Институт проблем управления им. В.А. Трапезникова РАН

УПРАВЛЕНИЕ БОЛЬШИМИ СИСТЕМАМИ

Выпуск 92
Июль 2021

СБОРНИК ТРУДОВ

ISSN 1819-2467

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

Москва – 2021

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

УПРАВЛЕНИЕ БОЛЬШИМИ СИСТЕМАМИ

СБОРНИК ТРУДОВ

Выпуск 92

Москва – 2021

УДК 519 ISSN 1819-2467
ББК 32.81
У 67

Управление большими системами / Сборник трудов. – Вып. 92. – М.: ИПУ РАН, 2021. – 109 с. **Дата опубликования:** 31.07.2021.

Адрес редакции: 117997, г. Москва, ул. Профсоюзная, д. 65.
Адрес в интернете: ubs.mtas.ru.

СОДЕРЖАНИЕ

ПРИНЦИП МИНИМУМА ДОХОДНОСТИ И CC-VAR ПРИ ЧАСТИЧНОМ ПРОГНОЗЕ РЫНКА

Агасандян Г. А.[1]
(*Вычислительный центр им. А.А. Дородницына ФИЦ ИУ РАН, Москва*)

Работа продолжает исследования, связанные с применением континуального критерия VaR (CC-VaR) на рынках опционов. Рассматривается ситуация, когда прогноз инвестора относительно будущих вероятностных свойств базового актива ограничивается лишь частичными суждениями. Неполнота прогноза моделируется введением в прогноз некоторых параметров, выбор значений которых инвестор доверяет рынку. Постулируется принцип минимума доходности (ПМД), согласно которому инвестору следует назначать свободные параметры из соображений минимизации доходности инвестиции. Тем самым инвестор приобретает определенную гарантию от возможных ошибок в прогнозе. Исследуются теоретические свойства введенного принципа, имеющие самостоятельный интерес, а в ряде случаев упрощающие анализ результатов его применения. Демонстрация его работы проводится аналитически на двусторонних экспоненциальных и равномерных распределениях и численными методами на двухпараметрических бета-распределениях. Результаты подтверждают адекватность принципа и алгоритмов расчетов.

Ключевые слова: континуальный критерий VaR (CC-VaR), функция рисковых предпочтений (ф.р.п.), частичность прогноза, процедура Неймана-Пирсона, диссонанта, функция упорядочения, доходность, принцип минимума доходности (ПМД), волатильность, регрессия.

1. Введение

В работе изучаются проблемы, связанные с применением континуального критерия VaR (CC-VaR) [1, 2] на рынках опционов. Рассматривается ситуация, когда прогноз инвестора относительно будущих вероятностных свойств базового актива лишь частичен. Неполнота прогноза моделируется введением некото-

---

[1] *Геннадий Аршавирович Агасандян, д.ф.-м.н.* (*agasang17@yandex.ru*).

рых параметров, выбор значений которых инвестор доверяет рынку.

Применяется *принцип минимума доходности* (ПМД), позволяющий в рамках обшей задачи связать, фактически, две подзадачи – оценивания неизвестных параметров прогноза и оптимизации по CC-VaR портфеля инвестора при известном параметре. Принцип минимума доходности предлагает инвестору выбирать такое значение параметра неполноты, которое минимизирует доходность инвестиции. Введение ПМД представляется вполне естественным, поскольку предоставляет инвестору определенную надежность при выборе портфеля.

Проводятся теоретические исследования принципа и устанавливаются некоторые его свойства, упрощающие анализ особенностей его применения в конкретных случаях. На примере двухпараметрического двустороннего экспоненциального распределения демонстрируется возможность находить результаты применения ПМД аналитическими средствами. Аналитически решаются задачи с ПМД и для равномерных распределений, хотя они для CC-VaR образуют экзотический случай. Наконец, двухпараметрические бета-распределения предоставляют пример распределений, для которых без привлечения методики численных расчетов, считающейся в рассматриваемых задачах универсальной, не удается обойтись.

2. Формализация принципа минимума доходности

Как и при изучении многих иных теоретических проблем с континуальным критерием, здесь решается базовая континуальная *задача СВ* [1, 2], в которой рассматривается одноперидный теоретический идеальный рынок, инвестиционная сумма A (> 0) не задается, а оптимальный портфель ищется из условия его регулярности и минимума стоимости. Сам критерий требует, чтобы портфельный доход q удовлетворял условию

$$\mathsf{P}\{q \ge \phi(\varepsilon)\} \ge 1-\varepsilon \quad \text{для всех } \varepsilon \in [0, 1],$$

где $\phi(\varepsilon)$ – *функция рисковых предпочтений* (ф.р.п.) инвестора, $\mathsf{P}\{S\}$ – вероятность множества S. При необходимости в расчетах принимается $\phi(\varepsilon; \lambda) = \varepsilon^{\lambda}$, или даже с $\lambda = 2$.

Частичность прогноза моделируется (возможно, векторным) параметрическим заданием семейства прогнозных плотностей вероятности $\{p(x; \theta), \theta \in \Theta\}$ (и порождаемых ими мер $\mathsf{P}\{\cdot; \theta\}$). Не следует путать параметр θ с возможным параметром риска λ у ф.р.п. $\phi(\varepsilon; \lambda)$, характеризующим готовность инвестора к риску и единым для инвестора во всех задачах с CC-VaR.

Алгоритм оптимизации по CC-VaR с применением ПМД состоит в том, что для каждого возможного значения θ проводится стандартная оптимизация по CC-VaR, континуальная или дискретная в зависимости от типа задачи [2–4], основанная на процедуре Неймана – Пирсона [5], и находится доходность $y(\theta)$. А затем согласно ПМД определяется значение параметра неопределенности, минимизирующее доходность инвестиции.

В условиях частичного прогноза инвестора привычные агрегаты задачи, такие как функция относительных доходов, прогнозная функция (относительных доходов), стоимостная функция (относительных доходов) и диссонанта, по необходимости снабжаются дополнительным аргументом $\theta \in \Theta$ и записываются соответственно соотношениями

(1) $\rho(x;\theta) = p(x;\theta)/c(x), \quad \theta \in \Theta,$

(2) $\mathsf{f}_{\mathsf{P}}(\tau;\theta) = \mathsf{P}\{\rho(x;\theta) \le \tau\}, \quad \mathsf{f}_{\mathsf{C}}(\tau;\theta) = \mathsf{C}\{\rho(x;\theta) \le \tau\}, \quad \tau \ge 0,$

(3) $\gamma(\varepsilon;\theta) = \mathsf{f}_{\mathsf{C}}\left(\mathsf{f}_{\mathsf{P}}^{\leftarrow}(\varepsilon;\theta);\theta\right), \quad \varepsilon \in [0,1].$

Функция рисковых предпочтений инвестора $\phi(\varepsilon)$, $\varepsilon \in [0, 1]$, разумеется, не зависит от θ. При фиксированном значении θ оптимизация по CC-VaR дает средний доход R (не зависящий от θ), стоимость портфеля $A(\theta)$ и среднюю доходность инвестиции $y(\theta)$, равные соответственно (с учетом (1)–(3))

(4) $R = \int_0^1 \phi(\varepsilon)\,d\varepsilon, \quad A(\theta) = \int_0^1 \phi(\varepsilon)\,d\gamma(\varepsilon;\theta) = \int_0^\infty \phi\left(\mathsf{f}_{\mathsf{P}}(\tau)\right)d\mathsf{f}_{\mathsf{C}}(\tau),$

(5) $y(\theta) = R/A(\theta) - 1 = R\Big/\int_0^1 \phi(\varepsilon)\,d\gamma(\varepsilon;\theta) - 1.$

Поэтому применение ПМД сводится к нахождению значения параметра $\theta = \theta_{min}$, *минимизирующего* доходность $y(\theta)$ или, что то же, *максимизирующее* инвестиционную сумму $A(\theta)$:

(6) $$\theta_{min} = \arg\min_\theta y(\theta) = \arg\max_\theta \int_0^1 \phi(\varepsilon) d\gamma(\varepsilon;\theta).$$

3. Вспомогательные теоретические результаты

Задача инвестора формально будет описываться парой $\pi = (\varphi(x), \psi(x))$, первая компонента которой означает стоимостную плотность, а вторая – прогнозную. Напомним [2], что пара плотностей (также порождаемых ими мер) и задача в целом называются регулярными (невырожденными), когда диссонанта непрерывна и строго возрастает на [0, 1]. Рассмотрим свойства, связанные с преобразованиями цены базового актива, которые интересны сами по себе, а в ряде случаев бывают полезными.

Лемма 1. Пусть задача $\pi_1 = (c(x), p(x))$ порождает диссонанту $\gamma_1(\varepsilon)$, $\varepsilon \in [0, 1]$, и функцию упорядочения $w_1(x)$, $x \in \Re$. Тогда при линейном преобразовании цены актива $X_2 = \mu + \nu X_1$, $\mu \in \Re$, $\nu > 0$, для задачи $\pi_2 = (c(x - \mu), p(x - \mu))$ имеет место

(i) $\gamma_2(\varepsilon) = \gamma_1(\varepsilon)$, $\varepsilon \in [0, 1]$;

(ii) $w_2(x) = w_1\left(\frac{x-\mu}{\nu}\right)$, $x \in \Re$.

Доказательство. Задаче π_1 в алгоритме оптимизации отвечает функция относительных доходов $\rho_1(x) \equiv p_1(x)/c_1(x)$, а система оптимальных множеств $Z_1(\tau)$ [2] определяется условием

$$Z_1(\tau) = \left\{x \middle| \rho_1(x) \le \tau\right\}, \quad \tau \ge 0.$$

Для задачи π_2 по свойствам плотности вероятности имеем

$$\rho_2(x) \equiv p_2(x)/c_2(x) \equiv p_1\left(\tfrac{x-\mu}{\nu}\right)\Big/c_1\left(\tfrac{x-\mu}{\nu}\right) \equiv \rho_1\left(\tfrac{x-\mu}{\nu}\right).$$

Очевидны эквивалентности

$$\rho_2(x) = \tau \quad \Leftrightarrow \quad \rho_1\left(\tfrac{x-\mu}{\nu}\right) = \tau, \qquad x \in Z_2(\tau) \quad \Leftrightarrow \quad \tfrac{x-\mu}{\nu} \in Z_1(\tau).$$

Поэтому

$$\mathsf{P}_1\{Z_1(\tau)\} = \mathsf{P}_2\{Z_2(\tau)\} = \varepsilon, \qquad \mathsf{C}_1\{Z_1(\tau)\} = \mathsf{C}_2\{Z_2(\tau)\} = \gamma(\varepsilon).$$

Второе из этих равенств фактически означает справедливость утверждения (i) леммы. Кроме того,

$$\varepsilon \equiv \mathsf{f}_{\mathsf{P};1}(\tau) = \mathsf{P}_1\{Z_1(\tau)\}, \quad \varepsilon \equiv \mathsf{f}_{\mathsf{P};2}(\tau) = \mathsf{P}_2\{Z_2(\tau)\},$$

и потому

$$\mathsf{f}_{\mathsf{P};1}(\tau) = \mathsf{f}_{\mathsf{P};2}(\tau) = \mathsf{f}_{\mathsf{P}}(\tau), \quad \mathsf{f}_{\mathsf{P}}(\rho_1(x)) = \mathsf{f}_{\mathsf{P}}(\rho_2(x)).$$

Согласно определению функции упорядочения [2] имеем

$$w_1(x) = \mathsf{f}_{\mathsf{P}}(\rho_1(x)),$$

$$w_2(x) = \mathsf{f}_{\mathsf{P}}(\rho_2(x)) = \mathsf{f}_{\mathsf{P}}\left(\rho_1\left(\tfrac{x-\mu}{\nu}\right)\right) = w_1\left(\tfrac{x-\mu}{\nu}\right),$$

что доказывает утверждение (ii) и лемму в целом. □

Как правило, если не оговорено противное, предполагается, что диссонанта на отрезке [0, 1] *непрерывна* и *строго монотонно возрастает*. Она вообще может претерпевать разрыв лишь при $\varepsilon = 0$, а сохранять постоянное значение – лишь при $\gamma(\varepsilon) = 1$, но эти случаи рассматриваются как экзотические, хотя они часто допускают осмысления.

