ такой, что при всех $\gamma < \gamma_{\max}$ у семейства (2) имеется общая квадратичная функция Ляпунова:

$$\gamma_{\max} = \sup\{\gamma: (A + F\Delta H)P(A + F\Delta H)^\top - P \prec 0 \text{ при некотором } P \text{ и всех } \Delta: \|\Delta\| \leq \gamma\}.$$

Соответствующий результат дается следующим утверждением, см. [7].

Лемма 1. Пусть $\hat{\rho}$ – решение задачи полуопределенного программирования

$$\max \rho$$

при ограничениях

$$\begin{pmatrix} P - \rho FF^\top & AP & 0 \\ PA^\top & P & PH^\top \\ 0 & HP & I \end{pmatrix} \succ 0, \quad P \succ 0,$$

относительно матричной переменной $P = P^T \in \mathbb{R}^{n \times n}$ и скалярной переменной ρ . Тогда радиус квадратичной устойчивости семейства (2) равен

$$\gamma_{\max} = \sqrt{\hat{\rho}}.$$

3. Задача синтеза

В этом разделе мы зададимся целью поиска стабилизирующей обратной связи для линейной дискретной системы управления. По сравнению с предыдущим разделом немного усложним постановку и рассмотрим систему управления вида

$$(5) \quad x_{k+1} = (A + F\Delta_1 H + M\Delta_2 N)x_k + Bu_k,$$

с заданными матрицами $A \in \mathbb{R}^{n \times n}$, $F \in \mathbb{R}^{n \times p}$, $H \in \mathbb{R}^{q \times n}$, $M \in \mathbb{R}^{n \times r}$, $N \in \mathbb{R}^{s \times m}$, $B \in \mathbb{R}^{n \times m}$, фазовым состоянием $x_k \in \mathbb{R}^n$, ненулевым начальным условием x_0 , управлением $u_k \in \mathbb{R}^m$ и матричными неопределенностями Δ_1 и Δ_2 такими, что

$$\Delta_1 \in \mathbb{R}^{p \times q}: \quad \|\Delta_1\|_2 \leq \gamma_1,$$

$$\Delta_2 \in \mathbb{R}^{r \times s}: \quad \|\Delta_2\|_2 \leq \gamma_2;$$

пара (A, B) предполагается управляемой.

Как показано в [7], размах неопределенностей Δ_1 и Δ_2 можно считать общим за счет масштабирования «обрамляющих» матриц F , H , M и N ; обозначим его γ :

$$\|\Delta_i\|_2 \leq \gamma, \quad i = 1, 2.$$

Итак, будем искать стабилизирующую статическую линейную обратную связь по состоянию

$$(6) \quad u_k = Kx_k, \quad K \in \mathbb{R}^{m \times n},$$

минимизирующую величину максимального отклонения в замкнутой системе.

Замкнув систему (5) обратной связью (6), приходим к замкнутой системе

$$(7) \quad x_{k+1} = (A + BK + F\Delta_1 H + M\Delta_2 N)x_k.$$

Построим квадратичную функцию Ляпунова для семейства (7) такую, что ее матрица минимальна по спектральной норме. При этом нам придется иметь дело с условием

$$(A + BK + F\Delta_1 H + M\Delta_2 N)P \times \\ \times (A + BK + F\Delta_1 H + M\Delta_2 N)^\top - P \prec 0,$$

поэтому нам понадобится модификация леммы Питерсена на случай нескольких неопределенностей, см. [7], справедливую в достаточной части. Приведем её в следующей формулировке.

Лемма 2 (Лемма Питерсена для нескольких неопределенностей).

Пусть $G = G^\top \in \mathbb{R}^{n \times n}$, $M_i \in \mathbb{R}^{n \times p_i}$, $N_i \in \mathbb{R}^{q_i \times n}$, $i = 1, \dots, \ell$.

Если существуют числа $\varepsilon_1, \dots, \varepsilon_\ell$ такие, что

$$\begin{pmatrix} G - \sum_{i=1}^{\ell} \varepsilon_i M_i M_i^\top & N_1^\top & \dots & N_\ell^\top \\ * & \varepsilon_1 I & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \dots & \varepsilon_\ell I \end{pmatrix} \succ 0,$$

то матричное неравенство

$$G + \sum_{i=1}^{\ell} (M_i \Delta_i N_i + N_i^\top \Delta_i^\top M_i^\top) \succ 0$$

справедливо для всех $\Delta_i \in \mathbb{R}^{p_i \times q_i}$: $\|\Delta_i\|_2 \leq 1$, $i = 1, \dots, \ell$.

Воспользовавшись леммой 2, приходим к условию

$$\begin{pmatrix} P - \gamma^2(\varepsilon_1 F F^\top + \varepsilon_2 M M^\top) & (A + BK)P & 0 & 0 \\ P(A + BK)^\top & P & P H^\top & P^\top N^\top \\ 0 & H P & \varepsilon_1 I & 0 \\ 0 & N P & 0 & \varepsilon_2 I \end{pmatrix} \succ 0.$$

Введя вспомогательную матричную переменную $Y = KP$ и следуя идее доказательства теоремы 1, окончательно приходим к задаче:

$$\min \|P\|_2$$

при ограничениях

$$\begin{pmatrix} P - \gamma^2(\varepsilon_1 FF^\top + \varepsilon_2 MM^\top) & AP + BY & 0 & 0 \\ PA^\top + Y^\top B^\top & P & PH^\top & P^\top N^\top \\ 0 & HP & \varepsilon_1 I & 0 \\ 0 & NP & 0 & \varepsilon_2 I \end{pmatrix} \succ 0,$$

$$P \succ I,$$

относительно матричных переменных P , Y и скалярных переменных $\varepsilon_1, \varepsilon_2$.

При этом в силу $P \succ 0$ матрица регулятора K восстанавливается единственным образом:

$$K = YP^{-1}.$$

Итак, получено следующее утверждение.

Теорема 2. Пусть \hat{P} , \hat{Y} — решение задачи выпуклой оптимизации

$$\min \|P\|_2$$

при

$$\begin{pmatrix} P - \gamma^2(\varepsilon_1 FF^\top + \varepsilon_2 MM^\top) & AP + BY & 0 & 0 \\ PA^\top + Y^\top B^\top & P & PH^\top & P^\top N^\top \\ 0 & HP & \varepsilon_1 I & 0 \\ 0 & NP & 0 & \varepsilon_2 I \end{pmatrix} \succ 0,$$

$$P \succ I,$$

относительно матричных переменных $P = P^\top \in \mathbb{R}^{n \times n}$, $Y \in \mathbb{R}^{m \times n}$ и скалярных переменных $\varepsilon_1, \varepsilon_2$.

Тогда для решений системы (5), замкнутой регулятором (6) с матрицей

$$\hat{K} = \hat{Y}\hat{P}^{-1},$$

справедлива оценка отклонения

$$\hat{\xi} \leq \sqrt{\|\hat{P}\|_2}$$

при всех допустимых значениях неопределенностей Δ_1 и Δ_2 .

Сделаем важное замечание: при используемом подходе, безусловно, необходимо обращать внимание и на иные показатели качества системы (накладывая на них при необходимости дополнительные ограничения). В рамках настоящей статьи такие ограничения не рассматриваются, однако авторы планируют исследовать этот аспект в последующих публикациях.

4. Пример

Рассмотрим задачу о цепочке поставок, аналогичную приведенной в [8]: в месяц с номером k ($k = 1, 2, \dots$) предприятие S закупает количество u_k продукции (сырья) типа 1, доля δ_1 которого затем отбрасывается как отходы, а доля α_1 отправляется на предприятие P для производства продукции типа 2. Далее P отправляет долю α_2 своей продукции предприятию R , выпускающему продукцию типа 3, а долю δ_2 составляют отходы. Предприятие R ежемесячно отправляет производителю P долю β_3 продукции типа 3, а долю γ_3 продает сторонним предприятиям.

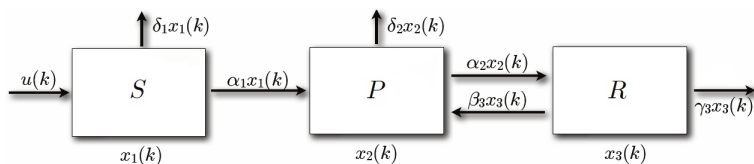


Рис. 1. Цепочка поставок

Пусть x_{ik} – объемы продукции типа i в месяц с номером k , $i = 1, 2, 3$, $k = 1, 2, \dots$ (см. рис. 1). Объемы продукции, имеющиеся на предприятиях S , P и R к началу расчетного периода, обозначим как x_{i0} , $i = 1, 2, 3$ соответственно. Тогда получаем мо-

дель поставок в виде дискретной линейной системы

$$\begin{aligned}x_{1,k+1} &= (1 - \alpha_1 - \delta_1)x_{1k} + u_k, \\x_{2,k+1} &= \alpha_1 x_{1k} + (1 - \alpha_2 - \delta_2)x_{2k} + \beta_3 x_{3k}, \\x_{3,k+1} &= \alpha_2 x_{2k} + (1 - \beta_3 - \gamma_3)x_{3k}.\end{aligned}$$

Предположим, что ежемесячные объемы отходов δ_1 и δ_2 соответственно в пунктах S и P известны неточно, и будем трактовать это в виде наличия неопределенностей Δ_1 и Δ_2 в описании системы:

$$(8) \quad \begin{aligned}x_{1,k+1} &= (1 - \alpha_1 - (\delta_1 + \Delta_1))x_{1k} + u_k, \\x_{2,k+1} &= \alpha_1 x_{1k} + (1 - \alpha_2 - (\delta_2 + \Delta_2))x_{2k} + \beta_3 x_{3k}, \\x_{3,k+1} &= \alpha_2 x_{2k} + (1 - \beta_3 - \gamma_3)x_{3k}.\end{aligned}$$

Обозначив вектор фазового состояния

$$x_k = (x_{1k} \quad x_{2k} \quad x_{3k})^\top,$$

получим матричную запись системы (8):

$$x_{k+1} = (A + F\Delta_1 H + M\Delta_2 N)x_k + Bu_k$$

с матрицами

$$A = \begin{pmatrix} 1 - \alpha_1 - \delta_1 & 0 & 0 \\ \alpha_1 & 1 - \alpha_2 - \delta_2 & \beta_3 \\ 0 & \alpha_2 & 1 - \beta_3 - \gamma_3 \end{pmatrix}, \quad B = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix},$$

$$F = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}, \quad H = (1 \quad 0 \quad 0),$$

$$M = \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix}, \quad N = (0 \quad 1 \quad 0),$$

где $x_k \in \mathbb{R}^3$ — состояние объекта, $u_k \in \mathbb{R}$ — управление, $\Delta_1, \Delta_2 \in \mathbb{R}$ — скалярные неопределенности. При этом

$$(9) \quad \begin{aligned} 0 &\leq 1 - \beta_3 - \gamma_3 \leq 1, \\ \alpha_i + \delta_i + \Delta_i &\leq 1, \quad i = 1, 2. \end{aligned}$$

Целью управления является минимизация резких скачков объемов продукции в пунктах S , P и R соответственно.

Примем $\alpha_1 = 0,7$, $\delta_1 = 0,1$, $\alpha_2 = 0,5$, $\delta_2 = 0,2$, $\beta_3 = 0,85$, $\gamma_3 = 0,15$. Тогда из (9) получаем

$$|\Delta_i| \leq \min(|1 - \alpha_i - \delta_i|, |\alpha_i + \delta_i|), \quad i = 1, 2,$$

т.е. $|\Delta_1| \leq 0,2$, $|\Delta_2| \leq 0,3$. Поэтому в качестве γ , максимального допустимого одновременно для Δ_1 и Δ_2 , выберем наименьшее из вычисленных ограничений: $\gamma = 0,2$.

Применяя теорему 2, с помощью пакета `cvx` [10] в среде `MATLAB` находим матрицу

$$\widehat{P} = \begin{pmatrix} 1,3863 & -0,1522 & -0,3593 \\ -0,1522 & 1,7217 & -0,1387 \\ -0,3593 & -0,1387 & 1,4529 \end{pmatrix}$$

квадратичной функции Ляпунова, соответствующую матрицу

$$\widehat{K} = (-0,2868 \quad -0,1667 \quad -0,1053)$$

регулятора (6) и верхнюю оценку

$$\widehat{\xi} = 1,3343$$

максимального отклонения. При этом

$$\max_k \|x_k\| = 1,2581,$$

что свидетельствует о низком консерватизме полученной оценки.

На рис. 2 жирной линией показана динамика величины $\|x_k\|$, а тонкими линиями — изменения фазовых координат замкнутой системы при некотором начальном условии x_0 из единичного шара и неопределенностях

$$\Delta_1^k = 0,2 \operatorname{sign}(\sin(k/2)), \quad \Delta_2^k = 0,2 \operatorname{sign}(\cos(2k)), \quad k = 1, 2, \dots$$

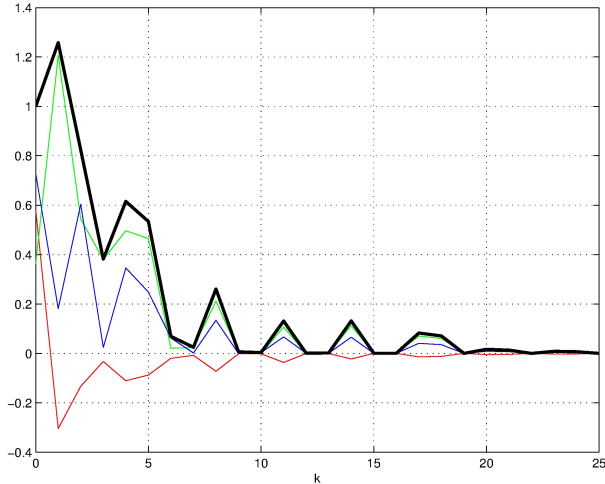


Рис. 2. Отклонения траекторий для системы с цепью поставок между тремя пунктами

5. Заключение

В статье продолжено изучение эффекта больших отклонений траекторий линейных систем с ненулевыми начальными условиями. Опираясь на ранее полученные результаты, установлены простые оценки отклонений для дискретных линейных систем при наличии структурированных матричных неопределенностей в матрице системы. Предложен подход к робастной минимизации отклонений в линейных дискретных системах управления при помощи статической линейной обратной связи по состоянию.

В дальнейшем авторы планируют рассмотреть и иные постановки задач, в частности, при наличии внешних ограниченных возмущений.

Литература

1. БАЛАНДИН Д.В., КОГАН М.М. *Синтез законов управления на основе линейных матричных неравенств*. – М.: ФИЗМАТЛИТ, 2007. – 280 с.
2. ДОРОФЕЕВ Ю.И. *Применение линейных матричных неравенств в задаче синтеза оптимального управления запасами при наличии структурных ограничений* // Міжвідомчий науково-технічний збірник «Адаптивні системи автоматичного управління». – 2015. – №1(26). – С. 13–25.
3. КВИНТО Я.И., ХЛЕБНИКОВ М.В. *Верхние оценки больших отклонений в линейных системах при наличии неопределенности* // Проблемы управления. – 2018. – № 3. – С. 2–7.
4. КОГАН М.М., КРИВДИНА Л.Н. *Синтез многоцелевых линейных законов управления дискретными объектами при интегральных и фазовых ограничениях* // Автоматика и телемеханика. – 2011. – №7. – С. 83–95.
5. ПОЛЯК Б.Т., ТОПУНОВ М.В., ЩЕРБАКОВ П.С. *Идеология инвариантных эллипсоидов в задаче о робастном подавлении ограниченных внешних возмущений* // Стохастическая оптимизация в информатике. – 2007. – Т. 3, №1-1. – С. 51–84.
6. ПОЛЯК Б.Т., ТРЕМБА А.А., ХЛЕБНИКОВ М.В., ЩЕРБАКОВ П.С., СМІРНОВ Г.В. *Большие отклонения в линейных системах при ненулевых начальных условиях* // Автоматика и телемеханика. – 2015. – № 6. – С. 18–41.
Англ.: POLYAK B.T., TREMBA A.A., KHLEBNIKOV M.V., SHCHERBAKOV P.S., SMIRNOV G.V. *Large Deviations in Linear Control Systems with Nonzero Initial Conditions* // Automation and Remote Control. – 2015. – Vol. 76, No. 6. – P. 957–976.

7. ПОЛЯК Б.Т., ХЛЕБНИКОВ М.В., ЩЕРБАКОВ П.С. *Управление линейными системами при внешних возмущениях: Техника линейных матричных неравенств.* – М.: ЛЕНАНД, 2014.
8. BEMPORAD A. *Discrete-Time Linear Systems. Lecture Notes on "Automatic Control". Part 1.* – University of Trento, 2010. – URL: http://cse.lab.imtlucca.it/~bemporad/teaching/ac/pdf/04a-TD_sys.pdf.
9. BOYD S., EL GHAOUI L., FERON E., BALAKRISHNAN V. *Linear Matrix Inequalities in Systems and Control Theory.* – Philadelphia: SIAM, 1994.
10. GRANT M., BOYD S. *CVX: Matlab Software for Disciplined Convex Programming, Version 2.1.* – URL: <http://cvxr.com/cvx/>.
11. PETERSEN I.R. *A Stabilization Algorithm for a Class of Uncertain Linear Systems* // *Systems and Control Letters.* – 1987. – Vol. 8. – P. 351–357.

UPPER BOUNDS OF LARGE DEVIATIONS IN LINEAR DISCRETE-TIME SYSTEMS: THE ROBUST STATEMENT

Yana Kvinto, V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Senior Researcher, Ph.D (yanakvinto@mail.ru).

Mikhail Khlebnikov, V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Laboratory Head, Doctor of Science (khlebnik@ipu.ru).

Abstract: The paper is devoted to the study of the important effect of large deviations in linear dynamical systems with nonzero initial conditions. The study of transients is actual and practically significant direction in the linear systems theory. The common Lyapunov quadratic function for the family of systems with uncertainties and the invariant ellipsoids approach are used in the article as main technical tools. All the results obtained are also applicable for non-stationary uncertainties: the only condition for an uncertainty is its spectral norm constraint. The analysis and design problems are considered, and the upper bounds of deviations for linear discrete-time systems with structured matrix uncertainties are obtained. The obtained results have the form of semi-definite programs, which are easy to solve numerically via standart software packages. Using the technique of linear matrix inequalities, the problem of minimization the magnitude of deviations while stabilizing the system via the linear static state feedback was investigated. Numerical simulations demonstrate the low degree of conservatism of the obtained approach. The results have a great potential for generalizations.

Keywords: linear discrete-time system, large deviations, structured matrix uncertainty, linear matrix inequalities, Lyapunov function.

УДК 519.7

ББК 32.965

DOI: <https://doi.org/10.25728/ubs.2019.77.4>

Статья представлена к публикации членом редакционной коллегии И.Б. Фуртатом.

Поступила в редакцию 28.08.2018.

Дата опубликования 31.01.2019.

СТАТИСТИЧЕСКИЕ МОДЕЛИ ПРОЦЕССОВ ПРИЁМА И УВОЛЬНЕНИЯ СОТРУДНИКОВ

Белов М. В.¹

(Компания IBS, Москва)

Работа продолжает исследования в области математического моделирования экономики человеческого капитала и жизненного цикла сотрудников, а также изучения зависимости экономических эффектов бизнеса от этих факторов. Исследуется движение сотрудников между фирмой и рынком труда. Рассматриваются характеристики процесса жизненного цикла сотрудника (ЖЦС), понимаемого как весь интервал времени с момента, когда сотрудник принял предложение о работе в фирме до момента его увольнения. Продемонстрирована существенность влияния продолжительности среднего времени работы в фирме, длительностей фазы адаптации сотрудника при приёме на работу и фазы, предшествующей увольнению, на экономику фирмы. Процесс увольнений сотрудников проанализирован непараметрическими методами. В результате показана невозможность высоко достоверного оценивания функции распределения вероятностей времени работы сотрудника в фирме на основании наблюдений. Предложены процедуры статического последовательного анализа для выявления момента изменения характеристик жизненного цикла сотрудников – вероятностей принятия предложений о приёме на работу и интенсивности увольнения по собственному желанию. Разработанные методы и полученные результаты проиллюстрированы численными примерами из практики автора в сфере экономики человеческого капитала и управления персоналом.

Ключевые слова: жизненный цикл сотрудника, процесс увольнений, цензурированные выборки, последовательный критерий, разладка.

1. Введение. Актуальность проблемы

Ключевым элементом любого бизнеса (фирмы, предприятия) являются сотрудники, поэтому вопросы управления человеческими ресурсами всегда находились и находятся в центре внимания и менеджеров-практиков, и исследователей. Значительное количество исследований посвящено не только работе с персоналом, вовлеченности, кадровому делопроизводству, но также экономике труда и учёту человеческих ресурсов (HR accounting) – см., например, [8, 17, 18, 19]. В последние десяти-

¹ Михаил Валентинович Белов, к.т.н. (mbelov@ibs.ru).

летия популярной стала проблематика «управления человеческим капиталом» [14, 16, 22, 23], подчёркивающая центральную роль персонала в любом бизнесе. Вместе с тем вопросы влияния характеристик деятельности сотрудников на экономический эффект¹, получаемый фирмой, изучены недостаточно полно.

Данная статья продолжает работы в области математического моделирования экономики человеческого капитала и жизненного цикла сотрудника [1, 2], а также изучения зависимости экономических эффектов бизнеса от этих факторов. Исследуется такой важный аспект бизнеса как движение сотрудников между фирмой и рынком труда, в частности, увольнения сотрудников по собственному желанию (в дальнейшем «увольнения», если не оговорено особо), предлагаются модели и методы исследования этого процесса.

Во втором разделе вводится математическая модель и доказывается утверждение о зависимости экономических эффектов бизнеса от характеристик ЖЦС, в частности, вероятности принятия предложений о работе и распределения времени работы в фирме. Третий раздел посвящён модели процесса увольнений и непараметрическим методам его статистического анализа. В четвёртом разделе предлагаются методы статистического оценивания и анализа характеристик жизненного цикла сотрудников. В заключении обсуждаются полученные методы и приводятся примеры практического использования предлагаемых подходов. В приложения П1–П5 вынесены выкладки, обосновывающие используемые соотношения.

¹ *Экономический эффект, следуя современным экономическим подходам [3, 5, 12], понимается как разность между результатом экономической деятельности и затратами, произведенными для его получения.*

2. Модель зависимости экономических эффектов бизнеса от характеристик жизненного цикла сотрудников

Будем представлять бизнес, как и в статьях [1, 2], в виде двух процессов с дискретным временем (дискретные единицы времени будем называть интервалами):

- процесса продаж продукции или услуг, порождающего потребность в персонале;
- процесса движения персонала.

Персонал рассматриваем как организованную совокупность специалистов, обладающих определёнными знаниями, навыками, квалификацией и мотивированных на решение ставящихся перед ними задач. Для описания процесса движения персонала (как и в [1, 2]) используем два понятия:

- «функциональных домов» (ФД) – групп специалистов, обладающих идентичными функциональными возможностями для решения бизнес-задач;
- жизненного цикла сотрудника, понимая под ним весь интервал времени с момента, когда сотрудник принял предложение о работе в фирме, до момента его увольнения из фирмы.

Проиллюстрируем значимость исследуемого в работе фактора – характеристик ЖЦС – на примере стационарного бизнеса для случая единственного ФД. Стационарность бизнеса будем понимать как постоянство характеристик процессов продаж и движения персонала, т.е. считаем среднюю потребность в персонале, интенсивность увольнений сотрудников и, соответственно, приёма постоянными. Единственность ФД предполагает однородность предмета бизнеса (продукта или сервиса) и, соответственно, однородность основного персонала. Запишем выражение для экономического эффекта V , получаемого за один интервал времени, сразу используя соотношение (5) из [1], полученное с учётом следующих ключевых предположений (сами предположения и адекватность их применения подробно обсуждены в [1]):

- стационарности условий бизнеса;
- статистической эквивалентности¹ и независимости поведения сотрудников друг от друга и на различных интервалах времени.

$$V = n_{\text{пред}} \pi \sum_{\theta=0}^{\infty} (S(\theta) \varphi_{\theta}) - n_{\text{пред}} S_{\text{пред}} - n_{\text{пред}} \pi S_{\text{найм}} + V^*,$$

где

$n_{\text{пред}}$ – среднее количество предложений о работе кандидатам на приём актуальных в течение одного интервала времени;

π – вероятность того, что кандидат примет предложение в течение одного интервала времени;

φ_{θ} – доля сотрудников, отработавших фирме ровно θ интервалов времени с момента приёма на работу до увольнения;

$S(\theta)$ – средний экономический эффект за всё время работы сотрудника, отработавшего в фирме θ интервалов времени, равный его экономическому результату за минусом издержек на его содержание – компенсационного пакета, налогов и социальных выплат, затрат на обучение и др.;

$S_{\text{пред}}$ – средние затраты на подготовку одного предложения о работе, включая поиск кандидатов;

$S_{\text{зам}}$ – средние затраты на замещение уволившегося сотрудника включая транзакционные издержки на оформление, начальную адаптацию и другие²;

V^* – часть экономического эффекта компании, не зависящая непосредственно от персонала (например, затраты на сырье, электроэнергию и т.д.).

Будем считать время работы сотрудника в фирме (от момента приёма до увольнения) случайной величиной, принимающей целочисленные неотрицательные значения. Набор

¹ Статистическая эквивалентность сотрудников здесь и далее понимается в смысле возможности описания их поведения одинаково распределёнными случайными величинами.

² В состав этих затрат также имеет смысл включать стоимость временного привлечения человеческих ресурсов на период поиска постоянных. Например, за счёт организации сверхурочных работ штатных сотрудников.

$\{\varphi_0; \varphi_1; \varphi_2; \varphi_3; \dots\}$ является распределением¹ вероятностей времени работы сотрудника в фирме.

Очевидно, что $n_{\text{пред}}\pi$ будет равно среднему числу принимаемых на работу в течение каждого интервала новых сотрудников. А оно, в свою очередь, равно среднему числу увольняющихся в предположении стационарности процессов.

$t_{\text{cp}} = \sum_{\theta=0}^{\infty} \theta \varphi_{\theta}$ – среднее время работы сотрудника в фирме,

$n_{\text{пред}} \pi t_{\text{cp}} = n_{\text{пред}} \pi \sum_{\theta=0}^{\infty} \theta \varphi_{\theta}$ – среднее число работающих сотрудников,

а $\sum_{\theta=0}^{\infty} (S(\theta) \varphi_{\theta})$ – средний эффект от сотрудника за весь период работы в фирме (усреднённый по длительности срока работы в фирме).

Важно заметить, что менеджмент фирмы, если он действует рационально, обеспечивает значение численности сотрудников как можно более близким к уровню, необходимому для выполнения всего объёма деятельности соответственно объёму продаж [1, 2]. Поэтому среднее число работающих сотрудников определяется потребностью со стороны продаж:

$$(1) \quad n_{\text{пред}} \pi t_{\text{cp}} = D,$$

где D – средний объём продаж за интервал времени, отнесённый к средней результативности одного сотрудника, т.е. объём продаж за интервал времени, выраженный в количестве сотрудников, необходимых для его реализации, или средняя численность ФД, необходимая для реализации бизнеса данного объёма.

Экономический эффект $S(\cdot)$ от сотрудника за всё время работы в фирме, за весь жизненный цикл, складывается из эффектов

на каждом из интервалов $S(\theta) = \sum_{\tau=1}^{\theta} s(\tau)$ и в общем случае

может быть представлен следующим образом. Жизненный цикл

¹ В силу естественной ограниченности срока человеческой жизни, случайная величина «время работы сотрудника в фирме» будет иметь конечные моменты любого порядка.

сотрудника логично разделить на три фазы: а) адаптации, б) продуктивной работы и в) предшествующей увольнению. Причём также логично предполагать, что экономический эффект сотрудника на фазах а) и в) ниже, чем на фазе продуктивной работы:

$$(2) \quad s(\tau) = \begin{cases} s_{\text{прд}} - \Delta s_{\text{ад}} & \text{при } 0 \leq \tau \leq t_{\text{ад}}, \\ s_{\text{прд}} & \text{при } t_{\text{ад}} < \tau \leq \theta - t_{\text{пр.ув}}, \\ s_{\text{прд}} - \Delta s_{\text{ув}} & \text{при } \max\{t_{\text{ад}}; \theta - t_{\text{пр.ув}}\} < \tau \leq \theta. \end{cases}$$

Здесь

$s_{\text{прд}}$ – средний экономический эффект на продуктивной фазе;

$\Delta s_{\text{ад}}$ – среднее снижение эффекта на фазе адаптации;

$t_{\text{ад}}$ – длительность фазы адаптации;

$\Delta s_{\text{ув}}$ – среднее снижение эффекта на фазе, предшествующей увольнению;

$t_{\text{пр.ув}}$ – продолжительность фазы, предшествующей увольнению.

Подставив (1) и (2) в выражение для экономического эффекта, получим (соответствующие выкладки приведены в приложении П1):

$$(3) \quad V = D s_{\text{прд}} - D \left(\Delta s_{\text{ад}} \frac{t_{\text{ад}}^*}{t_{\text{ср}}} + \Delta s_{\text{ув}} \frac{t_{\text{пр.ув}}^*}{t_{\text{ср}}} + S_{\text{зам}} \frac{1}{t_{\text{ср}}} + \frac{S_{\text{прел}}}{\pi} \frac{1}{t_{\text{ср}}} \right) + V^*,$$

где $t_{\text{ад}}^* = \sum_{\theta=0}^{t_{\text{ад}}} \theta \varphi_{\theta} + t_{\text{ад}} \sum_{\theta=t_{\text{ад}}+1}^{\infty} \varphi_{\theta}$ и $t_{\text{пр.ув}}^* = \sum_{\theta=t_{\text{ад}}+1}^{t_{\text{ад}}+t_{\text{пр.ув}}} \theta \varphi_{\theta} + t_{\text{пр.ув}} \sum_{\theta=t_{\text{ад}}+t_{\text{пр.ув}}+1}^{\infty} \varphi_{\theta}$.

Проанализируем с помощью (3), как характеристики деятельности сотрудников влияют на экономический эффект фирмы V .

Третий элемент выражения (3) V^* не зависит от персонала и в дальнейшем рассматриваться не будет.

Первый элемент выражения (3) $D s_{\text{прд}}$ пропорционален объему продаж – требуемому количеству рабочих мест, замещённых сотрудниками, D , и среднему экономическому эффекту $s_{\text{прд}}$ от сотрудника в течение одного интервала времени на продуктивной фазе. Требуемое количество рабочих мест D является

внешним по отношению к деятельности сотрудников и потому не зависит от характеристик их деятельности. Средний экономический эффект $s_{\text{прд}}$ зависит от результативности сотрудников, однако в рамках принятой концепции ФД как групп специалистов, обладающих идентичными функциональными возможностями, $s_{\text{прд}}$ имеет смысл также считать постоянным в рамках данной модели.

Второй элемент выражения (3) пропорционален D и сумме четырёх слагаемых, рассмотрим их.

Величины снижения эффекта $\Delta s_{\text{ад}}$ и $\Delta s_{\text{ув}}$ в силу сделанного выше замечания об идентичности сотрудников также будем считать постоянными. Издержки на приём каждого сотрудника $S_{\text{найм}}$ и подготовку одного предложения о работе $S_{\text{пред}}$ не зависят от деятельности сотрудников. Фактически все эти параметры являются априорными нормативными характеристиками функциональных домов, которые в свою очередь определяются потребностями и характеристиками бизнеса.

Таким образом, в рамках принятой концепции ФД экономический эффект V зависит от вероятности π и временных характеристик жизненного цикла сотрудников (ЖЦС) $t_{\text{ад}}^*$, $t_{\text{пр.ув}}^*$ и $t_{\text{ср}}$, зависящих от $t_{\text{ад}}$, $t_{\text{пр.ув}}$ и распределения $\left\{ \varphi_{\theta}; \theta=0, 1, \dots, \infty; \sum_{\theta=0}^{\infty} \varphi_{\theta} \equiv 1 \right\}$.

Будем обозначать зависимость экономического эффекта от характеристик ЖЦС как $V(\pi, \{ \varphi_{\theta} \}, t_{\text{ад}}, t_{\text{пр.ув}})$. В приложении П2 выполнен анализ функции $V(\pi, \{ \varphi_{\theta} \}, t_{\text{ад}}, t_{\text{пр.ув}})$ от своих аргументов. Проинтерпретируем полученные свойства с точки зрения бизнеса.

Во-первых, $\frac{\partial}{\partial \pi} V(\pi, \{ \varphi_{\theta} \}, t_{\text{ад}}, t_{\text{пр.ув}}) > 0$. Причём при уменьшении π экономический эффект становится отрицательным, начиная с некоторых π . А при стремлении π к нулю V может неограниченно уменьшаться, будучи отрицательным $\lim_{\pi \rightarrow 0} \left\{ V(\pi, \{ \varphi_{\theta} \}, t_{\text{ад}}, t_{\text{пр.ув}}) \right\} = -\infty$. Содержательно это означает, что низкая эффективность процедур поиска и найма новых сотруд-

ников (низкая доля кандидатов, принимающих предложение по сравнению с рассматриваемыми и приглашаемыми) – малые значения π – негативно сказывается на экономическом эффекте бизнеса V . Хотя, конечно же, неограниченное – неконтролируемое – уменьшение экономического эффекта, в том числе из-за неэффективного найма, является скорее исключением из правил.

$$\text{Во-вторых,} \quad \frac{\partial}{\partial t_{\text{ад}}} V(\pi, \{\varphi_\theta\}, t_{\text{ад}}, t_{\text{пр.ув}}) < 0 \quad \text{и}$$

$$\frac{\partial}{\partial t_{\text{пр.ув}}} V(\pi, \{\varphi_\theta\}, t_{\text{ад}}, t_{\text{пр.ув}}) < 0. \text{ С точки зрения бизнеса монотонное}$$

убывание экономического эффекта при росте продолжительности интервала адаптации и/или фазы снижения производительности, предшествующей увольнению, является интуитивно очевидным, так как это «вынужденные» фазы ЖЦС, основной эффект от сотрудника фирма получает не в течение них.

Вид распределения вероятностей времени работы сотрудников в фирме $\{\varphi_\theta\}$ влияет на экономический эффект более сложным образом, для исследования этого влияния использованы «малые изменения» $\delta\varphi_\theta(x, i)$. В приложении П2 показано, что влияние на экономический эффект «малых изменений» $\delta\varphi_\theta(x, i)$ в области $0 \leq i < t_{\text{ад}} + t_{\text{пр.ув}}$ проявляется неоднозначно. При длительности работы менее $t_{\text{ад}} + t_{\text{пр.ув}}$ сотрудник не «успевает» достичь продуктивной фазы и, следовательно, произвести требуемый фирме эффект. Наличие таких сотрудников свидетельствует о неэффективных процессах и найма, и адаптации, и удержания. То есть сотрудников, увольняющихся ранее $t_{\text{ад}} + t_{\text{пр.ув}}$ интервалов должно быть как можно меньше, «в идеале» – ни одного. Поэтому детальный анализ экономики жизненного цикла таких сотрудников не имеет смысла – достаточно оценивать их долю и стремиться уменьшить эту долю.

Однако для «малых изменений» в области времени работы больше $t_{\text{ад}} + t_{\text{пр.ув}}$ справедливо $\frac{\partial}{\partial t_{\text{сп}}} V(\pi, \{\varphi_\theta\}, t_{\text{ад}}, t_{\text{пр.ув}}) > 0$ (вообще говоря, независимо от вида распределения вероятностей време-

ни работы в фирме $\{\varphi_\theta\}$), что также соответствует логике бизнеса: увеличение среднего времени работы в этом случае увеличивает относительную продолжительность продуктивной фазы по отношению к общей продолжительности ЖЦС.

Выполненный анализ выражения (3) для экономического эффекта легко расширяется на случай нескольких функциональных домов, а также на случай известного нестационарного процесса $D(t)$ изменения потребности в персонале, необходимом для реализации всех продаж.

Таким образом, на основании выполненного анализа можно сформулировать **утверждение 1** «О зависимости экономического эффекта фирмы от характеристик жизненного цикла сотрудников в рамках концепции функциональных домов»:

Экономический эффект фирмы (независимо от вида распределения вероятностей времени работы сотрудников фирмы до увольнения по собственному желанию) монотонно возрастает при:

- увеличении вероятности того, что кандидат принимает предложение о приёме на работу;
- уменьшении длительности фазы адаптации и длительности фазы, предшествующей увольнению;
- увеличении среднего времени работы в фирме (за счет изменения вероятностей увольнения позднее суммарной длительности фазы адаптации и фазы, предшествующей увольнению).

Сформулированное утверждение 1 иллюстрирует важность оперативного контроля характеристик ЖЦС для обеспечения эффективного функционирования фирмы, то есть актуальность задачи моделирования и контроля характеристик ЖЦС, поставленной во введении.

3. Модель процесса увольнения и непараметрические методы его анализа

Рассмотрим более детально процесс увольнений сотрудников из функционального дома с точки зрения отвечающего за него менеджера.

ФД рассматривается как пул эквивалентных и статистически независимых сотрудников, численность которых необходимо поддерживать на требуемом (в общем случае меняющемся со временем) уровне. То есть в случае увольнения сотрудника по собственному желанию его место восполняется в течение того же интервала времени, в котором он уволился¹, что в целом пренебрежимо мало сказывается на значении численности. Адекватной метафорой в данном случае является следующая: ФД состоит из одинаковых функциональных ролей, каждая из которых постоянно заполнена сотрудниками, поведение которых статистически эквивалентно и независимо друг от друга. Тогда события увольнений сотрудников на каждой из ролей образуют независимые процессы, заключающиеся в том, что если предыдущее увольнение (и восполнение) произошло на t_d -м интервале времени, то следующее увольнение произойдет на t -м интервале с вероятностью φ_{t-t_d} . Так как менеджер ФД знает, в какой момент t_d произошло предыдущее увольнение на каждой из функциональных ролей, можно, используя распределение $\{\varphi_\theta\}$, определять вероятность следующего увольнения на t -м интервале как $\varphi_{t-t_d} \left(\sum_{\theta=t-t_d}^{\infty} \varphi_\theta \right)^{-1}$ (если между интервалами t_d и t увольнений на данной роли не было).

Использование вычисленных таким образом вероятностей² для всех ролей ФД удобно для получения и аналитического исследования различных зависимостей. Однако их использование затрудняется тем, что распределение вероятностей $\{\varphi_\theta\}$ времени работы сотрудника априори не известно и основывается не на

¹ Такая модель замещения увольняющихся корректна, например, при: а) неограниченно большом пуле кандидатов; б) возможности фирмы формировать также неограниченно большое количество предложений о приёме на работе; в) после чего отбирать требуемое количество новых сотрудников.

² Вычисленные таким образом вероятности отвечают случаю невозможности более одного увольнения на каждой роли в течение одного интервала времени, то есть $\varphi_0 = 0$. Если это не так, значения вероятностей одного и более увольнений будут другими, что, однако, никак не сказывается на всех остальных результатах работы и выводах.

объективных фундаментальных закономерностях, а на активном выборе субъектов-сотрудников. То есть распределение $\{\varphi_\theta\}$ (или его характеристики) необходимо оценивать на основании наблюдений. Кроме того, есть основания считать, что бизнес-условия отличаются определённой изменчивостью: можно говорить о существовании сменяющих друг друга периодов стационарности с характерной продолжительностью в нескольких лет, что принципиально ограничивает доступные интервалы наблюдений и объёмы получаемых данных.

Поэтому рассмотрим методики оценивания характеристик процесса увольнения и использования оценок для анализа экономического эффекта фирмы. Распределение вероятностей $\{\varphi_\theta\}$ наиболее полно характеризует процесс увольнений, поэтому рассмотрим возможность непосредственно оценивания его значений.

Известно, что в общем случае выборочная функция распределения случайной величины сходится почти наверное к теоретической функции распределения (если её четвёртый момент конечен), а также является её асимптотически нормальной оценкой со скоростью убывания величины доверительного интервала обратно пропорционально квадратному корню объёма выборки (теоремы Гливленко–Кантелли и Колмогорова, см. например, §13 главы III в [9]).

Однако для оценивания распределения вероятностей времени работы в фирме целесообразно использовать процедуру, предложенную в классической работе [20] для анализа «выживаемости» на основе цензурированных данных, широко применяемую и известную сейчас как «процедура Каплана–Мейера» [7]. Процедура [7, 20] позволяет получить оценку «функции выживания», равной вероятности того, что увольнение сотрудника произойдёт позже t интервалов времени с момента прихода в фирму ($\Phi(t) = \sum_{\theta=t+1}^{\infty} \varphi_\theta$).

В приложении ПЗ проанализированы свойства процедуры «Каплана–Мейера» и зависимость от t и N (количества сотрудников, работающих и работавших в ФД) относительного дове-

рительного интервала оценки функции выживания в сравнении с относительным доверительным интервалом оценки интенсивности $\lambda(t)$. Показано, что оба интервала имеют характер зависимости от t и N вида $(N\varphi_t)^{-1/2}$, т.е. заметно растут по t – длительности срока работы сотрудников – даже при значительных объёмах выборки N (потому что на практике φ_t существенно убывает с ростом t). Например, для геометрического распределения времени работы $\varphi_t = \varphi(1-\varphi)^{t-1}$ зависимость относительных доверительных интервалов от t и N будут иметь характер $O\left(N^{-1/2}(1-\varphi)^{-t/2}\right)$ для обеих рассматриваемых оценок – и для функции выживания, и для интенсивности.

Из этого следует, что оценить распределение вероятностей $\{\varphi_\theta\}$ с приемлемой точностью для больших θ времён работы сотрудников в ФД в принципе невозможно: отмеченная выше изменчивость бизнес-условий не позволит «накопить» настолько большие N (количества сотрудников, работающих и работавших в ФД) которые обеспечили бы приемлемо малые значения $(N\varphi_t)^{-1/2}$. Данный тезис иллюстрируется графиками оценок интенсивностей увольнения сотрудников двух российских фирм («фирма А» и «фирма Б») из различных индустрий (рис. 1 и 2).

На обоих графиках вертикальная ось соответствует интенсивности увольнений в течение одного интервала времени, интервалы времени отложены по горизонтальной оси. Штриховой линией показаны оценки интенсивности, более тонкими сплошными линиями – верхняя и нижняя границы доверительных интервалов, принятых в иллюстративных целях равными плюс-минус три среднеквадратичных отклонения.

На рис. 1 «фирма А», интервал времени соответствует 1 неделе, а численность сотрудников рассматриваемого ФД имеет порядок 50–70 тыс. человек.

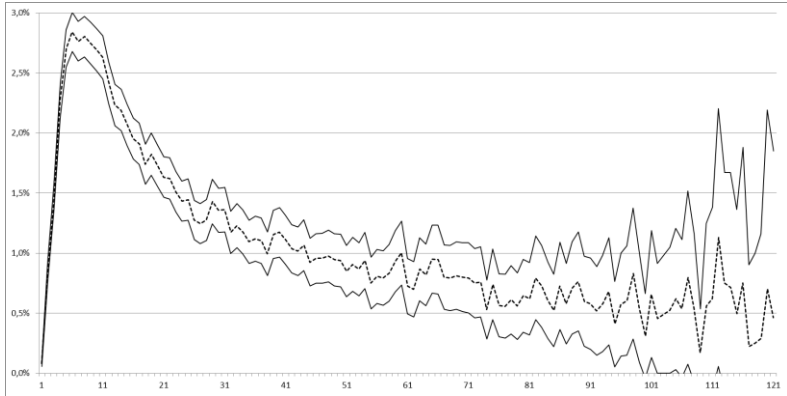


Рис. 1. Оценка интенсивности увольнений и её доверительный интервал для «фирмы А»

Количество увольняющихся на начальных интервалах имеет порядок несколько тысяч человек в течение одного интервала, спадает до сотен и далее до десятков человек. Это позволяет получить относительно точную оценку для первых 20–30 интервалов. Начиная со значений стажа 90–100 интервалов количество увольнений в выборке уменьшается до десяти-двадцати, что приводит к тому, что доверительный интервал становится вдвое больше самого значения оценки. Такую оценку вряд ли можно считать достоверной.

Во втором случае («фирма Б») интервал времени соответствует одному месяцу, численность сотрудников имеет порядок 8–10 тыс. человек. В результате существенно меньшее количество увольняющихся в течение одного интервала времени значительно ухудшает точность оценки даже для малых значений стажа.

Это означает, что даже для фирм с численностью однородных сотрудников в несколько десятков тысяч человек возможности точного оценивания распределение вероятностей времени работы сотрудников являются ограниченными: относительно достоверно могут быть оценены интенсивности увольнений только для малых значений стажа сотрудников, когда доля та-

ких сотрудников относительно велика, а для больших значений стажа оценки становятся мало достоверными.

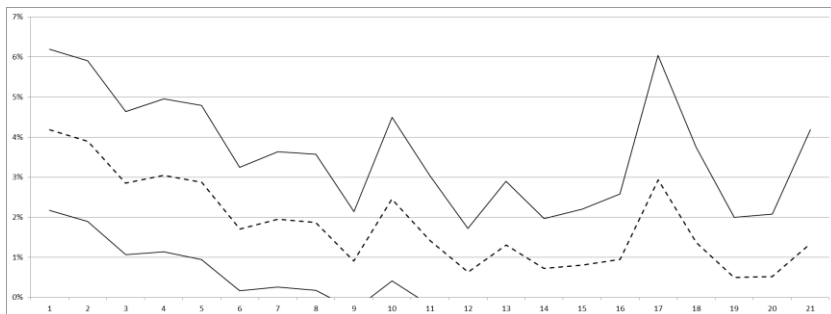


Рис. 2. Оценка интенсивности увольнений и её доверительный интервал для «фирмы Б»

Практическое следствие этого заключается в том, что непосредственно распределение вероятностей (или функция интенсивности) увольнений сотрудников может использоваться для анализа и принятия управленческих решений только для относительно малых длительностей стажа работы сотрудников. Но в разделе 2 выше было показано, что количество сотрудников, увольняющихся при малых значениях стажа, должно быть как можно меньше, в идеальном случае их вообще не должно быть. Поэтому достоверное оценивание и детальный анализ распределение вероятностей для малых значениях стажа может использоваться для анализа именно этой группы сотрудников и принятия управленческих решений с целью уменьшения их доли. В частности уместно использование для этого регрессионных методик Кокса [15] для анализа факторов, влияющих на «ранние» увольнения.

Во всех иных случаях целесообразно использовать другие статистические инструменты, рассмотрим их в следующем разделе.

4. Последовательная задача выявления изменения характеристик жизненного цикла сотрудников

Конечной целью деятельности практического менеджера, отвечающего за ФД, является обеспечение его экономической эффективности на определённом промежутке времени. Данная задача была решена в постановке поиска равновесных оптимальных решений [1] и для стационарных условий [2]. Для обеспечения этих решений необходимо контролировать нарушение стационарности процессов ЖЦС сотрудников; рассмотрим эту задачу в последовательной постановке. Выше было отмечено, что бизнес-условия на длительном периоде не могут считаться стационарными: для бизнес-среды свойственны периодические смены ключевых факторов – переходы от одного периода стационарности к другому. Поэтому, безусловно, важной с практической точки зрения является задача как можно более оперативного и в то же время достоверного выявления таких переходов.

Рассмотрим задачу оптимального обнаружения момента изменения («разладки») характеристик жизненного цикла сотрудников (вероятности π и распределения вероятностей $\{\varphi_\theta\}$) в рамках моделей, представленных в разделах 1 и 2. Считаем, что до разладки характеристики ЖЦС имели одни постоянные значения, а после – другие, также постоянные. Характеристики ЖЦС ни до, ни после разладки априори неизвестны, а их значения должны быть оценены на основании наблюдений за процессами ζ_i^1 – приема новых сотрудников и ζ_i^2 – увольнения сотрудников по собственному желанию. Сохраним предположение о статистической эквивалентности и независимости поведения сотрудников друг от друга и на различных интервалах времени. Априори имеются наблюдения в течение предварительного промежутка времени, про который известно, что весь он относится к единому периоду стационарности условий бизнеса. На основании накопленных в течение предварительного промежутка наблюдений характеристики ЖЦС должны быть оценены, а начиная с текущего момента необходимо на основании

оперативных наблюдений последовательно принимать решение о наличии или отсутствии разладки.

Учитывая вывод, сделанный в заключении раздела 3, о невозможности достоверной идентификации распределение вероятностей $\{\varphi_\theta\}$, процесс увольнений ζ_t^2 будем рассматривать не детально, по каждому сотруднику, а агрегировано – по всему функциональному дому в целом. В качестве характеристик ЖЦС будем использовать интенсивность потока увольнений в целом $\lambda(t)$, идентифицировать стационарные значения интенсивности и выявлять их изменения. Очевидно, при таком подходе – преднамеренном «загрублении модели» – теряется некоторая информация, которая теоретически могла бы быть учтена, если бы распределение вероятностей $\{\varphi_\theta\}$, было достоверно известно, однако на практике эта потеря неизбежна.