Лемма 2. Пусть $\gamma(\varepsilon)$ и $w(x)$ – соответственно диссонанта и функция упорядочения для регулярной задачи $\pi = (c(x), p(x))$. Тогда функции $\gamma^*(\varepsilon)$ и $w^*(x)$, порождаемые *сопряженной* задачей $\pi^* = (p(x), c(x))$, определяются равенствами

(i) $\gamma^*(\varepsilon) = 1 - \gamma^{\leftarrow}(1-\varepsilon), \quad \varepsilon \in [0, 1];$

(7) (ii) $w^*(x) = 1 - \gamma(w(x)), \quad x \in \mathsf{X}.$

Доказательство. Каждая из функций относительных доходов порождает свое семейство множеств по $\tau \geq 0$:

$$Z(\tau) = \{x \mid \rho(x) \leq \tau\}, \quad Z^*(\tau) = \{x \mid \rho^*(x) \leq \tau\}.$$

Очевидно,

$$\rho(x) \leq \tau \Leftrightarrow \rho^*(x) \geq 1/\tau, \quad Z^*(\tau) = \overline{Z(1/\tau)} \cup \{\rho(x) = 1/\tau\},$$

и потому для регулярной задачи

(8) $\mathsf{P}\{Z^*(\tau)\} = \mathsf{P}\{\overline{Z(1/\tau)}\} = 1 - \mathsf{P}\{Z(1/\tau)\},$

(9) $\mathsf{C}\{Z^*(\tau)\} = \mathsf{C}\{\overline{Z(1/\tau)}\} = 1 - \mathsf{C}\{Z(1/\tau)\}.$

Также

$$(10)\quad \varepsilon = \mathsf{C}\left\{Z^*(\tau)\right\} = \mathsf{C}\left\{\overline{Z(1/\tau)}\right\} = 1 - \mathsf{C}\left\{Z(1/\tau)\right\},$$

$$\gamma^*(\varepsilon) = \mathsf{P}\left\{Z^*(\tau)\right\} = \mathsf{P}\left\{\overline{Z(1/\tau)}\right\} = 1 - \mathsf{P}\left\{Z(1/\tau)\right\}.$$

Из определения диссонанты

$$\gamma\left(\mathsf{P}\left\{Z(1/\tau)\right\}\right) = \mathsf{C}\left\{Z(1/\tau)\right\}, \quad \tau > 0,$$

или, с учетом (8) и (9),

$$\gamma\left(1 - \gamma^*(\varepsilon)\right) = 1 - \varepsilon,$$

откуда и следует утверждение (i) леммы.

Обратимся к доказательству утверждения (ii). Имеем

$$w(x) = \mathsf{f}_{\mathsf{P}}\left(\rho(x)\right).$$

Функция упорядочения для задачи π^* находится из (10):

$$w^*(x) = \mathsf{f}_{\mathsf{C}}\left(\rho^*(x)\right) = \mathsf{C}\left\{Z^*\left(\rho^*(x)\right)\right\} = 1 - \mathsf{C}\left\{\rho(x) \le 1/\rho^*(x)\right\} =$$

$$= 1 - \gamma\left(\mathsf{f}_{\mathsf{P}}(1/\tau)\right)\Big|_{\tau=\rho^*(x)} = 1 - \gamma\left(\mathsf{f}_{\mathsf{P}}\left(\rho(x)\right)\right) = 1 - \gamma\left(w(x)\right).$$

Таким образом, при сопряжении задачи действует правило

$$(11)\quad w^*(x) = 1 - \gamma\left(w(x)\right).$$

Его инвертированием определяется и функция упорядочения

$$(12)\quad w(x) = \gamma^{\leftarrow}\left(1 - w^*(x)\right).$$

Некоторая асимметрия этих формул устраняется, если подставить в формулу (7) вместо аргумента ε функцию $w^*(x)$:

$$\gamma^*\left(w^*(x)\right) = 1 - \gamma^{\leftarrow}\left(1 - w(x)\right).$$

В сочетании с (12) получается симметричная к (11) формула

$$w(x) = 1 - \gamma^*\left(w^*(x)\right).$$

Подобные логические построения, основанные на взаимной симметрии задач, могут сами по себе служить доказательством утверждения (ii). Лемма полностью доказана. □

Лемма 3. Если $\gamma_1(\varepsilon) \le \gamma_2(\varepsilon)$, $\varepsilon \in [0, 1]$, то для любой неубывающей и ограниченной функции $\varphi(\varepsilon)$

$$\int_0^1 \varphi(\varepsilon)\, d\gamma_1(\varepsilon) \ge \int_0^1 \varphi(\varepsilon)\, d\gamma_2(\varepsilon).$$

Доказательство. Поскольку диссонанты в нуле и единице равны соответственно нулю и единице, интегрированием по частям получается неравенство

$$\int_0^1 \varphi(\varepsilon)\, d\left(\gamma_1(\varepsilon)-\gamma_2(\varepsilon)\right) = -\int_0^1 \left(\gamma_1(\varepsilon)-\gamma_2(\varepsilon)\right) d\varphi(\varepsilon) \geq 0$$

(внеинтегральное слагаемое обращается в нуль). Тем самым утверждение леммы доказано. □

Очевидным следствием леммы 3 является

Лемма 4. Если существует значение $\theta' \in \Theta$, при котором для всего семейства диссонант $\{\gamma(\varepsilon;\theta), \theta \in \Theta\}$ выполняется неравенство

$$\gamma(\varepsilon;\theta') \leq \gamma(\varepsilon;\theta), \quad \varepsilon \in [0, 1],$$

то θ' удовлетворяет (6) при определениях (4), (5) с $\theta' = \theta_{min}$.

Доказательство. Достаточно применить лемму 3 к функции $\varphi(\varepsilon) \equiv \phi(\varepsilon)$ и положить $\gamma_1(\varepsilon) = \gamma(\varepsilon;\theta')$, $\gamma_2(\varepsilon) = \gamma(\varepsilon;\theta)$. □

Таким образом, если в семействе диссонант $\{\gamma(\varepsilon;\theta), \theta \in \Theta\}$ существует минимальный элемент, то соответствующее этому элементу значение θ' параметра θ является оптимальным по ПМД (и не зависит от функции $\phi(\varepsilon)$. В отсутствие минимального элемента это, вообще говоря, не так, и следует обращаться к самой формуле (6).

Леммы 3, 4 можно переформулировать в терминах прогнозной и стоимостной функций:

$$\int_0^\infty \varphi\left(\mathsf{f}_{\mathsf{P};1}(\tau)\right) d\mathsf{f}_{\mathsf{C}}(\tau) \geq \int_0^\infty \varphi\left(\mathsf{f}_{\mathsf{P};2}(\tau)\right) d\mathsf{f}_{\mathsf{C}}(\tau),$$

$$\mathsf{f}_{\mathsf{P}}(\tau;\theta_{min}) \geq \mathsf{f}_{\mathsf{P}}(\tau;\theta), \quad \tau \in [0,\infty).$$

4. Двустороннее экспоненциальное распределение

Рассматривается пример задачи оптимизации по CC-VaR для δ-рынка, в которой прогнозная $p(\cdot)$ и стоимостная $c(\cdot)$ плотности подчинены билатеральному (двустороннему) экспоненциальному распределению с параметрами μ, α, $\mu \in \Re$, $\alpha > 0$:

$$\mathrm{Exp}(\mu,\alpha): \quad \tfrac{1}{2\alpha}\exp\left(-\tfrac{|x-\mu|}{\alpha}\right), \quad x \in \mathsf{X} = \Re.$$

при этом $p(x) \sim \mathrm{Exp}(\mu, \alpha)$, $\mu > 0$, $c(x) \sim \mathrm{Exp}(0, \beta)$, т.е.

$$p(x)=\tfrac{1}{2\alpha}\exp\left(-\tfrac{|x-\mu|}{\alpha}\right),\quad c(x)=\tfrac{1}{2\beta}\exp\left(-\tfrac{|x|}{\beta}\right).$$

Здесь среднее ν второго распределения приравнивается нулю для простоты (и без ограничения общности). Функция относительных доходов

$$\rho(x)=p(x)/c(x)=\tfrac{\beta}{\alpha}\exp\left(-\tfrac{|x-\mu|}{\alpha}+\tfrac{|x|}{\beta}\right),\quad x\in\Re.$$

4.1. АНАЛИЗ ДИССОНАНТЫ ПРИ $\mu > \nu$, $\alpha < \beta$

Пусть теперь $\kappa = \alpha/\beta < 1$. Это условие означает, что инвестор считает рынок более волатильным, чем об этом свидетельствуют рыночные цены. В таком случае функция $\rho(x)$ унимодальна, принимает максимальное значение $\tau_{max} = \exp(\mu/\beta)/\kappa$ при $x_{max} = \mu$ с изломом и претерпевает дополнительный излом в точке $x = 0$ со значением $\tau_{br} = \exp(-\mu/\alpha)/\kappa$. При $x \to -\infty$ функция стремится к минимальному значению $\tau_{min} = 0$.