Наблюдаемые процессы определяют изменения фактической численности N_t (также наблюдаемой) ФД во времени (некоторые свойства N_t приведены в приложении П4):

$$(4) \quad N_{t+1} = N_t + \zeta_t^1 - \zeta_t^2,$$

где ζ_t^1 – последовательность независимых случайных величин, распределённых биномиально с параметрами $n_{\text{пред}}(t)$ и $\pi(t)$ и отражающих процесс найма сотрудников; ζ_t^2 – последовательность независимых случайных величин с пуассоновским распределением с параметром $\Lambda_t = \lambda(t)N_t$, отражающих процесс увольнений.

Тогда формально задача выявления разладки может быть поставлена следующим образом: ненаблюдаемый двумерный процесс $x(t) = (\pi(t); \lambda(t))^T$ изменяет свои значения в априори неизвестные моменты времени, сохраняя значения постоянными между моментами изменений. Необходимо обнаруживать момент изменения значений процесса $x(t)$ на основании последовательных наблюдений процесса $\zeta_t = (\zeta_t^1; \zeta_t^2)^T$, связанного с $x(t)$ закономерностями, описываемыми соотношениями (4). Начальное значение процесса $x(t)$ должно быть идентифицировано на основании ζ_t в течение предварительного периода наблюдений

длительностью $T_{\text{апр}}$ (условно в моменты времени $-T_{\text{апр}}$; $-T_{\text{апр}} + 1$; $-T_{\text{апр}} + 2$; ...; 0)

Для решения будем использовать оценки максимального правдоподобия.

Введём обозначения:

$$\zeta_{[u,v]}^1 = \sum_{\tau=u}^v \zeta_{\tau}^1; \quad \zeta_{[u,v]}^2 = \sum_{\tau=u}^v \zeta_{\tau}^2; \quad n_{\text{пред}[u,v]} = \sum_{\tau=u}^v n_{\text{пред}}(\tau); \quad N_{[u,v]} = \sum_{\tau=u}^v N_{\tau}.$$

Оценка $(\hat{\pi}_0; \hat{\lambda}_0)^T$ начального значения процесса $x(t)$ с учётом (4) имеет следующий вид и характеристики:

$$\begin{aligned} \hat{\pi}_0 &= \left(n_{\text{пред}[-T_{\text{апр}};0]} \right)^{-1} \zeta_{[-T_{\text{апр}};0]}^1, \quad E[\hat{\pi}_0] = \pi_0, \\ \text{var}[\hat{\pi}_0] &= \pi_0 (1 - \pi_0) \left(n_{\text{пред}[-T_{\text{апр}};0]} \right)^{-1}, \quad \hat{\lambda}_0 = \left(N_{[-T_{\text{апр}};0]} \right)^{-1} \zeta_{[-T_{\text{апр}};0]}^2, \\ E[\hat{\lambda}_0] &= \lambda_0, \quad \text{var}[\hat{\lambda}_0] = \lambda_0 \left(N_{[-T_{\text{апр}};0]} \right)^{-1}, \quad \text{cov}[\hat{\pi}_0; \hat{\lambda}_0] = 0. \end{aligned}$$

Имеет смысл отметить, что в практически интересных случаях величины $n_{\text{пред}[-T_{\text{апр}};0]}$; $\zeta_{[-T_{\text{апр}};0]}^1$; $N_{[-T_{\text{апр}};0]}$ и $\zeta_{[-T_{\text{апр}};0]}^2$ имеют порядок нескольких сотен, поэтому распределения обеих величин $\zeta_{[-T_{\text{апр}};0]}^1$ и $\zeta_{[-T_{\text{апр}};0]}^2$ хорошо аппроксимируются нормальными, а оценка $(\hat{\pi}_0; \hat{\lambda}_0)^T$ является несмещённой и состоятельной.

Решение о выявлении разладки в каждый момент времени t фактически заключается в проверке семейства гипотез \mathbf{H}_{τ} о том, что разладка произошла в момент времени $t - \tau$ ($\tau \geq 1$) против гипотезы \mathbf{H}_0 , что разладки не было.

Тогда последовательный критерий выявления разладки на основе относительного правдоподобия может быть построен в следующем виде:

$$(5) \quad \begin{cases} H_0: \max_{\theta} \left\{ \ln \left(\Pr \left(\left(\zeta_{[-T_{\text{анп}}; t]}^1; \zeta_{[-T_{\text{анп}}; t]}^2 \right); t; \theta \right) \right) \right\} \leq C; \\ H_{\tau}: \max_{\theta} \left\{ \ln \left(\Pr \left(\left(\zeta_{[-T_{\text{анп}}; t]}^1; \zeta_{[-T_{\text{анп}}; t]}^2 \right); t; \theta \right) \right) \right\} > C; \\ \tau = \arg \left\{ \max_{\theta} \left\{ \ln \left(\Pr \left(\left(\zeta_{[-T_{\text{анп}}; t]}^1; \zeta_{[-T_{\text{анп}}; t]}^2 \right); t; \theta \right) \right) \right\} \right\}; \end{cases}$$

где $\ln \left(\Pr \left(\left(\zeta_{[-T_{\text{анп}}; t]}^1; \zeta_{[-T_{\text{анп}}; t]}^2 \right); t; \theta \right) \right)$ – логарифмическая функция относительного правдоподобия (л.ф.о.п.) вычисленная в момент времени t на основании наблюдений на интервале $[-T_{\text{анп}}; t]$ для пары гипотез: а) что разладка произошла в момент времени $t - \theta$, и б) разладка до текущего момента не произошла. Л.ф.о.п. может быть записана также в виде суммы $\ln(\Pr(\cdot)) = \sum_{\omega=t-\theta}^t \Delta l(\omega)$,

где $\Delta l(\omega)$ – частные л.ф.о.п. на каждом интервале времени ω .

Используя функции $\Delta l(\omega)$, образуем критерий, эквивалентный (5), но имеющий иной вид, традиционно используемый в задачах о разладке (см. например, [6, 10, 11, 13, 21]):

$$(6) \quad \begin{cases} L(t) = \max \{0; L(t-1) + \Delta l(t)\}; \quad L(0) = 0; \\ H_0: L(t) \leq C; \\ H_{\tau}: L(t) > C; \quad \tau = \arg \left\{ \min_{\theta} \{L(t-\theta) = 0\} \right\}. \end{cases}$$

Исследование характеристик последовательных критериев (ошибок первого и второго рода в последовательной интерпретации) всегда является трудной задачей, что отмечается всеми исследователями, начиная с основополагающей работы Вальда [4]. К текущему моменту накоплено значительное количество результатов в области последовательного анализа и предложены различные методики, имеющие практическое значение, например, в работах [6, 10, 11, 13, 21]. В частности, в [6] разбирается проблематика получения аналитических выражений характеристик критериев, аналогичных (6), и отмечается невозможность получения для них точных аналитических выражений. Вместе с

тем, в [6] предлагаются (стр. 67, (4.2.14–16)) граничные оценки для среднего времени между ложными выявлениями разладки и среднего времени от момента наступления разладки до момента её выявления. Упомянутые граничные оценки основываются на значениях первых и вторых моментов приращений решающих функций $\Delta l(t)$ до и после разладки, когда $\Delta l(t)$ являются независимыми нормально распределёнными случайными величинами.

При отсутствии разладки средние значения процесса $\Delta l(t)$ отрицательны, а после наступления разладки – положительны, соответственно, до разладки процесс $L(t)$ образует супермартинал, а после – субмартинал. В этих условиях для границ каждой из характеристик критерия $T_{лт}$ – среднего времени между ложными выявлениями разладки и $T_{в.р.}$ – среднего времени от момента наступления разладки до момента её выявления могут быть получены аналитические выражения.

Введём (используя выражения (4.2.14–16) из [6], стр. 67) функцию $a(m, h)$, зависящую от нормированного среднего значения m и порога критерия h , имеющую вид

$$a(m, h) = \frac{1}{2m^2} (e^{-2mh} - 1 + 2mh) + \frac{\varphi(m)}{m\Phi(m)} + 1, \text{ где } \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$\text{и } \Phi(x) = \int_{-\infty}^x \varphi(\chi) d\chi.$$

Функция $a(m, h)$ позволяет получить следующие граничные значения характеристик каждого из критериев, подставляя моменты процесса $\Delta l(t)$:

$$(7) \quad \begin{aligned} T_{лт} &> a \left(\frac{E[\Delta l(t)|H_0]}{\sqrt{\text{var}[\Delta l(t)|H_0]}}; \frac{C}{\sqrt{\text{var}[\Delta l(t)|H_0]}} \right) \\ T_{в.р.} &< a \left(\frac{E[\Delta l(t)|H_\tau]}{\sqrt{\text{var}[\Delta l(t)|H_\tau]}}; \frac{C}{\sqrt{\text{var}[\Delta l(t)|H_\tau]}} \right) \end{aligned}$$

Выражения (7) позволяют получить оценки характеристик критериев и выбрать параметры критерия – порог принятия ре-

шения C и уровень γ ожидаемого изменения характеристик ЖЦС.

Следует заметить, что решение задачи происходит в условиях, когда значения изменений процесса $x(t)$ априори неизвестны. Практически значимым является выявление увеличения интенсивности увольнений $\hat{\lambda}_0 \rightarrow \hat{\lambda}_0 + \Delta\lambda$ и снижение вероятности принятия предложений $\hat{\pi}_0 \rightarrow \hat{\pi}_0 - \Delta\pi$. При этом не всегда имеются априорные основания, рассматривать одновременные изменения обеих компонент $\Delta\pi$ и $\Delta\lambda$, в общем случае следует считать, что они могут меняться независимо. Поэтому рассмотрим несколько сценариев изменения характеристик ЖЦС – в одном или нескольких функциональных домах, независимое или одновременное изменение $\hat{\lambda}_0 \rightarrow \hat{\lambda}_0 + \Delta\lambda$ и $\hat{\pi}_0 \rightarrow \hat{\pi}_0 - \Delta\pi$, соответственно, несколько вариантов построений приращений решающей функции $\Delta I(t)$.

Сценарий 1. Выявление независимых изменений характеристик ЖЦС $\hat{\lambda}_0 \rightarrow \hat{\lambda}_0 + \Delta\lambda$ и $\hat{\pi}_0 \rightarrow \hat{\pi}_0 - \Delta\pi$ одного функционального дома, причём детектировать относительно существенные изменения не менее чем, скажем, на $\gamma = 10\%$. В этом случае задача естественным образом распадается на две параллельные задачи – выявления изменения $\{\hat{\pi}_0 \rightarrow (1 - \gamma)\hat{\pi}_0; \gamma \geq 0, 1\}$ на основании наблюдений за $\zeta_{[-T_{\text{анп}}, t]}^1$ и $\{\hat{\lambda}_0 \rightarrow (1 + \gamma)\hat{\lambda}_0; \gamma \geq 0, 1\}$ – на основании $\zeta_{[-T_{\text{анп}}, t]}^2$. Тогда вместо критерия (6) используются два параллельных и независимых, имеющих вид (6), но отличающихся значениями порогов C_1 и C_2 , а также видом функций частных л.ф.о.п.:

$$\Delta I^1(t) = \zeta_t^1 \ln(1 - \gamma) + (n_{\text{пред}, t} - \zeta_t^1) \ln\left(1 + \frac{\gamma \hat{\pi}_0}{1 - \hat{\pi}_0}\right);$$

$$\Delta I^2(t) = \zeta_t^2 \ln(1 + \gamma) - N_t \hat{\lambda}_0 \gamma.$$

Каждый из процессов $\Delta I^1(t)$, $\Delta I^2(t)$ образует последовательность независимых случайных величин, распределение которых адекватно аппроксимируется нормальным с параметрами:

$$E[\Delta l^1(t)] = E[\zeta_t^1] \ln\left(\frac{(1-\gamma)(1-\hat{\pi}_0)}{1-\hat{\pi}_0 + \gamma\hat{\pi}_0}\right) + n_{\text{пред},t} \ln\left(\frac{1-\hat{\pi}_0 + \gamma\hat{\pi}_0}{1-\hat{\pi}_0}\right);$$

$$\text{var}[\Delta l^1(t)] = \left(\ln\left(\frac{(1-\hat{\pi}_0)(1-\gamma)}{1-\hat{\pi}_0 + \gamma\hat{\pi}_0}\right)\right)^2 \text{var}[\zeta_t^1];$$

$$E[\Delta l^2(t)] = E[\zeta_t^2] \ln(1+\gamma) - N_t \hat{\lambda}_0 \gamma;$$

$$\text{var}[\Delta l^2(t)] = (\ln(1+\gamma))^2 \text{var}[\zeta_t^2].$$

Сценарий 2. Выявление одновременных изменений характеристик ЖЦС $\hat{\lambda}_0 \rightarrow \hat{\lambda}_0 + \Delta\lambda$ и $\hat{\pi}_0 \rightarrow \hat{\pi}_0 - \Delta\pi$ одного функционального дома также на $\gamma = 10\%$. Если есть веские основания полагать, что изменения $\Delta\pi$ и $\Delta\lambda$ произойдут одновременно $\{\hat{\lambda}_0 \rightarrow (1+\gamma)\hat{\lambda}_0; \hat{\pi}_0 \rightarrow (1-\gamma)\hat{\pi}_0; \gamma \geq 0, 1\}$, то решение принимается на основании единственного критерия, в котором функция частных л.ф.о.п. имеет вид:

$$\begin{aligned} \Delta l(t) &= \Delta l^1(t) + \Delta l^2(t) = \\ &= \zeta_t^1 \ln(1-\gamma) + (n_{\text{пред},t} - \zeta_t^1) \ln\left(1 + \frac{\gamma\hat{\pi}_0}{1-\hat{\pi}_0}\right) + \zeta_t^2 \ln(1+\gamma) - N_t \hat{\lambda}_0 \gamma, \end{aligned}$$

где первые два слагаемые детектируют уменьшение вероятности $\hat{\pi}_0 \rightarrow \hat{\pi}_0 - \Delta\pi$, а вторые два – увеличение интенсивности $\hat{\lambda}_0 \rightarrow \hat{\lambda}_0 + \Delta\lambda$.

Процесс $\Delta l(t)$ образует последовательность независимых случайных величин, распределение которых адекватно аппроксимируется нормальным с параметрами:

$$E[\Delta l(t)] = E[\Delta l^1(t)] + E[\Delta l^2(t)];$$

$$\text{var}[\Delta l(t)] = \text{var}[\Delta l^1(t)] + \text{var}[\Delta l^2(t)].$$

Сценарий 3. Практически интересными является случай, когда фирма состоит из существенного числа независимых ФД, причём есть бизнес-основания считать, что в нормальном состоянии характеристики ЖЦС всех ФД идентичны. А разладкой считается появление аномалий в поведении сотрудников от-

дельных ФД независимо по $\hat{\lambda}_0 \rightarrow \hat{\lambda}_0 + \Delta\lambda$ и $\hat{\pi}_0 \rightarrow \hat{\pi}_0 - \Delta\pi$, в то время как в нормальном состоянии такое поведение стационарно по времени и идентично для всех ФД. То есть для каждого i -го ($i = 1, 2, \dots, I$) функционального дома необходимо проверить семейство гипотез \mathbf{H}_τ о том, что начиная с момента времени $t - \tau$ ($\tau \geq 1$) характеристики ЖЦС данного ФД стали отличаться от характеристик ЖЦС всех остальных ФД против гипотезы \mathbf{H}_0 , что характеристики ЖЦС данного ФД равны характеристикам ЖЦС всех остальных ФД. В этом случае для всех ФД целесообразно сформировать частные л.ф.о.п. $\Delta l_i^{1,2}(t)$ на каждом интервале времени и применять критерий вида (6). Рассмотрим вид частных л.ф.о.п. $\Delta l_i^1(t)$ и $\Delta l_i^2(t)$ в этом случае, аналогично сценарию 1.

$$\Delta l_i^1(t) = \zeta_{i,t}^1 \ln(1 - \gamma) + (n_{\text{пред},i,t} - \zeta_{i,t}^1) \ln\left(1 + \frac{\gamma \hat{\pi}_{0,i}}{1 - \hat{\pi}_{0,i}}\right);$$

$$\Delta l_i^2(t) = \zeta_{i,t}^2 \ln(1 + \gamma) - N_{i,t} \hat{\lambda}_{0,i} \gamma;$$

где

$$\hat{\pi}_{0,i} = \left(\sum_{j \neq i} n_{\text{пред},j,[-T_{\text{анп}},0]} \right)^{-1} \sum_{j \neq i} \zeta_{j,[-T_{\text{анп}},0]}^1; \quad E[\hat{\pi}_{0,i}] = \pi_0;$$

$$\text{var}[\hat{\pi}_{0,i}] = \pi_0 (1 - \pi_0) \left(n_{\text{пред},i,[-T_{\text{анп}},0]} \right)^{-1};$$

$$\hat{\lambda}_{0,i} = \left(\sum_{j \neq i} N_{j,[-T_{\text{анп}},0]} \right)^{-1} \sum_{j \neq i} \zeta_{j,[-T_{\text{анп}},0]}^2; \quad E[\hat{\lambda}_{0,i}] = \lambda_0;$$

$$\text{var}[\hat{\lambda}_{0,i}] = \lambda_0 \left(N_{i,[-T_{\text{анп}},0]} \right)^{-1}; \quad \text{cov}[\hat{\pi}_0, \hat{\lambda}_0] = 0.$$

Выражения для моментов $\Delta l_i^1(t)$ и $\Delta l_i^2(t)$ также аналогичны выражениям, приведённым для сценария 1.

Сценарий 4. Множество независимых ФД и одновременное изменение $\hat{\lambda}_0 \rightarrow \hat{\lambda}_0 + \Delta\lambda$ и $\hat{\pi}_0 \rightarrow \hat{\pi}_0 - \Delta\pi$. Для этого случая

частные л.ф.о.п. равны $\Delta l_i(t) = \Delta l_i^1(t) + \Delta l_i^2(t)$, а их моменты аналогичны сценарию 2.

$$\begin{aligned} \Delta l(t) &= \Delta l^1(t) + \Delta l^2(t) = \\ &= \zeta_t^1 \ln(1-\gamma) + (n_{\text{пред},t} - \zeta_t^1) \ln\left(1 + \frac{\gamma \hat{\pi}_0}{1 - \hat{\pi}_0}\right) + \zeta_t^2 \ln(1+\gamma) - N_t \hat{\lambda}_0 \gamma. \end{aligned}$$

Покажем теперь, что для каждого из сценариев решающие функции до разладки образуют супермартингалы, а после неё – субмартингалы.

Если условия в действительности отвечают условиям гипотезы \mathbf{H}_0 , то есть $x(u) = (\hat{\pi}_0; \hat{\lambda}_0)^T$ при всех $-T_{\text{анп}} \leq u \leq t$ тогда моменты процессов $\Delta l_1(t)$ и $\Delta l_2(t)$ принимают значения $(\Delta l_i^1(t)$ и $\Delta l_i^2(t)$ – аналогично):

$$E[\Delta l^1(u)|H_0] = \left(\ln(1-\gamma) \hat{\pi}_0 + \ln\left(1 + \frac{\gamma \hat{\pi}_0}{1 - \hat{\pi}_0}\right) (1 - \hat{\pi}_0) \right) n_{\text{пред},u};$$

$$\text{var}[\Delta l^1(u)|H_0] = \left(\ln\left(\frac{(1-\gamma)(1-\hat{\pi}_0)}{1-\hat{\pi}_0 + \gamma \hat{\pi}_0}\right) \right)^2 \hat{\pi}_0 (1 - \hat{\pi}_0) n_{\text{пред},u};$$

$$E[\Delta l^2(u)|H_0] = (\ln(1+\gamma) - \gamma) N_u \hat{\lambda}_0;$$

$$\text{var}[\Delta l^2(u)|H_0] = (\ln(1+\gamma))^2 N_u \hat{\lambda}_0.$$

И среднее значение любого из процессов $\Delta l^1(t)$, $\Delta l^2(t)$, $\Delta l(t)$ отрицательно при любых γ , $\hat{\pi}_0$ и $\hat{\lambda}_0$ (см. приложение П5), а каждая из решающих функций $L^1(t)$, $L^2(t)$, $L(t)$ образует супермартингал ($\Delta l_i^1(t)$, $\Delta l_i^2(t)$ и $\Delta l_i(t)$ – аналогично).

Если условия отвечают условиям гипотезы \mathbf{H}_t , $x(u) = (\hat{\pi}_0; \hat{\lambda}_0)^T$ при $-T_{\text{анп}} \leq u \leq t - \tau$ и $x(u) = (\pi_p; \lambda_p)^T$ при $t - \tau \leq u \leq t$, причём $\pi_p \leq (1 - \gamma) \hat{\pi}_0$ и $\lambda_p \geq (1 + \gamma) \hat{\lambda}_0$, тогда моменты процессов $\Delta l^1(t)$, $\Delta l^2(t)$ принимают значения $(\Delta l_i^1(t)$ и $\Delta l_i^2(t)$ – аналогично):

$$E[\Delta l^1(u)|H_\tau] = \left(\hat{\pi}_0(1-\gamma)\ln(1-\gamma) + (1-\hat{\pi}_0(1-\gamma))\ln\left(1 + \frac{\gamma\hat{\pi}_0}{1-\hat{\pi}_0}\right) \right) n_{\text{пред},u};$$

$$\text{var}[\Delta l^1(u)|H_\tau] = \left(\ln\left(1 - \frac{\gamma}{1-\hat{\pi}_0 + \gamma\hat{\pi}_0}\right) \right)^2 \hat{\pi}_0(1-\gamma)(1-\hat{\pi}_0(1-\gamma))n_{\text{пред},u};$$

$$E[\Delta l^2(u)|H_\tau] = ((1+\gamma)\ln(1+\gamma) - \gamma) N_u \hat{\lambda}_0;$$

$$\text{var}[\Delta l^2(u)|H_\tau] = (1+\gamma)(\ln(1+\gamma))^2 N_u \hat{\lambda}_0.$$

В этом случае все средние значения $\Delta l^1(t)$, $\Delta l^2(t)$, $\Delta l(t)$ положительны при любых γ , $\hat{\pi}_0$ и $\hat{\lambda}_0$ и монотонно возрастают по γ (см. приложение П5), а решающие функции $L^1(t)$, $L^2(t)$, $L(t)$ образует субмартиггал ($\Delta l_i^1(t)$, $\Delta l_i^2(t)$ и $\Delta l_i(t)$ – аналогично).

5. Примеры практического применения разработанных процедур

Проиллюстрируем применение предложенных методик на практических примерах «фирмы А» и «фирмы Б», которые были кратко описаны в разделе 2 при обсуждении достоверности оценивания распределения вероятностей $\{\varphi_0\}$.

Пример 1, «фирма А».

Для «фирмы А», как видно из иллюстративного рис. 1, характерна существенная доля сотрудников, увольняющихся в течение первых 6 недель после приёма на работу. На рис. 3 представлена гистограмма распределения времени работы сотрудников до увольнения. Устойчивое увольнение в течение таких малых сроков работы в фирме свидетельствует о явных проблемах в процедурах найма и адаптации сотрудников. Это означает, что, начав работать в фирме, сотрудник обнаруживает, что условия его не устраивают, или это является результатом неудачной адаптации на рабочем месте, в частности, линейный менеджер выясняет, что сотрудник не соответствует требуемому профилю. Экономическая модель (1)–(3) позволяет оценить, какие потери несёт «фирма А» из-за проблем в процедурах найма и адаптации. На основании оценок все сотрудники разбиты

на группы по стажу работы (на момент увольнения), в таблице 1 представлены доли сотрудников и среднее время работы внутри групп. В столбце «фактическое» показано среднее время работы в неделях, полученное в результате оценок по фактическим данным. В столбцах «возможное» проставлены ожидаемые менеджментом экспертные оценки возможного среднего времени после улучшения процедур найма и адаптации. Соотношения (1)–(3) позволяют получить в таком случае численные финансовые оценки влияния улучшения бизнес-процессов на финансовый результат фирмы. В случае «фирмы А» изменение среднего времени для рассматриваемой группы с 3,9 недель до 7 приводит к повышению среднего эффекта от одного сотрудника на 0,7%, а с 3,9 до 12,8 недель – на 2%. При численности сотрудников порядка нескольких десятков тысяч эти незначительные относительные изменения влекут за собой существенные суммарные эффекты. И, что очень важно, эти эффекты достигаются за счёт улучшения вспомогательных бизнес-процессов, т.е. не требуют затрат в области основных процессов, которые всегда велики.

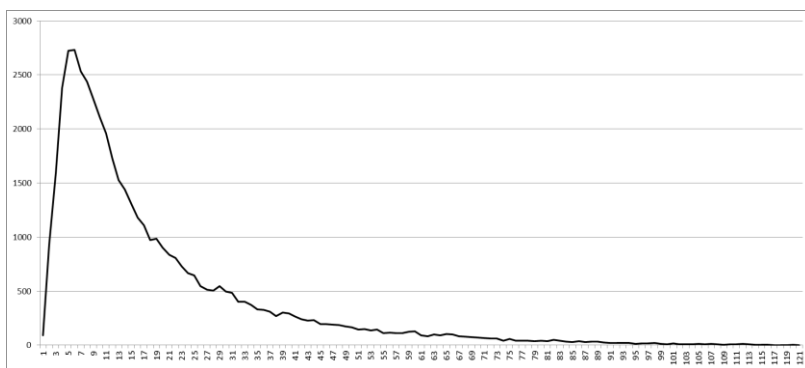


Рис. 3. Гистограмма распределения времени работы сотрудников до увольнения (среди уже уволившихся) «фирмы А»

Таблица 1. Распределения времени работы сотрудников до увольнения

Группы: (недели)	Доля	Среднее время работы (недели)		
		Факт	Возможное 1	Возможное 2
0–6	27%	3,9	7,0	12,8
7–22	46%	12,8	12,8	12,8
23 и более	26%	41,6	41,6	41,6
Всего:	100%	18,5	19,3	20,9

Пример 2, «фирма Б».

Для «фирмы Б» не свойственно большое количество увольнений на ранних стадиях, для неё актуальной является задача оперативного контроля характеристик ЖЦС, чтобы управленческими решениями устранять угрозы потери эффективности при возникновении системных изменений интенсивности увольнений λ и вероятности принятия предложений новыми сотрудниками π .

Особенностью «фирмы Б» является наличие большого количества функциональных домов – более тысячи, по которым необходимо оперативно контролировать характеристики ЖЦС. То есть группа менеджеров, отвечающая за бизнес «фирмы Б» в целом, должна на каждом интервале времени оперативно сопоставлять более тысячи кортежей:

- численность N_i ;
- количество уволившихся ζ_i^2 ;
- количество актуальных предложений $n_{\text{пред}}(t)$;
- количество принятых предложений ζ_i^1 .

Поставленная задача решается с помощью предложенной в разделе 3 методики, применяемой параллельно ко всем ФД в варианте сценария 3.

Результаты выявления служат основанием для более детального анализа ситуации в функциональных домах, по которым сформировано решение об обнаружении разладки. Для наглядности результаты отображаются не только в численной, но также и в графической форме (рис.4).

На рис. 4 представлен пример результатов выявления в виде трёхмерной поверхности. По вертикали отложено относительное правдоподобие наличия разладки против его отсутствия, одна из горизонтальных осей соответствует функциональным домам, а другая – текущему времени. Пики на трёхмерной поверхности показывают значительное превышение правдоподобия разладки над её отсутствием и служат индикаторами для менеджмента для детального рассмотрения ситуации в соответствующих функциональных домах.

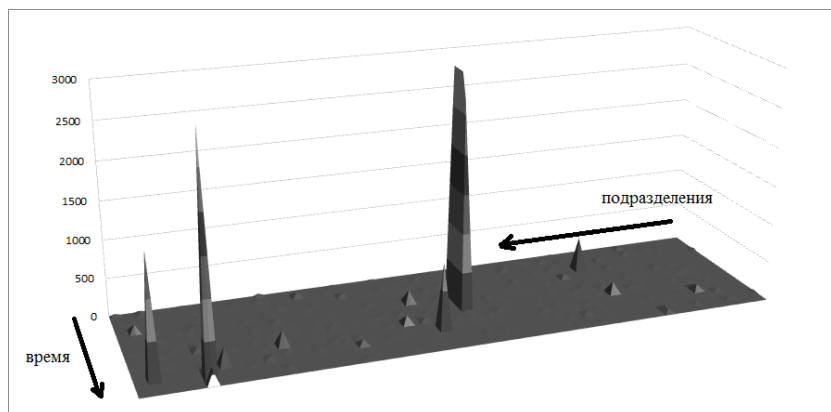


Рис. 4. Пример отображения результатов выявления изменений характеристик ЖЦС на множестве функциональных домов «фирмы Б»

6. Заключение

В первом разделе рассмотрена актуальность задачи оценивания и анализа таких характеристик жизненного цикла сотрудников, как распределение вероятностей $\{\varphi_{\theta}\}$ времени работы до увольнения и вероятность π принятия предложений о работе в ходе найма.

Во втором разделе проанализирована зависимость экономической эффективности бизнеса от характеристик жизненного цикла сотрудников. Показано наличие существенного влияния

рассматриваемых характеристик ЖЦС на эффективность бизнеса, отражённое в утверждении 1.

Третий раздел посвящён более детальному рассмотрению процесса увольнения сотрудников по собственному желанию. Показана невозможность оценивания распределения $\{\varphi_\theta\}$ с высокой достоверностью и предложено использовать агрегированный параметр интенсивности увольнений λ вместо распределения $\{\varphi_\theta\}$. Поэтому на практике следует использовать агрегированные характеристики ЖЦС сотрудников, не рассчитывая на детальные модели, а также использовать «экспертные» модели, сочетать параметры, значения которых известно из отчётности фирм, с оцениваемыми и задаваемыми экспертно.

В четвёртом разделе предложены процедуры последовательного выявления моментов изменения (разладки) характеристик ЖЦС π , λ для четырёх сценариев таких изменений, а также на основании известных результатов получены граничные значения характеристик последовательных критериев обнаружения разладки.

В пятом разделе приведены примеры практического применения разработанных процедур.

Приложения

П.1. ПОЛУЧЕНИЕ ВЫРАЖЕНИЯ ДЛЯ ЭКОНОМИЧЕСКОГО ЭФФЕКТА РАЗДЕЛА 2

$$s(\tau) = \begin{cases} s_{\text{прд}} - \Delta s_{\text{ад}} & \text{при } 0 \leq \tau \leq t_{\text{ад}}, \\ s_{\text{прд}} & \text{при } t_{\text{ад}} < \tau \leq \theta - t_{\text{пр.ув}}, \\ s_{\text{прд}} - \Delta s_{\text{ув}} & \text{при } \max\{t_{\text{ад}}; \theta - t_{\text{пр.ув}}\} < \tau \leq \theta. \end{cases}$$

$$S(\theta) = s_{\text{прд}} \theta - \begin{cases} \Delta s_{\text{ад}} \theta & \text{при } 0 \leq \theta \leq t_{\text{ад}}, \\ \Delta s_{\text{ад}} t_{\text{ад}} + \Delta s_{\text{ув}} (\theta - t_{\text{ад}}) & \text{при } t_{\text{ад}} < \theta \leq t_{\text{ад}} + t_{\text{пр.ув}}, \\ \Delta s_{\text{ад}} t_{\text{ад}} + \Delta s_{\text{ув}} t_{\text{пр.ув}} & \text{при } t_{\text{ад}} + t_{\text{пр.ув}} < \theta. \end{cases}$$

$$\sum_{\theta=0}^{\infty} (S(\theta)\varphi(\theta)) = s_{\text{прд}} \sum_{\theta=0}^{\infty} (\theta\varphi(\theta)) - \Delta S_{\text{ад}} \sum_{\theta=0}^{t_{\text{ад}}} (\theta\varphi(\theta)) - \Delta S_{\text{ад}} t_{\text{ад}} \sum_{\theta=t_{\text{ад}}+1}^{\infty} \varphi(\theta) -$$

$$- \Delta S_{\text{ыв}} \sum_{\theta=t_{\text{ад}}+1}^{t_{\text{ад}}+t_{\text{пр.ыв}}} (\theta\varphi(\theta)) - \Delta S_{\text{ыв}} t_{\text{пр.ыв}} \sum_{\theta=t_{\text{ад}}+t_{\text{пр.ыв}}+1}^{\infty} \varphi(\theta).$$

Обозначим
$$t_{\text{ад}}^* = \sum_{\theta=0}^{t_{\text{ад}}} (\theta\varphi(\theta)) + t_{\text{ад}} \sum_{\theta=t_{\text{ад}}+1}^{\infty} \varphi(\theta) \quad \text{И}$$

$$t_{\text{пр.ыв}}^* = \sum_{\theta=t_{\text{ад}}+1}^{t_{\text{ад}}+t_{\text{пр.ыв}}} (\theta\varphi(\theta)) + t_{\text{пр.ыв}} \sum_{\theta=t_{\text{ад}}+t_{\text{пр.ыв}}+1}^{\infty} \varphi(\theta).$$

Заметим, что $t_{\text{ад}}^* < t_{\text{ад}}$; $t_{\text{ад}}^* < t_{\text{ср}}$; $t_{\text{пр.ыв}}^* < t_{\text{пр.ыв}}$; $t_{\text{пр.ыв}}^* < t_{\text{ср}}$.

П.2. АНАЛИЗ ЗАВИСИМОСТИ ФУНКЦИИ $V(\pi, \{\varphi_{\theta}\}, t_{\text{ад}}, t_{\text{пр.ыв}})$ ОТ АРГУМЕНТОВ

Проанализируем зависимость функции $V(\pi, \{\varphi_{\theta}\}, t_{\text{ад}}, t_{\text{пр.ыв}})$ от всех своих аргументов.

$$V(\pi, \{\varphi_{\theta}\}, t_{\text{ад}}, t_{\text{пр.ыв}}) = DS_{\text{прд}} - D \left(\Delta S_{\text{ад}} \frac{t_{\text{ад}}^*}{t_{\text{ср}}} + \Delta S_{\text{ыв}} \frac{t_{\text{пр.ыв}}^*}{t_{\text{ср}}} + S_{\text{зам}} \frac{1}{t_{\text{ср}}} + \frac{S_{\text{пред}}}{\pi t_{\text{ср}}} \right) + V^*,$$

где $t_{\text{ад}}^* = \sum_{\theta=0}^{t_{\text{ад}}} \theta\varphi_{\theta} + t_{\text{ад}} \sum_{\theta=t_{\text{ад}}+1}^{\infty} \varphi_{\theta}$; $t_{\text{пр.ыв}}^* = \sum_{\theta=t_{\text{ад}}+1}^{t_{\text{ад}}+t_{\text{пр.ыв}}} \theta\varphi_{\theta} + t_{\text{пр.ыв}} \sum_{\theta=t_{\text{ад}}+t_{\text{пр.ыв}}+1}^{\infty} \varphi_{\theta}$ и

$$t_{\text{ср}} = \sum_{\theta=0}^{\infty} \theta\varphi_{\theta}.$$

Во-первых, вероятность π влияет на значение функции $V(\cdot)$, только входя в элемент $-D \frac{S_{\text{пред}}}{t_{\text{ср}} \pi}$. Поэтому очевидно, что

функция $V(\cdot)$ монотонно возрастает по π , т.е.

$$\frac{\partial}{\partial \pi} V(\pi, \{\varphi_{\theta}\}, t_{\text{ад}}, t_{\text{пр.ыв}}) > 0.$$

Напротив, при уменьшении π становится отрицательным, начиная с некоторых малых π , и при стремлении π к нулю может неограниченно уменьшаться, будучи отрицательным

$$\lim_{\pi \rightarrow 0} \left\{ V(\pi, \{\varphi_{\theta}\}, t_{\text{ад}}, t_{\text{пр.ыв}}) \right\} = -\infty.$$

Во-вторых, длительность периода адаптации $t_{ад}$ влияет на функцию $V(\cdot)$ только через элемент $-D\Delta_{s_{ад}} \frac{t_{ад}^*}{t_{ср}}$. Причём $t_{ср}$ не за-

висит от $t_{ад}$, а из $t_{ад}^* = \sum_{\theta=0}^{t_{ад}} \theta \varphi_{\theta} + t_{ад} \sum_{\theta=t_{ад}+1}^{\infty} \varphi_{\theta}$ непосредственно следует,

что $t_{ад}^*$ монотонно растёт с ростом $t_{ад}$. Поэтому

$\frac{\partial}{\partial t_{ад}} V(\pi, \{\varphi_{\theta}\}, t_{ад}, t_{пр.ув}) < 0$. Аналогичные соображения также

справедливы и для продолжительности фазы, предшествующей

увольнению, следовательно $\frac{\partial}{\partial t_{пр.ув}} V(\pi, \{\varphi_{\theta}\}, t_{ад}, t_{пр.ув}) < 0$. С точки

зрения бизнеса монотонное убывание функции $V(\cdot)$ при росте продолжительности интервала адаптации и/или фазы снижения производительности, предшествующей увольнению является интуитивно очевидным.

Вид распределения вероятностей времени работы сотрудников в фирме $\{\varphi_{\theta}\}$ влияет на функцию $V(\cdot)$ более сложным образом. Изучим это влияние, для чего используем «изменения» функции $\{\varphi_{\theta}\}$ следующего вида:

$$\delta\varphi_{\theta}(x, i) = x\varepsilon_{\theta}(i); \text{ где } \varepsilon_{\theta}(i) = \begin{cases} 1 & \text{при } \theta=i \\ -1 & \text{при } \theta=i+1 \\ 0 & \text{при } \theta < i \text{ или } \theta > i+1 \end{cases}$$

Несложно показать, что разность между любыми распределениями вероятностей $\{\varphi_{\theta}\}$ и $\{\tilde{\varphi}_{\theta}\}$ может быть представлена в виде суммы счётного числа «изменений» $\delta\varphi_{\theta}(x, i)$.

Рассмотрим влияние «малых изменений» $\delta\varphi_{\theta}(x, i)$ ($x \rightarrow 0$) при различных i на функцию $V(\cdot)$, т.е. оценим

$$\frac{\partial}{\partial x} V(\pi, \{\varphi_{\theta} + \delta\varphi_{\theta}(x, i)\}, t_{ад}, t_{пр.ув}).$$

Из (3) следует: $\frac{\partial}{\partial x} V(\pi, \{\varphi_{\theta} + \delta\varphi_{\theta}(x, i)\}, t_{ад}, t_{пр.ув}) =$

$$\begin{aligned}
 &= -D \frac{\partial}{\partial x} \left(\Delta s_{ад} \frac{t_{ад}^*}{t_{ср}} + \Delta s_{уе} \frac{t_{пр.ув}^*}{t_{ср}} + S_{зам} \frac{1}{t_{ср}} + \frac{S_{перед}}{\pi} \frac{1}{t_{ср}} \right) = \\
 &= -D \frac{\partial}{\partial x} \left(a \frac{t_{ад}^*}{t_{ср}} + b \frac{t_{пр.ув}^*}{t_{ср}} + c \frac{1}{t_{ср}} \right),
 \end{aligned}$$

где $a = \Delta s_{ад}$, $b = \Delta s_{уе}$ и $c = S_{зам} + S_{перед} / \pi^+$ не зависят от x .

Также из выражений для $t_{ср}$, $t_{ад}^*$ и $t_{пр.ув}^*$ следует:

$$t_{ср} \left(\{ \varphi_{\theta} + \delta \varphi_{\theta}(x, i) \} \right) = \sum_{\theta=0}^{\infty} \theta \varphi_{\theta} + xi - x(i+1) = t_{ср} \left(\{ \varphi_{\theta} \} \right) - x \quad \text{для}$$

всех i .

$$\begin{aligned}
 t_{ад}^* \left(\{ \varphi_{\theta} + \delta \varphi_{\theta}(x, i) \} \right) &= \begin{cases} t_{ад}^* \left(\{ \varphi_{\theta} \} \right) - x & \text{для } 0 < i < t_{ад} \\ t_{ад}^* \left(\{ \varphi_{\theta} \} \right) & \text{для } t_{ад} \leq i \end{cases} \\
 t_{пр.ув}^* \left(\{ \varphi_{\theta} + \delta \varphi_{\theta}(x, i) \} \right) &= \begin{cases} t_{ад} x & \text{для } i = t_{ад} \\ t_{пр.ув}^* \left(\{ \varphi_{\theta} \} \right) - x & \text{для } t_{ад} < i < t_{ад} + t_{пр.ув} \\ t_{пр.ув}^* \left(\{ \varphi_{\theta} \} \right) & \text{для } t_{ад} + t_{пр.ув} \leq i \text{ или } i < t_{ад} \end{cases}
 \end{aligned}$$

Откуда видно, что при $i < t_{ад} + t_{пр.ув}$ знак производной $\frac{\partial}{\partial x} V(\pi, \{ \varphi_{\theta} + \delta \varphi_{\theta}(x, i) \}, t_{ад}, t_{пр.ув})$ может быть различным в зависимости от соотношений a , b , и c , то есть влияние «малых изменений» $\delta \varphi_{\theta}(x, i)$ на функцию $V(\cdot)$ в области $0 \leq i < t_{ад} + t_{пр.ув}$ проявляется неоднозначно.

Однако при $t_{ад} + t_{пр.ув} \leq i$ справедливо

$$\begin{aligned}
 \frac{\partial}{\partial x} V(\pi, \{ \varphi_{\theta} + \delta \varphi_{\theta}(x, i) \}, t_{ад}, t_{пр.ув}) &= -D \frac{\partial}{\partial x} \left(a \frac{t_{ад}^*}{t_{ср}} + b \frac{t_{пр.ув}^*}{t_{ср}} + c \frac{1}{t_{ср}} \right) = \\
 &= D \frac{(at_{ад}^* + bt_{пр.ув}^* + c)}{t_{ср}^2} \frac{\partial t_{ср}}{\partial x}
 \end{aligned}$$

Откуда

$$\frac{\partial/\partial x V(\pi, \{\varphi_\theta + \delta\varphi_\theta(x, i)\}, t_{ад}, t_{пр.ув})}{\frac{\partial t_{ср}}{\partial x}} = \frac{\partial V(\pi, \{\varphi_\theta\}, t_{ад}, t_{пр.ув})}{\partial t_{ср}} =$$

$$= D \frac{(at_{ад}^* + bt_{пр.ув}^* + c)}{t_{ср}^2} > 0$$

ПЗ. СВОЙСТВА ОЦЕНОК ФУНКЦИИ ВЫЖИВАНИЯ И ПРОЦЕДУРЫ «МЕЙЕРА-КАПЛАНА»

Процедура [7, 20] позволяет получить оценку «функции выживания», равной вероятности того, что увольнение сотрудника произойдёт позже t интервалов времени с момента прихода в фирму ($\Phi(t) = \sum_{\theta=t+1}^{\infty} \varphi_\theta$).

Процедура заключается в следующем:

На временной оси своими границами задаются интервалы $(0, u_1); (u_1, u_2); \dots; (u_{j-1}, u_j); \dots; (u_{J-1}, u_J)$.

На основании наблюдений формируется выборка $\{n_1; n_2; \dots; n_j; \dots; n_J\}$, где n_j – количество сотрудников, проработавших в фирме время, не меньшее u_{j-1} . Обозначим N – общее число сотрудников, работавших и работающих в фирме. Среди n_j будут и сотрудники продолжающие работать в фирме, стаж которых на момент наблюдений не меньше u_{j-1} , и уже уволившиеся, но стаж которых на момент увольнения был не меньше u_{j-1} . Аналогично формируется выборка $\{m_1; m_2; \dots; m_j; \dots; m_J\}$, где m_j – количество сотрудников, у которых стаж работы на момент увольнения принадлежал интервалу (u_{j-1}, u_j) . Тогда оценка «функции выживания» будет иметь вид $\hat{\Phi}(t) = \prod_{j=1}^k \left(1 - \frac{m_j}{n_j}\right)$, где

$k = \max_{u_j \leq t} \{u_j\}$. Данная оценка является состоятельной и несмещённой, с дисперсией $var[\hat{\Phi}(t)] = \Phi^2(t) \sum_{j=1}^k \left(\frac{1}{n_j - m_j} - \frac{1}{n_j}\right)$ [20],

т.е. относительный доверительный интервал оценки $\hat{\Phi}(t)$ равен

$$\sqrt{\sum_{j=1}^k \left(\frac{1}{n_j - m_j} - \frac{1}{n_j} \right)} = \sqrt{\sum_{j=1}^k \left(\frac{m_j}{n_j(n_j - m_j)} \right)}.$$

Так как средние выборочных значений $E[n_k] = N\Phi(t-1) = N\sum_{\theta=t}^{\infty} \varphi_{\theta}$ и $E[m_k] = N\varphi_t$, то характер зависимости относительного доверительного интервала оценки $\hat{\Phi}(t)$ может быть аппроксимирован

$$\hat{\Phi}(t) \approx \sqrt{\frac{1}{N} \sum_{j=1}^t \left(\frac{\varphi_j}{\sum_{\theta=j}^{\infty} \varphi_{\theta} \sum_{\theta=j+1}^{\infty} \varphi_{\theta}} \right)} \approx (N\varphi_t)^{-1/2}.$$

«Функция выживания» связана с интенсивностью соотношением:

$$\lambda(t) = \varphi_t \left(\sum_{\theta=t}^{\infty} \varphi_{\theta} \right)^{-1} = \varphi_t \Phi(t-1)^{-1} = 1 - \Phi(t)\Phi(t-1)^{-1}.$$

Поэтому выражение для оценки интенсивности $\lambda(t)$, которое может быть получено с помощью процедуры Каплана-Мейера, совпадает с «прямой» оценкой интенсивности $\hat{\lambda}(t) = 1 - \hat{\Phi}(t)\hat{\Phi}(t-1)^{-1} =$

$$= 1 - \prod_{j=1}^k \left(1 - \frac{m_j}{n_j} \right) \left(\prod_{j=1}^{k-1} \left(1 - \frac{m_j}{n_j} \right) \right)^{-1} = \frac{m_k}{n_k}.$$

Оценка интенсивности $\lambda(t) = \frac{m_k}{n_k}$ также является несмещённой и состоятельной с дисперсией $var[\lambda(t)] = \lambda(t)(1 - \lambda(t))n_k$. Соответственно, относительный доверительный интервал аппроксимируется $(1 - \lambda(t))^{1/2} (\lambda(t)n_k)^{-1/2} \approx (\lambda(t)n_k)^{-1/2} \approx (N\varphi_t)^{-1/2}$.

П4. СВОЙСТВА ПРОЦЕССА N_T ИЗ РАЗДЕЛА 4

Приведём некоторые свойства процесса N_t (5).

Среднее значение численности будет изменяться как

$$E[N_{t+1}] = (1 - \lambda(t))E[N_t] + \pi(t)n_{\text{пред}}(t),$$

или

$$E[N_{t+T}] = E[N_t] \prod_{\omega=1}^T (1 - \lambda(t + \omega)) + \sum_{\tau=1}^T \pi(t + \tau)n_{\text{пред}}(t + \tau) \prod_{\omega=\tau}^T (1 - \lambda(t + \omega)).$$

Для стационарного случая, когда $\lambda(t) = \lambda = \text{const}$; $n_{\text{пред}}(t) = n_{\text{пред}} = \text{const}$ и $\pi(t) = \pi = \text{const}$, имеем

$$E[N_{t+T}] = N_{\text{ст}} + (1 - \lambda)^T (E[N_t] - N_{\text{ст}}) \quad \text{и} \quad \lim_{T \rightarrow \infty} E[N_{t+T}] = N_{\text{ст}}, \quad \text{где}$$

$N_{\text{ст}} = \pi n_{\text{пред}} \lambda^{-1}$ – стационарное значение численности.

П5. ВСПОМОГАТЕЛЬНЫЕ РЕЗУЛЬТАТЫ К РАЗДЕЛУ 4

Покажем, что если условия бизнеса в действительности отвечают условиям гипотезы \mathbf{H}_0 , то средние значения $\Delta l^1(t)$, $\Delta l^2(t)$, $\Delta l(t)$ отрицательны при любых γ , $\hat{\pi}_0$ и $\hat{\lambda}_0$:

$$\text{Рассмотрим } E[\Delta l^1(u) | H_0] = \ln(1 - \gamma)\hat{\pi}_0 + \ln\left(1 + \frac{\gamma\hat{\pi}_0}{1 - \hat{\pi}_0}\right)(1 - \hat{\pi}_0).$$

Обозначим $F(\hat{\pi}_0; \gamma) = \ln(1 - \gamma)\hat{\pi}_0 + \ln\left(1 + \frac{\gamma\hat{\pi}_0}{1 - \hat{\pi}_0}\right)(1 - \hat{\pi}_0)$ и исследуем поведение функции $F(x; y)$

Во-первых, $F(0; y) = 0$. Вычислим производную по x .

$$\begin{aligned} \frac{\partial}{\partial x} F(x; y) &= \ln(1 - y) - \ln\left(1 + \frac{xy}{1 - x}\right) + (1 - x) \frac{\partial}{\partial x} \ln\left(1 + \frac{xy}{1 - x}\right) = \\ &= \ln(1 - y) - \ln\left(1 + \frac{xy}{1 - x}\right) + (1 - x) \left[\frac{\partial}{\partial x} \ln(1 - x + xy) - \frac{\partial}{\partial x} \ln(1 - x) \right] = \\ &= \ln(1 - y) - \ln\left(\frac{1 - x + xy}{1 - x}\right) + (1 - x) \left[-\frac{1 - y}{1 - x + xy} + \frac{1}{1 - x} \right] = \\ &= \ln(1 - y) - \ln\left(\frac{1 - x + xy}{1 - x}\right) + 1 - \frac{(1 - x)(1 - y)}{1 - x + xy} = \\ &= \ln(1 - y) - \ln\left(\frac{1 - x + xy}{1 - x}\right) + \frac{1 - x + xy - 1 + x + y - xy}{1 - x + xy} = \end{aligned}$$

$$\begin{aligned}
 &= \ln(1-y) - \ln\left(\frac{1-x+xy}{1-x}\right) + \frac{y}{1-x+xy} = \\
 &= \ln\left(\frac{(1-y)(1-x)}{1-x+xy}\right) + \frac{y}{1-x+xy} = \ln\left(1 - \frac{y}{1-x+xy}\right) + \frac{y}{1-x+xy} < 0
 \end{aligned}$$

То есть $F(x; y)$ монотонно убывает по x при любых y . Также найдём предел $F(x; y)$ при $x \rightarrow 1$:

$$\begin{aligned}
 \lim_{x \rightarrow 1} \left\{ \ln(1-y)x + \ln\left(1 + \frac{yx}{1-x}\right)(1-x) \right\} &= \ln(1-y) + \lim_{x \rightarrow 1} \left\{ \ln\left(\frac{1}{1-x}\right)(1-x) \right\} = \\
 &= \ln(1-y) + \lim_{x \rightarrow \infty} \left\{ \frac{\ln(x)}{x} \right\} = \ln(1-y) < 0
 \end{aligned}$$

Таким образом, $E[\Delta l^1(u)|H_0]$ всегда отрицательно.

Рассмотрим $E[\Delta l^2(u)|H_0] = N_u \hat{\lambda}_0 (\ln(1+\gamma) - \gamma)$: для любых $\gamma > 0$ справедливо $\gamma > \ln(1+\gamma)$, поэтому $E[\Delta l^2(u)|H_0]$ второе слагаемое также всегда отрицательно.

Так как $E[\Delta l^1(u)|H_0] < 0$ и $E[\Delta l^2(u)|H_0] < 0$, тогда и $E[\Delta l(u)|H_0] < 0$.

Если условия бизнеса отвечают условиям гипотезы \mathbf{H}_τ – $x(u) = (\hat{\pi}_0; \hat{\lambda}_0)^T$ при $-T_{\text{анп}} \leq u \leq t - \tau$ и $x(u) = (\pi_p; \lambda_p)^T$ при $t - \tau \leq u \leq t$, причём $\pi_p \leq (1-\gamma)\hat{\pi}_0$ и $\lambda_p \geq (1+\gamma)\hat{\lambda}_0$, тогда

$$E[\Delta l^1(u)|H_\tau] = \left(\hat{\pi}_0(1-\gamma)\ln(1-\gamma) + (1-\hat{\pi}_0(1-\gamma))\ln\left(1 + \frac{\gamma\hat{\pi}_0}{1-\hat{\pi}_0}\right) \right) n_{\text{прод}};$$

$$E[\Delta l^2(u)|H_\tau] = ((1+\gamma)\ln(1+\gamma) - \gamma) N_u \hat{\lambda}_0;$$

Рассмотрим первое среднее значение $E[\Delta l^1(u)|H_\tau]$.

Обозначим

$$G(\hat{\pi}_0; \gamma) = \hat{\pi}_0(1-\gamma)\ln(1-\gamma) + (1-\hat{\pi}_0(1-\gamma))\ln\left(1 + \frac{\gamma\hat{\pi}_0}{1-\hat{\pi}_0}\right) \quad \text{и ис-}$$

следуем поведение функции $G(x; y)$.

Во-первых, $G(0; y) = 0$. Вычислим производную по x .

$$\begin{aligned} \frac{\partial}{\partial x} G(x; y) &= (1-y)\ln(1-y) + \frac{\partial}{\partial x}(1-x(1-y))\ln\left(1 + \frac{xy}{1-x}\right) = \\ &= (1-y)\ln(1-y) - (1-y)\ln\left(1 + \frac{xy}{1-x}\right) + (1-x(1-y))\frac{\partial}{\partial x}\ln\left(\frac{1-x+xy}{1-x}\right) = \\ &= (1-y)\left(\ln(1-y) - \ln\left(\frac{1-x+xy}{1-x}\right)\right) + (1-x(1-y))\left(-\frac{1-y}{1-x+xy} + \frac{1}{1-x}\right) = \\ &= (1-y)\ln\left(\frac{(1-y)(1-x)}{1-x+xy}\right) + (1-x(1-y))\frac{1-x+xy - (1-y)(1-x)}{(1-x(1-y))(1-x)} = \\ &= (1-y)\ln\left(\frac{(1-y)(1-x)}{1-x+xy}\right) + \frac{y}{(1-x)} \end{aligned}$$

Итак,

$$\frac{\partial}{\partial x} G(x; y) = (1-y)\ln(1-x) + (1-y)\ln(1-y) - (1-y)\ln(1-x+xy) + \frac{y}{(1-x)}$$

Покажем, что производная положительна.

$$\begin{aligned} \frac{\partial}{\partial x} G(x; y) \Big|_{x=0} &= (1-y)\ln(1-y) + y = \\ &= -(1-y)\sum_{n=1}^{\infty} \frac{y^n}{n} + y = y - \sum_{n=1}^{\infty} \frac{y^n}{n} + \sum_{n=2}^{\infty} \frac{y^n}{n-1} = \\ &= y - y - \sum_{n=2}^{\infty} \frac{y^n}{n} + \sum_{n=2}^{\infty} \frac{y^n}{n-1} = \sum_{n=2}^{\infty} \left(\frac{1}{n-1} - \frac{1}{n}\right) y^n = \\ &= \sum_{n=2}^{\infty} \frac{y^n}{(n-1)n} > 0 \end{aligned}$$

Рассмотрим вторую производную:

$$\begin{aligned} \frac{\partial^2}{\partial x^2} G(x;y) &= \frac{\partial}{\partial x} \{ (1-y)\ln(1-y) + (1-y)\ln(1-x) \} + \\ &+ \frac{\partial}{\partial x} \left\{ -(1-y)\ln(1-x(1-y)) + \frac{y}{(1-x)} \right\} = \\ &= \frac{1-y}{1-x} + \frac{(1-y)^2}{1-x+xy} + \frac{y}{(1-x)^2} > 0 \end{aligned}$$

Первая производная в точке $x = 0$ положительна, а вторая – положительна при $x \geq 0$, откуда следует, что $G(x; y) > 0$ при $x > 0$ и первое среднее значение $E[\Delta l^1(u)|H_\tau]$ всегда больше нуля.

Рассмотрим второе среднее значение $E[\Delta l^2(u)|H_\tau]$.

Функция $(1+\gamma)\ln(1+\gamma) - \gamma$ обращается в 0 при $\gamma = 0$. Вычислим её производную

$$\frac{d}{d\gamma} (1+\gamma)\ln(1+\gamma) - \gamma = \ln(1+\gamma) + \frac{(1+\gamma)}{(1+\gamma)} - 1 = \ln(1+\gamma) > 0. \text{ Тогда}$$

$(1+\gamma)\ln(1+\gamma) - \gamma$ при всех $\gamma > 0$, следовательно второе среднее значение $E[\Delta l^2(u)|H_\tau] > 0$.

Откуда $E[\Delta l(u)|H_\tau] = E[\Delta l^1(u)|H_\tau] + E[\Delta l^2(u)|H_\tau] > 0$

Литература

1. БЕЛОВ М.В. *Модель управления человеческим капиталом фирмы* // Проблемы управления. – 2016. – №5. – С. 24–34.
2. БЕЛОВ М.В. *Модели управления численностью сотрудников предприятия* // Проблемы управления. – 2017. – №1 – С. 19–30.
3. БОРИСОВ А.Б. *Большой экономический словарь* / А.Б. Борисов. – М.: Книжный мир, 2003. – 895 с.
4. ВАЛЬД А. *Последовательный анализ*. – М.: Физматлит, 1960. — 328 с.

5. ЛОПАТНИКОВ Л.И. *Экономико-математический словарь: Словарь современной экономической науки*. – М.: Дело, 2003. – 520 с.
6. НИКИФОРОВ И.В. *Последовательное обнаружение изменения свойств временных рядов*. – М.: Наука, 1983. – 199 с.
7. *Процедура Каплана-Мейера. MachineLearning.ru. Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных*. – URL: http://www.machinelearning.ru/wiki/index.php?title=Процедура_Каплана-Мейера (дата обращения: 12.12.2017).
8. СМИРНОВ В.Т., СОШНИКОВ И.В., РОМАНЧИН В.И., СКОБЛЯКОВА И.В. *Человеческий капитал: содержание и виды, оценка и стимулирование: монография* / Под ред. В.Т. Смирнова. – М.: Машиностроение-1; Орел: ОрелГТУ, 2005. – 513 с.
9. ШИРЯЕВ А.Н. *Вероятность*: В 2-х кн. – М.: МЦНМО, 2007.
10. ШИРЯЕВ А.Н. *Вероятностно-статистические методы в теории принятия решений*. – М.: МЦНМО, 2014. — 144 с.
11. ШИРЯЕВ А.Н. *Стохастические задачи о разладке*. – М.: МЦНМО, 2016. – 392 с.
12. *Экономический словарь* / Е.Г. Багудина и др. – М.: ТК Велби; Проспект, 2004. – 624 с.
13. BARTROFF J., FINKELMAN M., TZE LEUNG LAI *Modern Sequential Analysis and Its Applications to Computerized Adaptive Testing* // *Psychometrika*. – September 2008. – Vol. 73, Iss. 3. – P. 473–486.
14. BLACKBURN K., VARVARIGOS D. *Human Capital Accumulation in a Stochastic Environment: Some New Results on the Relationship Between Growth and Volatility*. – URL: <http://www.socialsciences.manchester.ac.uk/medialibrary/cgbcr/discussionpapers/dpcgbc74.pdf> (дата обращения: 15.02.2016).
15. COX D.R. *Regression models and life tables (with discussion)* // *J. of the Royal Statistical Society. Series B (Methodological)*. – 1972. – Vol. 34, No. 2. – P. 187–220.

16. FAN X., SESHADRI A., TABER C. *Estimation of a Life-Cycle Model with Human Capital, Labor Supply and Retirement*. University of Chicago. Department of Economics. Workshops. – URL: <http://www.ssc.wisc.edu/~aseshadr/WorkingPapers/FST.pdf> (дата обращения: 12.09.2016).
17. FLAMHOLTZ E. *Human resource accounting: advances in concepts, methods, and applications*. – N.-Y.: Springer Science+Business Media, 1999. – 390 p.
18. FLAMHOLTZ E., BULLEN M., HUA W. *Human resource accounting: a historical perspective and future implications* // Management Decision. – 2002. – Vol. 40, Iss. 10. – P. 947–954.
19. KAHN W.A. *Psychological Conditions of Personal Engagement and Disengagement at Work* // Academy of Management Journal. – 1990. – Vol. 33, No. 4. – P. 692–724.
20. KAPLAN E.L., MEIER P. *Nonparametric Estimation from Incomplete Observations* // J. of the American Statistical Association. – June 1958. – Vol. 53, No. 282. – P. 457–481.
21. NIKIFOROV I.V. *Sequential detection/isolation of abrupt changes* // Sequential Analysis. Design Methods and Applications. – 2016. – Vol. 35, Iss. 3. – P. 268–301.
22. PALACIOSY M. *Human Capital as an Asset Class Implications From a General Equilibrium Model*. – Owen Vanderbilt University. Graduate School of Management. – URL: http://www2.owen.vanderbilt.edu/miguel.palacios/index_files/Palacios_Human_Capital_as_an_Asset_Class.pdf (дата обращения: 12.02.2016).
23. SMITH D., SILVERSTONE Y., LAJTHA A. *A new lens on business advantage: Human capital strategy and the drive for high performance* // Accenture, 2013.

STATISTICS OF HIRING AND ATTRITION PROCESSES

Mikhail Belov, IBS, Moscow, Cand.Sc. (mbelov@ibs.ru).

Abstract: The paper presents research in the field of mathematical modeling of the economy of human capital and the life cycle of employees, as well as studying the dependency of the economic effects of business on these factors. The traffic of employees between the firm and the labor market is studied. The characteristics of the life cycle process of the employee (LCE), considered as the total time from the moment when the employee accepted the job offer from the company until he leaves the company. The materiality of the influence of the duration of the average work time in the firm, the duration of the employee's adaptation phase in hiring and the phase preceding the dismissal on the firm's economy is demonstrated. The process of dismissal of employees is analyzed by non-parametric methods. As a result, the impossibility of a highly reliable assessment of the probability distribution function of the employee's time in the firm based on observations was shown. The procedures of statistic sequential analysis are proposed to identify the moment of change in the characteristics of the life cycle of employees - the probability of accepting job offers and the intensity of dismissal at will. The developed methods and the results obtained are illustrated by numerical examples from the author's practice in the field of human capital economics and personnel management.

Keywords: attrition process, poison estimation, censored sampling, sequential analysis, fault detection.

УДК 519.21; 681.518

ББК 22.18

DOI: <https://doi.org/10.25728/ubs.2019.77.5>

*Статья представлена к публикации
членом редакционной коллегии Д.А. Новиковым.*

Поступила в редакцию 25.12.2017.

Опубликована 31.01.2019.

МНОГОАЛЬТЕРНАТИВНОСТЬ: ЭВОЛЮЦИОННАЯ СТРАТЕГИЯ БИОЛОГИЧЕСКИХ СИСТЕМ

Подвальный С. Л.¹, Васильев Е. М.²

(Воронежский государственный
технический университет, Воронеж)

Излагаются результаты математического моделирования и количественного анализа биологического свойства разнообразия. Целью работы является методологическое обобщение указанного свойства живой природы в виде кибернетической концепции многоальтернативности как эволюционной стратегии функционирования сложных систем. Исследование добиологического этапа возникновения жизни проведено на моделях квазивидов и сайзеров; анализ процесса молекулярно-генетической эволюции на уровне популяций – на модели марковского случайного процесса; моделирование эволюционных процессов в сложной иерархической экосистеме с замкнутым трофическим циклом – на основе дифференциальных уравнений массового баланса. В итогах работы сформулирован ряд содержательных принципов, составляющих концепцию многоальтернативности: принципы многоуровневости, разделения функций и дискретности. Конструктивный характер этих принципов открывает возможность практической реализации стратегии многоальтернативности при построении сложных систем управления различного назначения.

Ключевые слова: механизмы эволюции, эволюционная кибернетика, концепция многоальтернативности.

1. Введение

В предлагаемой статье обсуждается на первый взгляд очевидное свойство живой природы – её разнообразие. Генетические причины разнообразия живого мира, подчиняющиеся законам расщепления Г. Менделя [33], вкупе со специализацией образа жизни у его представителей [18], в настоящее время не вызывают сомнений. Ключевая роль биоразнообразия для устойчивости экосистем в изменяющихся условиях существования отмечается в работах [3, 11, 30, 31, 37, 66, 67, 70]. Однако ниже с помощью соответствующих математических моделей мы

¹ Семён Леонидович Подвальный, д.т.н., профессор (spodvalny@yandex.ru).

² Евгений Михайлович Васильев, к.т.наук, с.н.с. (vgtu-aits@yandex.ru).

сделаем попытку показать, что это свойство не является обычным следствием указанных выше причин, но несёт в себе более глубокий смысл и предназначение для возникновения и развития жизни, являясь эволюционной стратегией многоальтернативности биологических систем на всех уровнях их существования – от молекулярного до экосообществ.

Целью работы является методологическое обобщение указанного свойства живой природы в виде кибернетической концепции многоальтернативности как эволюционной стратегии функционирования сложных систем.

Формирование концепции многоальтернативности в современных условиях важно не только для построения политики природопользования, но необходимо для понимания и использования её кибернетических принципов в прикладных областях человеческой деятельности – социально-экономических и информационно-технических системах широкого назначения [38].

Проведение прямой биологической аналогии и применение принципов многоальтернативности при построении сложных систем управления открывает возможность искусственного воспроизведения в этих системах высоких адаптационных свойств, присущих биологическим системам, и преодоления основной проблемы управления большими системами – проблемы их высокой размерности.

Настоящая статья является продолжением предыдущей работы [39] и распространяет содержание принципов многоальтернативности на процессы эволюции.

С гносеологических позиций обращение к свойству многоальтернативности эволюции соответствует общей диалектике научного познания: от констатации явления – к теоретическому обобщению закономерностей происходящих в нём процессов, позволяющему встроить рассматриваемое явление в единую систему знаний о природе.

2. Многоальтернативность в добиологической эволюции

Современные гипотезы возникновения жизни (добиологический период её организации) [19, 45, 46, 55, 59, 62–64, 68] во многом сходны в том, что этот процесс является процессом упорядочения и самокопирования относительно устойчивых молекулярных элементов – репликаторов, в роли которых могут выступать короткие цепочки из некоторого набора элементарных компонентов.

Рассмотрим значение многоальтернативности в указанном процессе.

В качестве исходного примера исследуем модель добиологической эволюции молекулярных соединений – модель квазивидов, предложенную М. Эйгеном [55]:

$$(1) \quad \begin{cases} \dot{x}_i = \left(w_i - \sum_{j=1}^n (w_j x_j) \right) x_i; \\ \sum_{j=1}^n x_j = 1; \quad i = \overline{1, n}, \end{cases}$$

где w_i – степень приспособленности (селективная ценность) макромолекулы вида i ; x_i – частота встречаемости молекул вида i в популяции; $\sum_{j=1}^n (w_j x_j)$ – средняя селективная ценность всех видов молекул; n – число различных видов молекул в популяции.

Общий вид этой модели, а также численное решение (1) для $w = [1 \ 0,995 \ 0,99 \ 0,985 \ 0,98]^T$, представленное на рис. 1, подтверждает теоретически предсказуемый результат: наиболее приспособленный вид молекул, обладающий степенью приспособленности выше средней, вытесняет все другие виды вплоть до их полного вымирания.

Однако если изменившиеся условия среды приводят к снижению продуктивности этого доминирующего вида (момент времени $t = 1000$ на рис. 1), то его место может занять очередной по производительности вид репликаторов. (Здесь и далее в

статье не используются конкретные единицы измерения времени, выбор которых не влияет на анализ хронологической последовательности событий и может быть произвольным).

Если же эти изменения среды произойдут на фоне единственного вида молекул, то рассматриваемый процесс репликации завершится безрезультатно, т.е. полной гибелью всех исходных молекул-репликаторов.

Если в уравнение (1) ввести составляющие $\sum_j^n (\psi_{ij}x_j)$, описывающие с помощью коэффициентов ψ_{ij} мутационные потоки в каждый из видов i со стороны других видов $j \neq i$ молекул [46, 55]:

$$(2) \quad \dot{x}_i = \left[w_i - \left(\sum_{j=1}^n (w_j x_j) + \sum_{\substack{k=1 \\ k \neq j}}^n \sum_{\substack{j=1 \\ j \neq k}}^n (\psi_{kj} x_j) \right) \right] x_i + \sum_{\substack{j=1 \\ j \neq i}}^n (\psi_{ij} x_j),$$

то даже при малых значениях мутационных потоков процесс самокопирования претерпевает качественные изменения (рис. 2).

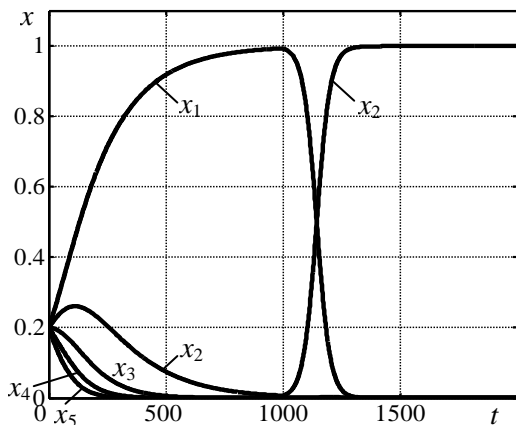


Рис. 1. Процесс неустойчивой эволюции (1) для популяции из пяти элементов с частотами встречаемости x_1, \dots, x_5

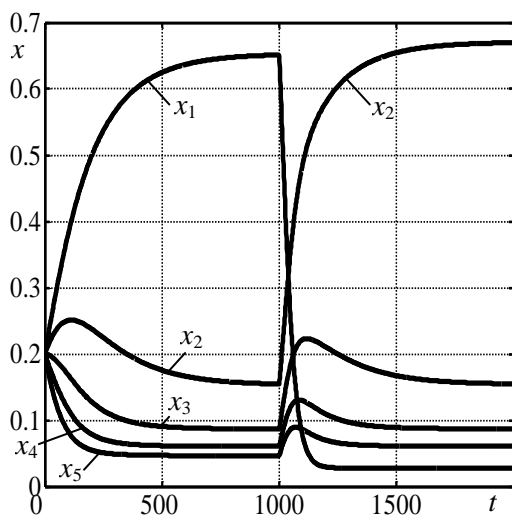


Рис. 2. Формирование популяции с устойчивым разнообразием (2) конкурирующих элементов с частотами x_1, \dots, x_5

На рис. 2, полученном для очень малого по сравнению с w значения параметра мутаций $\psi_{ij} = 0,001$, показано, что процесс эволюции совершается с некоторым распределением концентраций всех начальных видов элементов популяции и образованием их устойчивого разнообразия, поддерживающего сохранение этих элементов при изменяющихся условиях самокопирования некоторых из них. Этот распределённый итог получил название квазивида, содержащего эволюционно устойчивое вероятностно-распределённое многообразие близких по организации макромолекул.

Таким образом, уже на уровне элементарных макромолекул многообразие результатов копирования, вызванное его малыми мутациями, обеспечивает прогрессивное протекание эволюционного процесса, т.е. является необходимым условием добиологической организации. Без мутационных вариаций селективной ценности устойчивая репликация макромолекул невозможна.

Рассмотрим теперь более сложную модель добиологической эволюции – модель сайзеров (от англ. SYSER – SYstem of SELF-Reproduction – самовоспроизводящаяся система) [45, 46, 51, 71], содержащую в своей структуре элементы биологических организмов: полинуклеотидную макромолекулу (матрицу), хранящую информацию о своём строении, ферменты (катализаторы) трансляции, обеспечивающие в соответствии с этой информацией синтез строительных элементов макромолекулы, и ферменты репликации, выполняющие функции самокопирования.

Процесс самовоспроизводства макромолекул в результате совместного взаимодействия указанных элементов в сайзерах с единственной полинуклеотидной матрицей (в другой терминологии – сайзерах со сцеплёнными матрицами [45]) описывается уравнениями [46]

$$(3) \quad \begin{cases} \dot{x}_i = x_i \sum_{j=1}^n (a_j y_j) - x_i \sum_{k=1}^n x_k \sum_{j=1}^n (a_j y_j); \\ \dot{y}_i = x_i \sum_{j=1}^n (b_j z_j) - y_i \cdot 2 \sum_{k=1}^n x_k \sum_{j=1}^n (b_j z_j); \\ \dot{z}_i = x_i \sum_{j=1}^n (b_j z_j) - z_i \cdot 2 \sum_{k=1}^n x_k \sum_{j=1}^n (b_j z_j), \end{cases}$$

в которых x_i , z_i , y_i – концентрации нуклеотидных матриц, ферментов трансляции и ферментов репликации в сайзере типа i , причём эти концентрации подчиняются условиям постоянства их общих суммы: $\sum_{i=1}^n x_i = 1$, $\sum_{i=1}^n (y_i + z_i) = 1$; a_j – коэффициенты, отражающие скорость синтеза матриц, а b_j – скорость синтеза ферментов y_j и z_j .

Численное решение системы уравнений (3) показано на рис. 3, 4 и свидетельствует об отсутствии эволюционных изменений: начальные концентрации матриц не изменяются. Причиной этого заключается в гомогенности молекулярной популяции, в которой скорость синтеза компонентов x_i недифференциро-

ванно определяется коэффициентами $\sum_{j=1}^n (a_j y_j)$, одинаковыми для любого типа сайзера, и условие $\sum_{i=1}^n x_i = 1$ приводит к равенству

$$x_i \sum_{j=1}^n (a_j y_j) = x_i \sum_{k=1}^n x_k \sum_{j=1}^n (a_j y_j),$$

вследствие которого $\dot{x}_i = 0$ [46]. Гомогенность популяции приводит к её эволюционному застою.

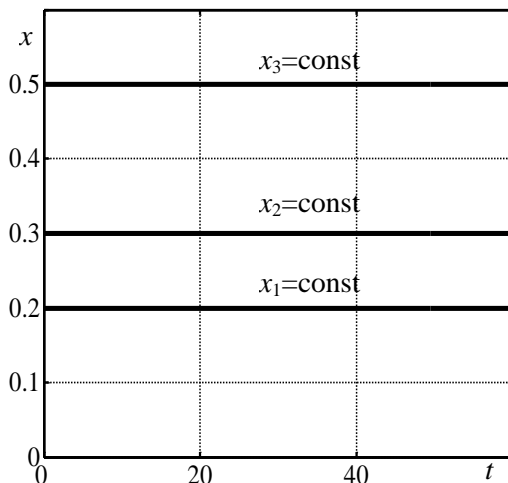


Рис. 3. Эволюционный застой в однородной ферментной среде (3) из трёх типов элементов с концентрациями матриц x_1, x_2, x_3

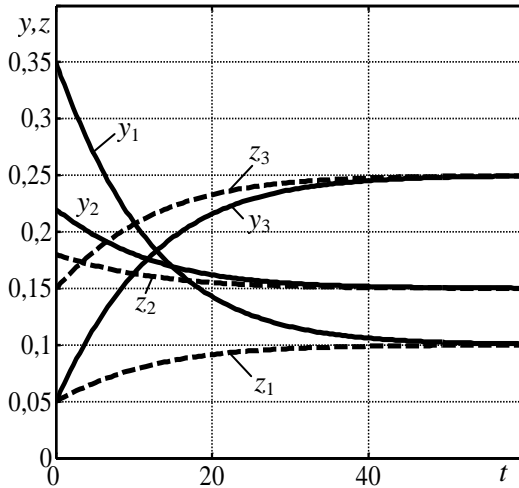


Рис. 4. Формирование однородной ферментной среды (3)
 (y – концентрация ферментов репликации, z – концентрация ферментов трансляции)

Усложнение модели (3) путём пространственного разделения сайзеров – их размещения в фазово-обособленные образования – коацерваты [51], в каждом из которых сосредоточены сайзеры только одного типа, позволяет для отдельно взятого коацервата переписать уравнения (3) в виде

$$(4) \quad \begin{cases} \dot{x} = aux - x(aux + bzx + dzx); \\ \dot{y} = bzx - y(aux + bzx + dzx); \\ \dot{z} = dzx - z(aux + bzx + dzx), \end{cases}$$

где коэффициенты a, b, d , характеризующие скорости синтеза, индивидуальны для каждого коацервата. Конкуренция между коацерватами описывается уравнениями [46]

$$(5) \quad \dot{v}_i = \left(w_i - \sum_{j=1}^n (w_j v_j) \right) v_i,$$

в которых v_i – доля коацерватов типа i в их общем количестве $\sum_{i=1}^n v_i = 1$; $w_i = a_i y_i x_i + b_i z_i x_i + d_i z_i x_i$ – коэффициент, характеризующий среднюю скорость синтеза коацерватов типа i , а $\sum_{j=1}^n (w_j v_j)$ – среднюю скорость синтеза всех n коацерватов.

Результат численного решения (4), (5) представлен на рис. 5 и указывает на возникновение естественного отбора наиболее приспособленных типов макромолекулярных матриц, при этом именно в силу доминирующего синтеза макромолекул одного вида устойчивого эволюционного процесса всей совокупности коацерватов не возникает. Однако введение в (5) даже малых поперёк мутационных потоков $\psi_{ij} = 0,001$, аналогично (2), позволяет получить процесс эволюции, устойчивый к изменению внешних условий (момент времени $t = 1000$ на рис. 6), т.е. подтверждается полученный при анализе квазивидов результат: эволюционный процесс устойчив не вопреки, а благодаря мутациям селективной ценности макромолекул.

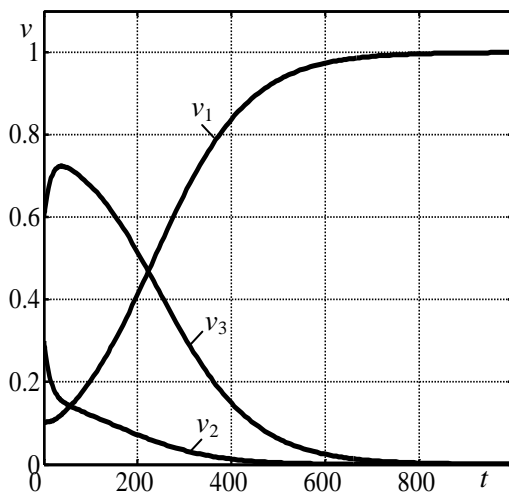


Рис. 5. Возникновение конкуренции и естественного отбора в результате перехода от однородной ферментной среды (рис. 3,4) к селективному синтезу (5)

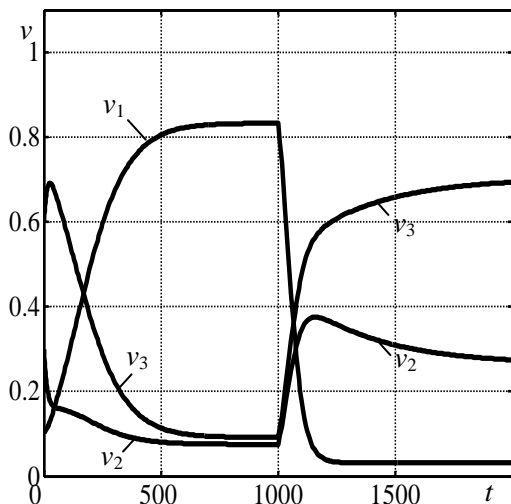


Рис. 6. Формирование популяции с устойчивым разнообразием конкурирующих элементов с частотами v_1 , v_2 , v_3 путём введения в (5) мутационных потоков

Таким образом, анализ моделей добиологического этапа эволюции жизни показывает, что для возникновения и устойчивого существования макромолекул, способных к саморепликации, необходимо выполнение ряда взаимосвязанных условий:

- наличие некоторого многообразия конкурирующих типов макромолекул, в том числе близких по строению;
- межмолекулярная избирательность (дифференцированность) приспособительных и репликативных механизмов разных типов молекул;
- совместность существования вариантов макромолекул, обеспечивающая перекрёстные мутационные потоки и общий эволюционный прогресс популяции в целом, при котором те или иные преимущества одного элемента популяции используются всеми её другими элементами.

Указанные механизмы входят составной частью в более широкое понятие, свойственное уже биологическим системам – многоальтернативную стратегию эволюции.

3. Стратегия многоальтернативности на биологическом уровне эволюции

Перейдём к анализу стратегии многоальтернативности на более высоком – молекулярно-генетическом уровне эволюции, определяющем процессы возникновения и развития биологических видов в популяциях.

В соответствии с выводами популяционной генетики [43, 49, 58] доминирующим и направляющим механизмом внутривидовой эволюции является естественный отбор особей, имеющих в текущих условиях внешней среды преимущественные приспособительные реакции, обусловленные разнообразием генетических структур внутри популяции. Основным источником этого разнообразия являются мутации генов. При этом сами мутации, как таковые, не носят направленного характера, т.е. нейтральны [22, 23], однако, накапливаясь в популяции, они образуют скрытый, мобилизационный резерв её генофонда, обеспечивающий биологическому сообществу высокую приспособительную и эволюционную гибкость [54].

К классическим теоремам приспособительной эволюции относят теорему Фишера [65], согласно которой средняя степень приспособленности w популяции возрастает только при наличии её вариаций (генной дисперсии) внутри популяции, причём средняя скорость изменения приспособленности пропорциональна этой дисперсии. Формальная запись теоремы Фишера может быть получена из эволюционной модели (1):

$$(6) \quad \begin{cases} \dot{x}_i = \left(w_i - \sum_{j=1}^n (w_j x_j) \right) x_i; \\ \sum_{j=1}^n x_j = 1; \quad i = \overline{1, n}, \end{cases}$$

содержание которой здесь изменилось: w_i – степень приспособленности особей с генотипом i ; x_i – частота встречаемости генотипа i в популяции;

$$(7) \quad \sum_{i=1}^n w_i x_i = \bar{w},$$

\bar{w} – средняя приспособленность всех генотипов в популяции;
 n – число различных генотипов в популяции.

В соответствии с (7) получим, что скорость изменения средней приспособленности равна

$$(8) \quad \dot{\bar{w}} = \sum_{i=1}^n w_i \dot{x}_i = \sum_{i=1}^n w_i (w_i - \bar{w}) x_i,$$

или, с учётом (7),

$$(9) \quad \dot{\bar{w}} = \left(\sum_{i=1}^n x_i w_i^2 \right) - \bar{w}^2,$$

где правая часть равенства представляет собою дисперсию степеней генной приспособленности w_i в популяции. Таким образом, наличие вариаций в приспособленности особей (наличие альтернативных вариантов генотипов) является необходимым условием эволюционных, в первую очередь приспособительных изменений в популяции.

Справедливость этого вывода неоднократно подтверждалось результатами длительной селекционной работы по выведению, например, высокоурожайных сортов ячменя, низкорослой пшеницы, крупноразмерных лососевых рыб. Во всех указанных случаях наблюдалось существенное снижение генетического разнообразия (числа аллелей на локус) в отобранных для промышленного производства сортах и видах и, как следствие, снижение их устойчивости к изменениям внешней среды и заболеваниям [4].

Модель (6) справедлива для гипотетического варианта эволюции с некоторым изначально зафиксированным многообразием генотипов, в пределах и средствами которого реализуется механизм приспособительной эволюции. Очевидно, что при бесконечном разнообразии условий внешней среды указанный генетический запас будет исчерпан, что соответствует, в рамках этой модели, вырождению популяции.

Рассмотрим эволюционную роль генетической многоальтернативности в более реальных условиях, когда неограниченному многообразию внешних воздействий среды противопоставляется незафиксированное многообразие генетических структур, постоянно накапливающихся и обновляемых в ре-

зультате мутаций в популяции с непостоянным количеством особей $n(t)$, не превышающим некоторого максимального размера N популяции: $0 \leq n(t) \leq N$ [27], определяемого ограниченностью необходимых для её существования ресурсов.

Для анализа этого случая представим процесс молекулярно-генетической эволюции в популяции дискретным одноуровневым марковским процессом, имеющим множество состояний S_{ij} , $i = 0, 1, 2, \dots; j = 0, 1, \dots, N-1, N$, среди которых:

$S_{0,0}$ – состояние, соответствующее вырожденной популяции, т.е. $n(t) = 0$; вероятность выхода эволюционного процесса из этого состояния равна нулю;

$S_{i,1}, \dots, S_{i,N}$ – состояния популяции с числом особей $n = 1, \dots, N$ и набором i генетических изменений в популяции.

Предполагается, что рост популяции происходит с интенсивностью η , а накопление генофонда – с интенсивностью $n \cdot \mu$. События внешней среды описываются потоком с интенсивностью λ , при этом будем считать, что потоки событий среды и накопления генофонда комплементарны, т.е. в процессе эволюционного приспособления при всяком очередном изменении среды используется («расходуется») соответствующая генетическая структура и разнообразие генофонда уменьшается на некоторую условную единицу.

Модель описанного процесса в виде графа показана на рис. 7.

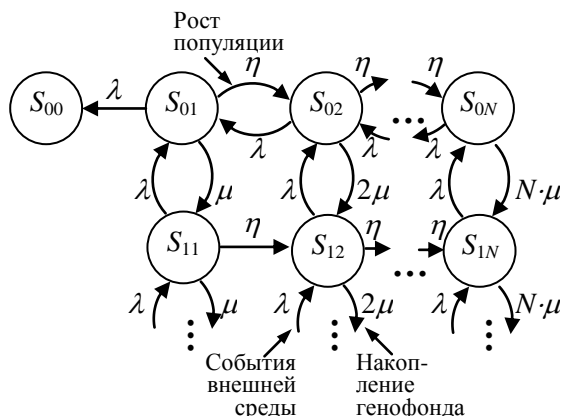


Рис. 7. Граф состояний популяции в процессе эволюции

Для количественного анализа этой модели воспользуемся тем обстоятельством, что скорость накопления генофонда популяции значительно превосходит скорость роста её численности, т.е. $\mu \gg \eta$. А.В. Яблоков в [58] отмечает, что каждая особь животных и растений в популяции обычно переменна по десятку тысяч генов, и генотип каждой особи должен содержать хотя бы одну вновь возникшую мутацию. Таким образом, ограничиваясь интересующим нас анализом влияния на процесс эволюции генетического разнообразия, можно перейти к одномерному варианту модели, представленному на рис. 8, на котором выделены три группы состояний популяции:

S_0 – поглощающее состояние, соответствующее вырожденной популяции, $n(t) = 0$;

S_1, \dots, S_N – состояния популяции, характеризующиеся одновременным изменением её численности $n(t)$ и генофонда, причём, в соответствии с приведённой выше нижней оценкой скорости мутагенеза, принято, что появление каждой новой особи эквивалентно новой генетической структуре;

S_{N+1}, \dots – состояния популяции, достигшей заданного ограничения по своему размеру N , но продолжающей накапливать с интенсивностью $N \cdot \mu$ генетические изменения, парирующие изменения внешней среды.

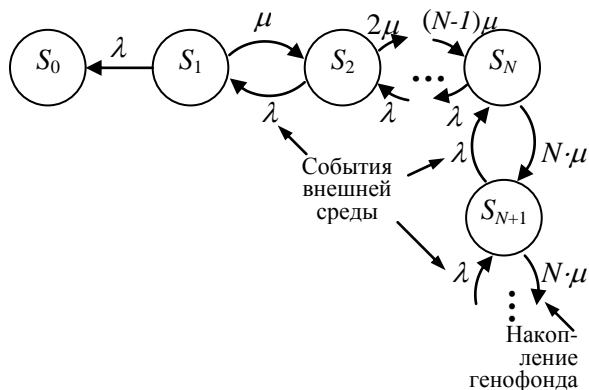


Рис. 8. Одномерная марковская модель накопления генофонда

В полученной одномерной марковской модели процесс изменения во времени вектора вероятностей $P(t) = [p_0(t) p_1(t) p_2(t) \dots]$ состояний S_0, S_1, S_2, \dots популяции описывается дифференциальным уравнением Колмогорова [12]:

$$(10) \frac{dP(t)}{dt} = P(t) \cdot \Lambda,$$

в котором Λ – матрица интенсивностей переходов (инфинитезимальная матрица) в рассматриваемом случае имеет нулевую первую строку:

$$(11) \Lambda = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & \dots \\ \lambda & -(\lambda + \mu) & \mu & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots \\ 0 & 0 & \lambda & -(\lambda + N \cdot \mu) & N \cdot \mu & \dots \\ 0 & 0 & 0 & \lambda & -(\lambda + N \cdot \mu) & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Общий вид решения уравнения (10):

$$(12) P(t) = P(0)e^{\Lambda t},$$

где $P(0) = [p_0(0) p_1(0) p_2(0) \dots]$ – вектор вероятностей начального состояния популяции.

С целью численного исследования эволюционных изменений в популяции на модели (10) были приняты следующие значения её параметров: максимальный размер популяции $N = 50$; число возможных генетических состояний $M = 500$, во много раз превышающее N , что позволяет экстраполировать полученные результаты на случай с неограниченным значением M ; интенсивность потока генетических изменений в одной особи $\mu = 1$; интенсивность событий внешней среды $\lambda = 25$.

Все рассматриваемые здесь потоки событий принимаются простейшими пуассоновскими (стационарными, ординарными и не имеющими последствия), т.е. интервал времени между соседними событиями в этих потоках имеет показательное распределение с плотностью $\varphi(t) = \kappa e^{-\kappa t}$, где κ – интенсивность потока. Заметим, что принятый простейший характер потоков предопределяет возможность применения в данном исследова-

нии марковской модели молекулярно-генетической эволюции [12] и, соответственно, уравнения (10).

Результаты моделирования показали, что при фиксированных значениях указанных параметров популяция имеет всего две альтернативы своего изменения: вырождение с вероятностью p_0 и неограниченного роста с вероятностью $p_\infty=1-p_0$, причём эти вероятности зависят от численности популяции $n_{\text{нач}}$ в момент условного времени начала исследования. На рис. 9 показана зависимость вероятности вырождения p_0 от соотношения интенсивности потока генетических изменений $n_{\text{нач}} \cdot \mu$ во всей популяции к интенсивности λ потока событий внешней среды. Из рис. 9 следует, что устойчивое существование популяции обеспечивается, по меньшей мере, при двукратном превышении скорости генетических изменений над скоростью вариаций внешней среды. Обратное соотношение практически всегда приводит к вырождению популяции.

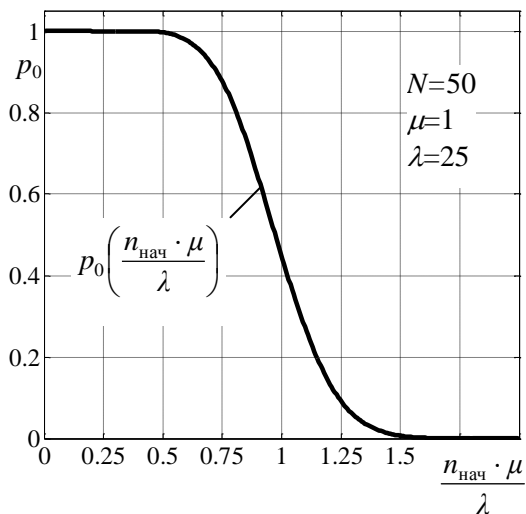


Рис. 9. Изменение вероятности p_0 вырождения популяции в зависимости от соотношения интенсивностей потока генетических изменений и потока событий внешней среды

На рис. 10, 11 иллюстрируется изменения в популяциях с различной начальной численностью: при $n_{\text{нач}} = 40$ наблюдается увеличение численности популяции до максимального значения $N = 50$, а затем неограниченное увеличение её генетических состояний S – от меньших номеров к большим – 60...70...80...100...; при $n_{\text{нач}} = 10$ изменения популяции имеют противоположную направленность: из состояния S_{10} происходит переход в состояния с меньшей численностью $S_8...S_6...S_1$ и, наконец, S_0 с вероятностью $p_0 = 1$, соответствующей вырождению популяции.

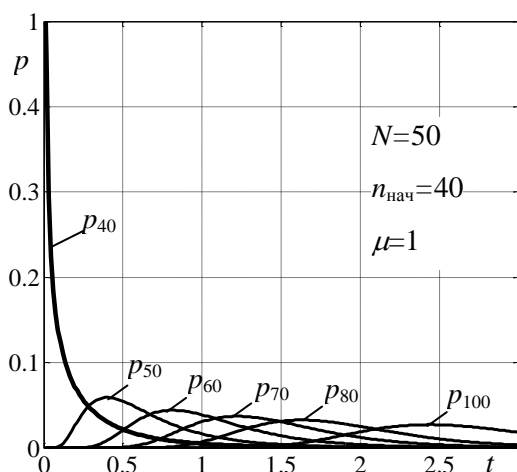


Рис. 10. Рост популяции и её генофонда при начальной численности $n_{\text{нач}} = 40$

Проведённое исследование процессов молекулярно-генетической эволюции свидетельствует о том, что устойчивое существование и развитие популяции реализуется не за счёт роста её численности, а за счёт увеличения её генетического многообразия со скоростью, превышающей скорость угрожающих существованию популяции изменений внешней среды, т.е. эволюция биосистем на уровне популяций необходимо обеспечивается многообразием, многоальтернативностью её молекулярно-генетических структур. Это многообразие тесно связано с

критической нижней численностью популяции, при переходе через которую скорость накопления генетических альтернатив становится меньше скорости появления неблагоприятных факторов среды, и популяция в статистическом смысле будет обречена на вырождение.

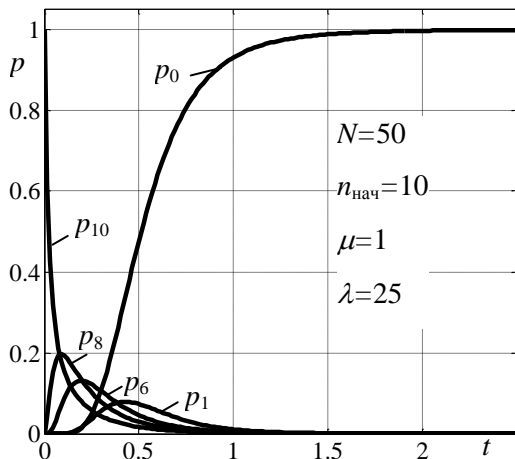


Рис. 11. Вырождение популяции при начальной численности $n_{\text{нач}} = 10$

Указанное обстоятельство учитывается на практике при восстановлении популяций, находящихся под угрозой исчезновения. Уникальным примером такого восстановления являются чатемские мухоловки (острова Новой Зеландии). Скорость накопления генофонда всего пяти сохранившихся мухоловок, обречённых на вымирание в естественной среде, но помещённых в стабильные, искусственно поддерживаемые благоприятные условия, оказалась достаточной для формирования полностью жизнеспособной популяции этих птиц из более чем 250 особей [7]. В естественных условиях существование малочисленных популяций также возможно, образуя так называемые эндемичные виды в ограниченных, как правило, географически изолированных ареалах с устойчивым и малоизменяемым составом экосистемы. Это наблюдение также подтверждает сде-

ланный вывод о роли соотношения молекулярно-генетического многообразия и факторов среды обитания в эволюции видов.

Следует обратить внимание на важное, зачастую упускаемое из виду свойство эволюционного процесса, принципиально определяющее саму возможность накопления генетических альтернатив и, как следствие, возможность естественного отбора наиболее приспособленных из них. Это свойство заключается в дискретности наследуемых признаков [49], вытекающей, в свою очередь, из дискретности генетических структур. Р. Докинз в [19] приходит к выводу о том, что именно дискретность форм гена определяет его высокую устойчивость – эволюционную «бессмертность». Свойство дискретности генетических структур в настоящее время обобщено до принципа блочно-модульной организации и эволюции молекулярно-генетических систем, согласно которому эволюция биологических образований любой степени сложности происходит в результате комбинирования уже существующих, устойчиво функционирующих блоков-модулей более низкого уровня сложности [22, 42, 44]. Указывая на эволюционное значение свойства дискретности биологических структур Э.М. Галимов в [16] отмечает, что «наиболее экономный способ производства низкоэнтропийного продукта состоит в комбинировании уже имеющихся низкоэнтропийных структур; эволюция в нашем понимании должна происходить не только и даже не столько путем малых изменений, сколько скачками, обусловленными новыми сочетаниями старых структур». Указанную общую черту эволюционных скачков к более высоким уровням организации системы путём накопления множества жизнеспособных подсистем низкого уровня и их последующей интеграции в метасистему с одновременной специализацией (разделением функций) внутри неё В.Ф. Турчин обобщил в теорию метасистемных переходов, составляющих кибернетический механизм эволюции [50].

Таким образом, принцип модульности в построении и развитии биологических структур обеспечивает возникновение многоальтернативных форм живой материи и является неотъемлемым свойством и условием эволюционного процесса.

4. Многоальтернативность в экосистемах

Рассмотрим стратегии многоальтернативности на уровне организации жизни, охватывающем множество совместно существующих популяций различных видов организмов – биоценозов, или, в более широком смысле, с учётом абиотических факторов – экосистем.

Исключительное внимание к эволюционным процессам на уровне экосистем объясняется наличием в большинстве из них, по выражению В.И. Вернадского [14], «нового геологического фактора» – человечества, активно изменяющего среду своего обитания, нарушая при этом естественное динамическое равновесие и разнообразие в биосообществах, т.е. нарушая процессы их эволюции [13]. Каноническими примерами нарушения устойчивости биосистем служат широко известные факты катастрофического для местной фауны и флоры роста численности кроликов, завезённых в XIX веке в Австралию, воробьёв – в Северную Америку, колорадского жука – в Европу [49, 58, 56]. Примечательно, что усилия по восстановлению нарушенного равновесия в экосистеме давали результаты только после применения соответствующего «биологического оружия»: так, борьба за снижение численности кроликов привела к успеху только после внедрения в популяцию этих животных специфического паразита, сдерживающего их размножение. Нарушение экологического равновесия, вызванное в начале прошлого века активным уничтожением американского серого волка как опасного хищника, было устранено искусственной реинтродукцией (повторным заселением) волков в Йеллоустонский национальный парк под контролем экологов [7]. Последующие наблюдения за изменениями биосферы этого заповедника подтвердили правильность принятого решения [29]. Вытекающие из подобных фактов актуальность и необходимость сохранения разнообразия в экосистемах осознаны достаточно давно и зафиксированы в Конвенции о биологическом разнообразии, принятой ООН в 1992 году [60].

В отличие от молекулярно-генетического уровня, процессы эволюции в биосистемах реализуются не только механизмом

конкуренции за ресурсы существования, но в результате более широкого спектра взаимодействий своих элементов: от совместной взаимопомощи – симбиоза – до отношений «хищник–жертва». Качественной особенностью этих отношений является появление в биоценозах биологически изолированных сообществ, образующих дополнительные трофические (пищевые) уровни системы – иерархическую структуру таксонов [49]. Общая схема взаимодействия трофических уровней, образующая замкнутый круговорот вещества и энергии в экосистеме, представлена на рис. 12 [21].