Алгоритм оптимизации по CC-VaR требует нахождения оптимальных по Нейману-Пирсону множеств $\{Z(\tau), \tau \ge 0\}$ [2, 5]. Это семейство определяется из неравенств

$$\rho(x)\le\tau,\quad 0<\tau\le\tau_{max}.$$

Разрешая их относительно x, получаем эквивалентную систему неравенств. При $x \le 0$, $0 < x \le \mu$, $\mu \le x$ соответственно

$$x\le\tfrac{\alpha\beta}{\beta-\alpha}\left(\tfrac{\mu}{\alpha}+\ln\left(\tfrac{\tau\alpha}{\beta}\right)\right),\ x\le\tfrac{\alpha\beta}{\beta+\alpha}\left(\tfrac{\mu}{\alpha}+\ln\left(\tfrac{\tau\alpha}{\beta}\right)\right),\ x\ge\tfrac{\alpha\beta}{\beta-\alpha}\left(\tfrac{\mu}{\alpha}-\ln\left(\tfrac{\tau\alpha}{\beta}\right)\right).$$

Производя замену $t = \alpha \ln(\tau\alpha/\beta)$ переменной τ, получаем ту же систему неравенств в терминах t:

$$x\le(\mu+t)/(1-\kappa),\quad x\le(\mu+t)/(1+\kappa),\quad x\ge(\mu-t)/(1-\kappa).$$

В терминах новой переменной $t_{min} = -\infty$, $t_{br} = -\mu$, $t_{max} = \mu\kappa$, а оптимальные по Нейману – Пирсону множества имеют вид:

$$Z(t)=\{x\le(\mu+t)/(1-\kappa)\}\cup\{x\ge(\mu-t)/(1-\kappa)\},\quad t\le-\mu,$$

$$Z(t)=\{x\le(\mu+t)/(1+\kappa)\}\cup\{x\ge(\mu-t)/(1-\kappa)\},\ -\mu<t\le\mu\kappa.$$

Множество $Z(t)$ при $t = \mu\kappa$ совпадает с $\Re_+$, а при $t\to-\infty$ имеет пределом пустое множество. Переменная t связывается с вероятностным уровнем $\varepsilon = \mathsf{f}(t)$ соотношением

$$\varepsilon = \mathsf{f}(t) = \mathsf{P}\{Z(t)\} = \tfrac{1}{2\alpha}\int_{Z(t)} \exp\left(\tfrac{x-\mu}{\alpha}\right)dx .$$

Интегрированием находится

$$\varepsilon = \begin{cases} \tfrac{1}{2}\exp\left(\tfrac{t+\kappa\mu}{\alpha-\kappa\alpha}\right) + \tfrac{1}{2}\exp\left(\tfrac{t-\kappa\mu}{\alpha-\kappa\alpha}\right), & t \le -\mu, \\ \tfrac{1}{2}\exp\left(\tfrac{t-\kappa\mu}{\alpha+\kappa\alpha}\right) + \tfrac{1}{2}\exp\left(\tfrac{t-\kappa\mu}{\alpha-\kappa\alpha}\right), & t > -\mu; \end{cases}$$

$$\left(\varepsilon_{\mathrm{br}} = \tfrac{1}{2}\exp\left(-\tfrac{\mu}{\alpha}\right)\left(1+\exp\left(\tfrac{1+\kappa}{1-\kappa}\right)\right), \quad t_{\mathrm{br}} = -\mu\right).$$

А диссонанта

$$\gamma(\varepsilon) = \mathsf{C}\{Z(t)\} = \tfrac{1}{2\beta}\int_{Z(t)} \exp\left(\tfrac{x}{\beta}\right)dx =$$

$$= \begin{cases} \tfrac{1}{2}\exp\left(\tfrac{t+\mu}{\beta-\kappa\beta}\right) + \tfrac{1}{2}\exp\left(\tfrac{t-\mu}{\beta-\kappa\beta}\right), & t \le -\mu, \\ 1-\tfrac{1}{2}\exp\left(-\tfrac{t+\mu}{\beta+\kappa\beta}\right) + \tfrac{1}{2}\exp\left(\tfrac{t-\mu}{\beta-\kappa\beta}\right), & t > -\mu; \end{cases}$$

$$\left(\gamma_{\mathrm{br}} = \gamma(\varepsilon_{\mathrm{br}}) = \tfrac{1}{2} + \tfrac{1}{2}\exp\left(-\tfrac{2\mu}{\beta-\kappa\beta}\right)\right).$$

Теперь проводится еще замена $t \to u$:

$$u = \upsilon\exp\left(\tfrac{t}{\alpha-\kappa\alpha}\right), \quad \upsilon = \exp\left(\tfrac{-\mu}{\beta-\kappa\beta}\right) < 1.$$

При этом $u_{min} = 0$, $u_{br} = \upsilon^{(1+\kappa)/\kappa}$, $u_{max} = 1$,

(13) $\varepsilon = \tfrac{1}{2}\tfrac{u}{\upsilon}\left(\upsilon^{-1}+\upsilon\right), \quad \gamma(\varepsilon) = \tfrac{1}{2}\tfrac{u^{\kappa}}{\upsilon^{\kappa}}\left(\upsilon^{-1}+\upsilon\right), \quad u \in \left(0, \upsilon^{(1+\kappa)/\kappa}\right],$,

(14) $\varepsilon = \tfrac{1}{2}\left(u+u^{\omega}\right), \quad \gamma(\varepsilon) = 1-\tfrac{\upsilon^{1-\kappa}}{2u^{\kappa\omega}}\left(1-u^{2\kappa/(1+\kappa)}\right), \quad u \in \left(\upsilon^{(1+\kappa)/\kappa}, 1\right],$,

где $\omega = (1-\kappa)/(1+\kappa)$, а $\varepsilon_{br} = \upsilon^{1/\kappa}(\upsilon^{-1}+\upsilon)/2$, $\gamma_{br} = (1+\upsilon^2)/2$.

Формулы (13) и (14) дают параметрическое задание диссонанты $\gamma(\varepsilon)$. Явное представление $\gamma(\varepsilon)$ удается получить лишь на первом интервале, тем не менее можно провести необходимое оценивание для применения ПМД. В связи с этим рассматриваются две задачи, в первой из которых в качестве свободного параметра распределения инвестора (именно он и находится из ПМД) выступает параметр μ, во второй – α.

4.2. СВОБОДНЫЙ ПАРАМЕТР μ

Рассмотрим в качестве свободного параметр μ. На основании (14) легко устанавливается, что при $\mu = 0$ и $\kappa < 1$

(15) $\gamma\left(\varepsilon;0,\kappa\right)=\varepsilon^{\kappa}$, $\varepsilon \in [0, 1]$.

Здесь в обозначение диссонанты дополнительно вводятся идентификационные параметры. Для доказательства оптимальности по ПМД значения $\mu = 0$ достаточно показать, что для всех α, β и $\kappa = \alpha/\beta < 1$

(16) $\gamma\left(\varepsilon;\mu,\kappa\right)\ge\gamma\left(\varepsilon;0,\kappa\right)$, $\mu>0$, $\kappa<1$, $\varepsilon \in [0, 1]$.

Из соотношений (13) и (14) переменная u выражается через ε, подставляется в (16) и находятся нужные оценки для двух интервалов (для второго используется неравенство $u^{\kappa\omega} \ge \upsilon^{1-\kappa}$):

$$\gamma\left(\varepsilon;\mu,\kappa\right)=2^{\kappa-1}\left(\upsilon^{-1}+\upsilon\right)^{1-\kappa}\varepsilon^{\kappa}\ge\varepsilon^{\kappa},\quad u\in\left(0,\upsilon^{(1+\kappa)/\kappa}\right],$$

$$\gamma\left(\varepsilon;\mu,\kappa\right)\ge\tfrac{1}{2}\left(1+u^{2\kappa/(1+\kappa)}\right),\quad u\in\left(\upsilon^{(1+\kappa)/\kappa},1\right].$$

Справедливость неравенства (16) (с учетом (15)) устанавливается, если будет доказано, что

$$\tfrac{1}{2}\left(1+u^{2\kappa/(1+\kappa)}\right)\ge\tfrac{1}{2^{\kappa}}\left(u+u^{(1-\kappa)/(1+\kappa)}\right)^{\kappa}\quad\left(=\varepsilon^{\kappa}\right).$$

После замены в последнем неравенстве $u \to z^{1+\kappa}$ и деления на z^{κ} получается эквивалентное неравенство

$$\tfrac{1}{2}\left(z^{-\kappa}+z^{\kappa}\right)\ge\left(\tfrac{1}{2}\left(z^{-\kappa}+z^{\kappa}\right)\right)^{\kappa},$$

которое справедливо, поскольку выражение слева не меньше единицы, а $\kappa < 1$. Неравенство (16) доказано.

4.3. СВОБОДНЫЙ ПАРАМЕТР α

Рассмотрим теперь задачу со свободным параметром α и найдем диссонанту $\gamma(\varepsilon)$ при $\mu > 0$ и $\kappa = 1$. Дальнейшему изложению предпошлем

Замечание. Соображения размерности позволяют уменьшить количество параметров в рассматриваемой задаче. В исходной постановке мы имеем наряду с переменной x набор параметров $(\mu, \alpha, \nu, \beta)$, и потому важные для задачи функции $\rho(x)$, $w(x)$ и $\gamma(\varepsilon)$ зависят, вообще говоря, еще от этих четырех параметров. Однако, как нетрудно видеть, существенными являются лишь комбинации (здесь могут возникать варианты) $x' = (x - \nu)/\beta$, $\mu' = (\mu - \nu)/\beta$, $\kappa = \alpha/\beta$, и потому

$$\rho(x;\mu,\alpha,\nu,\beta)=\widehat{\rho}\left(\tfrac{x-\nu}{\beta};\tfrac{\mu-\nu}{\beta},\tfrac{\alpha}{\beta}\right)=\widehat{\rho}(x';\mu',\kappa),$$

$$w(x;\mu,\alpha,\nu,\beta)=\widehat{w}\left(\tfrac{x-\nu}{\beta};\tfrac{\mu-\nu}{\beta},\tfrac{\alpha}{\beta}\right)=\widehat{w}(x';\mu',\kappa),$$

$$\gamma(\varepsilon;\mu,\alpha,\nu,\beta)=\widehat{\gamma}\left(\varepsilon;\tfrac{\mu-\nu}{\beta},\tfrac{\alpha}{\beta}\right)=\widehat{\gamma}(\varepsilon;\mu',\kappa).\ \square$$

Для упрощения записи формул полагаем $\beta = 1$ (иначе в них нужно произвести замены $x\to x' = x/\beta$, $\mu \to \mu' = \mu/\beta$), и тогда

$$c(x)=\tfrac{1}{2}\exp(-|x|),\quad p(x)=\tfrac{1}{2}\exp(-|x-\mu|),$$

$$\rho(x)=\exp(-|x-\mu|+|x|).$$

Очевидно,

$$\rho(x)=\left\{\mathrm{e}^{-\mu},\ x\le x_1;\ \mathrm{e}^{2x-\mu},\ x_1<x\le x_2;\ \mathrm{e}^{\mu},\ x_2<x\right\}=\tau,$$

$$x_1=0,\ x_2=\mu,\ \tau\in\left[\mathrm{e}^{-\mu},\mathrm{e}^{\mu}\right].$$

Параметризация $\rho(x) = \tau$ и надлежащее интегрирование по мерам $\mathsf{P}\{\cdot\}$ и $\mathsf{C}\{\cdot\}$ дает

$$\varepsilon_1=\mathsf{P}\{x\le x_1\}=\mathrm{e}^{-\mu}/2,\quad \varepsilon_2=1-\mathsf{P}\{x>x_2\}=1/2,$$

$$\varepsilon=\mathrm{f}(\tau)=\tfrac{1}{2}\mathrm{e}^{-\mu/2}\sqrt{\tau},\quad \tau\in\left(\mathrm{e}^{-\mu},\mathrm{e}^{\mu}\right),\quad \tau=\mathrm{f}^{\leftarrow}(\varepsilon)=4\varepsilon^2\,\mathrm{e}^{\mu},$$

$$\gamma_1=\mathsf{C}\{x\le x_1\}=1/2,\quad \gamma_2=1-\mathsf{C}\{x>x_2\}=1-\mathrm{e}^{-\mu}/2,$$

$$\gamma(\varepsilon)=1-\tfrac{1}{2}\mathrm{e}^{-\mu/2}\Big/\sqrt{\tau}\,\Big|_{\tau=\mathrm{f}^{\leftarrow}(\varepsilon)},\quad \tau\in\left(\mathrm{e}^{-\mu},\mathrm{e}^{\mu}\right).$$