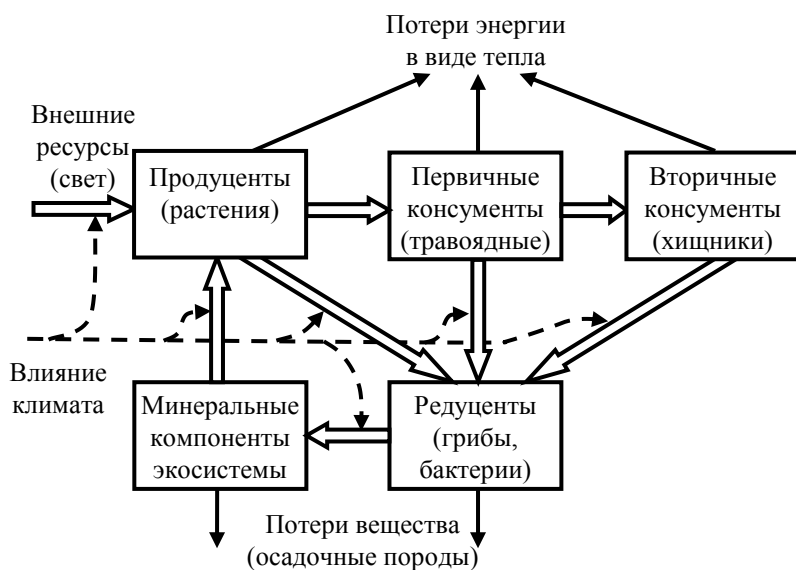


Рис. 12. Укрупнённая схема взаимодействия элементов экосистемы

Как уже отмечалось, в рамках каждого трофического уровня борьба за пищевые и пространственные ресурсы приводит к разделению экологических ниш между конкурирующими видами – их специализации [49], причём виды, имеющие в текущих равновесных условиях низкую конкурентоспособность, не исчезают, а сохраняются в малых нишах как биологический

резерв экосистемы [28, 69]. Ограниченность ресурсов каждой из ниш играет роль естественного ограничения в накоплении общей биомассы экосистемы, однако не исключает возможности освоения новых экологических ниш, т.е. роста разнообразия компонентов экосистемы. В целом, устойчивое, сбалансированное развитие экосистемы характеризуется стремлением к наиболее полному освоению доступных энергетических ресурсов и в освоении новых ресурсов, например, в нишах с загрязнениями или экстремальными условиями. Иными словами, в периоды устойчивого развития разнообразие в экосистеме носит не вытесняющий, а аддитивный – накопительный характер с нарастающей дифференциацией (разделением) функций преобразования вещества и энергии [17].

В кризисах, вызванных изменением климата или деятельностью человека, в экосистеме происходит нарушение равновесия: ранее доминирующие виды утрачивают свою ведущую роль в круговороте веществ в экосистеме вплоть до вымирания, но равновесие в системе восстанавливают альтернативные виды из малых ниш, заполняя освободившиеся или вновь появившиеся ниши. В работе [5] приводится пример скачкообразного увеличения биомассы ранее малочисленных солоноустойчивых двустворчатых моллюсков по мере высыхания Аральского моря и повышения его солёности. В результате общая биомасса грунтовых организмов, несмотря на исчезновение их пресноводных видов, была восстановлена. Характерно, что процесс освоения изменившейся среды обитания сопровождался быстрым образованием новых видов моллюсков. E.D. Core, указывая в [61] на рост многообразия таксонов экосистемы в кризисные периоды её существования, определяет это явление как общее эволюционное правило.

Наличие иерархической структуры экосистемы (трофической пирамиды) оказывает дополнительное многоуровневое регулирующее действие, которое можно рассматривать как внутренний стабилизирующий фактор, выработанный в результате эволюции биоценозов с целью сбалансированного развития всех своих компонентов и более полного использования и накопления (концентрации) поступающей извне энергии. Уве-

личение запасов энергии в системе способствует её стабилизации и прогрессивной эволюции [6].

В равновесном, точнее – сбалансированном состоянии экосистема может существовать сколь угодно долго, постепенно увеличивая объём концентрируемой в своём замкнутом контуре энергии путём освоения новых экологических ниш и наращивания числа трофических уровней, т.е. путём роста биоразнообразия [17, 20, 36, 41].

Таким образом, процесс эволюции на уровне экосистемы может рассматриваться в виде чередующихся периодов стабильного, сбалансированного роста разнообразия и биомассы её компонентов с периодами существенного, кризисного изменения условий существования, во время которых сокращается разнообразие видов на всех уровнях экосистемы (в первую очередь на верхних уровнях трофической цепи). Возникающее, как следствие, ослабление конкуренции открывает возможность преимущественного развития ранее угнетённым или даже скрытым на генетическом уровне форм, появление которых обусловливает скачкообразное изменение состава всей трофической цепи, т.е. наиболее существенные эволюционные изменения в экосистеме.

Изложенное содержание основных принципов эволюции в экосистемах даёт возможность перейти к составлению математической модели, позволяющей исследовать влияние видового многообразия и иерархичности структуры на эволюционные процессы в экосистеме.

В качестве базовой теоретической предпосылки составления модели примем законы сохранения и круговорота вещества, перерабатываемого каждым элементом экосистемы [8, 9, 15, 26, 48]. Рассматриваемые потоки вещества в экосистеме представлены на рис. 13.

На рис. 13 обозначены:

x^0, x^1, x^2, x^3 – значения биомассы редуцентов, продуцентов и консументов первого и второго уровней соответственно (здесь и далее под биомассой подразумевается масса сухого вещества в $\text{кг}/\text{м}^2$, определяемая в конце очередного (текущего) годового интервала времени);

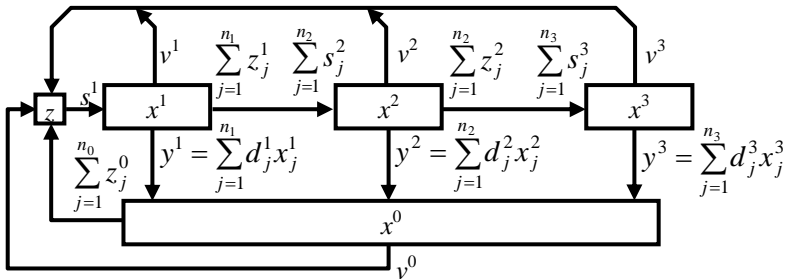


Рис. 13. Рассматриваемые потоки вещества в экологической системе

n_0, n_1, n_2, n_3 – число биологических видов, составляющие указанные элементы системы;

v^0, v^1, v^2, v^3 – годовой расход биомассы, эквивалентный затратам энергии на дыхание, выделение тепла или совершению работы для движения в элементах экосистемы, $\text{кг}/(\text{м}^2 \cdot \text{год})$. Эти затраты полностью компенсируются энергией, поступающей из внешней среды и, в силу этого, включены в замкнутый круговорот в виде эквивалентного вещества;

y^1, y^2, y^3 – биомасса продуктов выделения и естественного отмирания организмов в течение рассматриваемого года, $\text{кг}/(\text{м}^2 \cdot \text{год})$;

z_j^i – биомасса вида $j, j = 1, \dots, n_i$, элемента i системы, $i = 0, \dots, 2$, отдаваемая за год всему очередному (последующему) элементу трофической цепи, $\text{кг}/(\text{м}^2 \cdot \text{год})$;

z – масса накопленных минеральных (неорганических) компонентов экосистемы, которая может быть использована продуцентами, $\text{кг}/\text{м}^2$;

s_j^i – биомасса, получаемая за год видом $j, j = 1, \dots, n_i$, элемента i системы, $i = 1, \dots, 3$, от всего элемента $(i-1)$, предшествующего по трофической цепи, $\text{кг}/(\text{м}^2 \cdot \text{год})$.

С учётом принятых обозначений дифференциальные уравнения массового баланса в каждом элементе системы приобретают вид:

$$(13) \dot{x}^0 = \sum_{i=1}^3 y^i - \sum_{j=1}^{n_0} v_j^0 - \sum_{j=1}^{n_0} z_j^0;$$

$$(14) \dot{x}^1 = s^1 - \sum_{j=1}^{n_1} v_j^1 - \sum_{j=1}^{n_1} z_j^1 - y^1;$$

$$(15) \dot{x}^2 = s^2 - \sum_{j=1}^{n_2} v_j^2 - \sum_{j=1}^{n_2} z_j^2 - y^2;$$

$$(16) \dot{x}^3 = s^3 - \sum_{j=1}^{n_3} v_j^3 - y^3;$$

$$(17) \dot{z} = \sum_{i=0}^3 v^i + \sum_{j=1}^{n_0} z_j^0 - s^1,$$

где $\dot{x}^i = \sum_{j=1}^{n_i} \dot{x}_j^i$, $i = 0, \dots, 3$, – годовой прирост биомассы элемента i экосистемы, кг/(м²·год); \dot{z} – годовой прирост массы минеральных компонентов, кг/(м²·год).

Условия межэлементного баланса массы переносимого вещества:

$$(18) \sum_{j=1}^{n_1} z_j^1 = \sum_{j=1}^{n_2} s_j^2;$$

$$(19) \sum_{j=1}^{n_2} z_j^2 = \sum_{j=1}^{n_3} s_j^3.$$

Кроме того, годовое потребление продуцентов s^1 в числовом выражении не может превышать текущих запасов z минеральных компонентов системы:

$$(20) (s^1) \leq z;$$

Составляющие модели экосистемы (13)–(20) определим следующими соотношениями.

Для продуцентов:

$$(21) s^1 = \sum_{j=1}^{n_1} s_j^1;$$

$$(22) \quad s_j^1 = x_j^1 a_j^1 z_j - x_j^1 \sum_{i=1}^{n_1} b_{ij}^1 x_i^1;$$

$$(23) \quad z_j = z \cdot \frac{a_j^1}{\sum_{j=1}^{n_1} a_j^1};$$

$$(24) \quad \sum_{j=1}^{n_1} z_j = z;$$

$$(25) \quad z_j^1 = x_j^1 \sum_{i=1}^{n_2} a_{ji}^{12} x_i^2 - \sum_{i=1}^{n_2} \sum_{j=1}^{n_2} b_{ij}^2 x_i^2 x_j^2;$$

$$(26) \quad v_j^1 = c_j^1 x_j^1;$$

$$(27) \quad y_j^1 = d_j^1 x_j^1.$$

Здесь и далее обозначены:

a_j^1 – коэффициент, характеризующий относительную скорость потребления минеральных компонентов видом j элемента 1 системы, $\text{м}^2/(\text{кг}\cdot\text{год})$;

b_{ij}^k , $k = 0, \dots, 3$, – коэффициент влияния на прирост или потребление биомассы вида j межвидовой ($i \neq j$) и внутривидовой ($i = j$) конкуренции за пищевые или пространственные ресурсы в пределах элемента k системы, $\text{м}^2/(\text{кг}\cdot\text{год})$. Например, слагаемое

$\left(-x_j^1 \sum_{i=1}^{n_1} b_{ij}^1 x_i^1 \right)$ в (22) отражает снижение скорости прироста

биомассы продуцента j вследствие конкуренции этого вида со всеми другими продуцентами видов i (включая и внутривидо-

вую конкуренцию), а слагаемое $\left(-\sum_{i=1}^{n_2} \sum_{j=1}^{n_2} b_{ij}^2 x_i^2 x_j^2 \right)$ в (25) – снижение

скорости потребления любого вида j продуцента консументом первого уровня вследствие конкуренции консументов;

$z_j, j = 1, \dots, n^1$, – масса минеральных компонентов доступная для переработки виду j элемента 1 системы в течение рассматриваемого года, кг/м²;

a_{ji}^{12} – коэффициент, характеризующий относительную скорость потребления вида j элемента 1 видом i элемента 2 системы, м²/(кг·год);

$c_j^i, i = 0, \dots, 3$, – коэффициент (относительная скорость) эквивалентных потерь биомассы на дыхание, выделение тепла или совершению работы видом j элемента i системы, 1/год;

$d_j^i, i = 1, \dots, 3$, – коэффициент образования продуктов выделения и естественного отмирания организмов вида j элемента i системы, 1/год.

Для консументов первого уровня:

$$(28) s^2 = \sum_{j=1}^{n_2} s_j^2;$$

$$(29) v_j^2 = c_j^2 x_j^2;$$

$$(30) s_j^2 = x_j^2 \sum_{i=1}^{n_1} a_{ij}^{12} x_i^1 - x_j^2 \cdot n^1 \cdot \sum_{i=1}^{n_2} b_{ij}^2 x_i^2;$$

$$(31) z_j^2 = x_j^2 \sum_{i=1}^{n_3} a_{ji}^{23} x_i^3 - \sum_{i=1}^{n_3} \sum_{j=1}^{n_3} b_{ij}^3 x_i^3 x_j^3;$$

$$(32) y_j^2 = d_j^2 x_j^2,$$

где a_{ji}^{23} – коэффициент, характеризующий относительную скорость потребления вида j элемента 2 видом i элемента 3 системы, м²/(кг·год).

Аналогично для консументов второго уровня:

$$(33) s^3 = \sum_{j=1}^{n_3} s_j^3;$$

$$(34) v_j^3 = c_j^3 x_j^3;$$

$$(35) s_j^3 = x_j^3 \sum_{i=1}^{n_2} a_{ij}^{23} x_i^2 - x_j^3 \cdot n^2 \cdot \sum_{i=1}^{n_3} b_{ij}^3 x_i^3;$$

$$(36) \quad y_j^3 = d_j^3 x_j^3.$$

Для редуцентов:

$$(37) \quad y_j = \sum_{i=1}^3 y^i \cdot \frac{\sum_{i=1}^3 a_{ij}^0}{\sum_{i=1}^3 \sum_{j=1}^{n_0} a_{ij}^0};$$

$$(38) \quad \sum_{i=1}^3 y^i = \sum_{j=1}^{n_0} y_j;$$

$$(39) \quad v_j^0 = c_j^0 x_j^0;$$

$$(40) \quad z_j^0 = x_j^0 \sum_{i=1}^3 a_{ij}^0 y^i - b_{jj}^0 x_j^0 x_j^0,$$

a_{ij}^0 , $i = 1, \dots, 3$, $j = 1, \dots, n^0$, – коэффициент, характеризующий относительную скорость переработки биомассы элемента i системы видом j редуцентов, $\text{м}^2/(\text{кг}\cdot\text{год})$;

y_j , $j = 1, \dots, n^0$, – биомасса продуктов выделения и естественного отмирания всех продуцентов и консументов, получаемая видом j редуцентов в течение рассматриваемого года, $\text{кг}/(\text{м}^2\cdot\text{год})$.

Отметим, что $b_{ij}^0|_{i \neq j} = 0$ принято в силу незначительной межвидовой конкуренции продуцентов, обусловленной узкой специализацией характера их питания и экологических ниш [8]. Кроме того, в z_j^0 неявно включена биомасса отмерших редуцентов, переработанная функционирующими редуцентами. Предполагается также, что биомасса y^1 , y^2 , y^3 продуктов выделения и естественного отмирания организмов полностью перерабатывается редуцентами, т.е. потери – неиспользованные органические компоненты в экосистеме – отсутствуют.

Численное исследование экосистемы по уравнениям (13)–(20) проводилось на имитационной модели, реализованной в среде MatLab. Общая структура этой модели представлена на рис. 14.

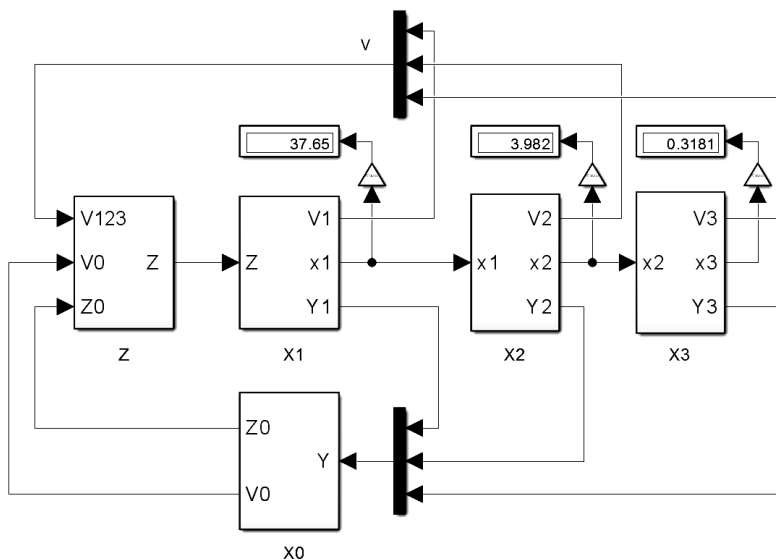


Рис. 14. Структурная схема имитационной модели замкнутой экологической системы

Для основных иерархических уровней экосистемы в её стационарном состоянии выбраны значения биомасс: $x^1 \approx 38 \text{ кг/м}^2$; $x^2 \approx 4 \text{ кг/м}^2$; $x^3 \approx 0,3 \text{ кг/м}^2$.

Начиная с момента $t = 300$ (рис. 15, 16) в систему, находящуюся в стационарном состоянии, начинает поступать поток внешней энергии (или эквивалентная масса неорганических компонентов) с интенсивностью $0,025 \text{ кг}$ за единицу времени. Общая биомасса всех компонентов системы начинает расти, однако при $t = 350$ в систему вводится критическое событие: из-за неблагоприятных условий среды на 40% снижается скорость поглощения третьим видом первичных консументов биомассы третьего вида продуцентов, т.е. уменьшается коэффициент a_{33}^{12} : при $t > 350$ третий столбец матрицы $a_{j3}^{12} = 10^{-3} \cdot [39,5 \quad 30,3 \quad 32,9]^T$ становится равным $a_{j3}^{12} = 10^{-3} \cdot [39,5 \quad 30,3 \quad 20,4]^T$.

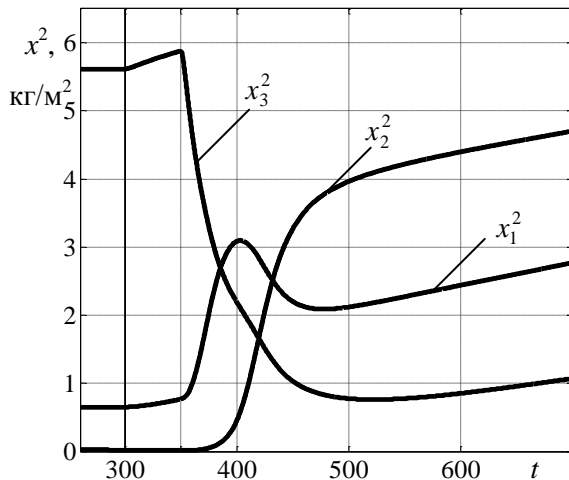


Рис. 15. Реакция двухуровневой ($x^3 = 0$) экосистемы на критическое событие для третьего вида консументов x_3^2

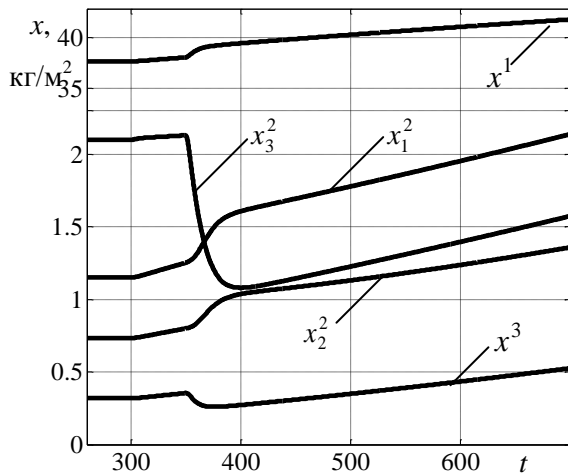


Рис. 16. Реакция трёхуровневой экосистемы на критическое событие для третьего вида консументов x_3^2

Реакция экосистемы на это событие для её усечённого варианта с двумя уровнями иерархии x^1 и x^2 ($x^3 = 0$ – вторичные консументы отсутствуют) показана на рис. 15 для биомассы x^2 . Исходное соотношение биомасс видов x_j^2 , которое до момента $t < 350$ характеризовалось доминирующим положением вида x_3^2 и угнетённым состоянием x_2^2 , изменилось: вызванное внешними причинами снижение биомассы x_3^2 третьего вида компенсируется ростом биомасс x_1^2 , x_2^2 первого и второго видов этого уровня системы. После колебательного переходного процесса длительностью около 150 единиц времени скорость роста общей биомассы системы восстанавливается. Характерно, что в новых условиях ранее угнетённый второй вид x_2^2 этого уровня системы стал занимать значимую долю в переработке биомассы, т.е. в результате критического события в функционирование экосистемы включились все имеющиеся три вида первичных консументов.

Исследование роли иерархического строения системы при наступлении в них критических событий проводилось на полном варианте экосистемы (введён третий уровень: консументы–хищники) при сохранении неизменными всех параметров элементов двухуровневой системы. Реакция трёхуровневой системы на снижение x_3^2 показана на рис. 16 и существенно отличается от предшествующего случая, показанного на рис. 15.

В первую очередь следует отметить, что в исходном, стационарном состоянии трёхуровневой экосистемы ($t < 300$) уже не наблюдается существенного различия биомасс x_3^2 и x_1^2 третьего и первого видов первичных консументов. Это свидетельствует о том, что несмотря на относительно небольшую биомассу x^3 третьего уровня (примерно 10% от биомассы x^2 второго уровня), существование вторичных консументов смягчило конкуренцию на предшествующем иерархическом уровне, и все три вида первичных консументов стали сопоставимыми по своей значимости в круговороте вещества.

Стабилизирующее влияние третьего уровня иерархии проявилось также при наступлении критического события ($t = 350$): переходный процесс приобрёл монотонный характер и завершился значительно быстрее – за 50 единиц времени (см. рис. 16).

Для более полной оценки эволюционной роли видового многообразия и иерархического строения экосистем рассмотрим фазовые траектории поведения системы в координатах x_1^2 и x_2^2 , т.е. в плоскости биомасс первого и второго видов первичных консументов, рис. 17.

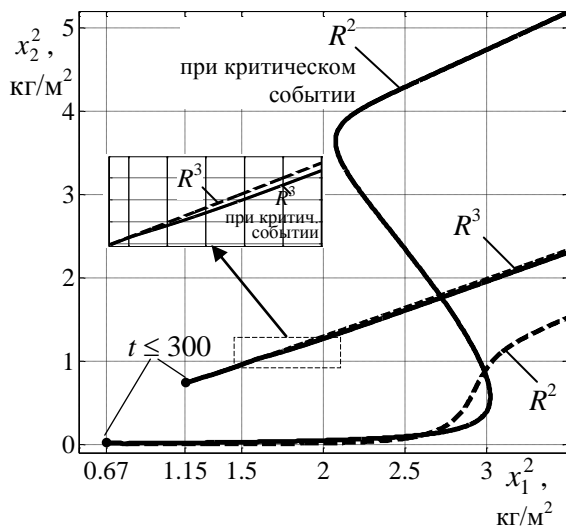


Рис. 17. Траектории развития экосистемы с двумя (R^2) и тремя (R^3) уровнями иерархии

Анализ рис. 17 показывает, что во всех режимах функционирования система стремится выйти на некоторую сбалансированную (линейную) траекторию своего развития, характеризующуюся постоянством таких пропорций между биомассами видов, которые обеспечивают максимальное усвоение поступающих в экосистему ресурсов (во всех рассмотренных выше

режимах система после завершения переходных процессов полностью поглощает поступающий поток внешней энергии (вещества) со скоростью, равной скорости поступления).

Штриховым линиям на рис. 17 соответствуют траектории роста биомасс x_1^2 и x_2^2 после начала поступления внешней энергии ($t > 300$) без последующего введения в систему описанного выше критического события. В двухуровневой системе траектория R^2 содержит достаточно протяжённый участок переходного процесса, в течение которого соотношение биомасс x_1^2 и x_2^2 приобретает новое установившееся значение (угол наклона R^2); в варианте с тремя уровнями иерархии система практически без переходного процесса начинает двигаться по траектории R^3 сбалансированного развития.

Сплошными линиями на рис. 17 показана реакция экосистемы на критическое событие (снижение скорости поглощения третьим видом первичных консументов биомассы третьего вида продуцентов). Переход траектории R^3 двухуровневой системы на новый участок сбалансированного роста носит ярко выраженный колебательный характер с существенным изменением абсолютных значений биомасс x_1^2 и x_2^2 . В экосистеме с тремя уровнями иерархии траектория R^3 изменилась незначительно, т.е. в многоуровневой экосистеме эффективность удержания системы на траектории сбалансированного роста существенно выше.

Таким образом, многообразие видов и развитое иерархическое строение экологической системы обеспечивают благоприятные условия для её развития – эволюции: многообразие видов за счёт перераспределения их биомасс выводит систему на траекторию сбалансированного роста, а наличие иерархической структуры в результате влияния межуровневых обратных связей значительно ускоряет этот процесс и делает его монотонным.

Проведённый выше анализ влияния видового многообразия и многоуровневости структуры на эволюционные процессы в экосистеме опирался на прямое воспроизведение – имитацию функционирования основных элементов системы. Вместе с тем,

представляет интерес количественная оценка результата этого влияния на более общем – информационном уровне: в какой мере многообразие стратегий поведения экосистемы в статистическом смысле способно противодействовать многообразию условий внешней среды? Для ответа на поставленный вопрос воспользуемся игровой моделью этого противодействия.

Рассмотрим игру G экосистемы S с природой (внешней средой) N , обладающих набором стратегий соответственно $S = \{S_1, S_2, \dots, S_n\}$ и $N = \{N_1, N_2, \dots, N_k\}$. Под эволюционно устойчивой стратегией S_i экосистемы будем понимать конкретный вариант её состава и структурной организации, обеспечивающий развитие экосистемы в текущих условиях внешней среды. В качестве числовой меры результата игры будем использовать степень приспособленности $h_{ij} = \{h_{ij} \in \mathbf{R}, 0 \leq h_{ij} \leq 1\}$; $i = \overline{1, n}$; $j = \overline{1, k}$, экосистемы в случае выбора ею стратегии S_i в условиях N_j среды.

Примем, что платёжная матрица $H = \{h_{ij}\}$ содержит значения h_{ij} , равномерно и непрерывно распределённые на отрезке $[0; 1]$ в каждом отдельно взятом элементе матрицы H . Указанный способ формирования платёжной матрицы в содержательном плане отражает способность экосистемы к параметрической адаптации, в результате которой каждая стратегия S_i может обеспечить примерно одинаковую степень приспособленности h_{ij} не только для одного значения j , но для некоторого множества $J \subset N$ стратегий внешней среды. Это обстоятельство, как будет показано ниже, позволяет экосистеме противодействовать неограниченному многообразию N условий среды с помощью ограниченного набора стратегий S структурного уровня. Отметим также, что нормирование h_{ij} до единицы позволяет интерпретировать степень приспособленности как вероятность сохранения экосистемой устойчивого состояния на рассмотренном многообразии N изменяющихся условий внешней среды.

Особенностью игры G является пассивность природы N , в отличие от активности экосистемы S , перестраивающей, как это было показано выше, внутренние потоки вещества и энергии так, чтобы сохранить устойчивое состояние. Поскольку предпо-

лагается, что варианты таких перестроек образуют множество стратегий S , то механизм естественного отбора приведёт к доминированию стратегии с максимальным значением $h_j^* = \max_{i=1, \dots, n} h_{ij}$ для каждого текущего состояния N_j внешней среды [10].

Очевидно, что если у экосистемы в наличии будет всего одна стратегия $n = 1$, то на всём многообразии N стратегий природы средняя степень приспособленности (цена v игры G) при достаточно больших $|N|$ составит $v(1) = 0,5$. При наличии у экосистемы нескольких ($n > 1$) альтернативных стратегий цена v игры определится выражением

$$(41) \quad v(n) = \frac{\sum_{j=1}^k h_j^*}{k}, \quad h_j^* = \max_i h_{ij}, \quad i = \overline{1, n}.$$

Результат вычислений $v(n)$ для $n = 1, \dots, 1000$ и $k = 1 \cdot 10^6$ представлен на рис. 18.

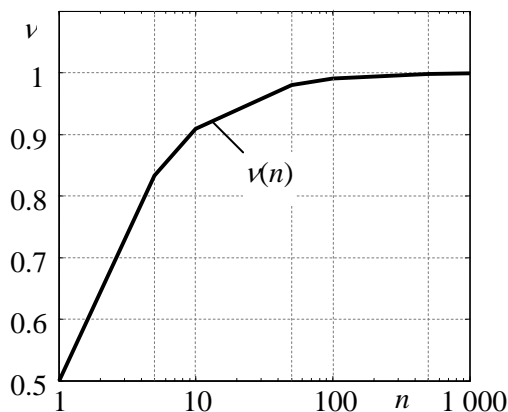


Рис. 18. Изменение вероятности v устойчивой эволюции с ростом числа n альтернативных стратегий экосистемы

Зависимость $v(n)$, показанная на рис. 18, позволяет сделать следующие выводы:

– с ростом числа альтернативных стратегий поведения экосистемы степень её приспособленности к изменяющимся условиям внешней среды возрастает и асимптотически приближается к некоторому уровню, обеспечивающему её устойчивое развитие;

– для обеспечения устойчивой эволюции в условиях неограниченного многообразия внешней среды экосистеме достаточно обладать ограниченным набором стратегий поведения структурного уровня, каждая из которых, согласно принципу неполной специализации [32], способна к параметрической адаптации в некотором диапазоне воздействий среды. Отметим, что выводы о достаточности ограниченного многообразия для устойчивости экосистемы можно найти в работах [2, 3].

Полученный результат был вполне предсказуем теоретически в силу равновероятного заполнения элементов платёжной матрицы H и условия существования доминируемых стратегий у активной экосистемы. Тем не менее использование в модели игры экосистемы с природой только наиболее существенных информационных свойств её участников – многообразия стратегий и степени приспособленности, позволяет выделить в явном виде и количественно оценить роль многоальтернативности в эволюционном поведении экосистем. Весьма уместно упомянуть здесь также о подтверждении известного информационного принципа необходимого разнообразия У. Эшби [57].

5. Эволюционная концепция многоальтернативности и задачи управления антропогенными системами

Проведенное выше исследование даёт основание утверждать, что способность биологических систем к приспособительной эволюции опирается на ряд достаточно простых принципов, составляющих концепцию многоальтернативности:

– принцип многоуровневости строения и функционирования системы;

– принцип многообразия и разделения функций, активируемых в зависимости от текущих условий среды;

– принцип дискретности (модульности), обеспечивающий реализацию перечисленных выше принципов в результате комбинаций ограниченного дискретного набора элементарных модулей.

Остановимся на взаимосвязи этих принципов с задачами управления антропогенными (созданными человеком) системами и покажем их конструктивный характер, т.е. возможность использования для решения соответствующих прикладных задач.

Основной практической проблемой управления большими системами является проблема их высокой размерности – «неупрощаемой сложности», приводящей к динамической неопределённости их моделей и снижению показателей качества регулирования [52]. Чем сложнее искусственная система, тем труднее обеспечить её надёжную работоспособность, особенно в граничных, критических режимах функционирования.

Вместе с тем, биологические системы не испытывают никаких трудностей с ростом сложности своей структуры и числа внутренних взаимосвязей [24, 34]. Более того, как было показано выше, именно многообразии строения и способов функционирования обеспечивает биосистемам устойчивое существование и развитие в изменяющихся условиях внешней среды.

Это обстоятельство указывает на целесообразность проведения биологической аналогии и целенаправленного воспроизведения принципов многоальтернативности в задачах управления сложными системами.

В частности, известным приёмом построения систем управления является их декомпозиция на несколько иерархических уровней, каждый из которых «вступает в работу» по мере исчерпания материальных или временных ресурсов нижестоящего уровня. Развитие этого приёма, начавшееся с простейших систем подчинённого регулирования, привело в настоящее время к созданию отдельных направлений в теории управления – гомеостатическому регулированию и магистральному управлению [1, 35, 53].

Принцип разделения функций нашёл своё применение в так называемых системах с переменной структурой, в которых

дискретное изменение закона управления осуществляется в соответствии с текущим состоянием системы в пространстве её координат. В системах обеспечения безопасности технологических процессов используется так называемый диверсный (от англ. diversity – разнообразие) подход к их построению, предусматривающий резервирование функций системы на физически различной элементной базе: программные средства резервируются аппаратными, энергозависимые – энергонезависимыми (пассивными) и т.п. [40].

Принцип модульности наиболее востребован в высоконадежных вычислительных и энергетических системах, обеспечивая иерархическую структуру и многообразие режимов функционирования этих систем [25].

С подробными примерами использования принципов многоальтернативности в технических системах управления можно познакомиться в уже упоминавшейся работе [39].

Таким образом, в широком смысле эволюционные принципы многоальтернативности являются составной частью общих кибернетических принципов построения открытых систем управления сложными объектами как естественного, так и искусственного происхождения [39, 40]. Можно констатировать, что кибернетическая сущность биологической эволюции в целом как процесса обработки информации привела в настоящее время к формированию нового научного направления – «эволюционная кибернетика» [46, 47, 50].

6. Заключение

Изложенный анализ кибернетических моделей эволюционного поведения самоорганизующихся систем позволяет установить, что в основе этого поведения лежит общая системообразующая стратегия – стратегия многоальтернативности.

На добиологическом этапе возникновения жизни эта стратегия выражается в перекрёстных мутационных вариациях селективной ценности макромолекул, исключающих полное доминирование только одного типа из них, т.е. обеспечивающих совместное, но в то же время конкурирующее существование

разнообразных типов макромолекул. В гомогенной среде устойчивая репликация этих молекул становится невозможной.

Для биосистем на уровне популяций определяющим условием эволюции является увеличение генетического многообразия популяции со скоростью, превышающей скорость потока критических событий внешней среды. Нарушение этого условия, обусловленное, например, падением численности популяции ниже некоторого критического значения, приводит к её вырождению.

Эволюционные процессы в крупных экосистемах принципиально осуществимы только в многоуровневой иерархической структуре системы при наличии многообразия видов с дифференцированными приспособительными функциями на каждом её уровне. Результаты моделирования показали, что с увеличением структурного и видового многообразия устойчивость экосистемы к кризисным ситуациям и её способность к сохранению сбалансированной траектории своего развития также возрастает. Эта уникальное эволюционное свойство «неусложняемой простоты» сложных биологических систем является следствием многоальтернативной стратегии их строения и функционирования.

Рассматривая результаты проведённого исследования с прикладных – кибернетических позиций, можно сформулировать ряд достаточно простых по отдельности конструктивных принципов, составляющих содержание эволюционной концепции многоальтернативности:

- принцип многоуровневости и иерархии строения и функционирования, обеспечивающий возникновение в сложных системах свойства гомеостаза;

- принцип многообразия и разделения функций подсистем, позволяющий реализовать адаптивную передачу управления между ними и высокую гибкость системы в целом в условиях её открытого взаимодействия с изменяющейся внешней средой;

- принцип модульности, порождающий комбинаторное многообразие вариантов построения и возможных режимов работы системы при ограниченном количестве отличающихся типовых модулей.

Реализация этих принципов на практике снимает остающееся в настоящее время актуальным «проклятие размерности» с управляющих систем высокой сложности, создаваемых человеком.

Литература

1. АБАКУМОВ А.И. *Модели Неймана-Гейла*. – Владивосток: ДВГУ, 2004. – 44 с.
2. АЛЕКСЕЕВ А.С., ДМИТРИЕВ В.Ю., ПОНОМАРЕНКО А.Г. *Эволюция таксономического разнообразия*. – М.: Геос, 2001. – 126 с.
3. АЛЕЩЕНКО Г.М., БУКВАРЕВА Е.Н. *Двухуровневая иерархическая модель оптимизации биологического разнообразия* // Известия РАН. Серия биологическая. – 2010. – №1. – С. 5–15.
4. АЛТУХОВ Ю.П. *Генетические процессы в популяциях*. – М.: Академкнига, 2003. – 431 с.
5. АНДРЕЕВА С.И., АНДРЕЕВ Н.И. *Эволюционные преобразования двустворчатых моллюсков Аральского моря в условиях экологического кризиса*. – Омск: Изд-во Омского государственного педагогического университета, 2003. – 382 с.
6. БАКАНОВ А.И. *О некоторых методологических вопросах применения системного подхода для изучения структур водных экосистем* / В кн. Количественные методы экологии и гидробиологии (сб. науч. тр.). – Тольятти: СамНЦ РАН, 2005. – С. 10–21.
7. БЕРНИ Д. *Большая иллюстрированная энциклопедия живой природы*. – М.: Махаон, 2006. – 320 с.
8. БИГОН М., ХАРПЕР ДЖ., ТАУНСЕНД К. *Экология. Особи, популяции и сообщества*: В 2-х т. – М.: Мир, 1989. – Т. 1. – 667 с.
9. БИГОН М., ХАРПЕР ДЖ., ТАУНСЕНД К. *Экология. Особи, популяции и сообщества*: В 2-х т. – М.: Мир, 1989. – Т. 2. – 477 с.

10. БОГДАНОВ А.В., ВАСИН А.А. *Модели адаптивно-подражательного поведения: I. Связь с равновесиями Нэша и решениями по доминированию* // Известия Российской академии наук. Теория и системы управления. – 2002. – №1. – С. 102–111.
11. ВАВИЛОВ Н.И. *Закон гомологических рядов в наследственной изменчивости*. – Л.: Наука, 1987. – 256 с.
12. ВЕНТЦЕЛЬ Е.С., ОВЧАРОВ Л.А. *Теория случайных процессов и её инженерные приложения*. – М.: Высш. шк., 2000. – 383 с.
13. ВЕРНАДСКИЙ В.И. *Биосфера и ноосфера*. – М.: Айрис-пресс, 2012. – 576 с.
14. ВЕРНАДСКИЙ В.И. *Научная мысль как планетное явление*. – М.: Наука, 1991. – 271 с.
15. ВОЛЬТЕРРА В. *Математическая теория борьбы за существование*. – М.: Наука, 1976. – 286 с.
16. ГАЛИМОВ Э.М. *Феномен жизни: между равновесием и нелинейностью. Происхождение и принципы эволюции*. – М.: Едиториал УРСС, 2006. – 256 с.
17. ГРИНИН Л.Е., МАРКОВ А.В., КОРОТАЕВ А.В. *Макроэволюция в живой природе и обществе*. – М.: Либроком, 2009. – 248 с.
18. ДАРВИН Ч. *Происхождение видов путём естественного отбора, или Сохранение благоприятных рас в борьбе за жизнь*. – С-Пб: Наука, 2001. – 568 с.
19. ДОКИНЗ Р. *Эгоистичный ген*. – М.: Изд. АСТ, 2013. – 512 с.
20. ЕСЬКОВ К.Ю. *Удивительная палеонтология: история Земли и жизни на ней*. – М.: ЭНАС, 2014. – 312 с.
21. ЗАСЛАВСКИЙ Б.Г., ПОЛУЭКТОВ Р.А. *Управление экологическими системами*. – М.: Наука, 1988. – 296 с.
22. ИОРДАНСКИЙ Н.Н. *Эволюция жизни*. – М.: Академия, 2001. – 425 с.
23. КИМУРА М. *Молекулярная эволюция: теория нейтральности*. – М.: Мир, 1985. – 394 с.

24. КОЛЕСНИКОВ А.А. *Синергетическая теория управления: концепция, методы, тенденции развития* // Известия Южного федерального университета. Технические науки. – 2001. – Т. 23, №5. – С. 7–27.
25. КОЛМОГОРОВ А.Н. *Три подхода к определению понятия «количество информации»* // Проблемы передачи информации. – 1965. – Т. 1, вып. 1. – С. 3–11.
26. КОЛМОГОРОВ А.Н. *Качественное изучение математических моделей популяций* / В кн. Проблемы кибернетики, Вып. 25. – М.: Наука, 1972. – С. 100–106.
27. КОРОСТЫШЕВСКИЙ М.А., ШТАБНОЙ М.Р., РАТ-НЕР В.А. *О некоторых стохастических закономерностях эволюционного процесса* / В кн. Вопросы математической генетики. – Новосибирск: Ин-т цитологии и генетики СО АН СССР, 1974. – С. 5-32.
28. КРАСИЛОВ В.А. *Нерешённые вопросы теории эволюции.* – Владивосток: ДВНЦ АН СССР, 1986. – 138 с.
29. КЭРОЛЛ Ш.Б. *Закон «джунглей». В поисках формулы жизни.* – С-Пб.: Питер, 2017. – 304 с.
30. ЛАМАРК Ж.Б. *Философия зоологии.* Т. 1. – М.: ОГИЗ-Биомедгиз, 1935. – 330 с.
31. ЛАМАРК Ж.Б. *Философия зоологии.* Т. 2. – М.: ОГИЗ-Биомедгиз, 1937. – 483 с
32. ЛЕВЧЕНКО В.Ф. *Биосфера: этапы жизни.* – С-Пб.: Своё издательство, 2012. – 264 с.
33. МЕНДЕЛЬ Г. *Опыты над растительными гибридами.* – М.: Наука, 1965. – 159 с.
34. НЕЙМАН ДЖ. ФОН. *Теория самовоспроизводящихся автоматов.* – М.: URSS, 2010. – 384 с.
35. НОВИКОВ Д.А. *Теория управления организационными системами.* – М.: МПСИ, 2005. – 584 с.
36. ОДУМ Ю.П. *Экология.* Т. 2. – М.: Мир, 1986. – 376 с.
37. ПАВЛОВ Д.С., БУКВАРЕВА Е.Н. *Биоразнообразие, экосистемные функции и жизнеобеспечение человечества* // Вестник РАН. – 2007. – Т. 77, №11. – С. 974–986.

38. ПОДВАЛЬНЫЙ С.Л. *Многоальтернативные системы: обзор и классификация* // Системы управления и информационные технологии. – 2012. – №2. – С. 4–13.
39. ПОДВАЛЬНЫЙ С.Л., ВАСИЛЬЕВ Е.М. *Многоальтернативное управление открытыми системами: концепция, состояние и перспективы* // Управление большими системами. – 2014. – Вып. 48. – С. 6–58.
40. ПОДВАЛЬНЫЙ С.Л., ВАСИЛЬЕВ Е.М. *Концепция многоальтернативности в живых и неживых структурах* // Онтология проектирования. – 2016. – Т. 6, №3(21). – С. 355–367.
41. ПОНОМАРЕНКО А.Г., ДМИТРИЕВ В.Ю. *Эволюция разнообразия и устойчивость экосистем* / В кн. Проблемы доантропогенной эволюции биосферы. – М: Наука, 1993. – С. 54–59.
42. РАТНЕР В.А. *Блочно-модульный принцип организации и эволюции молекулярно-генетических систем управления* // Генетика. – 1992. – Т.28, №2. – С. 5–23.
43. РАТНЕР В.А. *Математическая популяционная генетика*. – Новосибирск: Наука, 1977. – 128 с.
44. РАТНЕР В.А. *Молекулярно-генетическая система управления* // Природа. – 2001. – №3. – С. 16–22.
45. РАТНЕР В.А., ШАМИН В.В. *Сайзеры: моделирование фундаментальных особенностей молекулярно-биологических организаций. Соответствие общих свойств и конструктивных особенностей коллективов макромолекул* // Журнал общей биологии. – 1983. – Т. 44, №1. – С. 51–61.
46. РЕДЬКО В.Г. *Эволюция, нейронные сети, интеллект. Модели и концепции эволюционной кибернетики*. – М.: Либроком, 2013. – 224 с.
47. РЕДЬКО В.Г. *Эволюционная кибернетика*. – М.: Наука, 2003. – 155 с.
48. СВИРЕЖЕВ Ю.М., ЛОГОФЕТ Д.О. *Устойчивость биологических сообществ*. – М.: Наука, 1978. – 352 с.

49. ТИМОФЕЕВ-РЕСОВСКИЙ Н.В., ВОРОНЦОВ Н.Н., ЯБЛОКОВ А.В. *Краткий очерк теории эволюции*. – М.: Наука, 1977. – 303 с.
50. ТУРЧИН В.Ф. *Феномен науки. Кибернетический подход к эволюции*. – М.: ЭТС, 2000. – 368 с.
51. ФАЙСТЕЛЬ Р., РОМАНОВСКИЙ Ю.М., ВАСИЛЬЕВ В.А. *Эволюция гиперциклов Эйгена, протекающих в коацерватах* // Биофизика. – 1980. – Т. 25, №5. – С. 882–887.
52. ФИЛИМОНОВ Н.Б. *Методологический кризис «всепобеждающей математизации» современной теории управления* // Мехатроника, автоматизация, управление. – 2016. – Т. 17, №5. – С. 291–300.
53. ЧЕРЕМНЫХ Ю.Н. *Качественное исследование оптимальных траекторий динамических моделей экономики. Вопросы магистральной теории*. – М.: Изд-во Моск. университета, 1975. – 183 с.
54. ЧЕТВЕРИКОВ С.С. *Проблемы общей биологии и генетики*. – Новосибирск: Наука, 1983. – 273 с.
55. ЭЙГЕН М. *Самоорганизация материи и эволюция биологических макромолекул*. – М.: Мир, 1973. – 224 с.
56. ЭЛТОН Ч. *Экология нашествий животных и растений*. – М.: Изд. иностр. лит., 1960. – 231 с.
57. ЭШБИ У.Р. *Введение в кибернетику*. – М.: КомКнига, 2005. – 432 с.
58. ЯБЛОКОВ А.В. *Популяционная биология*. – М.: Высш. шк., 1987. – 303 с.
59. ANDERSON P.W. *Suggested model for prebiotic evolution: the use of chaos* // Proc. of the National Academy of Sciences of the USA. – 1983. – Vol. 80, No. 11. – P. 3386–3390.
60. *Convention on Biological Diversity // Treaty Series*. New York: United Nations. – 2001. – Vol. 1760. – I. Nos. 30690. – P. 79–307.
61. COPE E.D. *The primary factors of organic evolution*. – Sydney: Wentworth Press, 2016. – 588 p.
62. CRICK F.H. *The origin of the genetic code* // J. of Molecular Biology. – 1968. – Vol. 38, No. 3. – P. 367–379.

63. DYSON F.J. *A model for the origin of life* // J. of Molecular Evolution. – 1982. – Vol. 18, No. 5. – P. 344–350.
64. EIGEN M., GARDINER W., SCHUSTER P., WINKLER-OSWATICH R. *The origin of genetic information* // Scientific American. – 1981. – Vol. 244, No. 4. – P. 88–118.
65. FISHER R.A. *The genetical theory of natural selection*. – Oxford: Oxford University Press, 1999. – 318 p.
66. HOOPER D.U. CHAPIN F.S. III, EWEL J.J. et al. *Effects of biodiversity on ecosystem functioning: a consensus of current knowledge and needs for future research* // Ecol. Monographs. – 2005. – Vol. 75. – P. 3–35.
67. LOREAU M., NAEEM S., INCHAUSTI P. et al. *Biodiversity and ecosystem functioning: current knowledge and future challenges* // Science. – 2001. – Vol. 294. – P. 804–808.
68. ROKHSAR D.S., ANDERSON P.W., STEIN D.L. *Selforganization in prebiological system: Simulations of model for the origin of genetic information* // J. of Molecular Evolution. – 1986. – Vol. 23, No. 2. – P. 119–126.
69. ROONEY N., MCCANN K., GELLNER G., MOORE J.C. *Structural asymmetry and the stability of diverse food webs* // Nature. – 2006. – Vol. 442. – P. 265–269.
70. TILMAN D. *Causes, consequences and ethics of biodiversity* // Nature. – 2000. – Vol. 405. – P. 208–211.
71. WHITE D.H. *A theory for the origin of a self-replicating chemical system / Natural selection of the autogen from short random oligomers* // J. of Molecular Evolution. – 1980. – Vol. 16, No. 2. – P. 121–147.

MULTI-ALTERNATIVITY: THE EVOLUTIONARY STRATEGY OF BIOLOGICAL SYSTEMS

Semen Podvalny, Voronezh State Technical University, Voronezh, Doctor of Science, professor (spodvalny@yandex.ru).

Eugeny Vasiljev, Voronezh State Technical University, Voronezh, Cand.Sc., assistant professor (vgtu-aits@yandex.ru).

Abstract: The article describes the results of mathematical modeling and quantitative analysis of the biological properties of diversity. The aim of the work is a methodological generalization of this property of living nature in the form of a cybernetic concept of multi-alternativeity as an evolutionary strategy for the functioning of complex systems. The study of the pre-biological stage of life origin, conducted on quasi-species and sisers models, showed that for the stable existence of macromolecules capable of self-replication, they must have cross mutational flows generating a variety of competing types of molecules. Analysis of the process of molecular genetic evolution at the level of populations was carried out on the model of the Markov random process with discrete states and allowed to quantitatively substantiate the condition of intra-species evolution: the growth rate of the genetic diversity of the population must exceed the rate of occurrence of environmental conditions that threaten its existence. To model evolutionary processes in a complex multi-level ecosystem with a closed trophic cycle, differential equations of mass balance were used for each trophic level of the hierarchy, taking into account the exchange of energy with the external environment as a mass of an equivalent substance. The results of ecosystem modeling showed that when a critical event occurs in it, the diversity of species due to the redistribution of their biomass again brings the system to the trajectory of balanced growth, and the presence of a hierarchical structure as a result of the influence of interlayer feedbacks significantly speeds up this process and makes it monotonous. The results of the work formulate a number of substantive principles that make up the concept of multi-alternative: the principles of the many-leveledness, separation of functions and discreteness. The constructive nature of these principles opens the possibility of practical implementation of the multi-alternative strategy in the construction of complex control systems for various purposes.