Образуя требуемую суперпозицию $\gamma(\varepsilon)$ и используя значения параметров ε_1, ε_2, γ_1, γ_2, получаем представление

$$\text{(17)}\quad \gamma(\varepsilon;\mu,1)=\begin{cases}\mathrm{e}^{-\mu}\,\varepsilon, & 0\le\varepsilon\le\mathrm{e}^{-\mu}/2;\\ 1-\tfrac{1}{4}\mathrm{e}^{-\mu}/\varepsilon, & \mathrm{e}^{-\mu}/2<\varepsilon\le 1/2;\\ 1-\mathrm{e}^{-\mu}(1-\varepsilon), & 1/2<\varepsilon\le 1.\end{cases}$$

Здесь для более точной идентификации запись диссонанты снабжается необходимыми параметрами. Нетрудно заметить, что $\gamma(\varepsilon)$ – «самосопряженная» функция, переходящая сама в себя при преобразовании $\varepsilon^* \leftrightarrow 1-\gamma$, $\gamma^* \leftrightarrow 1-\varepsilon$.

Для доказательства оптимальности значения $\kappa = \alpha/\beta = 1$ при $\mu > 0$, $\kappa \le 1$ достаточно показать, что для всех α, β, $\kappa < 1$

$$\text{(18)}\quad \gamma(\varepsilon;\mu,\kappa)\ge\gamma(\varepsilon;\mu,1),\quad \mu>0,\quad \kappa<1,\quad \varepsilon\in[0,1],$$

где $\gamma(\varepsilon;\mu, 1)$ задается формулой (17).

На первом из интервалов u выражается через ε и подставляется в формулу для $\gamma(\varepsilon)$. При этом, поскольку некоторые из входящих в выражение функций при $\kappa = 1$ становятся неопределенными, следует рассмотреть их предельные значения, именно

$$\lim_{\kappa\to 1}\varepsilon_{\mathrm{br}} = \mathrm{e}^{-\mu}/2\,(=\varepsilon_1);\ \lim_{\kappa\to 1}\gamma_{\mathrm{br}} = 1/2\,(=\gamma_1).$$

В результате этих манипуляций из (13) и (14) сразу получается оценка. Для интервала $(0, \upsilon^{(1+\kappa)/\kappa}]$

$$\gamma(\varepsilon;\mu,1) = 2^{\kappa-1}\left(\upsilon^{-1}+\upsilon\right)^{1-\kappa}\varepsilon^{\kappa} \ge 2^{\kappa-1}\upsilon^{-1+\kappa}\varepsilon^{\kappa} =$$

$$= \mathrm{e}^{\mu}\,2^{\kappa-1}\varepsilon^{\kappa} = (2\varepsilon)^{\kappa}\,\tfrac{\mathrm{e}^{\mu}}{2} \ge (2\varepsilon)\tfrac{\mathrm{e}^{\mu}}{2} = \mathrm{e}^{\mu}\,\varepsilon,$$

для интервала $(\upsilon^{(1+\kappa)/\kappa}, 1]$ (используется также замена $u \to z^{1+\kappa}$)

$$\gamma(\varepsilon;\mu,1) = 1-\tfrac{\upsilon^{1-\kappa}}{2u^{\kappa\varpi}}\left(1-u^{\frac{2\kappa}{1+\kappa}}\right) > 1-\tfrac{\mathrm{e}^{-\mu}}{4\varepsilon} = 1-\tfrac{\mathrm{e}^{-\mu}}{2}\left(u+u^{\frac{1-\kappa}{1+\kappa}}\right)^{-1}$$

$$\Leftrightarrow\ \left(1-u^{\frac{2\kappa}{1+\kappa}}\right)\left(u+u^{\frac{1-\kappa}{1+\kappa}}\right)u^{\kappa\frac{1-\kappa}{1+\kappa}}\ \Leftrightarrow\ z^{1-\kappa} < z^{(1-\kappa)\kappa} + z^{1+3\kappa}.$$

Последнее неравенство справедливо, поскольку $z \in (0, 1]$, $\kappa < 1$, и потому

$$z^{1-\kappa} < \left(z^{1-\kappa}\right)^{\kappa}.$$

Соотношение (18) окончательно доказано.

4.4. ВЗАИМНАЯ СВОДИМОСТЬ ЗАДАЧ

Итак, полностью разобран случай $\mu \ge 0,\ \alpha \le \beta\ (\kappa \le 1)$, и найдено, что значение $\mu = 0$ является оптимальным для задачи со свободным параметром μ, а значение $\alpha = \beta$ $(\kappa = 1)$ является оптимальным для задачи со свободным параметром α.

Этот случай дает решение обеих задач на ПМД для случая (i) $\mu > \nu,\ \alpha < \beta\ (\kappa < 1)$. Остается рассмотреть еще три случая (ii) $\mu < \nu,\ \alpha < \beta\ (\kappa < 1)$; (iii) $\mu > \nu,\ \alpha > \beta\ (\kappa > 1)$; (iv) $\mu < \nu,\ \alpha > \beta$ $(\kappa > 1)$. Однако специальных построений для этого можно не проводить, а воспользоваться правилами сведения одних задач к другим родственным.

Напомним, что по лемме 2 (отчасти, по лемме 1) при линейном преобразовании $X' = \mu + \nu X$ цены базового актива

(19) $\gamma'(\varepsilon) = \gamma(\varepsilon), \quad w'(x) = w\left(\frac{x-\mu}{\nu}\right),$

а по лемме 3 при переходе к сопряженной задаче ($p(x) \leftrightarrow c(x)$)

(20) $\gamma^*(\varepsilon) = 1 - \gamma^{\leftarrow}(1-\varepsilon), \quad w^*(x) = 1 - \gamma(w(x)).$

Это означает, что в случаях:

(ii) функции $\rho(x)$, $w(x)$ и $\gamma(\varepsilon)$ получаются из аналогичных функций для случая (i) по правилу (19) при $\mu = 0$, $\nu = -1$;

(iii) функции $\rho(x)$, $w(x)$ и $\gamma(\varepsilon)$ получаются из аналогичных функций для случая (i) по правилу *сопряжения* (20), при этом оптимальность вытекает из утверждения леммы 3.

(iv) функции $\rho(x)$, $w(x)$ и $\gamma(\varepsilon)$ получаются из аналогичных функций для случая (iii) по правилу (19) при $\mu = 0$, $\nu = -1$.

Тем самым можно считать доказанной возможность замены (16) и (18) соответственно соотношениями

$$\gamma(\varepsilon;\mu,\kappa) \ge \gamma(\varepsilon;0,\kappa), \quad \mu \in \Re, \quad \kappa \in \Re_+, \quad \varepsilon \in [0, 1],$$

$$\gamma(\varepsilon;\mu,\kappa) \ge \gamma(\varepsilon;\mu,1), \quad \mu \in \Re, \quad \kappa \in \Re_+, \quad \varepsilon \in [0, 1],$$

что завершает анализ.

5. Равномерные распределения

Рассматривается ситуация, когда обе функции $c(x)$ и $p(x)$ являются плотностями равномерного распределения – $\mathsf{R}(0, 2)$ и $\mathsf{R}(\mu, 2\alpha)$, x, $\mu \in \Re$, $\alpha > 0$. Выбор первого из них, очевидно, не ограничивает общности, а симметрия распределений позволяет изучать лишь случай $\mu \ge 0$. Анализируются две задачи на ПМД, в первой из которых свободным будет параметр μ, во второй – α. Необходимые вычисления проводятся при $\phi(\varepsilon) = \varepsilon^\lambda$, $\lambda > 0$, хотя бóльшая часть выводов справедлива для произвольных ф.р.п.

Очевидно, эта пара плотностей не удовлетворяет условиям регулярности. Относительный доход может принимать лишь значения 0, $1/\mu$, ∞, а диссонанта состоит не более чем из трех отрезков – вертикального, наклонного и горизонтального, и потому обе задачи следует отнести к разряду экзотических. Но простота распределений позволяет построить решение и для них – строго аналитическими средствами и с сохранением сути ПМД. Нужно лишь проявить аккуратность. И в данном разделе, не вдаваясь в подробности анализа, ограничимся лишь формулированием решений, которые пытливый читатель сможет, надеемся, без труда проверить.

В первой задаче фиксируется параметр α, а μ свободен. При любом α оптимальны все значения $\mu^{*} \in [0, |\alpha - 1|)$ (по лемме 4), и среди них естественный и ожидаемый нуль. Для них выполняются условия минимальности леммы 4, кроме того:

при $\alpha \leq 1$ $\gamma(0; \mu', \alpha) = 0$, $\gamma(\varepsilon; \mu', \alpha) = 1 - \alpha + \varepsilon\alpha$, $\varepsilon \in (0, 1]$;

при $\alpha > 1$ $\gamma(\varepsilon; \mu', \alpha) = \min[\varepsilon\alpha, 1]$, $\varepsilon \in [0, 1]$.

Во второй задаче фиксируется параметр μ, а свободен α. В ней тип решения различается по μ: если $\mu \leq 1$, условия минимальности леммы 4 не выполняются и решение зависит от рискового параметра λ. Так, при $\lambda < \lambda_{cr} = \ln(1 - \mu)/\ln(1 + \mu)$ оптимальное значение $\alpha' = 1 + \mu$, при $\lambda > \lambda_{cr}$ – $\alpha' = 1 - \mu$.

Если $\mu > 1$, оптимально $\alpha' = 1 - \mu$. Такой выбор α минимизирует прогнозное множество цен из покрывающих отрезок $[-1, 1]$ и вовсе не зависит от рисковых предпочтений инвестора.

6. Бета-распределения – численные методы

В данном разделе исследуется применение принципа минимума доходности в задачах, в которых обе функции $p(\cdot)$ и $c(\cdot)$ являются плотностями бета-распределения, известного из теории вероятности и математической статистики. В примерах с этим распределением для рассматриваемых задач не удается провести полного аналитического исследования, и потому применяются численные методы.

6.1. *ОПРЕДЕЛЕНИЯ И РЕЛЕВАНТНЫЕ СВОЙСТВА*

Для обеих плотностей $p(\cdot)$ и $c(\cdot)$ используется двухпараметрическое бета-распределение:

$$\mathrm{Be}(\alpha,\mu):\quad x^{\alpha-1}(1-x)^{\mu-1}\big/\mathrm{B}(\alpha,\mu),\quad \alpha,\mu>0,\ \text{где}$$

$$\mathrm{B}(\alpha,\mu)=\int_0^1 x^{\alpha-1}(1-x)^{\mu-1}\,dx=\Gamma(\alpha)\Gamma(\mu)/\Gamma(\alpha+\mu)\quad \text{и}$$

$$\Gamma(\alpha)=\int_0^\infty x^{\alpha-1}\exp(-x)\,dx$$

– соответственно бета- и гамма-функции.

Его математическое ожидание и дисперсия

(21) $\mathsf{E}X = \alpha/(\alpha+\mu), \quad \mathsf{D}X = \alpha\mu\Big/\Big((\alpha+\mu)^2(1+\alpha+\mu)\Big).$

Далее полагаем $p(x) \sim \mathrm{Be}(\alpha, \mu)$, $c(x) \sim \mathrm{Be}(\beta, \nu)$, $\alpha, \mu, \beta, \nu > 1$, и тогда функция относительного дохода

(22) $\rho(x) = p(x)/c(x) = x^{\alpha-\beta}(1-x)^{\mu-\nu}\,\mathrm{B}(\beta,\nu)/\mathrm{B}(\alpha,\mu)$.

Ее производная имеет вид

$$\rho'(x) = u(x)l(x), \quad l(x) = (1-x)(\alpha-\beta) - x(\mu-\nu),$$

где функция $u(x)$ на интервале $(0, 1)$ строго положительна, может обращаться в нуль лишь на его границах, а $l(x)$ – линейная функция. Поэтому знаки $\rho'(x)$ и $l(x)$ совпадают. Кроме того, $l(0) = \alpha - \beta$, $l(1) = \nu - \mu$, а обращается в нуль $l(x)$ в точке

(23) $x^* = (\alpha-\beta)/(\alpha+\mu-\beta-\nu)$.

При этом $x^* \in (0, 1)$, если (i) $\alpha > \beta$, $\mu < \nu$ и (ii) $\alpha < \beta$, $\mu > \nu$. В этих двух случаях функция $\rho(x)$ унимодальна и достигает в $x^* \in (0, 1)$ соответственно минимума и максимума.

Если $x^* \notin (0, 1)$, производная $\rho'(x)$ на $(0, 1)$ сохраняет знак и

$$\mathrm{sgn}[\rho'(x)] \equiv \mathrm{sgn}[\alpha-\beta] = \mathrm{sgn}[-\mu+\nu].$$

В случаях (iii) $x^* < 0$ и (iv) $x^* > 1$, как это следует из (23), функция $\rho(x)$ на интервале $(0, 1)$ монотонно возрастает и убывает при $\alpha > \beta$ и $\mu > \nu$ соответственно. Или более детально: если $\Delta = \alpha - \beta + \mu - \nu$, то случаю (iii) отвечают условия 1) $\mu < \nu$, $\Delta > 0$ и 2) $\mu > \nu$, $\Delta < \min(\alpha - 1, 0)$, а случаю (iv) – 1) $\alpha < \beta$, $\Delta > 0$ и 2) $\alpha > \beta$, $\Delta < \min(\mu - 1, 0)$.

В разделе 6.2 рассматриваются игры на тренде и волатильности базового актива с использованием составных параметров. В разделе 6.3 рассматриваются игры на хвосте распределения уже с непосредственным заданием своими параметрами, но представляющими также и самостоятельный интерес.

6.2. ИГРЫ НА ТРЕНДЕ И ВОЛАТИЛЬНОСТИ БАЗОВОГО АКТИВА

При рассмотрении игр на тренде и волатильности базового актива выделяются два «чистых» случая:

(i) $\alpha > \beta, \alpha + \mu < \beta + \nu$; (ii) $\alpha < \beta, \alpha + \mu > \beta + \nu$;

(iii) $\mu < \nu, \alpha + \mu > \beta + \nu$; (iv) $\alpha > \beta, \alpha + 1 < \beta + \nu, \alpha + \mu < \beta + \nu$.

В случае (i) функция $\rho(x)$ на интервале (0, 1) монотонно возрастает, в случае (ii) – монотонно убывает. В случаях (iii) и (iv) функция $\rho(x)$ унимодальна и принимает в точке $x^* \in (0, 1)$ соответственно минимальное и максимальное значения.

Случаи (i) и (ii) обычно ассоциируются с негативным и позитивным трендами соответственно, а (iii) и (iv) – с покупкой и продажей волатильности базового актива. Хотя на рынках в чистом виде тренд и волатильность, как правило, не встречаются, но вполне могут уживаться.

Для упрощения выбора параметров распределений вместо среднего и дисперсии будем оперировать составными параметрами $m = \alpha/\mu$ и $s = \alpha + \mu$. В соответствии с (21) среднее однозначно определяется отношением $m = \alpha/\mu$, и чем оно *больше*, тем *больше* среднее. Соотношение дисперсий неплохо, во всяком случае, в широком множестве случаев характеризуется суммой $s = \alpha + \mu$, и чем она *меньше*, тем *больше* дисперсия.

Тем не менее для повышения надежности результатов выбора свободных параметров и в отсутствие однозначного определения волатильности (да и тренда) для реального рынка не мешало бы при оптимизации по ПМД сопоставлять s со стандартным отклонением σ, или с дисперсией, с учетом обратной зависимости s от σ (21), если исключить крайние, малопригодные, значения параметров в качестве показателей степени.

В новых составных параметрах *позитивному* (*негативному*) *тренду* соответствует неравенство $m_p > m_c$ ($m_p < m_c$). В примере 1 для позитивного тренда принимается $\beta = 1{,}5$, $\nu = 2{,}5$, что дает $m_c = 0{,}6$, $s_c = 4{,}0$ (для сравнения, $\sigma_c = 0{,}216506$), и требуется, чтобы $m_p = 0{,}8$ ($> m_c$). Оптимум для параметра μ будет определяться в соответствии с ПМД варьированием параметра s_p.

С этой целью изменяются оба параметра α и μ, но так, чтобы $\alpha/\mu = m_p = 0{,}8$. При $m_p = 0{,}8$ параметр $s_p = \alpha + \mu = \alpha\,(1 + 1/m_p)$ меняется в пределах от $2{,}25\omega_1$ до $2{,}25\omega_2$, где (ω_1, ω_2) – интервал изменения параметра α, подбираемый так, чтобы минимум доходности достигался в его пределах. Для этого подходят границы интервала $\omega_1 = 1{,}0$, $\omega_2 = 3{,}0$.

В этом интервале задается $k = 50$ значений параметра α: $\alpha_j = \omega_1 + j\Delta$, $j \in J = \{1, \ldots, k\}$, $\Delta = (\omega_2 - \omega_1)/k = 0{,}04$. Таким значениям отвечают прогнозные плотности

$p_j(x) \sim \mathrm{Be}(\alpha_j, \alpha_j/m_p)$, $j \in J$.

Наконец, дискретный алгоритм оптимизации по CC-VaR дает k-мерный вектор средних доходностей, а наименьшей среди них оказывается 29-я компонента, для которой $\alpha_{29} = 2{,}16$, и потому $s_p = s_{p,opt} = 4{,}86$ ($\sigma_p = 0{,}205269$).

На рис. 1 слева представлены графики плотности $c(\cdot)$ и найденной оптимальной плотности $p_{29}(\cdot)$ (прерывистая и сплошная толстая линии соответственно). Дополнительно сплошными тонкими линиями изображены 20 из 50 тестируемых плотностей $p(\cdot)$, следующих друг за другом на одинаковых по параметру μ расстояниях.

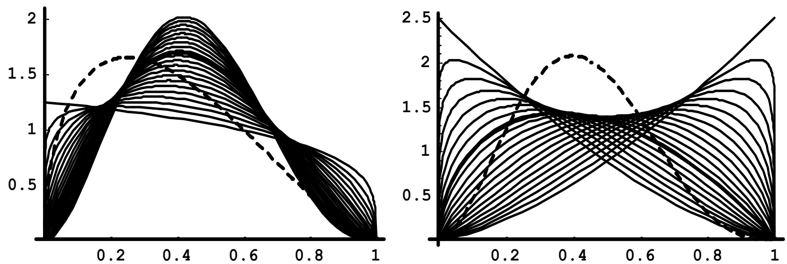

Рис. 1. Графики позитивного тренда и покупки волатильности

Аналогично *покупке* (*продаже*) *волатильности* соответствует неравенство $s_p < s_c$ ($s_p > s_c$). В примере 2 для *покупки волатильности* $\beta = 3{,}0$, $\nu = 4{,}0$, и потому $m_c = 0{,}75$, $s_c = 7{,}0$ ($\sigma_c = 0{,}174964$). Требуется, чтобы $s_p = 3{,}5 < s_c$ (при этом $\sigma_p = 0{,}259355 > \sigma_c$), а оптимум ищется варьированием m_p.

В распоряжении инвестора остается параметр s_p, который ему надлежит выбирать согласно ПМД. Его варьированием получается семейство плотностей $p(x)$ для тестирования.

Вновь изменяются оба параметра α и μ, но на этот раз так, чтобы $\alpha + \mu = \alpha\,(1 + 1/m_p) = s_p = 3{,}5$. Поскольку $m_p = \alpha/\mu$, то при $s_p = 3{,}5$ параметр m_p меняется в пределах от $\omega_2/(3{,}5 - \omega_2)$ до $\omega_1/(3{,}5 - \omega_1)$, где (ω_1, ω_2) – диапазон изменения параметра α

(зависимость m_p от α – обратная!). Интервал (ω_1, ω_2) для α подбирается аналогично задаче 1. Принимаются $\omega_1 = 1{,}0$, $\omega_2 = 2{,}5$.

В этом интервале задается $k = 50$ значений параметра α: $\alpha_j = \omega_1 + j\Delta$, $j \in J = \{1, \ldots, k\}$, $k = 50$, $\Delta = (\omega_2 - \omega_1)/k = 0{,}03$. Им отвечают прогнозные плотности

$p_j(x) \sim \mathrm{Be}(\alpha_j, s_p - \alpha_j)$, $j \in J$.

Применением к ним алгоритма оптимизации по CC-VaR находится k-мерный вектор средних доходностей. Наименьшую доходность доставляет 18-я компонента, для нее $\alpha_{18} = 1{,}54$, и потому $m_p = m_{p,opt} = 0{,}785714$. Соответствующие этой *покупке волатильности* графики изображены на рис. 1 справа.

Для рассмотрения *негативного тренда* достаточно в примере 1 произвести замену $x \leftrightarrow 1 - x$, а *продажи волатильности* – в примере 2 поменять ролями $p \leftrightarrow c$.

6.3. ИГРЫ НА ХВОСТЕ РАСПРЕДЕЛЕНИЯ

Здесь изучаются игры на одном из хвостов распределения, прогнозный параметр которого непосредственно задан, а другого хвоста – свободен. Желательно также выяснить, как они связаны с играми (i)–(iv) предыдущего раздела. Вновь речь идет о бета-распределенных цене базового актива и прогнозе.

В силу внутренней зеркальной симметрии семейства бета-распределений достаточно рассмотреть задачи о *продаже* и *покупке*, например, левого хвоста распределения; тогда аналогичные задачи для правого хвоста решаются простой переменной ролей параметров α и β с параметрами μ и ν соответственно.