Keywords: mechanisms of evolution, evolutionary cybernetics, concept of multi-alternative.

УДК 007:575.85

ББК 22.18 28.02

DOI: <https://doi.org/10.25728/ubs.2019.77.6>

*Статья представлена к публикации
членом редакционной коллегии Я.И. Квинто.*

Поступила в редакцию 15.09.2017.

Опубликована 31.01.2019.

МОДЕЛИ РАЗРАБОТКИ И ОСВОЕНИЯ ТЕХНОЛОГИИ КОМПЛЕКСНОЙ ДЕЯТЕЛЬНОСТИ¹

Белов М. В.²

(Компания ИБС, Москва)

Новиков Д. А.³

(ФГБУН Институт проблем управления
им. В.А. Трапезникова РАН, Москва)

Представленные математические модели и задачи основываются на результатах и продолжают анализ проблем управления организационно-техническими системами и их комплексной деятельностью, выполненный в предыдущих работах авторов. Проблема разработки и/или освоения технологии комплексной деятельности формализована в виде математической модели, являющейся обобщением вероятностных моделей научения. Исследованы свойства процесса разработки и/или освоения технологии (научения), показана сходимости процесса к состоянию полного освоения технологии, получены аналитические выражения характеристик моделей – среднего времени достижения заданного уровня освоения. Предложены модели научения, описывающие комплексирование элементов технологии – конъюнктивное, дизъюнктивное и параллельное освоение технологии. Разработаны и изучены модели научения в процессе работы и группового научения, в том числе модели «научения научению» - когда интенсивность процесса научения зависит от достигнутого уровня научения. Для всех моделей комплексирования получены аналитические выражения для уровней научения. Исследован асимптотический случай моделей при переходе к непрерывному времени. Показано, частными случаями предложенной модели являются модели экспоненциальных, гиперболических и логистических кривых научения, которые широко распространены в теории научения, теории испытаний систем, тестирования программного обеспечения и смежных отраслях знаний.

Ключевые слова: технология, комплексная деятельность, кривая научения.

¹ Работа выполнена при частичной финансовой поддержке Российского научного фонда (грант № 16-19-10609).

² Михаил Валентинович Белов, к.т.н. (mbelov59@mail.ru).

³ Дмитрий Александрович Новиков, д.т.н., чл.-корр. РАН, профессор (novikov@ipu.ru).

1. Введение

Рассматриваемые в настоящей статье математические модели и задачи основываются на результатах анализа проблем управления организационно-техническими системами (ОТС) и их комплексной деятельностью (КД), проведённого в статьях [4, 5] и монографии [6].

В монографии [6] исследуется *комплексная деятельность* (КД), определяемая как *деятельность* (целенаправленная активность человека), обладающая нетривиальной внутренней структурой, с множественными и/или изменяющимися субъектом, технологией, ролью предмета деятельности в его целевом контексте. Методология КД представлена в [6] в виде интегрированной системы формальных утверждений и моделей, описывающих общесистемные свойства КД, прежде всего – её структуру, неопределённость и жизненный цикл (ЖЦ).

Работа [4] посвящена определению средств управления организационно-техническими системами и их комплексной деятельностью, основным среди которых является управление технологией комплексной деятельности (см. определение ниже).

Проблема управления технологией КД ОТС рассмотрена и формализована в [5], для чего в указанной работе проанализированы наиболее важные особенности КД ОТС, предложены формальные модели и сформулированы математические задачи управления технологией КД.

Настоящая работа имеет следующую структуру. В разделе 2 на основании результатов [5] проанализированы общесистемные проблемы разработки, освоения, оптимизации и модернизации технологии КД и сформулированы в общем виде соответствующие математические задачи. Раздел 3 посвящён обзору известных моделей и методов из смежных областей знаний. В разделе 4 проанализированы свойства процесса разработки/освоения технологии. В разделе 5 приведены аппроксимации кривой научения при различных распределениях вероятности состояний внешней среды, в разделе 6 – оценки среднего времени достижения требуемого уровня научения. Раздел 7 содержит описание моделей комплексирования компонентов технологии. Заключение включает краткое перечисление основных

результатов и перспективных направлений дальнейших исследований.

2. Концептуализация проблемы управления технологией КД

Технология КД определяется в [4–6] как система условий, критериев, форм, методов и средств последовательного достижения поставленной цели. *Управлением технологией* в [5] условно названа деятельность по созданию компонентов технологии в виде соответствующих информационных моделей, их комплексированию и поддержанию в состоянии, адекватном условиям внешней среды в ходе реализации жизненного цикла КД; для описания этого процесса предложена интегрированная модель в BPMN-нотации [25] (см. рис. 1). Содержательно технология представляет собой сценарии деятельности субъекта в зависимости от внешних условий (состояний *внешней среды*).

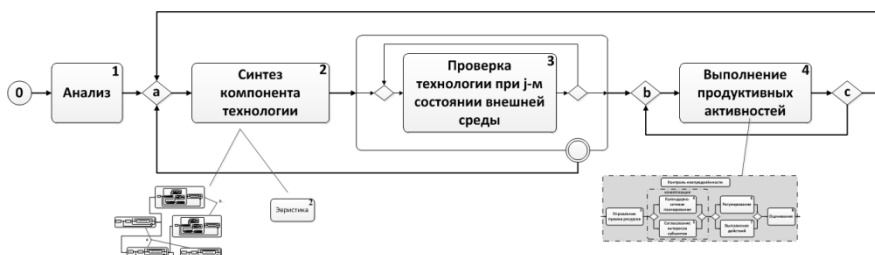


Рис. 1. Интегрированная модель управления технологией КД [4]

Уточним определение управления технологией КД, для чего расширим и проанализируем семантику интегрированной модели (рис. 1), которая представляет различные виды КД:

- *разработку* новой технологии КД или её *модернизацию* – блоки 1, 2 и цикл, включающий блок 3;
- последующее *применение* технологии КД – блок 4;
- выявление необходимости модернизации технологии - переход от *c* к *a*;
- тестирование или *освоение* технологии КД - объективное (когда технология новая) или субъективное (когда технология

существует, а субъект её осваивает/обучается) – цикл, включающий блок 3.

Несмотря на разнообразие перечисленных видов деятельности, их объединяет единое основание: их предметом является технология КД, все они направлены на изменение технологии КД, на изменение состояний технологии или ее отношений с субъектом КД. Поэтому деятельность всех этих видов является управлением технологией КД (*управление* согласно [6] - комплексная деятельность, обеспечивающая воздействие субъекта управления на управляемую систему (объект управления), призванное обеспечить ее (его) поведение, приводящее к достижению целей субъекта управления).

В общем случае оптимизация процесса управления технологией может выполняться за счёт:

- выбора разбиения состояний внешней среды на непересекающиеся подмножества, внутри которых все состояния внешней среды считаются эквивалентными;
- выбора последовательности «перебора» состояний внешней среды при проверках;
- перераспределения проверок между уровнем создания компонента технологии (блок «2» на рис. 1) и уровнем интеграции (блок «3»);
- разделения ограниченных ресурсов между отдельными операциями управления технологией;
- определения допустимого (с точки зрения рисков) объёма привлекаемого ресурса для обеспечения всех компонентов технологии.

Каждая из операций-«эвристик» создания технологии (блоки «1» и «2» на рис. 1) и проверки технологий (блок «3») являются специфическими – зависят от предметной области - и на общесистемном уровне не могут являться объектами управления и оптимизации. Поэтому в интересах рассматриваемой задачи можно считать, что они характеризуются известными временами выполнения и расходуемыми ресурсами (в частном случае – случайными величинами с известными законами распределения). Также считаем, что эти операции выполняются независимо, а связи между ними реализуются через заданные

причинно-следственные зависимости между процессами создания отдельных элементов технологии; фиксированными считаем перечни элементов КД, логические и причинно-следственные модели [6].

Выполнение различных видов КД, описываемых интегрированной моделью (рис. 1), будем представлять, следуя [5], как процесс с дискретным временем, когда на каждом шаге выполняется один элемент КД (один из блоков «1-2-3-4» модели рис. 1), при этом состояние внешней среды принимает одно и только одно значение из конечного множества *возможных состояний внешней среды* (МВСВС).

Будем считать, что если состояние внешней среды на каком-то шаге впервые принимает некоторое значение, то возникает событие неопределённости, требующее затрат на создание или адаптацию технологии применительно к этим условиям. Когда состояние внешней среды повторно принимает это значение на одном из более поздних шагов, затрат на создание технологии не требуется. Неопределённость внешней среды предполагает, что субъект не может влиять на выбор её очередного состояния; для описания неопределённости используем вероятностные инструменты.

Предположим, что МВСВС состоит из K различных состояний, одно и только одно из которых внешняя среда принимает на каждом шаге дискретного времени, независимо от принятых на предыдущих шагах значений. Обозначим через p_k вероятность того, что состояние внешней среды примет k -е значение (очевидно, $\sum_{k=1}^K p_k = 1$).

Состояния процесса реализации различных фаз жизненного цикла технологии в момент времени $t \in \{0, 1, \dots\}$ будем описывать K -мерным вектором-строкой $x_t = (x_{1t}, x_{2t}, \dots, x_{kt}, \dots, x_{Kt})$, каждый из элементов x_{kt} которого принимает значения 0 или 1 соответственно, если внешняя среда ещё не принимала или хотя бы один раз принимала k -е значение. В рамках рассматриваемой модели k -й элемент вектора x_t может переходить из состояния 0 в 1, но не наоборот.

В рамках рассматриваемой модели процесс реализации различных фаз жизненного цикла технологии описывается тем, сколько и какие значения из МВСВС принимало состояние внешней среды, а какие ещё нет.

Уровень зрелости, готовности технологии к использованию в момент времени t будем характеризовать величиной

$$L_t = \sum_{k=1}^K x_{kt} p_k, \quad 0 \leq L_t \leq 1 \text{ (или, наоборот, } 1 - L_t\text{)}. \text{ Величина } L_t \text{ соот-}$$

ветствует доле состояний внешней среды, для которых технология проверена или адаптирована в течение предшествующих t шагов, или же вероятности того, что на следующем шаге ($t + 1$) состояние внешней среды примет одно из значений, которые уже принимало ранее. Для обозначения L_t будем использовать термин *уровень разработанности* технологии или, следуя традициям моделирования процессов научения [15], *уровень научения*, а последовательность значений уровня научения называть *кривой научения*. Следует отметить, что термин «кривая научения» в схожих смыслах чрезвычайно распространен в современной науке, начиная с «кривых забывания» Г. Эббингауза [28], психологии всего XX века (см., например, классические статьи [31, 38, 39, 40] и монографии [2, 9]), моделей Т. Райта [42] и его последователей [27, 29] (описывающих снижение временных затрат на производство единицы продукции с ростом осуществлённых объёмов производства) до моделей обучения искусственных нейронных сетей.

Процесс разработки/проверки технологии в данной постановке может быть рассмотрен с другой точки зрения: он представляет собой последовательное наблюдение серий уже известных состояний внешней среды, прерывающихся наступлениями новых (впервые наблюдаемых) состояний. Длительность таких серий (от одного впервые наблюдаемого состояния до следующего впервые наблюдаемого) отвечает схеме Бернулли и описывается геометрическим распределением с параметром, равным уровню научения. В течение каждой такой серии этот параметр геометрического распределения постоянен и скачкообразно меняется в конечной точке серии – в момент наблюдения нового состояния. Тогда в каждый момент

времени t матожидание длины текущей серии (до ближайшего момента наблюдения нового состояния внешней среды, не включая его) равно $L_t(1 - L_t)^{-1}$. Длина серии соответствует количеству повторов, необходимых для увеличения уровня разработанности/научения, что, в свою очередь, отвечает величине «издержек» (например, временных), требуемых для получения очередного приращения уровня научения, фактически, для получения порции новых знаний. Поэтому в некоторых случаях будем использовать матожидание *длины серии* наряду с уровнем научения для описания процесса научения и обозначать его $N_t = L_t(1 - L_t)^{-1}$ для $L_t < 1$.

Естественной является постановка задачи управления с целью как можно более быстрого или наименее ресурсоёмкого достижения требуемого уровня разработанности технологии или научения; оптимизируемыми и/или ограничивающими параметрами могут быть ресурсы или время.

В зависимости от знаний субъекта КД о внешней среде, в [5] сформулированы две задачи управления компонентами технологии.

Первая задача (условно назовём её *базовой*) характеризуется условиями, когда перечень и вероятности реализации состояний внешней среды постоянны и известны субъекту, т.е. K и все $\{p_k\}$ известны субъекту, независимы и не меняются во времени. Эта задача возникает всегда и является в этом смысле базовой: первоначальный синтез технологии КД и процесс проверок (блок 3 на рис. 1) субъект всегда выполняет, ориентируясь на определённый перечень состояний внешней среды. Базовая задача формулируется как получение зависимости УРТ от времени и её оптимизация в зависимости от ресурсов.

Вторая задача характеризуется *неизвестными свойствами внешней среды*, т.е. тем, что перечень и вероятности реализации состояний внешней среды K и/или хотя бы некоторые из вероятностей $\{p_k\}$ неизвестны субъекту или могут изменяться. Данная задача с необходимостью возникает при сколько-нибудь продолжительном повторяемом использовании технологии, когда вследствие естественной изменчивости внешней среды разработанная ранее технология перестаёт соответствовать новым условиям, что обнаруживается в результате выполнения

продуктивных активностей (блок 4 на рис. 1). Так как вторая задача (с неизвестными свойствами внешней среды) решается на основе закономерностей, характерных для первой, то в данной работе всё внимание уделим первой задаче.

Комплексные компоненты технологии формируются путём интеграции компонентов согласно логической и причинно-следственным структурам КД [6]. В [5] показано, что на общесистемном уровне логические модели любых структурных элементов КД (обладающих внутренней структурой) эквивалентны – им соответствует верная структура, в то время как причинно-следственная модель описывается графом определённого вида – так называемой «бинарной сетью» [5]. Набор свойств бинарной сети позволяет говорить о необходимости и достаточности перечня задач комплексирования компонентов технологии в составе а) последовательного, б) параллельного конъюнктивного и в) параллельного дизъюнктивного комплексирования компонентов технологии. В качестве сложного варианта комплексирования следует также рассмотреть случай г) создания компонента технологии одновременно с созданием собственно технологии его создания, возникающий при «пионерных» инновационных разработках. В каждом из вариантов комплексирования следует оптимизировать уровень научения комплексного компонента технологии в зависимости от времени и ресурсов.

Таким образом, рассмотрение процессов разработки/освоения технологий КД порождает две задачи (базовую и с неизвестными свойствами внешней среды) управления компонентами технологии КД и четыре а)-г) задачи комплексирования компонентов технологии КД, причём все они вписываются в интегрированную модель (рис. 1).

3. Обзор моделей и методов из смежных областей знаний

В предыдущем разделе показано, что процесс управления технологией формально представляет собой последовательное повторение циклов интегрированной модели на рис. 1. Такие задачи встречаются во многих отраслях знаний: испытание

сложных систем, анализ и тестирование программного обеспечения и др., где получено большое количество соответствующих решений. Рассмотрим примеры известных моделей и методы описания и решения задач, аналогичных сформулированной:

- испытания сложных систем и проверка их характеристик [1, 12, 13,30];
- тестирование программного обеспечения [18, 21, 22, 23, 24, 32, 35,44];
- управление знаниями и извлечение/приобретение знаний [8, 10, 26, 35, 36, 37,43];
- итеративное научение в педагогике, психологии, физиологии человека и животных [15, 31, 33, 39,40];
- тестирование знаний обучаемых в педагогике [11, 17, 19, 41].

Все перечисленные задачи, как и процесс создания технологии, отличаются неопределённостью, поэтому для их описания практически во всех случаях используются вероятностные модели и/или аппарат случайных процессов.

Испытания сложных систем (в частности, авиационных комплексов) в [1] представляются в виде иерархической структуры, узлы которой описывают испытания элементов, блоков и систем. Постулируется экспоненциальная (или логистическая) зависимость эффективности компонентов изделия от продолжительности испытаний, при этом, как правило, скорость роста эффективности пропорциональна обнаруженной ненадёжности в данный момент времени. Вычисляются средние значения времени испытаний на каждом уровне иерархии, необходимого для достижения заданного уровня эффективности, а также соответствующие затраты. Среднее время испытаний системы (изделия) в целом принимается равным сумме времён испытаний на каждом уровне иерархии.

На основании известной аппроксимации распределений случайных величин, через математические ожидания и дисперсии записывается задача оптимизации процесса испытаний. Постулирование зависимостей эффективности компонентов системы от длительности испытаний выглядит не очень обоснованным, а комплексирование испытаний компонентов системы путём

простого суммирования средних времён – сильно упрощённым, хотя общая схема постановки задачи заслуживает того, чтобы быть взятой за основу для дальнейшего развития и корректного уточнения.

Задачи испытания изделий и тестирования математических моделей также часто ставятся как задачи проверки гипотез и планирования эксперимента, например [12] и [13].

Весьма распространённым является метод Model Checking и его вариант Statistical Model Checking [32], которые применяются для тестирования и комплексных систем, и сложного программного обеспечения. Данный метод используется для комплексных систем с конечным множеством состояний, чьи количественные свойства специфицируются логическими выражениями, что позволяет оценивать меру соответствия свойств системы требуемым значениям. Задача тестирования стохастической системы формулируется как проверка гипотезы, удовлетворяют ли свойства заданным требованиям. Наличие описаний системы в логических выражениях позволяет использовать эти выражения для комплексирования элементарных тестов в сложные.

Другим популярным методом тестирования программных систем являются различные вариации регрессионного тестирования (Regression Testing) [22]. Регрессионное тестирование выполняется для верификации эксплуатируемого программного обеспечения и подтверждения его качества после внесения изменений и модернизаций. Наборы тестов всегда растут в объёмах, соответствуя развитию программного обеспечения, что делает чрезмерно дорогостоящим выполнение всего набора тестов после каждого изменения. В рамках регрессионного тестирования используются такие приёмы как минимизация, отбор и установление приоритетов тестов, реализуемые в виде формальных процедур. Минимизация набора тестов направлена на устранение избыточных тестовых примеров. Отбор тестового примера направлен на выделение тестов, непосредственно связанных с последними изменениями. Приоритезация направлена на упорядочивание последовательности тестов таким образом, чтобы максимально быстро обнаруживать ошибки.

Также среди популярных методов необходимо отметить тестирование на основе моделей (Model-Based Testing) [18], различные последовательные процедуры тестирования [41], широкое использование последовательного анализа [20], имитационных моделей [34], специальные методы тестирования кибер-физических систем [23] и другие.

Общие подходы к тестированию изложены в достаточно большом количестве ставших классическими работ – см., например, [3], а также основном профессиональном он-лайн-источнике программной индустрии SWEBOK V3 [24].

В настоящее время разработано достаточно большое количество моделей, описывающих процесс извлечения/приобретения знаний, все они в той или иной степени реализуют последовательные процессы анализа предметной области и улучшения моделей; приведём несколько примеров.

Так, в [43] предложен общий алгоритм поддержки принятия решений на основании анализа большого количества смешанных данных (непрерывных и порядковых). В качестве основного элемента данных рассматриваются некоторые события и ассоциированные с ними области в пространстве признаков. Алгоритм выявляет статистически значимые паттерны, которые в дальнейшем используются для решения конкретных задач. Паттерны поддаются классификации, для них можно оценить функции плотности распределения вероятностей и создать правила/модели поддержки принятия решений.

В работе [34] рассматривается проблема формирования согласованных планов задач для группы автономных агентов – подвижных роботов, которые находятся на качающейся плоскости и задача которых стабилизировать плоскость. Для решения задачи предлагаются методика группового обучения автоматов, заключающаяся в циклическом генерировании обучающих сигналов и статистическом закреплении «навыков».

Работа [26] посвящена методам обобщения знаний, полученных из эмпирических наблюдений. Знание синтезируется алгоритмом кластеризации, основываясь на обнаружении статистически значимых событий. Алгоритм, использующий вероятностную информационную меру, группирует упорядоченные и неупорядоченные дискретные данные в ходе двух фаз:

Во время инициирования кластера выполняется анализ распределения расстояний между ближайшими соседями для выбора критерия объединения образцов в кластеры. Во время уточнения кластера образцы перегруппируются с использованием метода покрытия событий, который выбирает подмножества статистически релевантных событий.

В статье [37] утверждается, что стохастические и детерминированные знания дополняют друг друга и улучшают друг друга, вводится стохастическая модель приобретённых знаний на основе диффузионной аппроксимации.

Работа [35] посвящена оптимальному байесовскому агенту – алгоритму, моделирующему процесс извлечения знаний в ходе последовательных наблюдений за стохастической средой со счётным множеством состояний, в основу алгоритма положена теория индуктивного вывода Соломонова (Solomonoff's Theory of Inductive Inference).

В работе [10] предложена и реализована в виде математической модели и программного комплекса концепция использования коэффициентов ценности для ранжирования ответов в универсальных гибридных вопросно-ответных системах, которая позволяет повысить эффективность формирования базы знаний системы.

Объектом исследования [8] являются базы данных систем автоматизированного проектирования, а предметом - процесс извлечения знаний из таких баз данных. Распределенность источников данных, гетерогенность представленных в них данных и вычислительная сложность анализа данных большого объёма обуславливают применение агентно-ориентированного подхода к достижению поставленной цели. Разработана организационная модель многоагентной системы извлечения знаний из распределенных гетерогенных баз данных. Описаны основные роли агентов и их взаимодействие между собой.

Модели итеративного научения [15] в педагогике, психологии, физиологии человека и животных описывают процесс *итеративного научения* – многократное повторение обучаемой системой (живой или неживой – технической или кибернетической) действий, проб, попыток и т.д. для достижения фиксированной цели при постоянных внешних условиях («*научение*» в

общем случае – процесс и результат приобретения индивидуального опыта», см. [14]). В обзорной работе [15] проанализированы десятки известных и распространённых моделей итеративного научения, и сформулирована общая модель, обобщающая свойства отдельных моделей. Обобщающая модель описывает систему, осуществляющую научение в стационарных условиях, состоящую из конечного набора элементов, каждый из которых характеризуется экспоненциальным законом научения. Научение системы в целом характеризуется некоторой функцией от частных экспонент. Ограничением и отдельных моделей научения, и обобщающей модели является, во-первых, постулирование законов научения, во-вторых, отсутствие обоснованных механизмов комплексирования моделей элементов в модель сложной системы, осуществляющей научение.

В педагогических измерениях получили широкое развитие методы современной теории тестов Item Response Theory (IRT) [19,41], предназначенной для оценивания латентных (ненаблюдаемых) параметров испытуемых, и заданий тестов на основе статистических моделей измерения. Модель взаимосвязи между значениями латентных переменных и наблюдаемых результатов выполнения теста в IRT определяется как условная вероятность правильного выполнения обучаемыми заданий теста. При этом условная вероятность задаётся логистической кривой или функцией нормального распределения. Наиболее распространёнными являются модели Раша [11] и Бирнбаума [17,19] в которых выбираются конкретные значения коэффициентов логистической функции.

Подводя итоги краткого обзора известных результатов, можно сказать, что рассмотренные модели представляют те или иные элементы «цикла проверок» в относительно простой форме: во-первых, без учёта рекурсивности и фрактальности таких циклов; во-вторых, постулируя базовые закономерности (например, экспоненциальная или логистическая зависимость эффективности компонентов изделия от продолжительности испытаний в [1] или экспоненциальная или логистическая зависимость уровня научения от времени в [15]), которые вообще говоря, являются следствием более сложных процессов, требующих моделирования и анализа. Результаты, изложенные в [4],

позволили обобщить особенности этого цикла и отразить их в единой общесистемной интегрированной модели (см. рис. 1), которая благодаря этому обобщает, детализирует и уточняет известные модели.

4. Анализ процесса разработки/освоения компонента технологии КД в базовой модели

Исследуем свойства процесса *разработки/освоения компонента технологии* и уровня научения L_t в случае, когда K и все $\{p_k\}$ известны субъекту КД.

Данный процесс (вектор x_t) является марковской цепью с конечным множеством состояний, номера которых y_t образуем из элементов вектора-строки x_t по следующему правилу:

$y_t = \sum_{k=1}^K x_{kt} 2^{k-1}$; тогда y_t также является марковской цепью и принимает любые целочисленные значения от 0 до

$$I = \sum_{k=1}^K 2^{k-1} = 2^K - 1 \text{ включительно.}$$

Сформируем матрицу $\Pi = \{\pi_{ij}; i = 0, 1, \dots, I; j = 0, 1, \dots, I\}$ переходных вероятностей процесса y_t . В начальный момент времени $t = 0$ процесс y_t находится 0-м состоянии $y_0 = 0$ ($x_{k0} = 0$ для всех k), с вероятностью 1. Из состояния «0» процесс может перейти только в состояния с номерами 2^{k-1} с вероятностями p_k , в частности, в состояние «1» - с вероятностью p_1 , в состояние «2» - с вероятностью p_2 , в состояние «4» - с вероятностью p_3 и т.д.; оставаться в состоянии «0» процесс не может.

Из состояния «1» процесс не может вернуться в состояние «0», с вероятностью p_1 он может остаться в состоянии «1», а с вероятностями p_k перейти в состояния $2^{k-1} + 1$ для каждого $1 < k \leq K$.

Вычислим элементы i -й строки матрицы Π при $0 < i \leq I$. Будем обозначать функцию, выделяющую значение k -го разряда в двоичном представлении числа I , как $b(i, k)$, причём самый младший разряд считаем первым, т.е. $i = \sum_{k=1}^K b(i, k) 2^{k-1}$.

Для i -го состояния вероятность сохранения состояния на следующем шаге равна $\pi_{ii} = \sum_{k=1}^K b(i,k) p_k$; переход в любое состояние с номером $j < i$ невозможен ($\pi_{ij} = 0$); и для каждого k , для которого $x_k = 0$, вероятность перехода в состояние $i + 2^{k-1}$ равна p_k , для остальных состояний с номерами, большими чем I , вероятности также равны 0.

Запишем выражение для элементов матрицы переходных вероятностей:

$$\pi_{ij} = \begin{cases} 0, & j \neq i + 2^n, n = 1, 2, \dots, K, \\ \sum_{k=1}^K b(i,k) p_k, & j = i, \\ p_n, & j = i + 2^n, n = 1, 2, \dots, K \text{ и } j \leq I; \end{cases} \quad i = 0, 1, \dots, I, j = 0, 1, \dots, I.$$

Тогда матрица Π переходных вероятностей марковской цепи y_t (верхне)треугольная, а состояние с максимальным номером $I = 2^K - 1$ - поглощающее: $\pi_{II} = 1$ и $\pi_{ij} = 0$; $j \neq I$.

Обозначим вероятности состояний цепи через $q_{it} = Pr(y_t = i)$, в векторной нотации: $q_t = (q_{0t}, q_{1t}, \dots, q_{it}, \dots, q_{It})$; в начальный момент времени $t = 0$ распределение цепи $q_0 = (1, 0, \dots, 0)$, тогда $q_t = q_{t-1} \Pi$ для любого $t > 0$ и $q_t = e_0 \Pi^t = (1, 0, 0, \dots, 0) \Pi^t$. Здесь и далее будем обозначать через e_i вектор-строку соответствующей размерности, у которой все элементы, кроме i -го равны 0, а i -й элемент равен 1.

Утверждение 1. Вероятность q_{it} любого состояния $0 < i < I$ из l -й группы при любых $t > 1$ удовлетворяет условию $q_{it} < l! t^{l-1} v^{t-l+1}$, где $0 < v < 1$ - некоторая положительная константа.

Доказательство утверждения 1. Рассмотрим состояния из 1-й группы, для каждого из них $q_{it} = p_k^t$, где k - номер соответствующего состояния природы, тогда справедливо $q_{it} < l! t^0 v_1^t$, где $v_1 = \max_k \{p_k\} < 1$.

Пусть $q_{ii} < l!t^{l-1}v_l^{t-l+1}$ для всех состояний l -й группы. Эволюция распределения вероятностей состояний описывается соотношением $q_{ii} = \sum_{j=0}^{i-1} \pi_{ji} q_{ji-1} + \pi_{ii} q_{ii-1}$. Тогда для любого состояния из

$(l+1)$ -й группы при $t < l$ $q_{ii} \equiv 0$, а при $t \geq l$.

$$q_{ii} = \sum_{\theta=0}^{t-l} \pi_{ii}^{\theta} \sum_{j=0}^{i-1} \pi_{ji} q_{ji-1-\theta} = \sum_{j=0}^{i-1} \pi_{ji} \sum_{\theta=0}^{t-l} \pi_{ii}^{\theta} q_{ji-1-\theta}.$$

Получим ограничение на q_{ii} .

$$\begin{aligned} q_{ii} &= \sum_{j=0}^{i-1} \pi_{ji} \sum_{\theta=0}^{t-l} \pi_{ii}^{\theta} q_{ji-1-\theta} < \sum_{j=0}^{i-1} \pi_{ji} \sum_{\theta=0}^{t-l} \pi_{ii}^{\theta} l!(t-1-\theta)^{l-1} v_l^{t-l-\theta} < \\ < l!(t-1)^{l-1} \sum_{j=0}^{i-1} \pi_{ji} \sum_{\theta=0}^{t-l} \pi_{ii}^{\theta} v_l^{t-l-\theta} < \\ < t^{l-1} (l+1)! (t-l) (\max\{\pi_{ii}; v_l\})^{t-l} < (l+1)! t^l v_{l+1}^{t-l}, \end{aligned}$$

где $v_{l+1} = \max\{v_l; \max_i \{\pi_{ii}\}\}$, максимум берётся по всем состояниям i , принадлежащим $(l+1)$ -й группе.

Следовательно, для всех состояний $(l+1)$ -й группы выполнено $q_{ii} < (l+1)! t^l v_{l+1}^{t-l}$.

Обозначим $\nu = \max_i \{\pi_{ii}\}$, тогда $v_l \leq \nu$ для всех l , следовательно $q_{ii} < l! t^l \nu^{t-l}$.

Тогда из метода математической индукции следует справедливость утверждения для всех i -х состояний цепи кроме l -го. Утверждение 1 доказано.

Содержательно утверждение 1 означает, что вероятности всех состояний цепи, кроме l -го, уменьшаются по времени не медленнее, чем $t^{K-1} \nu^t$, и, наоборот, вероятность l -го состояния стремится к 1 не медленнее, чем $\alpha t^{K-1} \nu^t$, т.е. $q_{ll} > 1 - \alpha t^{K-1} \nu^t$, где $\nu = 1 - \min_k \{p_k\}$.

Из утверждения 1 непосредственно следует также, что стационарное распределение марковской цепи y_t существует и

единственно: $s = (0, 0, \dots, 0, 1) = e_t$, и оно является единственным решением матричного уравнения $s = s\Pi$ ($s_i = \sum_{j=0}^I \pi_{ji} s_j$).

Следовательно, рассматриваемая модель создания технологий имеет единственное устойчивое состояние – состояние, когда все возможные состояния внешней среды проверены, т.е. на длительном периоде можно достичь любого наперёд заданного уровня (сколь угодно близкого к единице) научения.

Исследуем *кривую научения* - поведение математического ожидания процесса $L_t = \sum_{k=1}^K x_{kt} p_k$. Используем символическую запись $E[\cdot]$ для обозначения математических ожиданий.

Прежде всего, из утверждения 1 следует сходимость по вероятности уровня научения к единице:

$$\forall \varepsilon > 0, \lim_{t \rightarrow \infty} \left\{ \Pr(|L_t - 1| > \varepsilon) \right\} = 0.$$

Во-первых, в силу определения процесса L_t его приращения всегда неотрицательны: $\Delta L_t = L_t - L_{t-1} \geq 0$, причём и значения L_t , и приращения ΔL_t неотрицательны и не больше единицы. Во-вторых, процесс L_t также является марковской цепью, т.е. для любого t L_t и ΔL_t - независимые случайные величины. Тогда⁴:

$$(1) \quad E[L_t] = \sum_{k=1}^K p_k E[x_{kt}] = \sum_{k=1}^K p_k \left(1 - (1 - p_k)^t \right) = 1 - \sum_{k=1}^K p_k (1 - p_k)^t.$$

С другой стороны, выражение (1) для $E[L_t]$ также можно получить на основании распределения состояний цепи q_t :

$$E[L_t] = \sum_{i=1}^I \left(\sum_{k=1}^K b(i, k) p_k \right) q_{it} = e_0 \Gamma^t \beta, \quad \text{где } \beta - \text{ вектор-столбец,}$$

элементами которого являются величины $\sum_{k=1}^K b(i, k) p_k$, кото-

⁴ Напомним, что кривая научения L_t соответствует вероятности того, что внешняя среда примет новое значение на $(t + 1)$ -м шаге; эта вероятность оценивается по наблюдениям в течение t шагов включительно.

рые, в свою очередь, также являются диагональными элементами матрицы Π .

Из (1) следует выражение для длины серии:

$$N_t = E[L_t](1 - E[L_t])^{-1} = \left(1 - \sum_{k=1}^K p_k (1 - p_k)^t\right) \left(\sum_{k=1}^K p_k (1 - p_k)^t\right)^{-1}.$$

Так как $\Delta L_t = L_t - L_{t-1} \geq 0$, первые разности последовательности $E[L_t]$ строго положительны для всех t . Получим выражение для m -х разностей, для этого сначала вычислим первые:

$$\Delta E[L_t] = 1 - \sum_{k=1}^K p_k (1 - p_k)^t - \left(1 - \sum_{k=1}^K p_k (1 - p_k)^{t-1}\right) = \sum_{k=1}^K p_k^2 (1 - p_k)^{t-1} > 0,$$

$$\begin{aligned} \Delta^2 E[L_t] &= \Delta E[L_t] - \Delta E[L_{t-1}] = \sum_{k=1}^K p_k^2 (1 - p_k)^{t-1} - \sum_{k=1}^K p_k^2 (1 - p_k)^{t-2} = \\ &= -\sum_{k=1}^K p_k^3 (1 - p_k)^{t-2} < 0. \end{aligned}$$

Продолжая, легко получить:

$$\Delta^m E[L_t] = (-1)^{m+1} \sum_{k=1}^K p_k^{m+1} (1 - p_k)^{t-m}.$$

Следует заметить, что разности значений кривой научения для любого момента времени t образуют знакопеременную последовательность, значения которой убывают по модулю ($|\Delta^m E[L_t]| > |\Delta^{m+1} E[L_t]|$), и, кроме того, для первых разностей справедливо неравенство $1 - E[L_t] > \Delta E[L_t]$.

Таким образом, справедливо следующее утверждение.

Утверждение 2. Кривая научения $E[L_t]$ обладает следующими свойствами:

- в начальный момент времени $t = 0$ её значение равно нулю: $E[L_0] = 0$;
- она монотонно возрастает: $\Delta E[L_t] > 0$;
- первые разности ограничены неравенством $1 - E[L_t] > \Delta E[L_t]$;

- скорость её роста монотонно убывает: $\Delta E^2[L_t] < 0$ и $\Delta E^3[L_t] > 0$;
- она асимптотически стремится к единице.

5. Аппроксимации кривой научения

Рассмотрим возможные аппроксимации кривой научения $E[L_t]$ (см. выражение (1)) в зависимости от вида распределения возможных состояний внешней среды

$$P = \left\{ p_k; k = \overline{1, K}; \sum_{k=1}^K p_k = 1 \right\}.$$

А) **Равномерное распределение** возможных состояний внешней среды P . Обозначим для упрощения выкладок $\delta = 1 / K$, тогда

$$(2) E[L_t] = 1 - \sum_{k=1}^K p_k (1 - p_k)^t = 1 - \sum_{k=1}^K \delta (1 - \delta)^t = 1 - (1 - \delta)^t = 1 - \exp(-\gamma t),$$

где $\gamma = \ln(1 + 1 / (K - 1))$ - скорость изменения уровня научения - *скорость научения* [12].

Экспоненциальная кривая научения (2) (ее разностный аналог: $E[L_t] = E[L_{t-1}] + \gamma (1 - E[L_{t-1}]) = \gamma + (1 - \gamma) E[L_{t-1}]$) является хрестоматийной для теории научения (см. обзор в [15] и пионерскую работу [31]). В то же время, для рассматриваемой модели она является частным случаем, соответствующим равномерному распределению состояний внешней среды.

Средняя длина серии в случае равномерного распределения экспоненциально растёт $N_t = \exp(\gamma t) - 1$, что интуитивно очевидно: с ростом уровня научения при всё большей доле «освоенных» состояний внешней среды получение новых знаний требует всё больших усилий для «поиска» новых состояний.

Разностное уравнение для N_t имеет простой вид $N_{t+1} = \exp(\gamma) N_t + (\exp(\gamma) - 1)$ - средняя длина серии мультипликативно растёт.

При $K \gg 1$ скорость научения $\gamma = \ln(1 + 1/(K - 1)) \approx 1/(K - 1) \approx 1/K$,

и

$$(3) \quad E[L_t] \approx 1 - \exp(-t/K).$$

Следует отметить, что экспоненциальные кривые научения типа (2) или (3) являются классическими для теории научения (см. обзор в [15]). Кроме того, как показано в [7], именно равномерное распределение состояний природы максимизирует ожидаемое значение уровня научения.

Б) Распределение (n, δ) - с несколькими (n) одинаково высоковероятными состояниями и остальными ($K - n$) одинаково маловероятными ($\delta \ll 1/K$) состояниями:

$$(4) \quad P = \left\{ p_k = (1 - \delta(K - n)) / n, k = \overline{1, n}; p_k = \delta, k = \overline{n+1, K} \right\}.$$

Имеет смысл рассматривать случай, когда вероятности $(1 - \delta(K - n)) / n$ первой группы состояний существенно больше вероятностей δ второй группы; случай, когда вероятности отличаются незначительно, аппроксимируется равномерным распределением (см. выше). То есть $\delta \ll (1 - \delta(K - n)) / n$, а это означает, что $\delta K \ll 1$. Найдём кривую научения для рассматриваемого случая:

$$\begin{aligned} E[L_t] &= 1 - \sum_{k=1}^K p_k (1 - p_k)^t = \\ &= 1 - \sum_{k=1}^n \frac{1 - \delta(K - n)}{n} \left(1 - \frac{1 - \delta(K - n)}{n} \right)^t - \sum_{k=n+1}^K \delta (1 - \delta)^t, \end{aligned}$$

т.е.

$$E[L_t] = 1 - (1 - \delta(K - n)) \left(1 - \frac{1}{n} + \frac{(K - n)}{n} \delta \right)^t - (K - n) \delta (1 - \delta)^t.$$

При больших t (так как $\delta = 1/K$ и $n < N$, то $(1 - 1/n) < (1 - \delta)$) случай (4) стремится к предыдущему (равномерному распределению):

$$E[L_t] \approx 1 - (K - n) \delta (1 - \delta)^t.$$

При малых t получается следующая аппроксимация:

$$E[L_t] \approx 1 - (1 - \delta(K - n)) \left(1 - \frac{1}{n} + \frac{(K - n)}{n} \delta \right)^t.$$

В) «Зашумлённое равномерное» распределение на «большом» множестве областей возможных состояний внешней среды:

$$(5) \quad P = \left\{ p_k; k = 1, 2, \dots, K; \sum_{k=1}^K p_k = 1; p_k \ll 1; K \gg 1 \right\}.$$

При малых t кривая научения аппроксимируется следующим образом:

$$(6) \quad E[L_t] = 1 - \sum_{k=1}^K p_k (1-p_k)^t \approx 1 - \sum_{k=1}^K p_k (1-tp_k) = 1 - \sum_{k=1}^K p_k + t \sum_{k=1}^K p_k^2 = t \sum_{k=1}^K p_k^2,$$

т.е. линейно растёт по времени t со скоростью $\sum_{k=1}^K p_k^2$.

При больших t возможны два варианта поведения кривой научения. Если все области возможных состояний внешней среды примерно эквивалентны и вероятности различаются незначительно, т.е. $p_k \approx 1/K; k = \overline{1, K}$, тогда справедлива оценка (3), полученная для равномерного распределения. Если имеется некоторое количество (n), существенно отличающихся областей, то адекватной является аппроксимация Б:

$$E[L_t] \approx 1 - \left(1 - \frac{n}{K}\right) \left(1 - \frac{1}{K}\right)^t.$$

Отметим, что аналитические выражения (2), (3) и (6), а также свойства кривой научения $E[L_t]$ в предложенной модели в некотором смысле соответствуют многим известным моделям научения (в частности рассмотренные в разделах 4.2 и 6.7 [15]). Однако известные модели постулируют сам закон изменения кривой научения или уравнения, его описывающие, в то время как предлагаемая модель описывает сам процесс научения-разработки, а уравнения и свойства кривой научения следуют из анализа модели.

6. Среднее время научения

Получим выражение для среднего времени достижения *требуемого уровня* $L_{\text{треб}} \in (0;1)$ научения (готовности, зрелости технологии), т.е. среднее значение момента времени t , когда

$$L_t = \sum_{k=1}^K x_{kt} p_k \geq L_{\text{треб}}.$$

Для этого исследуем поведение марковской цепи y_t , её распределения состояний $q_{it} = Pr(y_t = i)$. Выше было получено, что начальное (при $t = 0$) распределение цепи $q_0 = (1, 0, \dots, 0)$ и $q_t = q_{t-1} \Pi$ для любого $t > 0$, $q_t = e_0 \Pi^t = (1, 0, 0, \dots, 0) \Pi^t$.

Матрица Π является (верхне)треугольной, откуда следуют несколько свойств матрицы Π^t :

- определитель матрицы Π (обозначим его через $\Delta \Pi$) является произведением диагональных элементов $\Delta \Pi = \prod_{i=1}^I \pi_{ii}$;
- Π^t - также (верхне)треугольная (непосредственно следует из формулы умножения матриц);
- диагональные элементы матрицы Π^t являются степенями диагональных элементов матрицы Π , т.е. $\pi_{ii}^t = (\pi_{ii})^t$;
- определитель матрицы Π^t является степенью определителя Π , т.е. $\Delta \Pi^t = (\Delta \Pi)^t = \left(\prod_{i=1}^I \pi_{ii} \right)^t$.

Сформируем «маску» - вектор-столбец r той же размерности, что и вектор-строка q_t , по следующему правилу:

$$r_i = \begin{cases} 1, & \text{если } \sum_{k=1}^K b(i, k) p_k < L_{\text{треб}}, \\ 0, & \text{если } \sum_{k=1}^K b(i, k) p_k \geq L_{\text{треб}}; \end{cases} \quad i = \sum_{k=1}^K 2^k b(i, k) = 0, 1, \dots, I.$$

Такой вектор-маска «выделяет» состояния процесса y_t , соответствующие уровням научения ниже требуемого. Тогда в каждый момент времени вероятность того, что уровень научения достиг или превысил заданный уровень будет равна $Pr(L_t \geq L_{\text{треб}}) = 1 - q_t r$ или $Pr(L_t < L_{\text{треб}}) = q_t r$. Вероятность того, что момент времени $t_{\text{дост}}$ достижения требуемого уровня научения превышает текущий, $Pr(t_{\text{дост}} > t) = q_t r$ или вероятность, что

уровень уже достигнут, $\Pr(t_{\text{дост}} \leq t) = 1 - q_t r$. Очевидно, $r_t = 0$, тогда из утверждения 1 следует, что вероятности $\Pr(L_t < L_{\text{треб}}) = \Pr(t_{\text{дост}} > t)$ мажорируются функцией $t^{K-1}v^t$ при $t \rightarrow \infty$. Получаем:

$$\Pr(t_{\text{дост}} > t) = \sum_{i=0}^t r_i q_i < \sum_{i=0}^t r_i \alpha^i t^{K-1} v^i = \left(\alpha \sum_{i=0}^t r_i \right) t^{K-1} v^t.$$

Используем выражение $q_t = e_0 \Pi^t$, тогда $\Pr(t_{\text{дост}} > t) = q_t r = e_0 \Pi^t r$.

Так как $\Pr(t_{\text{дост}} > t-1) = \Pr(t_{\text{дост}} = t) + \Pr(t_{\text{дост}} > t)$, то $\Pr(t_{\text{дост}} = t) = \Pr(t_{\text{дост}} > t-1) - \Pr(t_{\text{дост}} > t)$.

Для среднего времени справедливо выражение

$$\bar{t}_{\text{дост}} = \sum_{t=0}^{\infty} t \Pr(t_{\text{дост}} = t) = \sum_{t=0}^{\infty} t (\Pr(t_{\text{дост}} > t-1) - \Pr(t_{\text{дост}} > t)).$$

Так как $\Pr(t_{\text{дост}} > t)$ мажорируются $t^{K-1}v^t$ при $t \rightarrow \infty$, то ряд $t \Pr(t_{\text{дост}} > t)$ сходится, а его сумма $\sum_{t=0}^{\infty} t \Pr(t_{\text{дост}} > t)$ конечна, тогда

$$\begin{aligned} \bar{t}_{\text{дост}} &= \sum_{t=0}^{\infty} t (\Pr(t_{\text{дост}} > t-1) - \Pr(t_{\text{дост}} > t)) = \\ &= \sum_{t=0}^{\infty} (t+1) \Pr(t_{\text{дост}} > t) - \sum_{t=0}^{\infty} t \Pr(t_{\text{дост}} > t) = \sum_{t=0}^{\infty} \Pr(t_{\text{дост}} > t) \end{aligned}$$

Итак, получаем:

$$(7) \quad \bar{t}_{\text{дост}} = \sum_{t=0}^{\infty} \Pr(t_{\text{дост}} > t) = \sum_{t=0}^{\infty} e_0 \Pi^t r = e_0 \left(\sum_{t=0}^{\infty} \Pi^t \right) r = e_0 (E - \Pi)^{-1} r,$$

где E - единичная матрица той же размерности, что и матрица Π . Существование обратной матрицы $(E - \Pi)^{-1}$ следует из того, что матрица $(E - \Pi)$ (верхне)треугольная, все её диагональные элементы могут быть найдены из выражения для π_{ij} выше, и их произведение больше нуля.

Вышеприведённые соображения обосновывают справедливость следующего утверждения.

Утверждение 3. Для любой марковской цепи среднее время первого достижения состояния из заданного множества равно

$$\bar{t}_{\text{дост}} = \sum_{t=0}^{\infty} \Pr(t_{\text{дост}} > t). \quad \text{Если данный ряд сходится, то}$$

$$\bar{t}_{\text{дост}} = e_0 (E - \Pi)^{-1} r.$$

Выражение (7) позволяет аналитически получать значения $\bar{t}_{\text{дост}}$ в общем случае в зависимости от распределения возможных состояний внешней среды $P = \{p_k; k = 1, 2, \dots, K\}$ (p_k входят в выражение (7) через матрицу Π) и от требуемого уровня научения ($L_{\text{треб}}$ учитывается через вектор m). Однако выражение (7) не является наглядным – ни вероятности p_k , ни уровень $L_{\text{треб}}$ в явном виде в него не входят. Кроме того, значительная размерность матрицы Π ($2^K \times 2^K$) затрудняет его практическое использование.

В ряде случаев можно получить более простые и наглядные выражения, рассмотрим один из них - равномерное распределение $P = \{p_k = 1/K = \delta; k = \overline{1, K}\}$ возможных состояний внешней среды. В этом случае вместо марковской цепи y_i , принимающей значения $i = 0, 1, \dots, I$, рассмотрим цепь y_i^* , значения которой отвечают количеству состояний внешней среды, для которых технология уже проверена, т.е. y_i^* принимает значения от 0 до K . Тогда $L_i = \delta y_i^*$, а переходные вероятности примут вид:

$$\pi_{ij} = \begin{cases} 0 & \text{при } j < i \text{ или } j > i + 1, \\ 1 - \delta i & \text{при } j = i + 1, \\ \delta i & \text{при } j = i; \end{cases} \quad \text{для } i = 0, 1, \dots, K.$$

Тогда матрица $(E - \Pi)$ является (верхне)треугольной, ленточной, с элементами:

$$\varepsilon_{ij} = \begin{cases} 0 & \text{при } j < i \text{ или } j > i + 1, \\ 1 - \delta i & \text{при } j = i, \\ \delta i - 1 & \text{при } j = i + 1; \end{cases} \quad \text{для } i = 0, 1, \dots, K.$$

Обратная матрица $(E - \Pi)^{-1}$ - также (верхне)треугольная с элементами

$$\varepsilon_{ij}^- = \begin{cases} 0 & \text{при } j < i, \\ (1 - \delta i)^{-1} & \text{при } i \leq j < K, \\ 1 & \text{при } j = K; \end{cases} \quad \text{для } i = 0, 1, \dots, K.$$

Операция $(1, 0, 0, \dots, 0)(E - \Pi)^{-1}r$ «вырезает» из матрицы $(E - \Pi)^{-1}$ первую строку и суммирует те элементы этой строки, для которых $\sum_{k=1}^K x_k P_k < L_{\text{треб}}$. В рассматриваемом случае равномерного распределения суммируются первые $L_{\text{треб}} / \delta = K L_{\text{треб}}$ элементов, тогда среднее время достижения требуемого уровня $L_{\text{треб}}$ равно:

$$\bar{t}_{\text{дост}} = \sum_{i=0}^{KL_{\text{треб}}} (1 - \delta i)^{-1} = \sum_{i=0}^{KL_{\text{треб}}} \frac{K}{K - i}.$$

Получим компактную аппроксимацию $\bar{t}_{\text{дост}}$ при $K \gg 1$:

$$\begin{aligned} \bar{t}_{\text{дост}} &= \sum_{i=0}^{KL_{\text{треб}}} \frac{K}{K - i} = K \sum_{i=0}^{KL_{\text{треб}}} \frac{1}{1 - K^{-1}i} K^{-1} \approx K \int_0^{KL_{\text{треб}}} \frac{1}{1 - K^{-1}x} K^{-1} dx = \\ &= -K \ln(1 - K^{-1}x) \Big|_0^{KL_{\text{треб}}} = -K \ln(1 - L_{\text{треб}}), \end{aligned}$$

т.е.

$$(8) \quad \bar{t}_{\text{дост}} \approx -K \ln(1 - L_{\text{треб}}).$$

Отметим, что выражение (8) совпадает с приближённым решением \hat{t} уравнения (см. выражение (2))

$$E[L_t] = 1 - (1 - \delta)^i = L_{\text{треб}}, \quad \text{для} \quad \text{которого}$$

$$\hat{i} = \frac{\ln(1 - L_{\text{треб}})}{\ln(1 - \delta)} \approx -\frac{\ln(1 - L_{\text{треб}})}{\delta} = -K \ln(1 - L_{\text{треб}}).$$

Получим выражение для среднего времени достижения «абсолютного» уровня научения: $L_t = 1$.