При $\alpha > \beta$ инвестор будет стремиться к *продаже* левого хвоста распределения, при $\alpha < \beta$ – к *покупке*. Интуитивные соображения, подкрепленные предварительными расчетами, подводят к тому, что продажа, например, левого хвоста (при $\alpha > \beta$) требует и продажи правого (когда $\mu_{\mathrm{opt}} > \nu$), что в итоге и дает продажу волатильности.

В сделках для левого хвоста распределения задаются три параметра α, β и ν, и требуется определить $\mu_{\mathrm{opt}} = h(\alpha, \beta, \nu)$. Будет проверяться гипотеза «тождественности прогноза со стоимостной картиной рынка», а именно, что

(i) $h(\beta, \beta, \nu) = \nu$.

(ii) при фиксированных параметрах β и ν функция $h(\alpha, \beta, \nu)$ монотонно *возрастает* по α.

Едва ли было бы оправдано применение ПМД при нарушении части (i) гипотезы. Но уже часть (ii) требует дополнительного осмысления, и если не в отношении монотонности, то, во всяком случае, в ее направленности.

Для подтверждения (или опровержения) гипотезы предлагается такая схема вычислительных процедур (экспериментов). Задаются тестовые дискретные множества параметров α, β, μ, ν в количествах I, J, M, N соответственно. Для каждого набора значений этих четырех параметров решается стандартная дискретная задача оптимизации по CC-VaR с K сценариями и находится доходность инвестиции.

Затем по результатам для каждой тройки параметров α, β, ν определяется значение μ_{opt} параметра μ с минимальной доходностью. Тем самым находится функция $\mu_{opt} = h(\alpha, \beta, \nu)$ и может проверяться гипотеза. В экспериментах, предназначенных для анализа проблемы, выбираются относительно небольшие объемы данных: $I = J = 12$, $M = 19$, $N = 5$, $K = 20$, именно

α, β = 1,25; 1,5; 1,75; 2; 2,25; 2,5; 2,75; 3; 3,25; 3,5; 3,75; 4;

ν = 1,5; 2; 2,5; 3; 3,5; $\mu_m = 1 + 10\,(m/M)^3$, $m = 1, 2, \ldots, M$.

Подобный выбор диапазонов для параметров определяется соображениями минимизации практической достаточности и сводится к их небольшим значениям, а для параметра μ еще и стремлением во избежание дополнительных систематических ошибок не допускать выхода μ_{opt} на границы диапазона.

Фактически цель обработки результатов состоит в построении функции регрессии, как это делается в математической статистике, но на основе детерминированных, а не случайных результатов. Просто для них нет точных аналитических представлений, а приближенная формула, как обычно и регрессия в статистике, находится методом наименьших квадратов.

В качестве функции регрессии принимается полиномиальная функция второй степени, притом фактически одной переменной $\alpha - \beta$, к чему склоняет сам вид функции $\rho(x)$ (22):

(24) $\mu^* = h(\nu; \alpha, \beta) = \nu + \vartheta_0 + \vartheta_1(\alpha - \beta) + \vartheta_2(\alpha - \beta)^2$.

Методом наименьших квадратов находятся:

(25) $\vartheta_0 = 0{,}00120266$, $\vartheta_1 = 0{,}973058$, $\vartheta_2 = 0{,}208499$ и
(26) $\mu^* = \nu + 0{,}00120266 - 0{,}973058\,(\alpha - \beta) + 0{,}208499\,(\alpha - \beta)^2$.

Среднеквадратическое отклонение, рассчитанное по остаточной сумме квадратов, $\sigma = 0{,}414601$, что представляется вполне приемлемым с учетом грубости дискретной модели и использования относительно высоких значений параметров бета-распределений, ответственных за излишне деликатное поведение их плотности в окрестности нуля и единицы.

Проводятся и тесты проверки на адекватность (скорее, на возможную неадекватность) ПМД-модели, и они подтверждают часть (i) гипотезы неизменным характером результатов. К типичным можно отнести расчеты, например, при $\nu = 2$ и всех тестируемых α и β, $\alpha = \beta$, от 0,25 до 2,75. При $K = 100$ и $M = 50$ оптимально по ПМД значение $\mu_{opt} = 1{,}95112$ при $m = 29$.

Хотя увеличение до $M = 100$ не показывает повышения точности, так как оптимум равен $\mu_{opt} = 2{,}05379$ и достигается также при $m = 59$, но уже при $M = 200$ и даже при пониженном объеме $K = 50$ оптимально $\mu_{opt} = 2{,}00201625$ (при $m = 117$).

Соотношения (24)–(26), оперирующие средними характеристиками, тем не менее с очевидностью говорят в пользу части (ii) гипотезы. Кроме того, в каждом эксперименте при прогонках по всем α, β и ν, не претендующих, разумеется, на теоретическую точность, неравенства $\alpha > \beta$ и $\mu_{\text{opt}} > \nu$ выполняются одновременно. То же самое наблюдается и в отношении противоположных неравенств, а это уже веское основание для признания гипотезы.

В завершение анализа игр на хвосте распределения приведем результаты подобных игр уже на новых данных, непосредственно не участвовавших в получении соотношений (25), (26). Кроме того, значения параметров распределений для них выходят за пределы интервалов, принятых в модели регрессии. Тем не менее результаты сопоставляются именно с моделью.

В тесте для *покупки* фиксируются $\alpha = 4$, $\beta = 3$, $\nu = 4$, программа оптимизации по CC-VaR с ПМД дает $\mu_{opt} = 6{,}49802$, по регрессии (26) $\mu^* = 6{,}18276$ и отклонение $\mu_{opt} - \mu^* = 0{,}31526$. Для *продажи* с $\alpha = 4$, $\beta = 3$, $\nu = 2$ также $\mu_{opt} = 3{,}47072$, согласно (26) $\mu^* = 4{,}23664$, и отклонение $\mu_{opt} - \mu^* = -0{,}76592$.

Оба эти результата вполне согласуются со среднеквадратическим отклонением в регрессии $\sigma = 0{,}414601$ (во всяком случае при правиле «трех сигм» из математической статистики). Соответствующие *покупке* и *продаже* левого хвоста графики изображены на рис. 2 слева и справа соответственно.

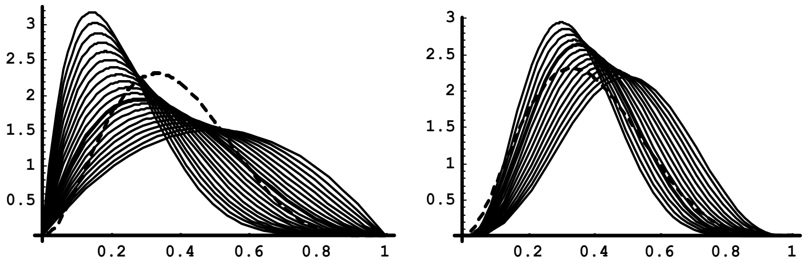

Рис. 2. Графики покупки и продажи левого хвоста

В графиках отчетливо усматривается также *покупка* и *продажа* волатильности (пунктирная линия отвечает стоимостной плотности, толстая сплошная – прогнозной). Однако обе операции показывают и некоторое смещение рынка (тренд).

7. Заключение

Работа исследует проблемы применения континуального критерия VaR на финансовых рынках в условиях частичного прогноза, когда инвестор формирует будущую плотность цены базового актива с точностью до некоторых параметров, оставляя рынку окончательный выбор их значений.

Вводится *принцип минимума доходности*, предлагающий инвестору выбирать значения свободных параметров прогноза, минимизирующие доходность инвестиции. Такой подход (идейно близкий критерию минимума χ^2 из математической статистики) дает инвестору определенную гарантию от ошибочных решений, связанных с частичностью прогноза.

Проблемы применения ПМД в отдельных важных для теории случаях удается решать аналитически на основе аппарата, развитого для методологии оптимизации по CC-VaR. Однако универсальным подходом следует рассматривать использование численных методов. В качестве примеров решаются задачи с бета-распределениями путем совмещения дискретного алгоритма оптимизации по CC-VaR с применением ПМД.

Специальные подходы предлагаются в задачах с бета-распределениями для игр на любом одном из хвостов распределений. Решения для них представляются любопытными и сводят их к играм на волатильности. Разумеется, ценность выводов может показаться несколько сниженной, поскольку решения находятся не аналитическими, а численными методами и потому не претендуют на абсолютную точность. Хотя и оснований доверять им также не достаточно.

Литература

1. АГАСАНДЯН Г.А. *Применение континуального критерия VaR на финансовых рынках*. – М. : ВЦ РАН, 2011. – 299 с.
2. АГАСАНДЯН Г.А. *Континуальный критерий VaR и оптимальный портфель инвестора* // Управление большими системами. – 2018. – Вып. 73. – С. 6–26.
3. АГАСАНДЯН Г.А. *Континуальный критерий VaR на сценарных рынках* // Информатика и ее применения. – М.: ИПИ ФИЦ ИУ РАН, 2018. – Т. 12, вып. 1. – С. 32–40.
4. АГАСАНДЯН Г.А. *Многомерные рынки опционов и оптимизация по CC-VaR* // Управление большими системами. – 2020. – Вып. 88. – С. 5–25.
5. КРАМЕР Г. *Математические методы статистики*. – М.: Наука, 1975. –750 с. (Перевод с англ.: Cramer. H. *Mathematical methods of statistics*. – Princeton University Press, 1946.)

MINIMUM YIELD PRINCIPLE AND CC-VAR UNDER PARTIAL MARKET FORECAST

Gennady Agasandyan, Dorodnicyn Computing Centre, FRC CSC RAS, Moscow, Doctor of Science, leading fellow (agasang17@yandex.ru).

Abstract: The work continues author's investigations connected with applying continuous VaR-criterion (CC-VaR) in option markets. A situation, when the investor's forecast about future probabilistic properties of an underlier is restricted by a partial view, is considered. The incompleteness of investor's forecast is modeled by introducing into forecast some parameters whose values are chosen by investor market properties considered. A minimum yield principle (MYP) that suggests minimizing the investment yield by the choice of parameters' values is postulated. By that the investor acquires some security for possible forecast mistakes. The theoretical properties of the principle introduced that has self-sufficient interest and simplifies the analysis in a variety of cases are investigated. Demonstration of the principle is realized analytically by two-sided exponential and uniform distributions and by numerical methods with beta-distributions. The results reassert the adequacy of the principle and of the computation algorithms.

Keywords: continuous VaR-criterion (CC-VaR), Newman-Pearson procedure, forecast incompleteness, risk-preferences function (r.p.f.), dissonance function, ordering function, yield, minimum yield principle (MYP), volatility, regression.

УДК 519.685
ББК 22.18
DOI: 10.25728/ubs.2021.92.1

Статья представлена к публикации членом редакционной коллегии Ф.И. Ерешко.

Поступила в редакцию 19.04.2021.
Опубликована 31.07.2021.

СИНТЕЗ СТАТИЧЕСКИХ РЕГУЛЯТОРОВ ПО ВЫХОДУ НА ОСНОВЕ РЕШЕНИЯ ЛИНЕЙНЫХ МАТРИЧНЫХ НЕРАВЕНСТВ

Мухин А. В.[1]
(*Нижегородский государственный университет им. Н.И. Лобачевского, Нижний Новгород*)

Одним из наиболее практически востребованных способов управления линейными объектами является управление в форме статического регулятора по выходу. Для реализации такого способа управления не требуется измерять все переменные состояния. Размерность замкнутой системы в таком случае совпадает с размерностью исходного объекта. Задача синтеза статических регуляторов по выходу сводится к поиску двух взаимно-обратных матриц, удовлетворяющих системе линейных матричных неравенств. Задача является невыпуклой и поэтому не может быть решена с помощью аппарата линейных матричных неравенств. В статье рассмотрен частный случай синтеза статических регуляторов по выходу, который может быть сведен к решению системы линейных матричных неравенств. Условие реализации такого случая выражается в том, что размерность пространства измерений должна быть на единицу меньше размерности пространства состояний. Рассмотрены две задачи синтеза статических регуляторов: синтез стабилизирующего регулятора и синтез оптимального регулятора с заданным квадратичным критерием. Полученные результаты применены к стабилизации электромагнитного подвеса, когда измеряемой переменной является вертикальное смещение ротора. Представлены графики переходных процессов в замкнутой системе с вычисленными статическими регуляторами.

Ключевые слова: линейные матричные неравенства, лемма Шура, статический регулятор по выходу, электромагнитный подвес.

1. Введение

Синтез законов управления линейными объектами на основе метода Ляпунова сводится к решению систем матричных неравенств [1, 7]. В зависимости от конкретного применяемого закона управления задача синтеза регулятора может быть либо выпуклой, либо невыпуклой и, соответственно, выражаться в терминах либо линейных, либо нелинейных матричных неравенств. Наиболее простым способом стабилизации линейного

[1] *Алексей Валерьевич Мухин, аспирант* (*muhin-aleksei@yandex.ru*).

объекта является статическое управление по состоянию. Синтез регулятора в таком случае сводится к решению системы линейных матричных неравенств, численная реализация которой не представляет трудностей и может быть выполнена с применением пакетов программ, например, таких как MATLAB [10] или SCILAB. Практическая реализация таких регуляторов предполагает возможность измерения всех переменных состояния. Очевидно, что такой тип управления не представляет существенного интереса и чаще всего не применяется на практике. Наиболее широко применяемым на практике подходом в задачах стабилизации является управление по измеряемому выходу в форме линейного динамического регулятора. В общем случае такая задача является невыпуклой, и синтез соответствующих регуляторов сводится к решению системы билинейных матричных неравенств. Вследствие отсутствия эффективных и надежных алгоритмов решения последних поиск таких регуляторов может представлять определенные трудности. Наиболее широко применяемые алгоритмы решения билинейных матричных неравенств приведены в [11–14]. При этом существующие алгоритмы не всегда могут давать решения, даже если таковые существуют. Частным случаем синтеза регуляторов по выходу, который может быть сведен к решению линейных матричных неравенств, является задача вычисления линейного динамического регулятора полного порядка [2]. Такой случай удобен и прост с теоретической точки зрения, однако приводит к заметному увеличению размерности замкнутой системы, а также к ее усложнению при последующей практической реализации.

Следует выделить еще один практически важный и востребованный случай управления линейными объектами – статический регулятор по выходу. Очевидным преимуществом такого подхода к управлению по сравнению с вышеописанными является то, что для его реализации не требуется измерять все переменные состояния. Недостаток синтеза таких регуляторов в общем случае также связан с невыпуклостью задачи. Решение указанной задачи сводится к поиску двух взаимно-обратных положительно определенных симметрических матриц, удовлетворяющих системе линейных матричных неравенств [1, 2]. Отметим, что сведение невыпуклой задачи к выпуклой имеет принципиальное значение в силу существенно более простой численной реализации последней.

Решению задачи синтеза статических регуляторов уделено большое внимание, в частности этой задаче посвящены работы [5, 6, 8, 9, 13]. Следует отдельно выделить работу [14]: в ней приведен обзор основных частных случаев синтеза стабилизирующих статических регуляторов, в которых задача является выпуклой. В первую очередь это вышеупомянутый случай управления по состоянию, а также некоторые другие случаи, подразумевающие те или иные ограничения либо совокупности ограничений.

В настоящей работе рассмотрен частный случай синтеза статических регуляторов по выходу, который может быть сведен к решению системы, состоящей из двух линейных матричных неравенств. Условие реализации такого случая выражается в том, что размерность пространства измерений должна быть на единицу меньше размерности пространства состояния. Рассмотрены две задачи синтеза статических регуляторов: синтез стабилизирующего регулятора и синтез оптимального статического регулятора с заданным квадратичным критерием качества. Закон управления во второй задаче можно рассматривать как минимаксный, поскольку он минимизирует относительное значение квадратичного функционала в наихудшем случае, когда начальное отклонение приводит к максимальному значению этого отношения [4]. На примере электромагнитного подвеса, в котором реализуется требуемое условие, вычислены стабилизирующий и оптимальный статические регуляторы. Предполагалось, что существует возможность измерения вертикального смещения ротора и последующего вычисления скорости.

Статья организована следующим образом. В первом разделе содержится формулировка задач управления. Второй раздел посвящен вопросам синтеза статических регуляторов с помощью матричных неравенств. Здесь же представлены условия, при которых изначально невыпуклая задача может быть преобразована к выпуклой, решение которой упрощается до решения двух линейных матричных неравенств. Третий раздел включает в себя результаты моделирования переходных процессов в электромагнитном подвесе с вычисленными статическими регуляторами.

2. Формулировка задач управления

Рассмотрим линейный управляемый объект стандартного вида:

(1) $x = Ax + Bu,\ x(0) = x_0,$

$y = C_2 x;$

где $x \in R^{n_x}$ – вектор состояния системы; $y \in R^{n_y}$ – измеряемый выход; $u \in R^{n_y}$ – управление; A, B, C_2 – заданные матрицы соответствующих порядков.

Применительно к объекту (1) рассмотрим две задачи управления: поиск стабилизирующего статического регулятора и поиск оптимального статического регулятора с заданным ниже квадратичным критерием качества.

2.1. ЗАДАЧА ПОИСКА СТАБИЛИЗИРУЮЩЕГО СТАТИЧЕСКОГО РЕГУЛЯТОРА

Задача управления сводится к поиску регулятора вида

(2) $u = \Theta y,$

обеспечивающего асимптотическую устойчивость замкнутой системы. Уравнение последней в данном случае имеет вид

(3) $x = (A + B\Theta C_2)x,$

где Θ – матрица параметров регулятора.

2.2. ЗАДАЧА ПОИСКА ОПТИМАЛЬНОГО СТАБИЛИЗИРУЮЩЕГО СТАТИЧЕСКОГО РЕГУЛЯТОРА

Для задания квадратичного критерия качества введем уравнение целевого выхода z:

(4) $z = C_1 x + Du$

где C_1, D – заданные матрицы соответствующих порядков.

Путем выбора матриц в уравнении (4) можно выделять те переменные состояния, по которым будет проводиться оптимизация. В качестве квадратичного критерия рассмотрим L_2-норму целевого выхода, учитывая при этом все переменные состояния и управления. Квадратичный функционал в таком случае примет следующий вид:

(5) $\|z(t)\|^2 = \int_0^\infty \left(\sum_{i=1}^{n_x} x_i^2(t) + u^2(t) \right) dt.$

Задача заключается в вычислении статического регулятора вида (2), обеспечивающего асимптотическую стабилизацию линейного объекта с выполнением следующего условия:

(6) $\inf_u \|z(t)\|^2 < \gamma^2 |x_0|^2, \ \forall x_0 \neq 0,$

где γ – минимально возможный параметр.

В отличие от линейно-квадратичного управления, неравенство (6) должно выполняться для всех ненулевых начальных состояний объекта [4].

Уравнения замкнутой системы в таком случае примут вид

(7) $x = A_c x$,

$z = C_c x.$

Матрицы в (7) записывается следующим образом

(8) $A_c = A + B\Theta C_2 \in R^{n_x \times n_x},$

(9) $C_c = (C_1 + D\Theta C_2) \in R^{n_z \times n_x}.$

3. Синтез статических регуляторов по выходу

Для решения поставленных выше задач применим метод Ляпунова и получим соотношения в форме линейных матричных неравенств. Будем рассматривать случай, когда размерность пространства измеряемого выхода n_y на единицу меньше размерности пространства состояний n_x, т.е. $n_y = n_x - 1$. Это равносильно тому, что ранг матрицы, столбцы которой образуют базис матрицы C_2 в уравнении измеряемого выхода, равен единице. Никаких дополнительных ограничений на матрицы системы не накладывается.

3.1. СИНТЕЗ СТАБИЛИЗИРУЮЩИХ СТАТИЧЕСКИХ РЕГУЛЯТОРОВ

Для асимптотической стабилизации объекта (1) должна существовать функция Ляпунова $V(x) = x^{\mathrm{T}}(t) X x(t)$, такая, что $V(x) < 0$. Задача поиска такой функции сводится к задаче вычисления матрицы $X = X^{\mathrm{T}} > 0$. Если расписать производную $V(x)$ в силу системы (1), то получим следующее неравенство:

(10) $AY + YA^{\mathrm{T}} + B\Theta C_2 Y + YC_2{}^{\mathrm{T}}\Theta^{\mathrm{T}}\mathrm{B}^{\mathrm{T}} < 0.$

Приведем неравенство (10) к эквивалентному линейному виду, выполнив соответствующее отображение в линейную область:

(11) $AY + YA^{\mathrm{T}} + BZ + Z^{\mathrm{T}}B^{\mathrm{T}} < 0.$

Матричная переменная Z в (11) определяется по формуле

(12) $Z = \Theta C_2 Y.$

Вследствие того что матрица C_2 имеет неполный ранг, разрешить (12) может оказаться невозможным, даже если искомый регулятор существует. Рассмотрим другой способ решения, заключающийся в сведении билинейного неравенства (10) к стандартному в теории управления линейному матричному неравенству следующего вида:

(13) $\Psi + P^{\mathrm{T}}\Theta^{\mathrm{T}}Q + Q^{\mathrm{T}}\Theta P < 0,$

где $\Psi = AY + YA^{\mathrm{T}}$; $P = B^{\mathrm{T}}$; $Q = C_2 Y$.

Согласно лемме исключения [1], неравенство (13) выполняется тогда и только тогда, когда разрешимы следующие матричные неравенства [1]:

(14) $W_P{}^{\mathrm{T}}\Psi W_P < 0,$

$W_Q{}^{\mathrm{T}}\Psi W_Q < 0,$

где столбцы W_P и W_Q образуют базисы ядер матриц P и Q соответственно.