Пусть в некоторый момент времени уже известны, проверены какие-то состояния внешней среды, и остались не проверенными несколько, а именно $I \leq K$, состояний $\{p_i; i = \overline{1, I}\}$, очевидно, что $\sum_{i=1}^I p_i \leq 1$.

Обозначим через $T(\{p_i; i = \overline{1, I}\})$ среднее время, в течение которого будут проверены все оставшиеся I состояний, т.е. достигнут уровень $L_t = 1$. Используя метод математической индукции, получим выражение:

$$(9) \quad T(\{p_i; i = \overline{1, I}\}) = \sum_{i=1}^I p_i^{-1} - \sum_{i;k} (p_i + p_k)^{-1} + \sum_{i;k;l} (p_i + p_k + p_l)^{-1} - \dots$$

Запишем (9) в иной, эквивалентной, форме:

$$(10) \quad T(\{p_i; i = \overline{1, I}\}) = \sum_{k=1}^I (-1)^{k+1} \sum_{i_1 i_2 \dots i_k} \left(\sum_{j=1}^k p_{i_j} \right)^{-1}.$$

То есть покажем, что среднее время равно сумме знакопеременных сумм всех k -к из $\{p_i; i = \overline{1, I}\}$, $n = \overline{1, I}$.

Обозначим сумму всех k -к из $\{p_i; i = \overline{1, I}\}$, $k = \overline{1, I}$ в выражении (10) для $T(\{p_i; i = \overline{1, I}\})$ как $\Theta(k; I)$:

$$(11) \quad \Theta(k; I) = \sum_{i_1 i_2 \dots i_k} \left(\sum_{j=1}^k p_{i_j} \right)^{-1} = \sum_{i_1 i_2 \dots i_k} \frac{1}{p_{i_1} + p_{i_2} + \dots + p_{i_k}}.$$

С учётом (11) выражение (10) примет вид:

$$(12) \quad T(\{p_i; i = \overline{1, I}\}) = \sum_{k=1}^I (-1)^{k+1} \sum_{i_1 i_2 \dots i_k} \left(\sum_{j=1}^k p_{i_j} \right)^{-1} = \sum_{j=1}^I (-1)^{j+1} \Theta(j; I).$$

Пусть осталось непроверенным одно состояние внешней среды, тогда $T(\{p_i; i = 1\}) = 1/p_1$.

Пусть выражения (9), (10), (12) справедливы для $I - 1$ состояний, т.е. $T(\{p_i; i = \overline{1, I-1}\}) = \sum_{k=1}^{I-1} (-1)^{k+1} \Theta(k; I-1)$.

Покажем, что (9), (10), (12) справедливы для I состояний. Событие, заключающееся в последовательной реализации $i = \overline{1, I}$ состояний, эквивалентно объединению I событий, каждое из которых заключается в реализации одного из I состояний и после этого в последовательной реализации оставшихся $I - 1$ состояний. Тогда

$$(13) T(\{p_i; i = \overline{1, I}\}) = \left(\sum_{i=1}^I p_i \right)^{-1} + \sum_{j=1}^I p_j \left(\sum_{i=1}^I p_i \right)^{-1} T(\{p_i; i = \overline{1, I}; i \neq j\}),$$

где первое слагаемое – среднее время до первого из реализовавшихся состояний $i=1, 2, \dots, I$. Каждое из j -х слагаемых под знаком суммы – вероятность того, что первым проверено j -е состояние и после этого $T(\{p_i; i = \overline{1, I}; i \neq j\})$ – среднее время проверки $I - 1$ оставшихся состояний.

Подставим (12) в правую часть (13):

$$(14) T(\{p_i; i = \overline{1, I}\}) = \left(\sum_{i=1}^I p_i \right)^{-1} + \left(\sum_{i=1}^I p_i \right)^{-1} \sum_{j=1}^I \left(p_j \sum_{k=1}^{I-1} (-1)^{k+1} \Theta(k; I-1) \right).$$

Преобразуем вторую сумму, подставив в неё (11).

$$\sum_{j=1}^I \left(p_j \sum_{k=1}^{I-1} (-1)^{k+1} \Theta(k; I-1) \right) = \sum_{j=1}^I \sum_{k=1}^{I-1} (-1)^{k+1} \sum_{i_1; i_2; \dots; i_k} \frac{p_j}{p_{i_1} + p_{i_2} + \dots + p_{i_k}}.$$

Так как каждая из вероятностей p_j в числителях не входит в сумму $p_{i_1} + p_{i_2} + \dots + p_{i_k}$ в знаменателях дробей, порядок суммирования можно поменять местами и получить:

$$\begin{aligned} \sum_{j=1}^I \left(p_j \sum_{k=1}^{I-1} (-1)^{k+1} \Theta(k; I-1) \right) &= \sum_{j=1}^I \sum_{k=1}^{I-1} (-1)^{k+1} \sum_{i_1; i_2; \dots; i_k} \frac{p_j}{p_{i_1} + p_{i_2} + \dots + p_{i_k}} = \\ &= \sum_{k=1}^{I-1} (-1)^{k+1} \sum_{i_1; i_2; \dots; i_k} \left(p_{i_1} + p_{i_2} + \dots + p_{i_k} \right)^{-1} \sum_{j=1; j \neq i_1; i_2; \dots; i_k}^I p_j. \end{aligned}$$

Суммирование по j осуществляется по всем j от 1 до I , но не совпадающим ни с одним из i_1, i_2, \dots, i_k , поэтому

$$p_j = \left(\sum_{i=1}^I p_i \right) - (p_{i_1} + p_{i_2} + \dots + p_{i_k}).$$

Подставим это соотношение в выражение для второй суммы

$$\begin{aligned} & \sum_{j=1}^I \left(p_j \sum_{k=1}^{I-1} (-1)^{k+1} \Theta(k; I-1) \right) = \sum_{k=1}^{I-1} (-1)^{k+1} \sum_{i_1, i_2, \dots, i_k} \frac{\left(\sum_{i=1}^I p_i \right) - (p_{i_1} + p_{i_2} + \dots + p_{i_k})}{p_{i_1} + p_{i_2} + \dots + p_{i_k}} = \\ & = \left(\sum_{i=1}^I p_i \right) \sum_{k=1}^{I-1} (-1)^{k+1} \sum_{i_1, i_2, \dots, i_k} (p_{i_1} + p_{i_2} + \dots + p_{i_k})^{-1} - \sum_{k=1}^{I-1} (-1)^{k+1} \sum_{i_1, i_2, \dots, i_k} 1 = \\ & = \left(\sum_{i=1}^I p_i \right) \sum_{k=1}^{I-1} (-1)^{k+1} \Theta(k; I) + \sum_{k=1}^{I-1} (-1)^k C_I^k. \end{aligned}$$

где C_I^k – число сочетаний из I по k .

Подставим это выражение в (14)

$$\begin{aligned} T(\{p_i; i = \overline{1, I}\}) &= \left(\sum_{i=1}^I p_i \right)^{-1} + \left(\sum_{i=1}^I p_i \right)^{-1} \left(\left(\sum_{i=1}^I p_i \right) \sum_{k=1}^{I-1} (-1)^{k+1} \Theta(k; I) + \right) = \\ &= \sum_{k=1}^{I-1} (-1)^{k+1} \Theta(k; I) + \left(\sum_{i=1}^I p_i \right)^{-1} \left(1 + \sum_{k=1}^{I-1} (-1)^k C_I^k \right). \end{aligned}$$

Соответственно определению $\Theta(I; I) = \left(\sum_{i=1}^I p_i \right)^{-1}$, а также

$$1 + \sum_{n=1}^{I-1} (-1)^n C_I^n + (-1)^I = (1-1)^I = 0, \text{ поэтому } 1 + \sum_{n=1}^{I-1} (-1)^n C_I^n = (-1)^{I+1}.$$

Получим окончательно требуемое выражение в форме (12):

$$T(\{p_i; i = \overline{1, I}\}) = \sum_{k=1}^I (-1)^{k+1} \Theta(k; I).$$

Итак, в данном и предыдущих разделах рассмотрены свойства процесса разработки/освоения компонента технологии, уровня научения, а также времени научения. Перейдем к анализу моделей комплексирования компонентов технологии.

7. Модели комплексирования компонентов технологии

Проанализируем свойства процессов *комплексирования частных компонентов технологии*, описываемых каждый в рамках базовой модели (раздел 4):

- А) последовательного;
- Б) параллельного конъюнктивного;
- В) параллельного дизъюнктивного;
- Г) параллельного с полным обменом информацией;
- Д) «научения научению».

Для этого рассмотрим процесс управления несколькими компонентами технологии с соответствующим объединением результатов. Состояния частных процессов образуют марковские цепи (и обладают свойствами, рассмотренными в разделе 3). Тогда поведение комплексного процесса также является марковской цепью на множестве состояний, равном произведению множеств состояний частных процессов.

Аналогично тому, как это было сделано выше, введём K^* -мерный процесс (где $K^* = \sum_{m=1}^M K^m$) $x_t^* = (x_{1t}^*, x_{2t}^*, \dots, x_{kt}^*, \dots, x_{K^*t}^*)$,

каждый элемент которого x_{kt}^* принимает значения 0 или 1 и отражает факт выполнения проверки соответствующего состояния внешней среды, а также процесс y_t^* , отражающий номер состояния процесса x_t^* . Оба процесса x_t^* и y_t^* являются марковскими цепями, матрица переходных вероятностей y_t^* является (верхне)треугольной, и для y_t^* справедливы утверждение 1 (об асимптотическом поведении распределения состояний) и 3 (о среднем времени достижения заданного уровня научения).

А. Если целью комплексного процесса является создание всех частных компонентов технологии (конъюнкция всех M частных компонентов), то уровень $L_t^{1..M}$ научения, разработанности комплексной технологии, равен вероятности того, что при очередном испытании ни в одном из частных процессов не будет получено ещё не проверенное состояние внешней среды, т.е. равен произведению уровней разработанности частных технологий. А эта вероятность, в свою очередь, равна произве-

дению вероятностей L_t^m всех частных компонентов технологии,

т.е. $L_t^{1...M} = \prod_{m=1}^M L_t^m$ и

$$(15) E[L_t^{1...M}] = \prod_{m=1}^M E[L_t^m] = \prod_{m=1}^M \left[1 - \sum_{k=1}^K p_k^m (1 - p_k^m)^t \right].$$

Формулу (7) для среднего времени достижения требуемого уровня научения одного компонента несложно расширить на случай M элементов:

$$(16) \bar{t}_{A \text{ дост}} = \sum_{t=0}^{\infty} \Pr(t_{A \text{ дост}} > t) = \sum_{t=0}^{\infty} \Pr\left(\max_m \{t_m \text{ дост}\} > t\right) = \\ = \sum_{t=0}^{\infty} \left(1 - \prod_{m=1}^M (1 - e_0 \Pi_m^t r_m) \right).$$

Если характеристики процессов разработки компонентов технологии одинаковы, то

$$\bar{t}_{A \text{ дост}} = \sum_{t=0}^{\infty} \left(1 - (1 - e_0 \Pi^t r)^M \right) = \sum_{t=0}^{\infty} \left(\sum_{m=1}^M C_M^m (-1)^{m-1} (e_0 \Pi^t r)^m \right) = \\ = M \sum_{t=0}^{\infty} e_0 \Pi^t r + \sum_{t=0}^{\infty} \left(\sum_{m=2}^{M-1} C_M^m (-1)^{m-1} (e_0 \Pi^t r)^m \right) + (-1)^{M-1} \sum_{t=0}^{\infty} (e_0 \Pi^t r)^M.$$

Рассматривая последовательность величин $\bar{t}_{A \text{ дост}}$ для различных возрастающих M , легко получить выражение для первых разностей $\Delta \bar{t}_M = \sum_{t=0}^{\infty} e_0 \Pi^t r (1 - e_0 \Pi^t r)^{M-1}$ и вторых разностей

$$\Delta^2 \bar{t}_M = - \sum_{t=0}^{\infty} (e_0 \Pi^t r)^2 (1 - e_0 \Pi^t r)^{M-2}.$$

Откуда видно, что $\bar{t}_{A \text{ дост}}$ растёт

с увеличением M , однако скорость роста по M убывает. Также первые разности ограничены сверху и снизу:

$$\bar{t}_{\text{дост}} - \sum_{t=0}^{\infty} (e_0 \Pi^t r)^2 = \sum_{t=0}^{\infty} e_0 \Pi^t r (1 - e_0 \Pi^t r) \leq \Delta \bar{t}_M < \sum_{t=0}^{\infty} e_0 \Pi^t r = \bar{t}_{\text{дост}}.$$

Б. Если все частные процессы реализуются независимо и параллельно, а целью комплексного процесса является создание не менее чем одного из частных компонентов (дизъюнкция M

частных компонентов), тогда доля «неразработанности» комплексной технологии равна доле непроверенных состояний «комплексной» внешней среды, соответственно

$$1 - L_t^{1-M} = \prod_{m=1}^M (1 - L_t^m) \text{ и}$$

$$(17) \quad E[L_t^{1-M}] = 1 - \prod_{m=1}^M \left[\sum_{k=1}^K p_k^m (1 - p_k^m)^t \right].$$

Среднее время достижения требуемого уровня научения M элементов в этом случае получится

$$(18) \quad \bar{t}_{B \text{ дост}} = \sum_{t=0}^{\infty} \Pr(t_{B \text{ дост}} > t) = \sum_{t=0}^{\infty} \Pr\left(\min_m \{t_{m \text{ дост}}\} > t\right) = \sum_{t=0}^{\infty} \left(\prod_{m=1}^M e_0 \Pi_m^t r_m \right).$$

Если характеристики частных процессов одинаковы, то

$$\bar{t}_{B \text{ дост}} = \sum_{t=0}^{\infty} \left(\prod_{m=1}^M e_0 \Pi_m^t r_m \right) = \sum_{t=0}^{\infty} (e_0 \Pi^t r)^M.$$

Для случаев А и Б параллельной реализации нескольких одинаковых частных процессов получим следующие соотношения:

$$\bar{t}_{A \text{ дост}} = M \bar{t}_{\text{дост}} + \sum_{t=0}^{\infty} \left(\sum_{m=2}^{M-1} C_M^m (-1)^{m-1} (e_0 \Pi^t r)^m \right) + (-1)^{M-1} \bar{t}_{B \text{ дост}},$$

где $\bar{t}_{\text{дост}}$ - среднее время завершения частного процесса, $\bar{t}_{A \text{ дост}}$ - среднее время завершения разработки всех M частных процессов, $\bar{t}_{B \text{ дост}}$ - среднее время завершения хотя бы одного из M частных процессов. Также получим оценки:

$$M \bar{t}_{\text{дост}} - (M-1) \bar{t}_{B \text{ дост}} \leq \bar{t}_{A \text{ дост}} < M \bar{t}_{\text{дост}}.$$

Отметим, что для случаев А и Б параллельной реализации двух одинаковых частных процессов $\bar{t}_{A \text{ дост}} = 2\bar{t}_{\text{дост}} - \bar{t}_{B \text{ дост}}$, т.е. средние времена соотносятся как $\bar{t}_{\text{дост}} = (\bar{t}_{A \text{ дост}} + \bar{t}_{B \text{ дост}}) / 2$.

В. Пусть последовательно выполняются два компонента технологии, причём второй начинается непосредственно после завершения первого. Данный случай описывается двумя независимыми марковскими цепями, вторая цепь стартует из известного состояния в тот момент, когда состояние первой достигло

заданной области. Распределение времени завершения создания такой комплексной технологии (состоящей из двух элементов, когда создание второго может быть начато только после завершения создания первого) - времени достижения заданной области второй цепью - равно свёртке распределений времени обеих цепей. Используя эту закономерность, можно вычислить «интегральную» кривую научения и получить среднее время разработки как сумму средних времён разработки частных технологий.

Г. Пусть создание одного компонента технологии выполняется параллельно и независимо в рамках нескольких (M) процессов с полным обменом информацией. Тогда за один шаг времени производится M независимых проверок, поэтому в этом случае

$$(19) E[L_t^{1/M}] = 1 - \sum_{k=1}^K p_k (1 - p_k)^{Mt}.$$

В этом случае среднее время достижения требуемого уровня научения M элементов равно:

$$(20) \bar{t}_{\text{дост}} = \sum_{t=0}^{\infty} \Pr(t_{\text{дост}} > t) = \sum_{t=0}^{\infty} e_0 \Pi^{Mt} r = e_0 (E - \Pi^M)^{-1} r.$$

Нетрудно показать, что для всех случаев кривых научения – (15), (17) и (19) - справедливы все позиции утверждения 2.

Д. «**Научение научению**». Рассмотрим процесс освоения технологии одновременно с самим ее созданием. То есть когда интенсивность процесса проверок состояний внешней среды меняется соответственно некоторой другой кривой научения.

Пусть процесс $L_t = \sum_{k=1}^K x_{kt} p_k$ характеризует разработку/освоение новой технологии, а процесс $L_t^* = \sum_{j=1}^J x_{jt}^* q_j$ описывает управление технологией создания этой новой технологии («научение научению»). Будем считать процессы L_t и L_t^* статистически независимыми друг от друга.

То есть в каждый момент времени t проверка состояний внешней среды производится с вероятностью L_t^* , а с вероятностью

стью $1 - L_t^*$ момент времени t оказывается «пропущен» для создания технологии: состояние не тестируется и процессы x_{kt} не изменяют состояния⁵:

$$E[x_{kt+1} | x_{kt}] = \begin{cases} x_{kt} & \text{с вероятностью } 1 - \sum_{j=1}^J x_{jt}^* q_j, \\ x_{kt} + (1 - x_{kt}) p_k & \text{с вероятностью } \sum_{j=1}^J x_{jt}^* q_j, \end{cases}$$

где символическая запись $E[\cdot | x_{kt}]$ означает условные математические ожидания при известных значениях x_{kt} .

Тогда $E[x_{kt+1} | x_{kt}] = x_{kt} + \Gamma_t p_k (1 - x_{kt})$, где

$$\Gamma_t = E\left[\sum_{j=1}^J q_j x_{jt}^*\right] = 1 - \sum_{j=1}^J q_j (1 - q_j)^t.$$

Переходя от условных математических ожиданий к безусловным, получим разностное уравнение, позволяющее последовательно вычислять $E[x_{kt}]$ для всех $t \geq 0$

$$(21) \quad E[x_{kt+1}] = E[x_{kt}] + \Gamma_t p_k (1 - E[x_{kt}]).$$

Обозначим через $\Psi_t = 1 - E[x_{kt}]$ (или $E[x_{kt}] = 1 - \Psi_t$), тогда с учётом введённых обозначений

$$E[x_{kt+1}] = 1 - \Psi_{t+1} = \Psi_t \Gamma_t p_k + 1 - \Psi_t, \quad \Psi_{t+1} = \Psi_t (1 - \Gamma_t p_k).$$

Так как $\Psi_0 = 1$, то получаем $\Psi_t = \prod_{\tau=0}^{t-1} (1 - \Gamma_\tau p_k)$, и в результате:

$$E[x_{kt}] = 1 - \prod_{\tau=0}^{t-1} (1 - \Gamma_\tau p_k) = 1 - \prod_{\tau=0}^{t-1} \left(1 - p_k + p_k \sum_{j=1}^J q_j (1 - q_j)^\tau \right).$$

⁵ Можно эту же модель интерпретировать как то, что проверки выполняются в каждый момент времени, а технологию для нового состояния вырабатывается с некоторой вероятностью, определяемой некоторым "мета-процессом". В логистической модели эта вероятность равна уровню научения в самом процессе, в гиперболической - вероятности «ошибки» в некоторой степени с k -том пропорциональности μ .

Тогда итоговая кривая научения:

$$(22) \quad E[L_t] = \sum_{k=1}^K p_k E[x_{kt}] = \sum_{k=1}^K p_k \left(1 - \prod_{\tau=0}^{t-1} (1 - \Gamma_{\tau} p_k) \right) = \\ = 1 - \sum_{k=1}^K p_k \prod_{\tau=0}^{t-1} \left(1 - p_k + p_k \sum_{j=1}^J q_j (1 - q_j)^{\tau} \right).$$

Изучим выражение (21). Вычислив первые разности: $\Delta E[x_{kt}] = \Gamma_t p_k \Psi_t$, заметим, что для любых значений p_k $\Delta E[x_{kt}]|_{t=0} = 0$, так как $\Gamma_0 = 0$, а также $\Delta E[x_{kt}] > 0$ для любых $t > 0$. То есть $E[x_{kt}]$ растёт при $t > 1$, что интуитивно очевидно.

Вычислим и исследуем вторые разности:

$$(23) \quad \Delta^2 E[x_{kt}] = \Delta E[x_{kt+1}] - \Delta E[x_{kt}] = \Gamma_{t+1} p_k (\Psi_t - \Psi_t \Gamma_t p_k) - \Gamma_t p_k \Psi_t = \\ = \Gamma_{t+1} p_k (1 - \Gamma_t p_k) \Psi_t - \Gamma_t p_k \Psi_t = p_k \Psi_t (\Gamma_{t+1} (1 - \Gamma_t p_k) - \Gamma_t).$$

Прежде всего, справедливо $\Delta^2 E[x_{kt}]|_{t=0} = p_k \Gamma_1 > 0$. Однако Γ_t монотонно растёт от 0, асимптотически приближаясь к 1 при $t \rightarrow \infty$. Тогда $\Delta^2 E[x_{kt}]|_{t \rightarrow \infty} = -p_k^2 \Psi_t < 0$ и $\Delta^2 E[x_{kt}]|_{t \rightarrow \infty} \rightarrow 0-$.

Так как кривая научения $E[L_t] = \sum_{k=1}^K p_k E[x_{kt}]$ является линейной комбинацией процессов $E[x_{kt}]$ со строго положительными коэффициентами, то для первых и вторых разностей кривой научения справедливы утверждения, сформулированные для $E[x_{kt}]$. А именно:

- $E[L_t]|_{t=0} = 0$;
- $\Delta E[L_t]|_{t=0} = 0$ и $\Delta E[L_t] > 0$ для всех $t > 0$, т.е. кривая научения растёт по t от нуля, асимптотически приближаясь к единице;
- $\Delta^2 E[L_t]|_{t=0} > 0$, $\Delta^2 E[L_t]|_{t \rightarrow \infty} < 0$ и $\Delta^2 E[L_t]|_{t \rightarrow \infty} \rightarrow 0-$, то есть для t , меньших некоторого значения (*точки перегиба*), кривая научения строго выпукла, а при больших t – строго вогнута.

Рассмотрим пример «научения научению». Пусть имеются равномерные распределения возможных состояний внешней

среды $P = \{p_k = 1/K; k = \overline{1, K}\}$, и $Q = \{q_j = 1/J; j = \overline{1, J}\}$.

Тогда, обозначив $\eta = J^{-1}$, получим:

$$G_t = E \left[\sum_{j=1}^J q_j \mathcal{R}_{jt} \right] = 1 - \sum_{j=1}^J J^{-1} (1 - J^{-1})^t = 1 - (1 - J^{-1})^t = 1 - (1 - \eta)^t.$$

Соответствующая кривая научения имеет вид:

$$(24) \quad E[L_t] = 1 - \prod_{\tau=0}^{t-1} (1 - \delta(1 - (1 - \eta)^\tau)).$$

Используя выражение (17) для вторых разностей, можно получить оценку точки перегиба t^* кривой научения (18) для «сложных» технологий (для которых $K \gg 1$, $J \gg 1$ и $K < J$):

$$t^* \approx \sqrt{KJ + 0,25K^2} - 0,5K.$$

Запишем (24) в виде $E[L_t] = 1 - \exp \left[\sum_{\tau=0}^{k-1} \ln(1 - \delta \exp(-\varphi\tau)) \right]$,

где $\varphi = \ln(1 + \frac{1}{J-1})$.

$$\begin{aligned} E[L_t] &= E[L_{t-1}] + \delta(1 - (1 - \eta)^{t-1}) \prod_{\tau=0}^{k-2} (1 - \delta(1 - (1 - \eta)^\tau)) = \\ &= E[L_{t-1}] + (1 - E[L_{t-1}]) \delta(1 - (1 - \eta)^{t-1}) = \beta_t + (1 - \beta_t) E[L_{t-1}], \end{aligned}$$

где $\beta_t = \delta(1 - (1 - \eta)^{t-1}) = \delta(1 - \exp(-\varphi(t-1)))$. Т.е. кривая научения (24) является «обобщением» кривой научения (2), в котором коэффициенты $\{\beta_t\}$ соответствующего разностного уравнения зависят от номера шага (времени).

Рассмотрим несколько частных случаев научения научению, а именно процессы, в которых вероятность успешной разработки технологии для каждого из впервые встретившихся состояний природы не равна тождественно единице и зависит не от состояния аналогичного «внешнего» процесса (см. выражение (22)), а от уже достигнутого уровня научения в самом процессе. Причём зависимость эта может быть как возрастающей (см. ниже модель логистической кривой научения), так и убывающей (см. ниже модель гиперболической кривой научения).

Соответствующий класс процессов научения назовем условно *автонаучением*.

Логистическая кривая научения. Рассмотрим важный частный случай «научения научению»: в достаточно большом числе практических ситуаций интенсивность процесса проверок состояний внешней среды пропорциональна уровню научения $\Gamma_t = \mu E[L_t]$. Действительно, практически всегда при испытаниях новой, например, авиакосмической или транспортной техники, или вводе в эксплуатацию новых производственных комплексов на первых этапах изделие или комплекс испытывается при ограниченном наборе режимов эксплуатации (стендовые и наземные испытания, работа на холостом ходу и т.д.). По мере накопления опыта диапазон режимов расширяется до полного множества всех возможных режимов и условий внешней среды при переходе к штатному использованию, что соответствует формальному предположению о пропорциональности «скорости научения» достигнутому уровню.

Такая ситуация соответствует частному случаю модели автонаучения - когда процесс разработки технологии L_t^* «совпадает» с процессом её освоения L_t .

Перепишем разностное уравнение (21), в несколько изменённой форме и проанализируем его:

$$(25) \Delta E[x_{kt+1}] = \Gamma_t p_k (1 - E[x_{kt}]).$$

Если все состояния внешней среды равновероятны ($p_k = \delta = 1/K$), то из (25) следует разностное уравнение для уровня научения:

$$\delta \Delta E[x_{kt+1}] = \Gamma_t \delta^2 (1 - E[x_{kt}]) = \Gamma_t \delta^2 - \Gamma_t \delta (\delta E[x_{kt}]).$$

Суммируя данные выражения по k , получим

$$(26) \Delta E[L_{t+1}] = \Gamma_t K \delta^2 - \Gamma_t \delta E[L_t] = \Gamma_t \delta (1 - E[L_t]).$$

В случае $\Gamma_t = \mu E[L_t]$ разностное уравнение для уровня научения примет форму (аналогичный результат для модели с непрерывным временем был получен в [33]):

$$(27) \Delta E[L_{t+1}] = \mu \delta E[L_t] (1 - E[L_t]).$$

Уравнение (27) является разностным аналогом дифференциального уравнения $dx/dt = \beta x (1 - x)$, где $\beta = \mu \delta$, решением

которого является классическая для теории научения (см. обзор в [15]) *логистическая кривая*, в дискретной форме логистическая кривая имеет вид:

$$(28) E[L_t] = \frac{1}{1 + \left(\frac{1}{\lambda} - 1\right) \exp(-\beta t)}.$$

Она монотонно возрастает от $\lambda > 0$ (при $t = 0$) до 1 (при $t \rightarrow +\infty$).

Так как решениями аналогичных разностных и дифференциальных уравнений, как правило, являются разные по виду функции, *логистическая кривая* в дискретной форме (28) не является решением (27) в общем случае. Поэтому исследуем функцию (28) и найдём условия, при которых, функция (28) корректно аппроксимирует решение уравнения (27).

Для упрощения локальных выкладок введём компактное обозначение функции (28) в виде $x_t = \frac{1}{1 + ba^t}$.

Во-первых, покажем, что при неограниченном уменьшении дискрета времени (обозначим его Δt) разностное уравнение, описывающее (28) переходит в дифференциальное вида $dx/dt = \beta x(1 - x)$.

С учётом обозначений $x_{t+\Delta t} = \frac{1}{1 + ba^{t+\Delta t}}$, преобразуем это выражение. При $\Delta t \rightarrow 0$ (на самом деле при $\ln(a)\Delta t = 1$) корректными будут следующие преобразования:

$$\begin{aligned} x_{t+\Delta t} &= \frac{1}{1 + ba^{t+\Delta t}} \approx \frac{1}{1 + ba^t(1 + \ln(a)\Delta t)} = \frac{1}{1 + ba^t + ba^t \ln(a)\Delta t} = \\ &= \frac{1}{1 + ba^t} \frac{1}{1 + ba^t(1 + ba^t)^{-1} \ln(a)\Delta t} \approx \\ &\approx \frac{1}{1 + ba^t} \left(1 - ba^t(1 + ba^t)^{-1} \ln(a)\Delta t\right) = \\ &= \frac{1}{1 + ba^t} - \frac{ba^t}{(1 + ba^t)^2} \ln(a)\Delta t = x_t + x_t(1 - x_t)\ln(a)\Delta t. \end{aligned}$$

То есть $x_{t+\Delta t} = x_t + x_t(1-x_t)\ln(a)\Delta t$ при условии $\ln(a)\Delta t = 1$.

Откуда непосредственно следует $\frac{x_{t+\Delta t} - x_t}{\Delta t} = \ln(a)x_t(1-x_t)$, что доказывает требуемое утверждение.

Очевидно, при $\ln(a) \ll 1$ и $\Delta t = 1$ все преобразования остаются корректными, поэтому разностное уравнение, описывающее логистическую кривую в дискретной форме (28), аппроксимируется выражением вида (27).

В силу введённых промежуточных обозначений $\ln(a) = \beta = \mu \delta = \mu/K$, тогда условию $\ln(a) \ll 1$ соответствует $\mu/K \ll 1$. Таким образом, при «большой» размерности K множества состояний внешней среды логистическая кривая (28) аппроксимируется разностным уравнением (27).

Логистическая кривая научения (28) является хрестоматийной для теории научения [15]. В то же время, для рассматриваемой модели научения научению она является частным случаем, соответствующим равномерному распределению состояний природы и пропорциональности интенсивности процесса проверок состояний внешней среды достигнутому уровню научения при «большой» размерности множества состояний внешней среды.

В случае, когда кривая научения является логистической – см. (28), средняя длина серии имеет вид $N_t = \lambda(1-\lambda)^{-1} \exp(\beta t)$ и порождается весьма компактным разностным уравнением $N_{t+1} = \exp(\beta)T_t$.

Гиперболическая кривая научения. В другом частном случае автонаучения интенсивность процесса проверок состояний внешней среды может уменьшаться по мере роста уровня научения: $G_t = \mu(1 - E[L_t])^a$, где $a > 0$. На практике это может иметь место, например, при ограниченных когнитивных (в том числе ограниченность кратковременной памяти) и/или «вычислительных» возможностях обучающегося субъекта.

Аналогично рассмотренному выше случаю логистической кривой научения для равновероятных состояний внешней среды получим разностное уравнение для уровня научения. Выраже-

ние (26) справедливо и в данном случае; подставив в него $\Gamma_t = \mu (1 - E[L_t])^a$, получим:

$$(29) \Delta E[L_{t+1}] = \Gamma_t \delta (1 - E[L_t]) = \mu \delta (1 - E[L_t])^{1+a}.$$

Уравнение (29) является разностным аналогом дифференциального уравнения $dx/dt = \beta (1 - x)^{1+a}$, где $\beta = \mu \delta$, решением которого является классическая для теории научения (см. обзор в [15] и пионерские работы [32, 39]) *гиперболическая кривая научения* как функция непрерывного времени.

Гиперболическая кривая в дискретной форме имеет вид

$$(30) E[L_t] = 1 - \frac{1}{(1 + a\beta t)^{1/a}}.$$

Она монотонно возрастает от нуля (при $t = 0$) до 1 (при $t \rightarrow +\infty$).

Как и в примере с логистической кривой, введя компактные обозначения $x_t = 1 - \frac{1}{(1 + a\beta t)^{1/a}}$, легко выполнить аналогичные выкладки при $\beta \ll 1$ и $a\beta \ll 1$ и в результате получить $x_{t+1} = x_t + \beta(1 - x_t)^{a+1}$.

Условие $\beta \ll 1$ равносильно $\mu \delta = \mu/K \ll 1$, поэтому гиперболическая кривая является решением разностного уравнения (29) при «большой» размерности K множества состояний внешней среды.

Средняя длина серии в этом случае равна $N_t = (1 + a\beta t)^{1/a} - 1$.

Разностное уравнение для средней длины серии:

$$N_{t+1} = \left[(N_t + 1)^a + a\beta \right]^{1/a} - 1.$$

В частности, при $a = 1$ уравнение имеет вид $N_{t+1} = N_t + \beta$.

Таким образом, гиперболическая кривая научения (30) (её разностный аналог: $E[L_t] = E[L_{t-1}] + \beta (1 - E[L_{t-1}]^{1+a})$) для рассматриваемой модели является частным случаем «научения научению», соответствующим равномерному распределению состояний природы и убыванию интенсивности процесса проверок состояний внешней среды с ростом достигнутого уровня научения.

Автонаучение. Рассмотрим модель автонаучения в непрерывном времени - дифференциальное уравнение для уровня научения $z(t) \in [0; 1], t \geq 0$:

$$(31) z'(t) = \gamma(1 - z)p(z)$$

с начальным условием $z(0) = \lambda \in [0; 1)$, где $\gamma > 0$, $p(\cdot) : [0; 1] \rightarrow (0; A]$ - непрерывная функция, где $0 < A < +\infty$ (если p интерпретируется как вероятность, то $A = 1$).

Из введённых предположений следует, что:

- а) решение уравнения (31) существует и единственно;
- б) зависимость $z(t)$ является строго монотонно возрастающей и $\forall t \geq 0 z'(t) \leq \gamma$;
- в) если $z(0) = 0$, то $\forall t \geq 0 z(t) \leq 1 - \exp(-\gamma A t)$;
- г) зависимость $z(t)$ является замедленно-асимптотической, т.е. $\lim_{t \rightarrow +\infty} z(t) = 1$ и $\lim_{t \rightarrow +\infty} z'(t) = 0$.

Варьируя $p(z)$, можно получать различные кривые автонаучения. Частными случаями являются многие рассмотренные выше классы кривых научения, а именно:

1) *Экспоненциальная кривая* («вырожденный случай» - автонаучение отсутствует: $p(z) \equiv 1$, имеет место обычное научение):

$$\lambda = 0; z'(t) = \gamma(1 - z); z(t) = 1 - \exp(-\gamma t).$$

2) *Логистическая кривая*:

$$p(z) = z, \lambda > 0; z'(t) = \gamma z(1 - z); z(t) = \frac{1}{1 + (\frac{1}{\lambda} - 1) \exp(-\gamma t)}.$$

3) *Гиперболическая кривая*:

$$p(z) \equiv (1 - z)^a, a > 0, \lambda = 0; z'(t) = \gamma(1 - z)^{1+a};$$

$$z(t) = 1 - \frac{1}{(1 + a \gamma t)^{1/a}}.$$

Модель автонаучения (31) допускает ряд расширений - переход к моделям научения в процессе работы и к моделям группового научения.

Рассмотрим модель *научения в процессе работы* (Learning-by-Doing), в рамках которой обучаемый субъект (агент) может выбирать интенсивность $w(t) \geq 0$ своей деятельности (объем работы, выполняемый в единицу времени, число состояний

природы, анализируемых в единицу времени, и т.д.). Объем выполненных работ $W(t) = \int_0^t w(\tau) d\tau$ можно условно считать тем «опытом», который накопил агент, его «эффективным внутренним временем» [16].

Подставляя в (31) вместо $p(z)$ интенсивность $w(t)$, получим дифференциальное уравнение

$$(32) \quad z'(t) = \gamma(1 - z)w(z),$$

решением которого является «экспоненциальная» кривая научения

$$(33) \quad z_w(t) = 1 - \exp(-W(t)).$$

Если считать, следуя [16], что (33) – вероятность достижения результата в момент времени t (доля успешных действий агента), то кумулятивный ожидаемый результат будет определяться следующим выражением

$$(34) \quad W_+(t) = \int_0^t z_w(\omega)w(\omega)d\omega = \int_0^t 1 - \exp\left(-\int_0^\omega w(\tau)d\tau\right)w(\omega)d\omega.$$

Если задан интервал времени $T \geq 0$ и максимальный объем работ W_0 , которые может выполнить агент, то из (32)–(34) получаем, что задача максимизации ожидаемого результата примет вид задачи динамического программирования

$$W_+(T) \rightarrow \max_{w(\cdot), W(T) \leq W_0}.$$

Аналогичные задачи (в том числе с учётом затрат агента и т.п.), трактуемые как задачи об оптимальной стратегии научения агента, рассматриваются в [7,16].

В заключение настоящего раздела опишем в терминах автонаучения процесс *группового научения* [16].

До сих пор при рассмотрении научения агента считалось, что агент учится только «на собственном опыте». Тем не менее в коллективах имеет место обмен опытом, и агенты, наблюдая за деятельностью других (их успехами и трудностями), могут также приобретать опыт (см. модели в [16]). Для того чтобы отразить этот эффект, можно считать, что «опыт» p , накопленный агентом, зависит от уровней научения других агентов.

Пусть имеется n агентов. Обозначая через i номер агента, через z_i – уровень его научения, через $\mathbf{z} = (z_1, z_2, \dots, z_n)$ – вектор уровней научения, запишем для каждого из агентов аналог уравнения (31):

$$(35) \quad z'_i(t) = \gamma_i(1 - z_i)p_i(\mathbf{z}), \quad i = 1, 2, \dots, n.$$

Влияние агентов друг на друга в модели (35) может быть различным:

- если $\frac{\partial p_i(\mathbf{z})}{\partial z_j} > 0$, то i -й агент перенимает опыт j -го агента;
- если $\frac{\partial p_i(\mathbf{z})}{\partial z_j} < 0$, то опыт j -го агента «мешает» i -му агенту;
- если $\frac{\partial p_i(\mathbf{z})}{\partial z_j} \equiv 0$, то приобретение опыта j -м агентом не влияет на i -го агента.

Для модели (35) задача о оптимальном совместном научении группы агентов может ставиться и решаться по аналогии с тем, как это делается для дискретной модели в [16].

8. Заключение

Таким образом, в настоящей работе рассмотрены модели формирования и освоения технологии комплексной деятельности. В том числе:

- получена базовая кривая научения в ходе разработки технологии (выражение (1)) и исследованы ее свойства (утверждения 1 и 2);
- для ряда практически важных частных случаев получены аппроксимации кривой научения (выражения (3) и (6));
- получена оценка (7) среднего времени достижения требуемого уровня научения и исследованы его свойства (утверждение 3);
- предложены и исследованы:
 - модели комплексирования частных компонентов технологии (выражения (15), (17), (19), (21), (25), (26) и (16), (18), (20), описывающие кривые научения и оценки средних времен соответственно);

- модели научения научению и автонаучения (раздел 7).

Следует отметить, что классические для теории научения (см. обзор в [15]) экспоненциальная кривая (2), а также логистическая (27) и гиперболическая (30) кривые являются частными (для рассматриваемых моделей) случаями предложенной модели автонаучения.

Дальнейших исследований и осмысления, вероятно, заслуживает тот факт, что введённый в работе показатель «средняя длина серии» в случаях канонических для теории научения кривых описывается линейными разностными уравнениями:

- для экспоненты $N_{t+1} = \exp(\gamma) N_t + (\exp(\gamma) - 1)$;
- для логисты $N_{t+1} = \exp(\beta) N_t$;
- для гиперболы первой степени $N_{t+1} = N_t + \beta$.

Перспективными представляются постановка и решение задач управления формированием и развитием технологий на базе предложенных моделей последних. Некоторые из таких задач управления рассматриваются в [7].

Литература

- 1 АЛЕКСАНДРОВСКАЯ Л.Н., КРУГЛОВ В.И., КУЗНЕЦОВ А.Г., ШОЛОМ А.М. *Теоретические основы испытаний и экспериментальная отработка сложных технических систем*. Учебное пособие. – М: Логос, 2003. – 735 с.
- 2 АТКИНСОН Р., БАУЭР Г., КРОТЕРС Э. *Введение в математическую теорию обучения*. – М.: Мир, 1969. – 468 с.
- 3 БЕЙЗЕР Б. *Тестирование чёрного ящика. Технологии функционального тестирования программного обеспечения и систем*. – СПб.: Питер, 2004. – 318 с.
- 4 БЕЛОВ М.В. *Проблемы управления жизненными циклами организационно-технических систем // Управление большими системами*. – 2018. Вып. 76. – С. 117–172.
- 5 БЕЛОВ М.В., НОВИКОВ Д.А. *Проблемы управления технологией комплексной деятельности организационно-технических систем // Проблемы теории и практики управления*. – 2018. – №7. – С. 26–42..

- 6 БЕЛОВ М.В., НОВИКОВ Д.А. *Методология комплексной деятельности*. – М.: Ленанд, 2017. – 320 с.
- 7 БЕЛОВ М.В., НОВИКОВ Д.А. *Модели управления технологией комплексной деятельности* // Управление большими системами. – 2019. – (принята к публикации).
- 8 БОНДАРЕНКО И.Б., ИВАНОВ А.И. *Организационная модель многоагентной системы извлечения знаний из распределённых гетерогенных баз данных систем автоматизированного проектирования* // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2015. – №4(36). – С. 54–63.
- 9 БУШ Р., МОСТЕЛЛЕР Ф. *Стохастические модели обучаемости*. – М.: Гос. изд-во физ.-мат. лит., 1962. – 483 с.
- 10 ВАЛЬЧУК А.С. *Разработка математической модели автоматического извлечения знаний для гибридной вопросно-ответной системы* // Вестник КРАУНЦ. Физ.-мат. науки. – 2017. – №2(18). – С. 76–80.
- 11 ДРОЗДОВ В.И., БОЙЦОВА Е.А., НОВИКОВ Ю.М. *Исследование качества тестов с использованием модели Раша*. – URL: <https://cyberleninka.ru/article/n/issledovanie-kachestva-testov-s-ispolzovaniem-modeli-rasha> (дата доступа: 23.01.2018).
- 12 КОВАРЦЕВ А.Н., ПОПОВА-КОВАРЦЕВА Д.А., СЕРПОВСКАЯ Е.Е. *Тестирование математических моделей вычислительных алгоритмов на основе метода глобальной оптимизации* // Сборник трудов конференции «Информационные технологии и нанотехнологии», Самара, 29 июня – 01 июля 2015 г. – С. 191-196.
- 13 КОЧУРА С.Г., КУЗНЕЦОВ Н.А., НОСЕНКОВ А.А. *О математическом моделировании электрических испытаний космических аппаратов связи* // ИЗВ. ВУЗОВ. ПРИБОРОСТРОЕНИЕ. – 2011. – Т. 54, №4. – С. 43–47.
- 14 *Краткий психологический словарь*. – М.: ИПЛ, 1985. – 201 с.
- 15 НОВИКОВ Д.А. *Закономерности итеративного научения*. – М.: Институт проблем управления РАН, 1998. – 96 с.

- 16 НОВИКОВ Д.А. *Модели обучения в процессе работы* // Управление большими системами. – 2007. – №19. – С. 5–22.
- 17 САМОЛЮК Н.Г. *Современные средства оценивания результатов обучения. Конспект лекций.* – ФГБОУ «Томский государственный педагогический университет». – URL: http://koi.tspu.ru/koi_books/samolyuk/ (дата доступа: 23.01.2018).
- 18 СТАРОЛЕТОВ С.М., КРЮЧКОВА Е.Н. *Проведение on-line тестирования программного обеспечения на основе построенной модели* // ПОЛЗУНОВСКИЙ ВЕСТНИК. – 2010. – №2. – С. 212–216.
- 19 ЧЕЛЫШКОВА М.Б. *Теория и практика конструирования педагогических тестов.* Учебное пособие. – М.: Логос, 2002. - 432 с.
- 20 ШИРЯЕВ А.Н. *Статистический последовательный анализ. Оптимальные правила остановки.* - М.: Физматлит, 1976. – 272 с.
- 21 AICHERNIG B.K., SCHUMI R, *Statistical Model Checking Meets Property-Based Testing* // IEEE Int. Conf. on Software Testing, Verification and Validation – 2017 (ICST-2017), Tokyo. – P. 390– 400. – DOI: 10.1109/ICST.2017.42.
- 22 ALAGOZ HERPEL T., GERMAN R. *A Selection Method for Black Box Regression Testing with a Statistically Defined Quality Level* // IEEE Int. Conf. on Software Testing, Verification and Validation – 2017 (ICST-2017), Tokyo. – P. 114–125. – DOI: 10.1109/ICST.2017.18.
- 23 ALI S., YUE T. *U-Test: Evolving, Modelling and Testing Realistic Uncertain Behaviours of Cyber-Physical Systems* // IEEE 8th Int. Conf. on Software Testing, Verification and Validation – 2015 (ICST-2015), Graz. – P. 1–2. – DOI: 10.1109/ICST.2015.7102637/
- 24 BOURQUE P., FAIRLEY R.E.(eds.) *Guide to the Software Engineering Body of Knowledge, Version 3.0* // IEEE Computer Society, 2014. – URL: www.swebok.org.
- 25 Business Process Model and Notation (BPMN), v2.0.2. – URL: <http://www.omg.org/spec/BPMN/2.0>.

- 26 CHUI D.K.Y., WONG A.K.C. *Synthesizing Knowledge: A Cluster Analysis Approach Using Event Covering* // IEEE Trans. on Systems, Man, and Cybernetics. – March 1986. – Vol. 16, No. 2. – P. 251–259. – DOI: 10.1109/TSMC.1986.4308945.
- 27 CRAWFORD J. *Learning Curve, Ship Curve, Ratios, Related Data* / Lockheed Aircraft Corporation. – 1944. – P. 122 – 128.
- 28 EBBINGHAUS H. *Über das Gedächtnis*. - Leipzig: Dunker, 1885. – 168 p.
- 29 HENDERSON B. *The Application and Misapplication of the Learning Curve* // J. of Business Strategy. – 1984. – Vol. 4. – P. 3–9.
- 30 HENRIQUES D., MARTINS J.G., ZULIANI P., PLATZER A., CLARKE E.M. *Statistical Model Checking for Markov Decision Processes* // Int. Conf. on Quantitative Evaluation of Systems, (QEST), London, United Kingdom, 2012. – P. 84–93. – DOI: 10.1109/QEST.2012.19.
- 31 HULL C.L. *Principles of Behavior and Introduction to Behavior Theory*. - New York: D. Appleton Century Company, 1943. – 422 p.
- 32 LEGAY A., DELAHAYE B., BENSALÉM S. *Statistical Model Checking: An Overview* // In: Barringer H. et al. (eds) Runtime Verification. RV 2010. Lecture Notes in Computer Science, Vol. 6418. – Springer, Berlin, Heidelberg, 2010.
- 33 LEIBOWITZ N., BAUM B., ENDEN G., KARNIEL A. *The Exponential Learning Equation as a Function of Successful Trials Results in Sigmoid Performance* // J. of Mathematical Psychology. – 2010. – Vol. 54. – P. 338–340.
- 34 MIKAMI S., KAKAZU Y. *Extended stochastic reinforcement learning for the acquisition of cooperative motion plans for dynamically constrained agents* // Proc. of IEEE Systems Man and Cybernetics Conference (SMC), Le Touquet, 1993. – Vol. 4. – P. 257–262. – DOI: 10.1109/ICSMC.1993.390719.
- 35 ORSEAU L., LATTIMORE T., HUTTER M. *Universal Knowledge-Seeking Agents for Stochastic Environments* // In: Jain S., Munos R., Stephan F., Zeugmann T. (eds) Algorithmic Learning Theory. ALT 2013. Lecture Notes in Computer Science, vol. 8139. – Springer, Berlin, Heidelberg, 2013.

- 36 PATRICK M., DONNELLY R., GILLIGAN C.A. *A Toolkit for Testing Stochastic Simulations against Statistical Oracles* // IEEE Int. Conf. on Software Testing, Verification and Validation (ICST), Tokyo, 2017. – P. 448–453. – DOI: 10.1109/ICST.2017.50.
- 37 STOCIA G., STACK B. *Acquired Knowledge as a Stochastic Process* // *Surveys in Mathematics and its Applications*. – 2017. – Vol. 12. – P. 65–70. – URL: <http://www.utgjiu.ro/math/sma>
<http://www.utgjiu.ro/math/sma>
- 38 TOLMAN E.C. *Theories of Learning* // *Comparative Psychology*. Ed. Moss F.A. Chapter 12. - New York: Prentice Hall, 1934. – P. 232 – 254.
- 39 THURSTONE L.L. *The Learning Curve Equation* // *Psychol. Monogr.* – 1919. – Vol. 26, No. 3. – P. 1–51.
- 40 THURSTONE L.L. *The Learning Function* // *J. of General Psychology*. – 1930. – No. 3. – P. 469–493.
- 41 VAN DER LINDEN W.J., HAMBLETON R.K. *Handbook of Modern Item Response Theory* – Springer Science & Business Media, 1996. – P. 512.
- 42 WRIGHT T. *Factors Affecting the Cost of Airplanes* // *J. of Aeronautical Sciences*. – 1936. – Vol. 3(4). – P. 122–128.
- 43 WONG A., WANG Y. *Pattern Discovery: A Data Driven Approach to Decision Support* // *IEEE Trans. on Systems, Man, and Cybernetics*. – February 2003. – Vol. 33, No. 1. – P. 114–124.
- 44 YIZHEN CHEN et al. *Effective Online Software Anomaly Detection* // *Proc. of the 26th ACM SIGSOFT Int. Symposium on Software Testing and Analysis (ISSTA-2017)*, Santa Barbara, CA, USA, July 10–14, 2017. – P. 136–146.

MODELS OF DESIGNING AND LEARNING A TECHNOLOGY OF COMPLEX ACTIVITY

Mikhail Belov, IBS, Moscow, Cand.Sc. (mbelov59@mail.ru),
Dmitry Novikov, Institute of Control Sciences, Moscow, Doct. of
Sc. (novikov@ipu.ru).

Abstract: Presented mathematical models are based on previous studies of the problems of managing organizational and technical systems and their complex activities, executed by the authors. The problem of developing and / or mastering the technology of complex activity is formalized in the form of a mathematical model, which generalizes of probabilistic learning models. The properties of the process of developing and / or mastering technology (learning) are studied, the convergence of the process to the state of full technology mastering is shown, analytical expressions of the characteristics of the models are obtained - the average time to reach a given level of mastering. The models of learning that describe the integration of elements of technology - conjunctive, disjunctive and parallel development of technology are proposed. Developed and studied models of learning in the process of work and group learning, including the "learning to learn" model - when the intensity of the learning process depends on the learning level achieved. For all models of integration, analytical expressions for learning levels are obtained. The asymptotic case of models during the transition to continuous time is investigated. It is shown that special cases of the proposed model are models of exponential, hyperbolic, and logistic learning curves, which are widespread in the theory of learning, theory of system testing, software testing, and related branches of knowledge.

Keywords: technology, complex activity, learning curve.

УДК 004.827

ББК 32.81 87 65.05 72.4

DOI: <https://doi.org/10.25728/ubs.2019.77.7>

*Статья представлена к публикации
членом редакционной коллегии Г.А. Угольником.*

*Поступила в редакцию 10.07.2018.
Опубликована 31.01.2019.*

НЕЧЁТКАЯ ОЦЕНКА ЭФФЕКТИВНОСТИ ТРАНСФЕРНОЙ СИСТЕМЫ АВИАПЕРЕВОЗОК

Гужа Е. Д.¹, Романенко В. А.², Скороход М. А.³

(Самарский национальный исследовательский
университет им. академика С.П. Королева, Самара)

Рассматривается трансферная система пассажирских воздушных перевозок в составе узлового аэропорта (хаба) и хабообразующей авиакомпании. Эффективность системы характеризуется уровнем её доходов от перевозки трансферных пассажиров при оптимально сформированном расписании. Предполагается, что некомфортное время ожидания пересадки способно вызвать отток с рейсов хабообразующей авиакомпании пассажиров, потенциально готовых к перевозке через рассматриваемый хаб, при этом минимизировать ущерб, причиненный оттоком пассажиров, позволяет оптимизация расписания хаба. Предлагаемая модель оценки эффективности проектируемой трансферной системы включает в качестве ключевого компонента задачу оптимизации расписания, предполагающую определение таких моментов прибытия и отправления рейсов в хабе, которые максимизируют доходы от перевозки трансферных пассажиров с учетом ограничений по пропускной способности производственных комплексов аэропорта. Неопределенность в исходных данных модели, характерная для этапа проектирования трансферной системы, представлена в нечетко-множественной форме. Модель может быть использована для поддержки принятия решения об экономической целесообразности и технической возможности формирования трансферной системы на базе определенных авиакомпании и аэропорта, для решения задач оптимизации расписания и формирования сбалансированного парка технологических ресурсов узлового аэропорта, рациональной организации работы его производственных комплексов, оперативного управления ресурсами. Рассмотрены примеры применения модели.

Ключевые слова: параметрическая оптимизация, нечеткие числа, узловой аэропорт, показатель эффективности.

1. Общая постановка задачи

Рассматривается трансферная система пассажирских воздушных перевозок, которая в качестве элементов включает узловой аэропорт (хаб), обеспечивающий массовые пересадки трансферных пассажиров, и хабообразующую авиакомпанию,

¹ Екатерина Дмитриевна Гужа, аспирант (e.guzha@ya.ru).

² Владимир Алексеевич Романенко, к.т.н., доцент (vla_rom@mail.ru).

³ Маргарита Артушевна Скороход, аспирант (skoro-margarita@yandex.ru).

организирующую и осуществляющую их подвоз в хаб. Формирование системы позволяет её организаторам повысить экономическую эффективность перевозочной деятельности за счет привлечения в хаб дополнительных пассажиропотоков. Трансферные системы активно действуют и развиваются на зарубежном авиатранспортном рынке. Предпринимаются отдельные попытки их создания отечественными авиаперевозчиками, в том числе на базе сравнительно небольших региональных аэропортов, характеризующихся ограниченной пропускной способностью.

Решение о формировании трансферной системы на базе определенных авиакомпании и аэропорта, действующих в рамках «традиционной» схемы (с преобладанием обслуживания прямых пассажиропотоков), должно приниматься по результатам предварительного технико-экономического анализа, содержащего оценку эффективности системы, проектируемой на основе аэропорта заданной, нередко невысокой, пропускной способности. Результаты такой оценки позволяют обоснованно выбрать наилучший вариант проектируемой системы; решить, можно ли при её формировании ограничиться не требующими серьезных расходов шагами по реформированию организации и технологии выполнения и обслуживания перевозок или необходимы более затратные меры; каков уровень этих затрат, является ли он приемлемым с точки зрения организаторов системы.

При выборе наиболее значимого показателя эффективности трансферной системы учтем, что привлекательность хаба в роли пересадочного пункта обеспечивается, если хабообразующей авиакомпании удаётся предложить пассажирам перевозку со сбалансированным временем пребывания в хабе, исключающем неприемлемо длительное ожидание, но при этом достаточным для гарантированной пересадки на стыковочные рейсы. Стыковка рейсов определяется как такое совмещение времени прибытия одного рейса и времени отправления другого, которое позволяет трансферному пассажиру определенной авиасвязи пересесть с прибывшего рейса на отправляющийся, успев пройти все необходимое обслуживание в хабе. Под авиасвязью понимается пара аэропортов, между которыми имеются регулярные воздушные линии, обеспечивающие возможность воздушной перевозки. Трансферная авиасвязь – это авиасвязь, предпо-

лагающая наличие стыковки рейсов в аэропорту трансфера. Для увеличения числа стыковок хабообразующими авиакомпаниями практикуется координация расписания движения самолетов, благодаря которой поток прибывающих и отправляемых рейсов в хабе принимает волнообразный характер [8, 12, 13]. Каждая волна начинается массовым прибытием рейсов в хаб с последующей пересадкой трансферных пассажиров на стыковочные рейсы, а завершается таким же массовым отправлением рейсов из хаба.

Логично предположить, что, несмотря на координацию расписания, некоторые из потенциальных пассажиров, принципиально готовых к трансферной перевозке по опубликованному тарифу, сочтут время пребывания в хабе некомфортным и поэтому откажутся от услуг рассматриваемой трансферной системы. По этой причине число действительных трансферных пассажиров, т.е. тех трансферных пассажиров, которые не откажутся от перевозки, найдя время пребывания в хабе приемлемым, будет в общем случае меньше числа потенциальных пассажиров. Это означает, что величина C^D доходов, полученных системой от перевозки действительных трансферных пассажиров в течение некоторой волны прибытий-отправлений, будет в общем случае меньше величины C^P доходов, которые могла бы получить система от перевозки всех потенциальных пассажиров в течение той же волны. Повысить величину C^D , приблизить её к C^P , позволяет оптимизация параметров расписания, состоящая в определении таких моментов времени прибытия и отправления рейсов в рамках каждой волны, которые максимизируют C^D . При оптимизации должны учитываться ограничения по пропускной способности хаба, влияющие на продолжительность наземного обслуживания воздушных судов (ВС), пассажиров, багажа, а значит, – и на время пребывания трансферных пассажиров в хабе, и на число действительных трансферных пассажиров, и на доходы C^D . Полагая, что организаторы трансферной системы безусловно заинтересованы в максимизации величины C^D и готовы к использованию в рамках системы оптимального по данному критерию расписания, далее под C^D будем понимать полученное в результате оптимизации максимальное значение доходов от перевозки действительных трансферных пассажиров

в течение рассматриваемой волны. Таким образом, правомерно считать, что эффективность проектируемой трансферной системы будет тем выше, чем выше будет величина C^D , чем ближе она будет к C^P .

Можно предположить, что использование показателя C^D будет целесообразно в условиях, когда авиакомпания планирует начать массовые трансферные перевозки через аэропорт, из которого ранее ей уже выполнялись перевозки по «традиционной» схеме. В этом случае сравнение планируемых доходов C^D с доходами предыдущих периодов позволит принять решение о целесообразности организации трансферной системы. Если же организация трансферных перевозок планируется авиакомпанией на базе малознакомого аэропорта, перевозки из которого ей ранее не выполнялись, то более разумным представляется выбор показателя в форме C^D/C^P , близость значения которого к единице будет свидетельствовать о высокой эффективности системы. Так или иначе, проблема оценки эффективности трансферной системы сводится к определению значений показателей C^D или C^D/C^P , характеризующих уровень доходов, на которые могут рассчитывать её организаторы в случае использования оптимально скоординированного расписания, формируемого с учетом ограниченных технических возможностей узлового аэропорта и предпочтений трансферных пассажиров в отношении времени пребывания в нём.

Из изложенного выше следует, что смысловым ядром проблемы оценки эффективности трансферной системы является задача оптимизации параметров расписания узлового аэропорта. Данная задача решена в [4] на основе чрезвычайно упрощенного технологического графика наземного обслуживания ВС и при игнорировании ограничений на пропускную способность производственных комплексов хаба, что затрудняет использование результатов для оценки возможности построения трансферной системы на основе аэропорта ограниченной пропускной способности. Чтобы сделать такую оценку возможной, в настоящей работе предложен подход, учитывающий численность и производительность ресурсов аэропорта, осуществляющих обслуживание ВС, пассажиров и багажа согласно технологическому графику, близкому к используемым в аэропортовой практике.

В исходных данных рассмотренного в [4] варианта постановки задачи были различимы два вида неопределенности – нечеткость и случайность. Громоздкость вычислительных операций над частотными распределениями стала причиной рассмотрения лишь очень упрощенного технологического графика. В настоящей работе все неопределенные величины сведены к единой, нечеткой, форме представления, что позволило детализировать модель аэропортовых процессов при существенном упрощении вычислительных процедур. Кроме того, наличие в модели только нечетких величин позволяет использовать её для решения задачи оценки эффективности трансферной системы в условиях существенной неопределенности исходных данных, присущих, например, случаю открытия хабообразующей авиакомпанией массовых трансферных перевозок через аэропорт, из которого ранее её самолетами рейсы не выполнялись.

Следует отметить явный недостаток исследований, посвященных формированию оптимального расписания трансферной системы. Опубликованные к настоящему времени работы, в основном англоязычные, касающиеся вопросов исследования расписания аэропортов, в частности, узловых, условно распределяются на две группы, первая из которых [8, 9, 10] посвящена методам анализа уровня координации заданного расписания хаба, в том числе с учетом влияния нарушений регулярности движения самолетов [17, 18]. В сравнительно многочисленных работах второй группы решаются задачи, относящиеся к категории задач теории расписаний с учетом ограничений по ресурсам, для отдельных технологических операций и типов аэропортового оборудования и механизации, например, заправки самолетов топливом [5], регистрации [6], таможенного и пограничного контроля [15] пассажиров, терминальной обработки грузов [16]. Методы теории нечетких множеств пока не находят широкого применения при решении теоретических авиатранспортных задач. В этой связи следует упомянуть работы [13, 14], в первой из которых в нечеткой постановке решена задача оптимизации пропускной способности элементов аэродрома, во второй описана модель оперативного управления производственным процессом одного из зарубежных аэропортов, использующая нечеткую информацию.

2. Учет неопределенности исходных данных

Набор основных исходных данных рассматриваемой задачи, являющихся неопределенными, включает уровни потенциальных пассажиропотоков, величины тарифов на авиасвязях, временные характеристики технологических операций наземного обслуживания и численность ресурсов, выделяемых аэропортом для их выполнения. Источниками неопределенности этих характеристик являются, во-первых, стохастичность перевозочных процессов и конъюнктуры авиатранспортного рынка и, во-вторых, недостаток знаний о проектируемой трансферной системе, принципиально непреодолимый априори, до начала её функционирования.

Неопределенность уровней пассажиропотоков и тарифов в первую очередь обусловлена причинами, связанными с недостатком знаний о системе, принципиальными трудностями в сборе статистических данных. Подробно эти причины разобраны в [4], где к ним отнесены динамичное изменение сети авиасвязей, влияние разнообразных краткосрочных маркетинговых акций и, главное, проблемы с прогнозированием на этапе составления расписания, т.е. до начала выполнения перевозок.

Неопределенность параметров технологических процессов, протекающих в условиях установившегося режима работы аэропорта, напротив, носит в большей степени стохастический характер. Причинами служат нарушения расписания по метеоусловиям, разнообразные отклонения и сбои в технологическом графике обслуживания рейсов, отказы техники и оборудования и др. Однако на этапе проектирования трансферной системы на первый план выходят факторы, связанные с недостатком знаний. Так, неопределенность знания о прогнозной численности пассажиров и багажа затрудняет или делает невозможным использование статистики, характеризующей технологические операции их обслуживания в аэропорту. К аналогичным последствиям приводит неоднозначность в тактике авиаперевозчика по отношению к трансферным пассажирам опаздывающих рейсов: авиакомпанией могут практиковаться задержки отправления рейсов с целью ожидания пассажиров, пересеживающихся с опаздывающих прибывающих рейсов, причем на одних авиасвя-

зях такое ожидание может осуществляться, а на других игнорироваться. Априори неопределенным и трудноформализуемым является подход авиакомпании к использованию собственного авиапарка при реализации расписания трансферной системы. На этапе её проектирования скорее всего еще не будет известно, насколько «напряжен» график оборота самолётов авиакомпании, насколько продолжительные простои самолетов в хабе он допускает. Отсутствие необходимых знаний станет наиболее явным в случае, когда авиакомпания начинает выполнять полеты из аэропорта, из которого ранее этого не делала. Таким образом, на этапе проектирования трансферной системы, когда и предполагается решение задачи оценки её эффективности, правомерным, а в ряде случаев и единственно возможным, источником любой исходной информации становится экспертный опрос. Нечеткость данных, полученных в результате экспертного опроса, будет отражать как неуверенность, сомнения эксперта в собственном прогнозе, так и его попытку выразить объективную изменчивость параметров системы.

Сделаем несколько предварительных замечаний, касающихся описания используемых далее нечетких величин и действий с ними. Под нечеткой величиной (НВ) будем понимать [3] произвольное нечеткое множество (НМ), заданное на множестве действительных чисел \mathcal{R} . Ограничимся тремя типами НВ с непрерывными нормальными кусочно-линейными функциями принадлежности (ФП), графики которых иллюстрирует рис. 1:

1) нечеткое число \tilde{A} с треугольной функцией ФП (ТНЧ), имеющей по одному интервалу линейного возрастания $[^L a, ^M a]$ и убывания $[^M a, ^R a]$;

2) НВ \tilde{B} с S-образной ФП, имеющей два интервала линейного возрастания: первый – $[^0 b, ^{0,5} b]$, на котором ФП меняется с 0 до 0,5, и второй – $[^{0,5} b, ^1 b]$, где она возрастает далее до 1,0;

3) НВ \tilde{C} с Z-образной ФП, имеющей два интервала линейного убывания: первый – $[^1 c, ^{0,5} c]$, на котором ФП уменьшается с 1 до 0,5, и второй – $[^{0,5} c, ^0 c]$, на котором она продолжает сокращаться до 0.

Перечисленные НВ для сокращения записи представим в виде кортежей, дополнив в случае необходимости обозначения

указанием на тип НВ: $\tilde{A} = \langle {}^L a, {}^M a, {}^R a \rangle^\Delta$, $\tilde{B} = \langle {}^0 b, {}^{0,5} b, {}^1 b \rangle^S$,
 $\tilde{C} = \langle {}^1 c, {}^{0,5} c, {}^0 c \rangle^Z$.

Основываясь на принципе обобщения Заде и поуровневом подходе, определим необходимые ниже операции с ТНЧ. Пусть даны два ТНЧ: $\tilde{A}_1 = \langle {}^L a_1, {}^M a_1, {}^R a_1 \rangle^\Delta$ и $\tilde{A}_2 = \langle {}^L a_2, {}^M a_2, {}^R a_2 \rangle^\Delta$. Результатами их нечеткого сложения (+), перемножения (\times), деления (/) и взятия максимума (max) будут ТНЧ, определяемые соответственно как:

$$\tilde{A}_1 + \tilde{A}_2 = \langle {}^L a_1 + {}^L a_2, {}^M a_1 + {}^M a_2, {}^R a_1 + {}^R a_2 \rangle^\Delta,$$

$$\tilde{A}_1 \times \tilde{A}_2 = \langle {}^L a_1 \times {}^L a_2, {}^M a_1 \times {}^M a_2, {}^R a_1 \times {}^R a_2 \rangle^\Delta,$$

$$\tilde{A}_1 / \tilde{A}_2 = \langle {}^L a_1 / {}^R a_2, {}^M a_1 / {}^M a_2, {}^R a_1 / {}^L a_2 \rangle^\Delta,$$

$$\max(\tilde{A}_1, \tilde{A}_2) = \langle \max({}^L a_1, {}^L a_2), \max({}^M a_1, {}^M a_2), \max({}^R a_1, {}^R a_2) \rangle^\Delta.$$

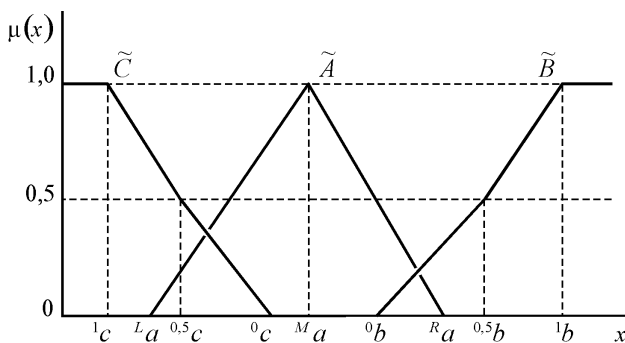


Рис.1. Функции принадлежности НВ различных типов

При выполнении операций взятия максимума и вычитания ($-$), если первая из величин \tilde{A}_1 является нечеткой, а вторая a_2 – нет, результатами в общем случае будут ТНЧ:

$$\max(\tilde{A}_1, a_2) = \langle \max({}^L a_1, a_2), \max({}^M a_1, a_2), \max({}^R a_1, a_2) \rangle^\Delta,$$

$$\tilde{A}_1 - a_2 = \langle {}^L a_1 - a_2, {}^M a_1 - a_2, {}^R a_1 - a_2 \rangle^\Delta.$$

При сравнении ТНЧ $\tilde{A} = \langle {}^L a, {}^M a, {}^R a \rangle^\Delta$ (${}^L a \leq {}^M a \leq {}^R a$) и четкого числа b будем считать, что

(1) $\tilde{A} \leq b$, если ${}^R a \leq b$,

(2) $\tilde{A} \geq b$, если ${}^L a \geq b$.

Решением нечеткого уравнения

(3) $\tilde{X} + \tilde{A}_1 = \tilde{A}_2$

будет ТНЧ $\tilde{X} = \langle {}^L a_2 - {}^L a_1, {}^M a_2 - {}^M a_1, {}^R a_2 - {}^R a_1 \rangle^\Delta$ [3].

Для дефаззификации ТНЧ используем метод центра тяжести [3], в соответствии с которым четкое число ${}^d a$ – результат дефаззификации ТНЧ $\tilde{A} = \langle {}^L a, {}^M a, {}^R a \rangle^\Delta$ – определяется как

$${}^d a = \frac{{}^L a + {}^M a + {}^R a}{3}.$$

3. Нечёткий технологический график процесса наземного обслуживания самолетов в хабе

Решение задачи оценки эффективности трансферной системы предполагает наличие нечеткой модели процесса наземного обслуживания в хабе, устанавливающей связи между параметрами расписания движения самолётов, величинами продолжительности технологических операций наземного обслуживания, численности и производительности задействованных при их выполнении ресурсов, общей численности и производительности ресурсов аэропорта.

Для описания технологических процессов в хабе используем модельный технологический график, содержащий операции, оказывающие наиболее существенное влияние на время наземной стоянки ВС. Введем множество $A = \{a_b, b = 1, \dots, B\}$ операций модельного графика в составе: $A = \{\text{ПТ, УТ, ВП, ПП, ВБ, ПБ, РНП, РТП, ВКБ, СНБ, СТБ, ТО, ЗВС, УС}\}$, где ПТ, УТ – соответственно подача и уборка пассажирского трапа; ВП, ПП – соответственно высадка из ВС и посадка в ВС пассажиров; ВБ, ПБ – соответственно выгрузка из ВС и погрузка в ВС багажа; РНП, РТП – регистрация соответственно начальных и

трансферных пассажиров; ВКБ – выдача багажа конечным пассажирам; СНБ, СТБ – сортировка багажа соответственно начальных и трансферных пассажиров, ТО – техническое обслуживание ВС; ЗВС – заправка ВС авиатопливом; УС – уборка салона и кабины ВС. Дополним модельный график промежуток ОЖ времени ожидания, не требующим выполнения каких-либо работ. Управление продолжительностью этого промежутка позволит изменять время начала операций, выполняемых перед вылетом (ПБ, ПП, УТ), и тем самым управлять моментом времени вылета. Введем множество $P = \{\rho_s, s = 1, \dots, B + 1\} = A \cup \{\text{ОЖ}\}$ элементов модельного графика, включающее, наряду с операциями, промежуток времени ОЖ. Модельный график в форме диаграммы Ганта отражен на рис. 2, где операции и промежутки времени обозначены сплошными линиями с приводимыми над линиями аббревиатурами их названий.

Чтобы установить связи между моментами наступления основных событий модельного графика, рассмотрим отдельную волну прибытий–отправлений, в течение которой в хаб должны прилететь, пройти обслуживание и вылететь K самолетов. Пусть продолжительность волны – величина промежутка времени от самого раннего момента прибытия рейса до самого позднего момента отправления – не должна превышать T . Введем определенные (не нечёткие) векторы $\bar{t}^a = (t_k^a)_{K \times 1}$ плановых (по расписанию) моментов времени прибытия и $\bar{\tau}^{\text{ОЖ}} = (\tau_k^{\text{ОЖ}})_{K \times 1}$ величин продолжительности промежутка ОЖ. Обозначим как ij трансферную авиасвязь, поддерживаемую двумя рейсами, первый из которых выполняет прилетающее ВС под номером i , а второй – вылетающее ВС под номером j , $i, j = 1, \dots, K$). Будем считать заданными в форме ТНЧ численности \tilde{v}_{ij}^T потенциальных трансферных пассажиров, т.е. лиц, которые имеют потребность и возможности совершить поездку на авиасвязи ij по установленному тарифу в течение рассматриваемой волны.

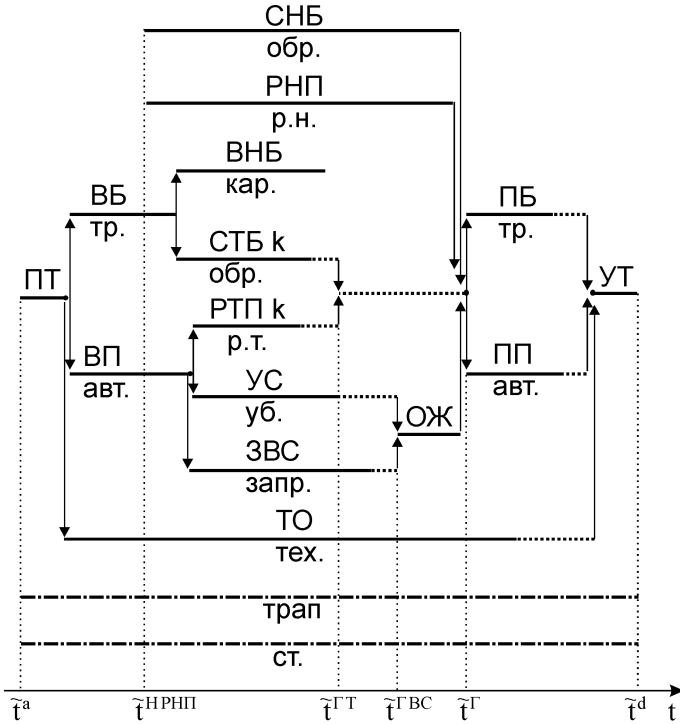


Рис. 2. Модельный график наземного обслуживания ВС

Для учета возможных отказов от перевозки тех потенциальных пассажиров, которых не устраивает длительность пребывания в хабе, введем НВ \tilde{w}_{ij}^T , выражающую число действительных пассажиров авиасвязи ij . Введем матрицы нечеткой численности потенциальных $\tilde{V}^T = \left\| \tilde{v}_{ij}^T \right\|_{K \times K}$ и действительных $\tilde{W}^T = \left\| \tilde{w}_{ij}^T \right\|_{K \times K}$ трансферных пассажиров. Будем считать ТНЧ отклонения $\Delta \tilde{t}_k^a$ фактического времени прибытия k -го ВС от t_k^a и величины продолжительности операций модельного графика.

Запишем формулы, описывающие связи между моментами наступления основных событий модельного графика, учитывая,

что моменты времени начала $\tilde{t}_k^{H\rho_s}$ и окончания $\tilde{t}_k^{K\rho_s}$ связаны с продолжительностью $\tilde{\tau}_k^{\rho_s}$ элемента ρ_s графика обслуживания k -го ВС соотношением

$$\tilde{t}_k^{K\rho_s} = \tilde{t}_k^{H\rho_s} + \tilde{\tau}_k^{\rho_s}, \quad k=1, \dots, K, \quad \rho_s \in P.$$

Момент времени \tilde{t}_k^a фактического прибытия ВС, определяемый с учетом возможных отклонений от расписания:

$$\tilde{t}_k^a = t_k^a + \Delta \tilde{t}_k^a, \quad k=1, \dots, K.$$

Момент времени готовности k -го ВС к посадке вылетающих пассажиров и погрузке багажа:

$$\tilde{t}_k^{\Gamma BC} = \tilde{t}_k^a + \tilde{\tau}_k^{\Pi\Pi} + \tilde{\tau}_k^{\text{БП}} + \max\{\tilde{\tau}_k^{\text{УС}}, \tilde{\tau}_k^{\text{ЗВС}}\}, \quad k=1, \dots, K.$$

Момент времени готовности трансферных пассажиров, прибывших k -м ВС, к посадке, а их багажа – к погрузке на вылетающее ВС:

$$\tilde{t}_k^{\Gamma T} = \tilde{t}_k^a + \tilde{\tau}_k^{\Pi\Pi} + \max\{\tilde{\tau}_k^{\text{ББ}} + \tilde{\tau}_k^{\text{СТБ}}, \tilde{\tau}_k^{\text{БП}} + \tilde{\tau}_k^{\text{РП\Pi\Pi}}\}, \quad k=1, \dots, K.$$

Момент времени готовности ВС, пассажиров, багажа к началу операций ПП и ПБ:

$$(4) \quad \tilde{t}_k^{\Gamma} = \max\left\{\tilde{t}_k^{\Gamma BC} + \tau_k^{\text{ОЖ}}, \max_{\substack{i=1, \dots, K, \\ \tilde{w}_k^i > 0}} \tilde{t}_i^{\Gamma T}\right\}, \quad k=1, \dots, K$$

Выражением (4) учитываются взаимосвязи между процессами обслуживания различных ВС, обусловленные наличием пересаживающихся между этими ВС пассажиров. Операции ПП и ПБ графика k -го ВС могут начаться только в случае готовности к ним трансферных пассажиров, пересаживающихся на k -е с других ВС, на которых такие пассажиры присутствуют. Учитывается также возможное присутствие на k -м ВС прилетевших пассажиров прямого транзита, следующих из хаба этим же ВС.

Время отправления k -го ВС:

$$(5) \quad \tilde{t}_k^d = \max\left\{\tilde{t}_k^{\Gamma} + \max\left\{\tilde{\tau}_k^{\text{ПП}}, \tilde{\tau}_k^{\text{ПБ}}\right\}, \tilde{t}_k^a + \tilde{\tau}_k^{\text{ПП}} + \tilde{\tau}_k^{\text{ТО}}\right\} + \tilde{\tau}_k^{\text{УТ}}, \quad k=1, \dots, K.$$

Заметим, что из выражений (4) и (5) следует указанная выше возможность управления моментами времени \tilde{t}_k^Γ и \tilde{t}_k^d посредством варьирования продолжительностью промежутка $\tau_k^{\text{ОЖ}}$.

ТНЧ \tilde{t}_k^d в силу своей неопределенности не может быть на практике непосредственно использовано в качестве планового времени отправления. Однако на этапе проектирования трансферной системы НВ \tilde{t}_k^d допустимо рассматривать в качестве предварительной модельной оценки планового времени, которая на последующих этапах может приниматься во внимание авиакомпанией в ходе разработки «реального» расписания.

Следует учитывать, что авиакомпания вольна выбирать ту или иную линию в отношении назначения планового времени отправления. Так, задание преимущественно раннего планового времени, близкого к наименьшему возможному, позволит авиакомпании не только придать расписанию большую формальную привлекательность с точки зрения трансферных пассажиров, оценивающих время своего предстоящего пребывания в хабе, ориентируясь именно на плановые моменты прибытия и отправления стыковочных рейсов, но и обеспечить меньшую среднюю продолжительность простоев ВС в аэропорту. К минусам такой линии относятся, во-первых, угроза систематических опозданий с вылетом, вызванных проблематичностью обслуживания ВС в столь сжатые сроки, и, как следствие, рост суммы штрафов за нарушения расписания, и, во-вторых, невозможность ожидания пассажиров, пересеживающихся с фактически опаздывающих рейсов, приводящая к затратам на отправку этих пассажиров другими рейсами или направление их в гостиницу, имиджевым потерям. Альтернативная линия поведения авиакомпании, предполагающая назначение преимущественно позднего планового времени отправления, позволяет сократить вероятность нарушения расписания и суммы сопутствующих штрафов, обеспечивает возможность ожидания пассажиров, пересеживающихся с опаздывающих рейсов, однако, приводит к чрезмерно длительным простоям ВС и слишком продолжительным пересадкам пассажиров.

На этапе проектирования затруднительно достоверно спрогнозировать, какую политику в отношении с клиентурой и стратегию по использованию авиапарка выберет авиакомпания. Чтобы учесть разнообразие возможных вариантов подхода к назначению планового времени отправления, будем рассматривать эту величину как нечеткую, равную \tilde{t}_k^d . При этом нет причин считать неопределенным ещё и плановое время прибытия t_k^a . Ничто не мешает ЛПР рационально определить его в чёткой форме в результате решения задачи оптимизации расписания и рекомендовать авиакомпании, и никакие соображения стратегии или тактики не препятствуют использованию этого решения в практике трансферной системы.

Принятие предположения о нечеткости планового времени отправления позволяет считать нечетким также и время $\tilde{t}_k^{\text{H PПН}}$ начала регистрации первоначальных пассажиров, поскольку в аэропортовой практике оно определяется авиаперевозчиком и доводится им до сведения пассажиров в прямой зависимости от планового времени вылета. ТНЧ $\tilde{t}_k^{\text{H PПН}}$ определяется из следующего нечеткого уравнения:

$$(6) \quad \tilde{t}_k^{\text{H PПН}} + \tilde{\tau}_k^{\text{PПН}} = \tilde{t}_k^{\Gamma}, \quad k = 1, \dots, K.$$

В соответствии с (6) величина $\tilde{t}_k^{\text{H PПН}}$ определяется таким образом, чтобы окончание регистрации совпадало бы с наступлением готовности самолета, пассажиров, багажа к началу операций ПП и ПБ, что позволяет минимизировать временные затраты на пребывание в аэропорту как пассажиров, и первоначальных, и трансферных, так и ВС.

Будем полагать, что обработка багажа должна производиться в те же сроки, что и обслуживание пассажиров:

$$(7) \quad \tilde{t}_k^{\text{H PПН}} \equiv \tilde{t}_k^{\text{H СНБ}}, \quad \tilde{\tau}_k^{\text{СНБ}} \equiv \tilde{\tau}_k^{\text{PПН}}, \quad k = 1, \dots, K.$$

Итак, получены формулы для нечётких моментов времени наступления всех основных событий модельного графика, за исключением \tilde{t}_k^{Γ} . Определение этой, а также целого ряда величин, характеризующих временные и ресурсные затраты на выполнение операций модельного графика, требует знания матри-

цы \tilde{W}^T . Численность \tilde{w}_{ij}^T действительных трансферных пассажиров авиасвязи ij будем определять в зависимости от их потенциальной численности \tilde{v}_{ij}^T и от времени пересадки в хабе:

$$(8) \tilde{w}_{ij}^T = \tilde{v}_{ij}^T \tilde{k}_{ij}^T, \quad i, j = 1, \dots, K,$$

где \tilde{k}_{ij}^T – нечёткая доля пассажиров, воспользовавшихся трансферной перевозкой, от числа потенциальных пассажиров авиасвязи ij , зависящая от времени Δt_{ij}^C пребывания трансферного пассажира авиасвязи ij в хабе:

$$(9) \Delta \tilde{t}_{ij}^C = \tilde{t}_j^d - t_i^a, \quad i, j = 1, \dots, K,$$

представляющего собой, так же как и \tilde{t}_j^d , ТНЧ.

Учитывая данные, приводимые в работах [8, 9, 10], предпочтения пассажиров в отношении времени пребывания в хабе Δt^C зададим в виде трапецеидальной функции желательности $f_T(\Delta t^C)$ с четырьмя опорными точками $t^{(1)}$, $t^{(2)}$, $t^{(3)}$, $t^{(4)}$, представленной на рис. 3. Сочетания прибывших и отправляющихся рейсов со значениями $\Delta t^C < t^{(1)}$ не рассматриваются ни одним из трансферных пассажиров в качестве возможных стыковок по причине слишком малого времени, отводимого на пересадку. Значения $\Delta t^C > t^{(4)}$ предполагают неприемлемо продолжительное ожидание пересадки. Наиболее комфортному времени пересадки соответствует промежуток $[t^{(2)}, t^{(3)}]$. Примем следующие значения параметров: $t^{(1)} = 45$ мин., $t^{(2)} = 75$ мин., $t^{(3)} = 90$ мин., $t^{(4)} = 180$ мин.

Согласно принципу обобщения Л. Заде для ТНЧ $\Delta \tilde{t}_{ij}^C = \langle {}^L \Delta t_{ij}^C, {}^M \Delta t_{ij}^C, {}^R \Delta t_{ij}^C \rangle^\Delta$ нечеткое значение функции

$\tilde{k}_{ij}^T = f_T(\Delta \tilde{t}_{ij}^C)$ определяется как ТНЧ:

$$(10) \tilde{k}_{ij}^T = \langle \min [f_T({}^L \Delta t_{ij}^C), f_T({}^R \Delta t_{ij}^C)], f_T({}^M \Delta t_{ij}^C), \max [f_T({}^L \Delta t_{ij}^C), f_T({}^R \Delta t_{ij}^C)] \rangle^\Delta.$$

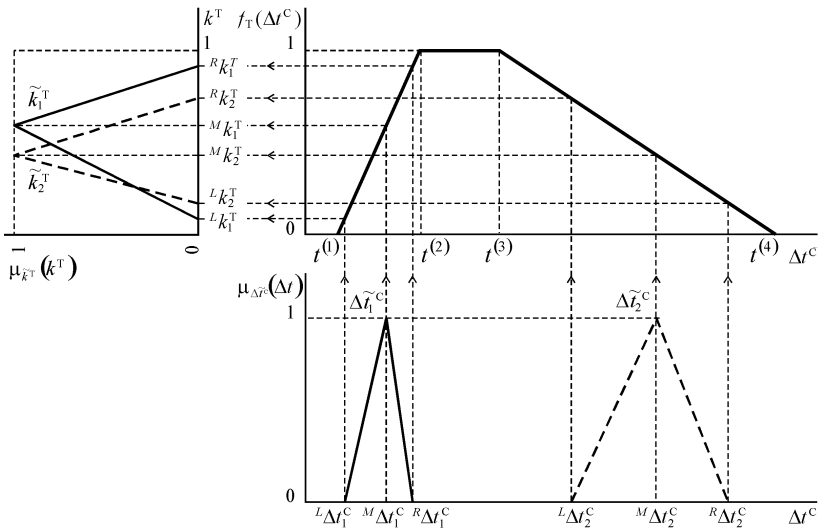


Рис. 3. Функция желательности $f_T(\Delta t^C)$ и ее использование для определения коэффициента \tilde{k}^T

Рис. 3 на двух примерах иллюстрирует процедуру определения \tilde{k}^T по заданному $\Delta \tilde{t}^C$ с использованием формулы (10). Поскольку \tilde{v}_{ij}^T и \tilde{k}_{ij}^T представляют собой ТНЧ, то \tilde{w}_{ij}^T , как результат их перемножения в соответствии с (8), также является ТНЧ.

Входящие в представленные выше выражения нечеткие продолжительности операций модельного графика либо задаются как ТНЧ, либо, если они зависят от уровня пассажиропотока, численности и производительности ресурсов, рассчитываются с использованием подходов, рассмотренных ниже.

4. Численность и производительность технологических ресурсов хаба

Для определения потребной численности и производительности технологических ресурсов аэропорта введем множество их типов $\Omega = \{\omega_h, h = 1, \dots, H\}$ в составе (в скобках – использу-

емые сокращения): $\Omega = \{ \text{«место стоянки ВС» (ст.), «пассажирский трап» (трап), «багажный погрузчик» (тр.), «пerrонный автобус» (авт.), «авиатопливозаправщик» (запр.), «средство выдачи багажа» (кар.), «место регистрации начальных пассажиров» (р.н.), «место регистрации трансферных пассажиров» (р.т.), «бригада ТО ВС» (тех.), «бригада уборки ВС» (уб.), «система обработки багажа» (обр.) } \}.$

Учитывая, что однотипные ресурсы могут использоваться для выполнения нескольких операций обслуживания одного ВС, будем наряду с понятием «операция» употреблять понятие «задание», которое определим как целенаправленный процесс использования одной или нескольких единиц однотипных ресурсов, не предполагающий перерывов. Введем множество $\Psi = \{ \psi^d, d = 1, \dots, D \}$ заданий модельного графика. Свяжем множество заданий с множествами операций и ресурсов, определив задание как кортеж

$$\psi^d = \langle \alpha_d^H, \alpha_d^K, \omega_d^3 \rangle, \alpha_d^H, \alpha_d^K \in A \subset P, \omega_d^3 \in \Omega,$$

где α_d^H, α_d^K – соответственно начальная и конечная операция d -го задания; ω_d^3 – тип ресурсов, используемых при отработке задания. На рис. 2 штрихпунктирной линией отмечены задания, охватывающие несколько операций. Задания с совпадающими начальной и конечной операциям отдельно не обозначены. Тип соответствующих заданиям ресурсов указан ниже линии. Обозначим $\tilde{\theta}_k^{Hd}, \tilde{\theta}_k^{Kd}$ – моменты времени начала и окончания отработки задания ψ^d , т.е. d -го задания модельного графика по обслуживанию k -го ВС, определяемые через нечеткие моменты времени начала и окончания соответствующих операций и являющиеся таким образом ТНЧ:

$$\tilde{\theta}_k^{Hd} = \tilde{t}_k^{H \alpha_d^H}, \tilde{\theta}_k^{Kd} = \tilde{t}_k^{K \alpha_d^K}, d = 1, \dots, D, k = 1, \dots, K.$$

Введем величину \tilde{Q}_k^d нечеткого интервала времени отработки задания ψ_k^d , которую определим как результат пересечения НМ S-типа \tilde{q}_k^{Hd} и НМ Z-типа \tilde{q}_k^{Kd} , заданных на универсальном множестве T :

$$\tilde{Q}_k^d = \tilde{q}_k^{Hd} \cap \tilde{q}_k^{Kd}, \quad d=1, \dots, D, \quad k=1, \dots, K.$$

НМ \tilde{q}_k^{Hd} будем трактовать как «момент времени, к наступлению которого отработка задания ψ_k^d уже начнется», НМ \tilde{q}_k^{Kd} – как «момент времени, к наступлению которого отработка того же задания еще не завершится». ФП $\mu_{\tilde{Q}_k^d}(t)$ нечеткого интервала \tilde{Q}_k^d может быть выражена через ФП $\mu_{\tilde{q}_k^{Hd}}(t)$, $\mu_{\tilde{q}_k^{Kd}}(t)$ указанных НМ как:

$$(11) \quad \mu_{\tilde{Q}_k^d}(t) = \mu_{\tilde{q}_k^{Hd}}(t) \wedge \mu_{\tilde{q}_k^{Kd}}(t), \quad t \in T, \quad d=1, \dots, D, \quad k=1, \dots, K.$$

Процедуру определения параметров ФП $\mu_{\tilde{q}_k^{Hd}}(t)$, $\mu_{\tilde{q}_k^{Kd}}(t)$ для $\forall t \in T$ иллюстрирует рис. 4.

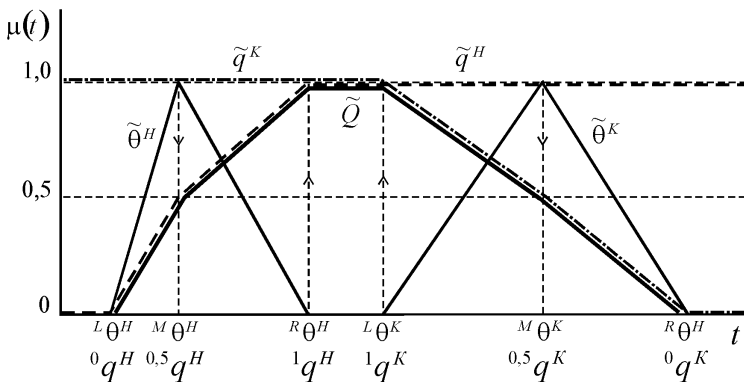


Рис. 4. К определению функций принадлежности

Сформируем функцию $\mu_{\tilde{q}_k^{Hd}}(t)$ на базе известной функции $\mu_{\tilde{\theta}_k^{Hd}}(t)$, воспользовавшись очевидной семантической связью между НВ \tilde{q}_k^{Hd} и $\tilde{\theta}_k^{Kd}$. Из вида $\mu_{\tilde{\theta}_k^{Hd}}(t)$ на участках $-\infty < t \leq {}^L\theta_k^{Hd}$ и ${}^R\theta_k^{Hd} \leq t < +\infty$ ясно, что до наступления времени $t = {}^L\theta_k^{Hd}$ выполнение задания не может начаться ни при каких обстоятельствах, а после наступления $t = {}^R\theta_k^{Hd}$ оно начнется обя-

зательно. Отсюда следует, что, если $t \leq L\theta_k^{Hd}$, то $\mu_{\tilde{q}_k^{Hd}}(t) = \mu_{\tilde{\theta}_k^{Hd}}(t) = 0$, и значит ${}^0q_k^{Hd} = L\theta_k^{Hd}$.

Аналогично, если ${}^R\theta_k^{Hd} \leq t$, то $\mu_{\tilde{q}_k^{Hd}}(t) = 1 - \mu_{\tilde{\theta}_k^{Hd}}(t) = 1$, и поэтому ${}^1q_k^{Hd} = {}^R\theta_k^{Hd}$. О промежутке $L\theta_k^{Hd} \leq t < {}^R\theta_k^{Hd}$ можно сказать лишь то, что в любой момент времени задание может как уже начаться, так и еще не начаться, причем возможность начала выполнения задания с течением времени возрастает. Не существует строгих нечётко-множественных правил, позволяющих «напрямую» преобразовать заданную треугольную ФП $\mu_{\tilde{\theta}_k^{Hd}}(t)$ в искомую ФП S-типа $\mu_{\tilde{q}_k^{Hd}}(t)$ на участке $L\theta_k^{Hd} \leq t < {}^R\theta_k^{Hd}$. Однако, чтобы не терять информацию, выражаемую функцией $\mu_{\tilde{\theta}_k^{Hd}}(t)$, введем два допущения. Во-первых, будем считать $\mu_{\tilde{q}_k^{Hd}}(t)$ кусочно-линейной. Во-вторых, примем, что модальному значению ${}^M\theta_k^{Hd}$ соответствуют равные возможности как того, что отработка задания к этому времени началась, так и того, что не началась, т.е. $\mu_{\tilde{q}_k^{Hd}}({}^M\theta_k^{Hd}) = 0,5$, и значит ${}^{0,5}q_k^{Hd} = {}^M\theta_k^{Hd}$. Таким образом, учитывая связи между координатами опорных точек НМ \tilde{q}_k^{Hd} и $\tilde{\theta}_k^{Hd}$, в окончательном виде имеем

$$(12) \quad \tilde{q}_k^{Hd} = \langle {}^0q_k^{Hd}, {}^{0,5}q_k^{Hd}, {}^1q_k^{Hd} \rangle^S = \langle L\theta_k^{Hd}, {}^M\theta_k^{Hd}, {}^R\theta_k^{Hd} \rangle^S, \\ d = 1, \dots, D, k = 1, \dots, K.$$

Опуская аналогичные рассуждения, запишем формулу, выражающую НМ Z-типа \tilde{q}_k^{Kd} через координаты опорных точек ТНЧ $\tilde{\theta}_k^{Kd}$:

$$(13) \quad \tilde{q}_k^{Kd} = \langle {}^1q_k^{Kd}, {}^{0,5}q_k^{Kd}, {}^0q_k^{Kd} \rangle^Z = \langle L\theta_k^{Kd}, {}^M\theta_k^{Kd}, {}^R\theta_k^{Kd} \rangle^Z, \\ d = 1, \dots, D, k = 1, \dots, K.$$

Численность и производительность ресурсов аэропорта определим исходя из объема работ, для выполнения которых предназначены ресурсы. Введем величину ρ_k^d скорости выпол-

нения работ [1] задания ψ_k^d , под которой будем понимать объем работ по обслуживанию ВС, пассажиров, багажа, производимый в единицу времени при выполнении задания ψ_k^d . Обозначим r_k^d расчетную скорость выполнения работ задания ψ_k^d , т.е. объем соответствующих заданию ψ_k^d работ, который необходимо производить в единицу времени, чтобы обеспечить соблюдение требований технологии аэропортового обслуживания и ограничений технологического графика. Примем, что r_k^d либо устанавливается производственной документацией аэропорта в зависимости от типа обслуживаемого ВС, в этом случае значение является вполне определенным, либо рассчитывается пропорционально уровню потоков пассажиров и багажа, при этом результат нечеткий. Чтобы излишне не усложнять модель, будем в любом случае считать величину r_k^d определенной; при ее нечеткости используем дефазсифицированное методом центра тяжести значение.

Наряду с расчетной, введем мгновенную скорость выполнения работ $\tilde{\rho}_k^d(t)$ задания ψ_k^d , т.е. скорость выполнения работ задания ψ_k^d в момент времени $t \in T$, нечеткость которой объясняется размытостью временного интервала \tilde{Q}_k^d . При определении ФП $\mu_{\tilde{\rho}_k^d}(\rho_k^d, t)$ примем во внимание, что заданию ψ_k^d в любой момент времени $t \in T$ может быть поставлено в соответствие одно из всего лишь двух значений скорости выполнения работ, первого – r_k^d , если в рассматриваемый момент времени задание выполняется, и второго – нуля, в противном случае. Таким образом, $\tilde{\rho}_k^d(t)$ определим как НМ дискретного типа с носителем, содержащим два элемента – нуль и r_k^d , или, в соответствии с одной из используемых форм записи, как множество двух упорядоченных пар:

$$(14) \tilde{\rho}_k^d(t) = \left\langle \left\langle \mu_{\tilde{\rho}_k^d}(0, t), 0 \right\rangle, \left\langle \mu_{\tilde{\rho}_k^d}(r_k^d, t), r_k^d \right\rangle \right\rangle, t \in T, \\ d = 1, \dots, D, k = 1, \dots, K.$$

Введенное выше допущение, согласно которому, если работы задания ψ_k^d в момент времени $t \in T$ выполняются, то выполняются только со скоростью $\rho_k^d = r_k^d$, приводит к очевидному тождеству:

$$(15) \mu_{\rho_k^d}(\rho_k^d = r_k^d, t) \equiv \mu_{\tilde{Q}_k^d}(t), \quad t \in T, \quad d = 1, \dots, D, \quad k = 1, \dots, K.$$

С другой стороны, если к моменту времени $t \in T$ выполнение задания либо еще не началось, либо уже завершилось, то скорость выполнения работ этого задания равна нулю ($\rho_k^d = 0$), и, следовательно:

$$(16) \mu_{\rho_k^d}(\rho_k^d = 0, t) \equiv \mu_{-\tilde{Q}_k^d}(t), \quad t \in T, \quad d = 1, \dots, D, \quad k = 1, \dots, K,$$

где $\mu_{-\tilde{Q}_k^d}(t)$ – ФП определенного на универсальном множестве T НМ $-\tilde{Q}_k^d = T \setminus \tilde{Q}_k^d$, представляющего собой множество моментов времени $t \in T$, когда задание ψ_k^d не выполняется. Учитывая, что $-\tilde{Q}_k^d$ является дополнением НМ \tilde{Q}_k^d , его ФП определим, следуя Л. Заде [19], как

$$\mu_{-\tilde{Q}_k^d}(t) = 1 - \mu_{\tilde{Q}_k^d}(t), \quad t \in T, \quad d = 1, \dots, D, \quad k = 1, \dots, K,$$

а выражение (16) перепишем в виде

$$(17) \mu_{\rho_k^d}(\rho_k^d = 0, t) \equiv 1 - \mu_{\tilde{Q}_k^d}(t), \quad t \in T, \quad d = 1, \dots, D, \quad k = 1, \dots, K.$$

Сводя тождества (15) и (17) в единое выражение, имеем

$$(18) \mu_{\rho_k^d}(\rho_k^d, t) = \left\{ \begin{array}{ll} 1 - \mu_{\tilde{Q}_k^d}(t), & \text{если } \rho_k^d = 0, \\ \mu_{\tilde{Q}_k^d}(t), & \text{если } \rho_k^d = r_k^d, \end{array} \right\}, \quad t \in T,$$

$$d = 1, \dots, D, \quad k = 1, \dots, K.$$

Если временные зависимости $\tilde{\rho}_k^d(t)$ для всех $d = 1, \dots, D$ и $k = 1, \dots, K$ известны, то суммарная мгновенная скорость работ $\tilde{\rho}_h(t)$, которая должна обеспечиваться всеми ресурсами типа $\omega_h \in \Omega$ при обслуживании всех K ВС в момент времени $t \in T$, определится суммированием по всем заданиям всех ВС, предполагающим использование этих ресурсов:

$$(19) \tilde{\rho}_h(t) = \sum_{k=1}^K \left(\sum_{d \in D_{kh}} \tilde{\rho}_k^d(t) \right), \quad h = 1, \dots, H,$$

где $D_{kh} = \{d \in \{1, \dots, D\} : \omega_d^3 = \omega_h\}$ – множество номеров заданий при обслуживании k -го ВС установленными средствами типа ω_h .

Вычисления по формуле (19) могут быть выполнены с использованием формализма принципа обобщения Л. Заде [19]. Сформулируем его с учетом специфики рассматриваемой задачи для некоторых заданных на дискретных универсальных множествах X_1, \dots, X_K ($\forall X_1, \dots, X_K \subset \mathfrak{R}$) НВ $\tilde{x}_1, \dots, \tilde{x}_K$ с известными ФП $\mu_{\tilde{x}_k}(x_k)$, $k = 1, \dots, K$. Согласно принципу обобщения, ФП $\mu_{\tilde{y}}(y)$

суммы этих НВ $\tilde{y} = \sum_{k=1}^K \tilde{x}_k$ задается выражением

$$(20) \mu_{\tilde{y}}(y) = \bigvee_{y = \sum_{k=1}^K x_k} \left(\bigwedge_{k=1}^K \mu_{\tilde{x}_k}(x_k) \right), \quad x_k \in X_k \subset \mathfrak{R},$$

$$k = 1, \dots, K, \quad y \in Y \subset \mathfrak{R},$$

где Y – универсальное множество выходного значения $y = \sum_{k=1}^K x_k$,

на котором определяется НМ \tilde{y} , также являющееся дискретным в силу дискретности X_1, \dots, X_K .

Чтобы сделать формулы (19) и (11) применимыми для расчетов, необходимо определиться с выбором способа выполнения операций объединения \vee и пересечения \wedge множеств из широкого ряда способов, предложенных к настоящему времени [3]. При выборе примем во внимание, что нечеткая форма представления результатов формируемой модели затрудняет их интерпретацию ЛПР, традиционно более подготовленного и «нацеленного» на работу с вероятностными, а не с нечеткими, данными. В еще большей степени, нежели подготовка ЛПР, делает желательным получение результатов в вероятностной форме тот факт, что требования к надежности и качеству наземного обслуживания задаются нормативной аэропортовой документацией именно в вероятностной, а не в нечеткой, форме. Чтобы иметь возможность вероятностной трактовки получаемых результатов, потре-

буем, чтобы значения ФП $\mu_{\tilde{y}}(y)$ были нормированы, т.е. для них было бы справедливо равенство

$$(21) \sum_n \mu_{\tilde{y}}(y_n) = 1.$$

Данное условие будет выполняться, если, во-первых, нормированными будут значения ФП всех суммируемых НВ $\mu_{\tilde{x}_k}(x_k)$, $k = 1, \dots, K$, и, во-вторых, для выполнения операций над множествами \vee и \wedge будут использоваться операторы сложения и умножения соответственно. В этом случае (20) будет полностью совпадать с выражением для вероятностного распределения случайной величины, представляющей собой сумму K случайных величин.

Отметим, что согласно (18) ФП $\mu_{\tilde{\rho}_k^d}(\rho_k^d, t)$ дискретных НВ $\tilde{\rho}_k^d(t)$ для $\forall t \in T$ удовлетворяют условию нормировки, что позволяет рассматривать их как субъективные вероятности выполнения в момент времени t работ задания ψ_k^d . Применяя для действий над множествами операторы сложения и умножения на основе формул (18), (19) получим расчетное выражение для ФП НВ $\tilde{\rho}_h(t)$:

$$(22) \mu_{\tilde{\rho}_h}(\rho_h, t) = \sum_{\rho_h = \sum_{k=1}^K \left(\sum_{d \in D_{kh}} \rho_k^d \right)} \left(\prod_{k=1}^K \prod_{d \in D_{kh}} \mu_{\tilde{\rho}_k^d}(\rho_k^d, t) \right), \rho_k^d \in \{0, r_k^d\},$$

$$\rho_h \in \mathfrak{R}, \forall t \in T, d \in D_{kh}, k = 1, \dots, K.$$

ФП $\mu_{\tilde{Q}_k^d}(t)$ нечеткого временного интервала \tilde{Q}_k^d также будем определять посредством умножения, при этом формула (11) примет вид

$$(23) \mu_{\tilde{Q}_k^d}(t) = \mu_{\tilde{q}_{kd}}(t) \cdot \mu_{\tilde{q}_k}(t), t \in T, d = 1, \dots, D, k = 1, \dots, K.$$

Нечеткая временная зависимость $\tilde{\rho}_h(t)$ сама по себе может представлять интерес для административных и диспетчерских служб аэропорта, однако значительно большую практическую ценность представляет ее четкий аналог, использование которого позволило бы аэропорту сделать вывод о достаточности име-

ющихся ресурсов, либо принять решение о формировании такого их парка, который обеспечил бы, при минимальных затратах на внедрение и эксплуатацию, возможность незамедлительного выполнения технологических операций в течение всего временного промежутка T (или значительной его части). Будем считать, что численность и производительность ресурсов должна быть минимальной, но достаточной для того, чтобы суммарная производительность комплекса ресурсов рассматриваемого типа в любой момент времени $t \in T$ с заданной надежностью превосходила суммарную скорость выполнения необходимых работ. Введем обоснованное выше допущение о возможности использования ФП $\mu_{\tilde{\rho}_h}(\rho_h, t)$ в качестве распределения вероятностей суммарной мгновенной скорости выполнения работ $\rho_h(t)$, потребной в момент $t \in T$. Все это позволит с надежностью не меньше заданной p_h^3 ($0 < p_h^3 < 1$) определять достаточную в момент $t \in T$ суммарную производительность $g_h(t)$ ресурсов типа $\omega_h \in \Omega$ как

$$(24) \quad g_h(t) = \min \left(g \in R : \sum_{\rho_h \leq g} \mu_{\tilde{\rho}_h}(\rho_h, t) \geq p_h^3 \right), \quad t \in T.$$

Здесь выполняется суммирование всех $\mu_{\tilde{\rho}_h}(\rho_h, t)$, соответствующих таким значениям ρ_h , принадлежащим носителю НМ $\tilde{\rho}_h(t)$, которые не превышают $g \in \mathfrak{R}$.

Достаточная для любого момента времени волны прибытий-отправлений производительность g_h ресурсов типа $\omega_h \in \Omega$ определится как наибольшая из $g_h(t)$:

$$(25) \quad g_h = \max_{\forall t \in T} [g_h(t)].$$

Представим набор формул (12), (13), (18), (22), (23), (24), (25) в виде оператора G , устанавливающего связь между, с одной стороны, величинами расчетной скорости, временными характеристиками и надежностью выполнения заданий, и, с другой стороны, суммарной потребной производительностью ресурсов, используемых для их выполнения:

$$(26) \quad g_h = G(R_h, \tilde{\Theta}_h^H, \tilde{\Theta}_h^K, p_h^3),$$

где $R_h, \tilde{\Theta}_h^H, \tilde{\Theta}_h^K$ – множества значений соответственно расчетной скорости выполнения работ, нечетких моментов времени начала и окончания тех заданий, для выполнения которых используются ресурсы типа $\omega_h \in \Omega$, т.е. такие множества $\{r_k^d\}, \{\tilde{\theta}_k^{Hd}\}, \{\tilde{\theta}_k^{Kd}\}$, для элементов которых справедлива запись: $d \in D_{kh}, k = 1, \dots, K$.

Уточним теперь подходы к расчету потребной численности и производительности для отдельных типов ресурсов. Примем, что аэропорт располагает единственным экземпляром ресурсов типа «обр.». Остальные типы ресурсов объединим в множество $\Omega' = \Omega \setminus \text{«обр.»}$. Допустим, что производительность единицы ресурсов из множества Ω' одинакова для всех единиц данного типа, благодаря чему расчетная скорость выполнения задания может считаться прямо зависящей только от расчетного числа единиц ресурсов, выделяемых аэропортом для отработки задания согласно требованиям технологической документации. При фиксированной производительности единицы суммарная производительность однотипных ресурсов, достаточная для обслуживания всех ВС в любой момент времени, будет обеспечиваться использованием ресурсов достаточной численности. Для всех ВС введем вектор $\bar{n}^d = (n_k^d)_{K \times 1}$ величин численности ресурсов типа $\omega_d^3 \in \Omega'$, выделяемых для отработки задания ψ_d на каждом из ВС. При известном \bar{n}^d потребная численность m_h ресурсов типа $\omega_h \in \Omega'$, достаточная с надежностью не меньше заданной p_h^3 ($0 < p_h^3 < 1$) для обслуживания всех K ВС в любой момент времени $t \in T$, определяется с использованием оператора G , для которого роль одной из входных переменных вместо расчетной скорости начинает играть расчетная численность ресурсов:

$$(27) \quad m_h = G(N'_h, \tilde{\Theta}_h^H, \tilde{\Theta}_h^K, p_h^3),$$

где $N'_h = \{n_k^d \in \bar{n}_d : d \in D_{kh}, k = 1, \dots, K\}$ – множество значений расчетной численности ресурсов, выделяемых для выполнения тех заданий, для которых используются ресурсы типа $\omega_h \in \Omega'$.

Выделив из множества Ω' элемент «р.т.», сформируем из оставшихся элементов подмножество $\Omega'' = \Omega \setminus \{\text{«обр.»}, \text{«р.т.»}\}$ типов ресурсов с заданной аэропортовой документацией численностью единиц, выделяемых для отработки задания. Объединим векторы установленной численности ресурсов, образующих множество Ω'' , в множество $N'' = \{\vec{n}^d : \omega_d^3 \in \Omega'', d = 1, \dots, D\} \subset N'$. Все значения $n_k^d \in \vec{n}^d \subset N''$ примем известными и не нечеткими.

В отношении ресурсов типа «р.т.», выделяемых для отработки задания (РТП, РТП, «р.т.») по обслуживанию трансферных пассажиров, прибывших k -м ВС, следует считать, что установленные численности $n_k^{\text{р.т.}}$ единиц ресурсов этого типа зависят от численности обслуживаемых действительных трансферных пассажиров. Необходимо обратить внимание на то, что выполнение расчета $n_k^{\text{р.т.}}$ усложняется взаимной зависимостью численности действительных трансферных пассажиров и характеристик модельного графика, в частности, $\tilde{t}^{\text{ГТ}}$ и \tilde{t}^d , требующей применения для вычислений метода последовательных приближений. В качестве альтернативы, позволяющей упростить модель, при подсчетах $n_k^{\text{р.т.}}$, как и некоторых других параметров процесса наземного обслуживания, может быть использовано допущение о совпадении числа действительных с числом потенциальных трансферных пассажиров, не зависящим от параметров графика:

$$(28) \tilde{w}_{ij}^{\text{T}} \approx \tilde{v}_{ij}^{\text{T}}, \text{ для всех } i, j = 1, \dots, K.$$

Следует подчеркнуть, что данное допущение применяется только в части расчета некоторых технологических параметров наземного обслуживания, и может рассматриваться как моделирование наиболее «жестких» условий работы службы обслуживания пассажирских перевозок хаба.

Используем в расчетах дефазифицированное по методу центра тяжести значение ${}^d v_{ij}^{\text{T}}$ ТНЧ $\tilde{v}_{ij}^{\text{T}}$. В этом случае, учитывая допущение (28), $n_k^{\text{р.т.}}$ приближенно определим как:

$$n_k^{p.t.} \approx \left\lceil \frac{1}{w^{1T}} \sum_{j=1}^K {}^d v_{kj}^T \right\rceil, \quad k=1, \dots, K,$$

где $\lceil \cdot \rceil$ – операция округления до ближайшего большего числа, w^{1T} – установленное нормативной документацией число трансферных пассажиров, для обслуживания которых выделяется единица ресурсов типа «р.т.»

Продолжительность $\tilde{\tau}_k^{PTP}$ операции РТП будем считать зависящей не только от численности трансферных пассажиров, которые прибыли k -м ВС, но также и от числа установленных для отработки задания ⟨РТП, РТП, «р.т.⟩» единиц ресурсов и средней продолжительности τ^{1T} регистрации одного пассажира. С учетом (28) имеем:

$$\tilde{\tau}_k^{PTP} = \frac{\tau^{1T}}{n_k^{p.t.}} \sum_{j=1}^K \tilde{v}_{kj}^T, \quad k=1, \dots, K.$$

При определении нечеткой продолжительности $\tilde{\tau}_k^{CTB}$ операции СТБ примем во внимание, во-первых, желательность синхронного завершения данной операции и операции РТП и, во-вторых, наличие технических ограничений на время пребывания места багажа в системе «обр.». Чтобы удовлетворить первому требованию, будем находить начальное приближение $\tilde{\tau}_k^{CTB0}$ в результате решения уравнения

$$\tilde{t}_k^{КВВ} + \tilde{\tau}_k^{CTB0} = \tilde{t}_k^{КРП}, \quad k=1, \dots, K.$$

Для выполнения второго требования искомую величину $\tilde{\tau}_k^{CTB}$ будем определять как

$$\tilde{\tau}_k^{CTB} = \max(\tilde{\tau}_k^{CTB0}, \tau_{\min}), \quad k=1, \dots, K,$$

где τ_{\min} – заданное (не нечеткое) время пребывания места багажа в системе «обр.».

Исключив из множества A операции, продолжительность которых рассчитывается (РТП, СТБ) или приравнивается продолжительности другой операции (СНБ), получим множество $A' = A \setminus \{РТП, СТБ, СНБ\}$. Введем вектор $\vec{\tau}^b = (\tilde{\tau}_1^b, \tilde{\tau}_2^b, \dots, \tilde{\tau}_K^b)$ НВ продолжительности операции $\alpha_b \in A$ графиков всех ВС. Примем

значения элементов векторов продолжительностей операций, которые образуют множество A' , заданными, а сами векторы объединим в множество $\tilde{T}' = \{\tilde{t}^b: \alpha_b \in A', b=1, \dots, B\}$.

Чтобы сделать формулы (12), (13), (18), (22), (23), (24), (25) пригодными для расчета потребной производительности системы типа «обр.», введем допущения, касающиеся НВ мгновенной интенсивности поступающих в систему потоков багажа $\tilde{\lambda}_k^{\text{СНБ}}(t)$ начальных пассажиров, вылетающих k -м ВС, и $\tilde{\lambda}_k^{\text{СТБ}}(t)$ трансферных пассажиров, пересеживающихся с k -го ВС, $k = 1, \dots, K$. Будем считать, что в течение всего времени выполнения операций СНБ и СТБ багаж соответственно начальных и трансферных пассажиров рейса поступает в систему «обр.» равномерно с расчетными интенсивностями $\tilde{\lambda}_k^{\text{СНБ}}$ и $\tilde{\lambda}_k^{\text{СТБ}}$, равными с учетом (28)

$$(29) \quad \tilde{\lambda}_k^{\text{СНБ}} = k^{\text{B}} \tilde{w}_k^{\text{H}} / \tilde{t}_k^{\text{СНБ}}, \quad \tilde{\lambda}_k^{\text{СТБ}} = \frac{k^{\text{B}}}{\tilde{t}_k^{\text{СТБ}}} \sum_{j=1}^K \tilde{v}_{kj}^{\text{T}}, \quad k = 1, \dots, K,$$

где k^{B} – среднее число мест зарегистрированного багажа одного пассажира.

При расчете потребной производительности $g_{\text{обр.}}$ системы типа «обр.» дефазифицированные методом центра тяжести значения ${}^d \lambda_k^{\text{СНБ}}$ и ${}^d \lambda_k^{\text{СТБ}}$ расчетных интенсивностей $\tilde{\lambda}_k^{\text{СНБ}}$ и $\tilde{\lambda}_k^{\text{СТБ}}$ логично принять в качестве величин расчетной скорости выполнения работ, связанных с сортировкой багажа. В этом случае оператором G устанавливается зависимость:

$$(30) \quad g_{\text{обр.}} = G(\Lambda_{\text{обр.}}, \bar{\Theta}_{\text{обр.}}^{\text{H}}, \bar{\Theta}_{\text{обр.}}^{\text{K}}, P_{\text{обр.}}^3),$$

где $\Lambda_{\text{обр.}} = \left\{ \left({}^d \lambda_k^{\text{СНБ}} \right)_{K \times 1}, \left({}^d \lambda_k^{\text{СТБ}} \right)_{K \times 1} \right\}$, $\bar{\Theta}_{\text{обр.}}^{\text{H}} = \left\{ \left(\tilde{t}_k^{\text{H СНБ}} \right)_{K \times 1}, \left(\tilde{t}_k^{\text{H СТБ}} \right)_{K \times 1} \right\}$, $\bar{\Theta}_{\text{обр.}}^{\text{K}} = \left\{ \left(\tilde{t}_k^{\text{K СНБ}} \right)_{K \times 1}, \left(\tilde{t}_k^{\text{K СТБ}} \right)_{K \times 1} \right\}$ – множества пар векторов величин расчетных интенсивностей багажных потоков, времени начала и окончания операций по обработке багажа.

Полученная в итоге нечеткая модель процесса наземного обслуживания в хабе, описанию которой посвящены пп.3, 4,

может быть в общем виде представлена оператором F , устанавливающим зависимость:

$$(31) (\tilde{t}^d, M', g_{\text{обп.}}) = F(\bar{t}^a, \Delta\tilde{t}^a, \tilde{V}^T, \tilde{w}^H, k^B, \tilde{T}', \tau^{1T}, N'', w^{1T}, \bar{\tau}^{\text{ОЖ}}, \tau_{\min}, \bar{p}^3),$$

где $\tilde{t}^d = (\tilde{t}_k^d)_{K \times 1}$ – вектор нечётких оценок времени отправления рейсов, $M' = \{\omega_h : \omega_h \in \Omega', h = 1, \dots, H\}$ – множество значений потребной численности ресурсов хаба, $\Delta\tilde{t}^a = (\Delta\tilde{t}_k^a)_{K \times 1}$ – вектор нечетких отклонений фактического времени прибытия от планового, $\tilde{w}^H = (\tilde{w}_k^H)_{K \times 1}$ – вектор численности начальных пассажиров, $\bar{p}^3 = (p_h^3)_{H \times 1}$ – вектор величин надежности выполнения условия достаточности суммарной производительности ресурсов для обслуживания всех ВС в любой момент времени $t \in T$.

5. Уточненная постановка задачи

Оценка эффективности трансферной системы, проектируемой или совершенствуемой в условиях неопределенности, предусматривает определение показателей C^{Π} и C^{Π}/C^{Π} . С этой целью введем нечеткую сумму \tilde{C}^{Π} доходов от перевозки всех потенциальных пассажиров по всем трансферным авиасвязям, которую при заданных матрицах трансферных пассажиропотоков \tilde{V}^T и тарифов $\tilde{C}^T = \|\tilde{c}_{ij}^T\|_{K \times K}$, состоящих из ТНЧ, будем определять как следующее ТНЧ:

$$(32) \tilde{C}^{\Pi} = \sum_{i=1}^K \sum_{j=1}^K \tilde{v}_{ij}^T \tilde{c}_{ij}^T.$$

Наряду с \tilde{C}^{Π} , введем нечеткую сумму \tilde{C} доходов от перевозки действительных трансферных пассажиров по всем авиасвязям, определяемую с учетом (8) и (9) как ТНЧ

$$(33) \tilde{C} = \sum_{i=1}^K \sum_{j=1}^K \tilde{w}_{ij}^T \tilde{c}_{ij}^T = \sum_{i=1}^K \sum_{j=1}^K [\tilde{v}_{ij}^T \tilde{k}^T (\tilde{t}_j^d - t_i^a)] \tilde{c}_{ij}^T.$$

Сравнение формул (32) и (33), различающихся только множителем \tilde{k}^T , по определению подчиненным условием $\tilde{k}^T(\cdot) \leq 1$, приводит к выводу о том, что величина \tilde{C} , при совпадении прочих исходных данных, не может превзойти величину \tilde{C}^{Π} . Таким образом, \tilde{C}^{Π} следует рассматривать в качестве максимального предельного значения \tilde{C} , соответствующего гипотетическому случаю перевозки всех потенциальных пассажиров. Из (33) следует, что организаторы трансферной системы перевозок в общем случае имеют возможность, рационально управляя параметрами расписания \tilde{t}^a и \tilde{t}^d , достичь повышения \tilde{C} , приблизив его к \tilde{C}^{Π} . Таким образом, величину \tilde{C} в записи (33) правомерно выбрать в качестве целевой функции задачи оптимизации расписания, назначив на роль оптимизируемых параметров векторы \tilde{t}^a и \tilde{t}^d , и учитывая ограничения:

1) на продолжительность волны прибытий-отправлений, которая не должна превышать T :

$$(34) \quad 0 \leq t_k^a \leq T, \quad 0 \leq \tilde{t}_k^d \leq T, \quad k = 1, \dots, K;$$

2) на потребную численность ресурсов хаба, которая не должна быть больше той, которой располагает аэропорт:

$$(35) \quad m_h \leq m_h^P, \quad m_h \in M', \quad m_h^P \in M'^P, \quad h \in \{1, \dots, H\}; \quad \omega_h \in \Omega',$$

где $M'^P = \{m_h^P : \omega_h \in \Omega', h = 1, \dots, H\}$ – множество величин располагаемой хабом численности m_h^P ресурсов, типы которых образуют множество Ω' ;

3) на потребную производительность ресурсов типа «обр.», которая не должна быть больше располагаемой производительности $g_{обр.}^P$, которой характеризуется установленная в аэропорту система «обр.»:

$$(36) \quad g_{обр.} \leq g_{обр.}^P.$$

Нечеткий характер оптимизируемых параметров, параметров целевой функции и ограничений приводит к необходимости решения оптимизационной задачи с бесконечным числом целевых функций и ограничений. По этой причине формулировку

задачи (33)–(36) следует рассматривать лишь как условную. Такого рода задачи требуют применения специальных подходов для придания определенности их постановкам. Одним из наиболее простых и часто используемых приемов, описанным, например, в [2], является дефаззификация целевой функции одним из методов, в числе которых метод центра тяжести. Следуя этому подходу, будем использовать в качестве целевой функции значение dC , полученное в результате дефаззификации указанным методом ТНЧ \tilde{C} .

Избежать нечёткости в описании оптимизируемых параметров позволяет установленная в п.3 и зафиксированная выражением (31) зависимость \tilde{t}^d от $\bar{\tau}^{\text{ОЖ}}$, благодаря которой нечеткий вектор \tilde{t}^d заменяется в составе оптимизируемых параметров четким вектором $\bar{\tau}^{\text{ОЖ}}$. Такая замена означает необходимость предварительного вычисления с помощью модели (31) элементов вектора \tilde{t}^d по некоторому определенному в процессе работы оптимизационного алгоритма вектору $\bar{\tau}^{\text{ОЖ}}$ и лишь последующую подстановку \tilde{t}^d в формулу (33) для подсчета значения целевой функции.

Дефаззификация ограничений (34) выполняется с использованием правил (1) и (2), позволяющих сравнивать ТНЧ и обычные числа. Рассматривая \tilde{t}_k^d как ТНЧ $\langle L_t^d, M_t^d, R_t^d \rangle^\Delta$, перепишем (34) в форме

$$(37) \quad 0 \leq t_k^a \leq T, \quad 0 \leq L_{t_k^d}, R_{t_k^d} \leq T, \quad k = 1, \dots, K.$$

Таким образом, задача оптимизации расписания сводится к определению таких векторов \bar{t}^a и $\bar{\tau}^{\text{ОЖ}}$, которые при заданных $\Delta \tilde{t}^a$, \tilde{V}^T , \tilde{w}^H , k^B , \tilde{T}' , τ^{1T} , N^n , w^{1T} , τ_{\min} , M'^P , $g_{\text{обр}}^P$, \bar{p}^3 обеспечивают максимум целевой функции dC и удовлетворяют ограничениям (35)–(37). Для определения входящих в состав целевой функции и ограничений параметров \tilde{t}^d , M' и $g_{\text{обр}}$ используется модель процесса наземного обслуживания, представленная в виде оператора F (31). Благодаря тому, что ни целевая

функция, ни ограничения не являются нечеткими, для решения задачи применимы известные методы математического программирования, реализованные в доступном программном обеспечении персональных ЭВМ, таком как надстройка «Поиск решения» табличного процессора Microsoft Excel, либо специализированный программный пакет IBM ILOG OPL.

Определившись с постановкой задачи оптимизации расписания, вернемся к определению показателей эффективности трансферной системы C^D и C^D/C^P . Представляется вполне справедливым отождествить показатель C^P величины доходов от перевозки всех потенциальных трансферных пассажиров волны прилетов-вылетов с нечеткой суммой \tilde{C}^P (33). Аналогичным образом, показатель C^D максимальной величины доходов, на которую могут рассчитывать организаторы трансферной системы в случае использования рационально составленного расписания, правомерно отождествить с нечеткой суммой \tilde{C} , но не с произвольным ее значением, а лишь с максимальным, полученным в результате решения задачи оптимизации расписания хаба. Предложенный подход приводит к определению показателей C^D и C^D/C^P в нечеткой форме, что может затруднять их применение в аэропортовой практике на этапе принятия решений об организации или реконструкции трансферной системы перевозок. Выход из затруднений, очевидно, может состоять в использовании дефазифицированных значений показателей.

Проиллюстрируем примером процедуру определения рассмотренных показателей.

6. Модельный пример оценки эффективности трансферной системы

Предполагается, что в рамках модельной волны в хаб будут прибывать 10 ВС. В таблице 1 приведены нечеткие данные по потенциальному трансферному и начальному пассажиропотоку на авиасвязях, в таблице 2 – по трансферным тарифам, выраженным в некоторых условных единицах. Прочерки соответствуют авиасвязям, устойчивый интерес пассажиров к которым не прогнозируется. Все ВС разделены на две категории, разли-

чающиеся пассажироместимостью и временными параметрами технологического графика. ВС I категории, к которым отнесены три последние ВС, считаются большими, требующими больших затрат на аэропортовое обслуживание. Нечёткие продолжительности технологических операций $\alpha \in A'$ для ВС обеих категорий сведены в таблицу 3. Предполагается, что ресурсы одинакового назначения способны обслуживать ВС обеих категорий. Расчетная численность ресурсов, типы которых образуют множество Ω'' , принята равной одной единице. Исключение составляет численность ресурсов «трап», «тр.», «авт.», «р.н.», выделяемых для обслуживания ВС I категории, равная двум единицам. Прочим исходным технологическим параметрам заданы следующие значения: $k^B = 1,0$, $\tau^{1T} = 1$ мин., $w^{1T} = 50$, $\tau_{\min} = 5$ мин. Согласно данным таблиц 1 и 2, суммарные доходы от перевозки потенциальных трансферных пассажиров могут составить $\tilde{C}^{\Pi} = (4001,8672,13996)$ у.е., или, в форме чёткого числа, ${}^d C^{\Pi} = 8890$ у.е.

Для оценки влияния технических возможностей хаба на эффективность трансферной системы задача определения показателей её эффективности была решена в двух постановках: без учета ограничений по численности (35) и производительности (36) ресурсов аэропорта («Задача I») и с их учетом («Задача II»). В задаче I рассматривалась гипотетическая ситуация, при которой мощности аэропорта заведомо достаточны для обслуживания ВС, пассажиров и багажа в рамках исследуемой волны прибытий-отправлений.

Таблица 1. Пассажиропотоки, пас.

$i \backslash j$	1	2	3	4	5	6	7	8	9	10
$(L_{ip}^T, M_{ip}^T, R_{ij}^T)$										
1	8, 16, 24	-	-	-	-	1, 4, 9	-	-	-	-
2	-	-	-	-	1, 6, 12	-	9, 18, 29	-	1, 2, 3	1, 3, 7
3	-	-	18, 26, 40	2, 9, 16	-	-	-	1, 2, 5	-	-
4	-	-	-	9, 18, 30	-	-	-	-	3, 16, 27	10, 19, 30
5	-	-	35, 48, 70	-	-	-	-	1, 6, 10	-	-
6	-	5, 11, 17	-	-	-	13, 21, 29	2, 5, 9	-	3, 19, 30	-
7	-	-	-	32, 53, 75	-	-	-	-	-	-
8	-	12, 24, 36	-	-	1, 3, 7	6, 10, 18	1, 2, 3	-	-	42, 82, 123
9	-	1, 2, 4	-	-	-	-	2, 11, 26	-	1, 3, 7	-
10	-	15, 35, 48	-	-	-	-	28, 45, 60	1, 4, 8	5, 35, 55	-
$(L_{ip}^H, M_{ip}^H, R_{ij}^H)$										
	40, 65, 95	8, 10, 15	5, 7, 10	1, 2, 3	65 72, 100	40, 46, 64	1, 2, 3	110, 124, 170	55, 62, 78	25, 32, 40

Таблица 2. Трансферные тарифы, $({}^L C_{ij}^T, {}^M C_{ij}^T, {}^R C_{ij}^T)$ (у.е.)

$i \backslash j$	1	2	3	4	5	6	7	8	9	10
1	20,3, 21,0, 21,6	-	-	-	-	13,6, 14,2, 14,7	-	-	-	-
2	-	-	-	-	13,2, 13,8, 14,3	-	13,2, 13,9, 14,3	-	13,2, 13,8, 14,3	13,1, 13,7, 14,2
3	-	-	17,8, 18,4, 19,0	15,8, 16,4, 17,0	-	-	-	24,1, 24,8, 25,5	-	-
4	-	-	-	11,6, 12,2, 12,7	-	-	-	-	20,5, 21,2, 21,8	20,2, 21,5, 21,7
5	-	-	17,0, 17,6, 18,2	-	-	-	-	23,3, 24,0, 24,7	-	-
6	-	13,2, 13,8, 14,3	-	-	-	13,2, 13,8, 14,3	13,2, 13,7, 14,3	-	13,2, 13,8, 14,3	-
7	-	-	-	15,0, 15,6, 16,2	-	-	-	-	-	-
8	-	12,2, 12,8, 13,3	-	-	12,2, 12,7, 13,3	12,2, 12,8, 13,3	12,2, 12,9, 13,3	-	-	11,9, 12,8, 13,0
9	-	20,5, 21,2, 21,8	-	-	-	-	20,5, 21,2, 21,8	-	20,5, 21,2, 21,8	-
10	-	14,2, 14,8, 15,4	-	-	-	-	14,2, 14,8, 15,4	13,6, 14,2, 14,7	14,2, 14,8, 15,4	-

В этом случае суммарные доходы от перевозки действительных трансферных пассажиров составили $\tilde{C}^{DI} = (1941, 8297, 13452)$ у.е., или, в чёткой форме, ${}^d C^{DI} = 7897$ у.е., что на 11% ниже потенциально возможного их уровня ${}^d C^{PI}$.

Полученная в результате решения задачи I минимально достаточная с надежностью $p_h^3 = 0,95, \forall h \in \{1, \dots, H\}$, численность ресурсов составила по типам: «ст.» – 10 единиц, «уб.» – 7 единиц, «авт.» – 7 единиц, «тр.» – 7 единиц, «кар.» – 6 единиц, «запр.» – 9 единиц, «р.т.» – 12 единиц, «р.н.» – 14 единиц. Пропускная способность системы «обр.» – $g_{обр.}^p = 37$ 1/мин.

Таблица 3. Продолжительность операций, (${}^L\tau^\alpha, {}^M\tau^\alpha, {}^R\tau^\alpha$), мин.

Кат.	Операция, α										
	ПТ	УТ	ВП	ПП	ВБ	ПБ	ВКБ	ТО	ЗВС	УС	РНП
I	2,	3,	10,	15,	12,	16,	10,	55,	31,	18,	45,
	3,	4,	15,	21,	16,	20,	20,	62,	35,	20,	50,
	4	5	25	28	20	28	25	76	39	28	65
II	1,	1,	7,	11,	8,	13,	8,	45,	30,	12,	45,
	2,	2,	13,	16,	14,	16,	15,	52,	32,	16,	50,
	3	3	17	22	18	24	21	65	34	24	60

При решении задачи II предполагалось, что аэропорт располагает ресурсами типов «авт.» и «запр.» численности меньшей («авт.» – 5 ед., «запр.» – 7 ед.), чем получена при решении задачи I. Ограничения по численности аэропортовых ресурсов вызвали сокращение планируемых доходов до уровня $\tilde{C}^{II} = (1617, 7401, 12400)$ у.е., или, в чёткой форме, ${}^dC^{II} = 7140$ у.е., что почти на 20% ниже ${}^dC^I$.

С целью оценки эффекта от оптимизации, был произведен расчет величины ${}^dC^D$ для исходных данных, имитирующих традиционное расписание не узлового аэропорта. Предполагалось, что описанные выше ВС прибывают в аэропорт в порядке возрастания номеров через каждые 10 мин., проходят обслуживание в течение 2 ч, а затем в той же последовательности через каждые 10 мин. вылетают из аэропорта. Определенное в этом случае значение ${}^dC^D$ оказалось более чем в 2 раза ниже ${}^dC^I$. Таким образом, следует констатировать, что проектируемая трансферная система вряд ли может быть признана эффективной без использования оптимизации расписания.

Варианты оптимального расписания, составляющие один из результатов решения задач I и II, представлены на рис. 5 в виде

набора трапеций, где горизонтальный размер каждой трапеции и ее смещение вдоль оси абсцисс отражают временные характеристики пребывания соответствующего ВС в хабе. Треугольниками отображены ФП нечёткого времени прибытия и отправления рейсов. Отмечено также плановое время прибытия, получаемое в чёткой форме.

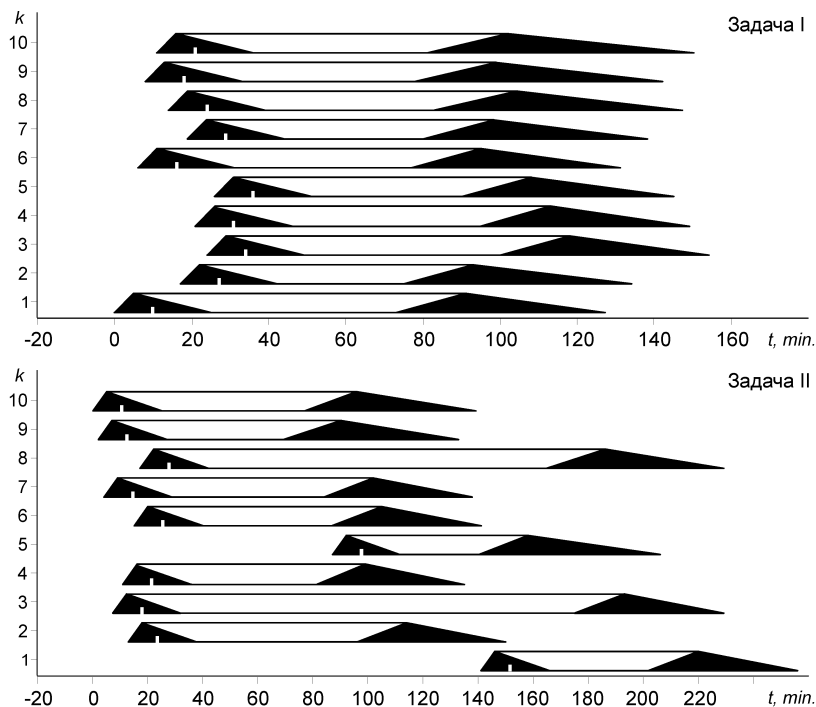


Рис. 5. Результаты оптимизации расписания

В отсутствие ограничений по ресурсам волновой характер движения ВС, являющийся одним из атрибутов хаба, проявляется вполне наглядно: ВС прибывают и отправляются тесными группами в течение непродолжительных промежутков времени. Наличие ограничений приводит к «размыванию» волны прибытий-отправлений, выражающемуся в значительно более позднем прибытии в хаб отдельных ВС по сравнению с основной их частью и большем разбросе моментов времени вылета ВС по

сравнению со случаем, когда ограничений на численность и производительность ресурсов нет. Так, позднее прибытие ВС 1 и 5 вызвано недостатком ресурсов хаба, не позволяющим выполнять одновременное обслуживание всех 10 ВС. Необычно долгая стоянка в аэропорту ВС 3 и 8, служащая причиной их более позднего, по сравнению с большинством других ВС, отправления, объясняется необходимостью ожидания многочисленных пассажиров, пересаживающихся с прибывающего позже остальных ВС 5, а также с самого ВС 3.

Рис. 6 позволяет оценить характер зависимости уровня загрузки ресурсов от времени в течение модельной волны.

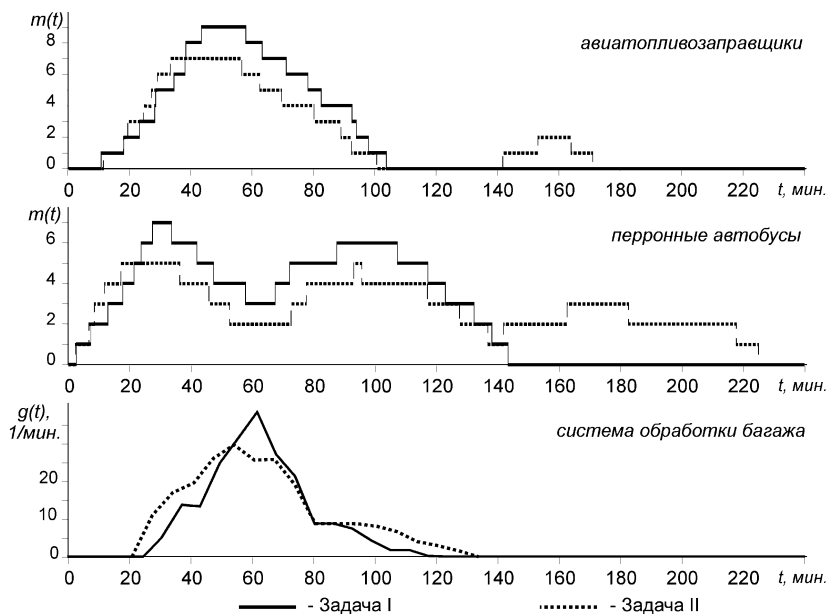


Рис. 6. Загрузка ресурсов аэропорта

В качестве примеров для вариантов расписания задач I и II приведены временные зависимости числа задействованных в обслуживании ВС авиатопливозаправщиков (запр.) и перронных автобусов (авт.), а также - интенсивности потока багажа, поступающего в СОБ. Введение ограничений на численность ресур-

сов обуславливает более равномерную их загрузку и сокращение потребности в тех ресурсах, на численность которых ограничения в явном виде не накладываются. Так, в задаче II уменьшение по сравнению с задачей I располагаемой численности ресурсов типов «авт.» и «запр.» привело к тому, что минимально достаточная численность ресурсов других типов так же сократилась и составила по типам: «ст.», «трап» – 8 ед., «уб.», «тр.» – 6 ед., «кар.» – 5 ед., «р.т.» – 10 ед., «р.н.» – 7 ед. Потребная пропускная способность системы «обр.» сократилась до $g_{обр.}^P = 29$ 1/мин. Таким образом, полученные результаты решения оптимизационных задач позволяют обосновать формирование сбалансированного по численности и производительности комплекса ресурсов аэропорта.

Продолжительность решения рассмотренных задач с использованием программного пакета IBM ILOG OPL на базе современной персональной вычислительной техники не превысила 10 мин.

7. Заключение

Представленные результаты свидетельствуют о работоспособности модели определения предложенных показателей, характеризующих эффективность трансферной системы пассажирских перевозок в плане доходности основной деятельности. Описанная модель может быть использована на этапе проектирования трансферной системы в условиях неопределенности исходной информации не только для поддержки принятия решения об экономической целесообразности и технической возможности выбора системой «авиакомпания–аэропорт» трансферной схемы перевозок, но также для решения ряда связанных задач, среди которых оптимизация расписания и формирование сбалансированного парка технологических ресурсов узлового аэропорта, рациональная организация работы его производственных комплексов и другие практически значимые задачи, вплоть до формирования сменного графика работы персонала и оперативного управления ресурсами.

Литература

1. ГОЛЕНКО Д.И. *Статистические методы сетевого планирования и управления*. – М.: Гл. ред. физ.-мат. лит-ры изд-ва «Наука», 1968. – 400 с.
2. ЛЮ Б. *Теория и практика неопределенного программирования*. – М.: БИНОМ, 2005. – 416 с.
3. ПЕГАТ А. *Нечёткое моделирование и управление*. 2-е изд., пер. с англ. – М.: БИНОМ Лаборатория знаний, 2013. – 798 с.
4. РОМАНЕНКО В.А. *Оптимизация параметров системы трансферных авиaperезовок с учетом нечеткой и стохастической неопределенностей* // Управление большими системами. – 2013. – №41. – С. 285–313.
5. ALVAREZ-VALDES R., CRESPO E., TAMARIT J.M. *Labour Scheduling at an Airport Refuelling Installation* // The J. of the Operational Research Society. – 1999. – Vol. 50(3). – P. 211–218.
6. ARAUJO G.E., REPOLHO H.M. *Optimizing the Airport Check-In Counter Allocation Problem* // J. of Transport Literature. – 2015. – Vol. 9(4). – P. 15–19.
7. BOOTSMA P.D. *Airline Flight Schedule Development: Analysis and Design Tools for European Hinterland Hubs*. – Utrecht, University of Twente – 1997. – 296 p.
8. BURGHOUWT G., REDONDI R. *Connectivity in Air Transport Networks: An Assessment of Models and Applications* // J. of Transport Economics and Policy. – 2013. – Vol. 47(1). – P. 35–53.
9. BURGHOUWT G., WIT J. *The Temporal Configuration of Airline Networks in Europe* // J. of Air Transport Management. – 2005. – Vol. 11(3). – P. 185–198.
10. DANESI A. *Measuring airline hub timetable co-ordination and connectivity: definition of a new index and application to a sample of European hubs* // European Transport. – 2006. – Vol. 34. – P. 54–74.
11. DENNIS N. *Airline Hub Operations in Europe* // J. of Transport Geography. – 1994. – Vol. 2(4). – P.219–233.

12. DENNIS N. *Scheduling Strategies for Airline Hub Operations* // J. of Air Transport Management. – 1994. – Vol. 1(3). – P. 131–144.
13. FITOURI-TRABELSI S., NUNES-CONSENZA C.A., MORA-CAMINO F.A.C. *Ground Handling Management at Airports with Fuzzy Information* // 6th IFAC Conf. on Management and Control of Production and Logistics MCPL, 2013 (Fortaleza, Brazil). International Federation of Automatic Control. – 2013. – Vol. 6 (1). – P. 373–378.
14. HANT C., CHUNGC C., LIANG G.-S. *Application of fuzzy critical path method to airports cargo ground operation systems* // J. of Marine Science and Technology. – 2006. – Vol. 14(3). – P. 139–146.
15. MASON A.J., RYAN D.M., PANTON D. M. *Integrated Simulation, Heuristic and Optimization Approaches to Staff Scheduling* // Operations Research. – 1998. – Vol. 46(2). – P. 161–175.
16. RONG A., GRUNOW M. *Manpower Planning at Air Cargo Terminals* // Proc. of the 9th Asia Pacific Industrial Engineering & Management Systems Conference and 11th Asia-Pacific Regional Meeting of the International Foundation for Production Research, 2008 (Nusa Dua, Bali). – 2008. – P. 666–677.
17. TRIETSCH D. *Scheduling Flights at Hub Airports* // Transportation Research Part B Methodological. – 1993. – Vol. 27(2). – P. 133–150.
18. WU C.-L. *Airline Operations and Delay Management — Insights from Airline Economics, Networks and Strategic Schedule Planning*. – Farnham: Ashgate, 2010. – 241p.
19. ZADEH L. A. *Fuzzy Sets* // Information and Control. – 1965. – Vol. 8. – P. 235– 350.

FUZZY ASSESSMENT OF PERFORMANCE OF AIR TRANSPORTATION TRANSFER SYSTEM

Ekaterina Guzha, Samara National Research University, Samara, postgraduate student (e.guzha@ya.ru).

Vladimir Romanenko, Samara National Research University, Samara, Candidate of Technical Sciences, Associate Professor (vla_rom@mail.ru).

Margarita Skorokhod, Samara National Research University, Samara, postgraduate student (skoro-margarita@yandex.ru).

Abstract: The transfer system of passenger air transportation as part of the hub airport and the hub-forming airline is considered. The efficiency of the system is characterized by the level of its revenues from the transportation of transfer passengers with an optimally formed timetable. It is assumed that uncomfortable waiting times for a transfer can cause an outflow of passengers potentially ready to be transported through the hub formed by flights of certain airline, while minimizing the damage caused by the outflow of passengers allows optimization of the hub's timetable. The proposed model for assessing the effectiveness of the projected transfer system includes, as a key component, the timetable optimization task, which involves determining such moments of arrival and departure of flights in the hub that maximize revenues from transportation of transfer passengers taking into account capacity constrained of airport production complexes. Uncertainty in the initial data of the model, characteristic for the design stage of the transfer system, is presented in a fuzzy set form. The model can be used to support decision making on the economic efficiency and technical capacity of forming a transfer system based on certain airlines and airports, for solving timetable optimization tasks and creating a balanced fleet of technological resources at the hub airport, rational organization of its industrial complexes, operational resource management. Some examples of application of the proposed model are considered.

Keywords: parametric optimization, fuzzy numbers, hub airport, performance criterion.

УДК 510.22:519.85:656.71

ББК 22.12 22.18 39.51

DOI: <https://doi.org/10.25728/ubs.2019.77.8>

Статья представлена к публикации членом редакционной коллегии М.В. Губко.

Поступила в редакцию 13.07.2018.

Опубликована 31.01.2019.