Неравенства (14) разрешимы тогда и только тогда, когда существует такая матрица $Y = Y^{\mathrm{T}} > 0$, которая удовлетворяет следующей системе матричных неравенств

(15) $W_{C_2}{}^{\mathrm{T}}\left(Y^{-1}A + A^{\mathrm{T}}Y^{-1}\right)W_{C_2} < 0,$

$W_{B^{\mathrm{T}}}{}^{\mathrm{T}}\left(AY + YA^{\mathrm{T}}\right)W_{B^{\mathrm{T}}} < 0,$

где W_{C_2} и $W_{B^{\mathrm{T}}}$ образуют ядра матриц C_2 и B^{T} соответственно.

Таким образом, задача вычисления статического регулятора по выходу сводится к решению системы матричных неравенств (15). Необходимо отметить, что полученная система не является линейной и рассматриваемое множество не является выпуклым. Следовательно, такие задачи не могут быть решены методами выпуклой оптимизации.

Для синтеза стабилизирующего регулятора в случае, когда область поиска является выпуклой, рассмотрим алгоритм, который заключается в совместном решении линейного и билинейного матричного неравенств относительно неизвестных матриц Y и Θ:

(16) $AY + YA^{\mathrm{T}} + BZ + Z^{\mathrm{T}}B^{\mathrm{T}} < 0,$

(17) $AY + YA^{\mathrm{T}} + B\Theta C_2 Y + YC_2{}^{\mathrm{T}}\Theta^{\mathrm{T}}B^{\mathrm{T}} < 0.$

Если существует матрица Y, удовлетворяющая (16), при которой неравенство (17) также разрешимо относительно Θ, то это означает, что статический регулятор существует. Таким образом, совместно решая (16) и (17) можно вычислять разные значения матрицы регулятора Θ.

3.2. СИНТЕЗ ОПТИМАЛЬНЫХ СТАТИЧЕСКИХ РЕГУЛЯТОРОВ

Для выполнения условия (6) передаточная матрица замкнутого объекта должна удовлетворять следующему неравенству:

(18) $\|H_c\|_2 < \gamma|x_0|\ , \forall x_0 \neq 0.$

В [1] показано, что для выполнения (18) необходимо и достаточно, чтобы выполнялась следующая система линейных матричных неравенств:

(19) $\begin{pmatrix} A_cY + YA_c^{\mathrm{T}} & YC_c^{\mathrm{T}} \\ C_cY & -\gamma I_{n_z} \end{pmatrix} < 0,$

$\begin{pmatrix} Y & I_{n_x} \\ I_{n_x} & \gamma I_{n_x} \end{pmatrix} > 0,$

где $Y = Y^{\mathrm{T}} > 0,\ Y \in R^{n_x \times n_x},\ \gamma > 0.$

Подставим (8), (9) в первое неравенство (19) и приведем получившееся выражение к стандартному линейному матричному неравенству вида (13), матрицы которого запишутся следующим образом

(20) $\Psi = \begin{pmatrix} AY + YA^{\mathrm{T}} & YC_1^{\mathrm{T}} \\ C_1Y & -\gamma I_{n_z} \end{pmatrix} \in R^{(n_x+n_z)\times(n_x+n_z)},$

(21) $P = (C_2Y \quad 0_{n_y \times n_z}) \in R^{n_y \times (n_x+n_z)},$

(22) $Q = (B^{\mathrm{T}}\ D^{\mathrm{T}}) \in R^{n_u \times (n_x+n_z)}.$

Согласно лемме исключения, которая уже упоминалась выше, неравенство (13) выполняется тогда и только тогда, когда разрешима система (14). Подстановка (20)–(22) в (14) даст следующую систему неравенств:

(23) $$W_P^{\mathrm{T}} \begin{pmatrix} AY + YA^{\mathrm{T}} & YC_1^{\mathrm{T}} \\ C_1 Y & -\gamma I_{n_z} \end{pmatrix} W_P < 0,$$
$$W_Q^{\mathrm{T}} \begin{pmatrix} AY + YA^{\mathrm{T}} & YC_1^{\mathrm{T}} \\ C_1 Y & -\gamma I_{n_z} \end{pmatrix} W_Q < 0.$$

Введем новую матрицу следующего вида

(24) $G = (C_2 \ \ 0_{n_y \times n_z}) \in R^{n_y \times (n_x + n_z)}$.

Тогда матрицу P можно представить в виде произведения двух матриц:

(25) $$P = G \begin{pmatrix} Y & 0_{n_x \times n_z} \\ 0_{n_z \times n_x} & I_{n_z} \end{pmatrix}.$$

Откуда следует, что

(26) $$W_P = \begin{pmatrix} Y^{-1} & 0_{n_x \times n_z} \\ 0_{n_z \times n_x} & I_{n_z} \end{pmatrix} W_G,$$

где столбцы W_G образуют базис ядра матрицы G.

Таким образом, задача поиска оптимального статического регулятора сводится к вычислению симметрической положительно определенной матрицы Y, удовлетворяющей системе матричных неравенств

(27) $$W_G^T \begin{pmatrix} Y^{-1}A + A^{\mathrm{T}}Y^{-1} & C_1^{\mathrm{T}} \\ C_1 & -\gamma I_{n_z} \end{pmatrix} W_G < 0,$$
$$W_Q^T \begin{pmatrix} AY + YA^{\mathrm{T}} & YC_1^{\mathrm{T}} \\ C_1 Y & -\gamma I_{n_z} \end{pmatrix} W_Q < 0,$$
$$\begin{pmatrix} Y & I_{n_x} \\ I_{n_x} & \gamma I_{n_x} \end{pmatrix} > 0,$$

где $Y = Y^{\mathrm{T}} > 0$, $Y \in R^{n_x \times n_x}$, $\gamma > 0$.

Необходимо отметить, что полученная система (27) также не является линейной и рассматриваемое множества поиска не является выпуклым. Следовательно, такая задача не может быть решена методами выпуклой оптимизации.

Преобразуем первое неравенство системы (27). Для этого рассмотрим матрицу G, базис ядра которой входит в данное неравенство. Ранг $\mathrm{rank}(G) = \mathrm{rank}(C_2) = n_y$. Базис ядра этой матрицы можно представить в виде следующей блочно-диагональной матрицы:

(28) $W_G = \begin{pmatrix} W_{C_2} & 0_{n_x \times n_z} \\ 0_{n_z \times (n_x - n_y)} & I_{n_z} \end{pmatrix} \in R^{(n_x + n_z) \times (n_x + n_z - n_y)}$,

где столбцы матрицы $W_{C_2} \in R^{n_x \times (n_x - n_y)}$ образуют базис ядра матрицы C_2.

Тогда с учетом (28) первое неравенство в (27) примет следующий вид:

(29) $\begin{pmatrix} {W_{C_2}}^{\mathrm{T}} (Y^{-1} A + A^{\mathrm{T}} Y^{-1}) W_{C_2} & {W_{C_2}}^{\mathrm{T}} {C_1}^{\mathrm{T}} \\ C_1 W_{C_2} & -\gamma I_{n_z} \end{pmatrix} < 0,$

откуда в силу леммы Шура имеем

(30) $-\gamma I < 0 \leftrightarrow \gamma > 0,$

(31) ${W_{C_2}}^{\mathrm{T}} (Y^{-1} A + A^{\mathrm{T}} Y^{-1}) W_{C_2} + \gamma^{-1} {W_{C_2}}^{\mathrm{T}} {C_1}^{\mathrm{T}} C_1 W_{C_2} < 0.$

В соответствии с рассматриваемым случаем, когда $\operatorname{rank}(W_{C_2}) = 1$, неравенство (31) можно представить в виде скалярного соотношения

(32) $\gamma > const.$

В результате выполненных преобразований приходим к выводу эквивалентности неравенств (30) и (31) относительно неизвестной скалярной переменной γ. Таким образом, в частном случае, когда $n_y = n_x - 1$, задача вычисления симметрической положительно определенной матрицы Y представляет собой задачу выпуклого программирования и может быть выражена в терминах линейных матричных неравенств. Сформулируем следующее утверждение

Утверждение. *Если размерность пространства измерений на единицу меньше размерности пространства переменных состояния, то для существования оптимального статического регулятора по критерию* (6) *необходимо и достаточно, чтобы существовала матрица* $Y = Y^{\mathrm{T}} > 0$, *удовлетворяющая линейным матричным неравенствам*

(33) ${W_Q}^{\mathrm{T}} \begin{pmatrix} AY + YA^{\mathrm{T}} & YC_1^{\mathrm{T}} \\ C_1 Y & -\gamma I_{n_z} \end{pmatrix} W_Q < 0,$

$\begin{pmatrix} Y & I_{n_x} \\ I_{n_x} & \gamma I_{n_x} \end{pmatrix} > 0,$

где $\gamma > 0$.

Если система (33) разрешима, то искомая матрица оптимального статического регулятора Θ определяется из решения стандартного линейного матричного неравенства (13).

4. Пример

Рассмотрим применение описанного выше подхода для стабилизации ротора в системе примере электромагнитного подвеса. Такая система является основной электромагнитных подшипников, получивших широкое распространение в самых разных областях техники [15, 16].

С физической точки зрения, электромагнитный подвес представляет собой механическую систему, состоящую из вывешиваемого ротора и расположенного сверху электромагнита. Ротор находится в поле действия двух сил: силы тяжести и силы магнитного притяжения. Согласно второму закону Ньютона при равенстве этих сил ротор будет находиться в неподвижном неустойчивом состоянии. Задача системы управления – обеспечить стабилизацию ротора.

Электромагнитный подвес является нелинейным объектом, описываемым системой дифференциальных уравнений вида [3]

(34) $x_1 = x_2,$

$$x_2 = \frac{1}{2}\left[\frac{(1+x_3)^2}{(1-x_1)^2} - 1\right],$$

$$x_3 = -\frac{(1+x_3)}{(1-x_1)}x_2 - a(1 - x_1)x_3 + (1 - x_1)u,$$

где $x = (x_1, x_2, x_3)^{\mathrm{T}} \in R^{n_x}$ – вектор состояния системы; $u \in R^{n_u}$ – управление; a – постоянная величина ($a = 7{,}5$).

Безразмерная переменная x_1 соответствует вертикальному перемещению ротора, x_2 соответствует скорости перемещения, а x_3 описывает ток в цепи электромагнита. Неустойчивым положением равновесия системы является точка $x = 0$. Для синтеза управлений линеаризуем (33) и получим соответствующую линеаризованную систему:

(35) $x_1 = x_2,$

$$x_2 = x_1 + x_3,$$

$$x_3 = -x_2 - ax_3 + u.$$

В соответствии с матрично-векторной канонической формой объекта (1) имеем матрицы

$$(36)\ A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & -1 & -a \end{pmatrix}, \qquad B = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \qquad C_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Отметим, что для получения скорости перемещения ротора достаточно измерять только смещение, а скорость можно получить путем дифференцирования полученных данных.

4.1. СТАБИЛИЗИРУЮЩИЙ СТАТИЧЕСКИЙ РЕГУЛЯТОР

В результате совместного решения неравенств (16) и (17) вычислен статический регулятор $\Theta = (-12{,}2126\,;\ -10{,}9900)$. Графики переходных процессов в замкнутой системе с вычисленным статическим регулятором Θ представлены на рис. 1.

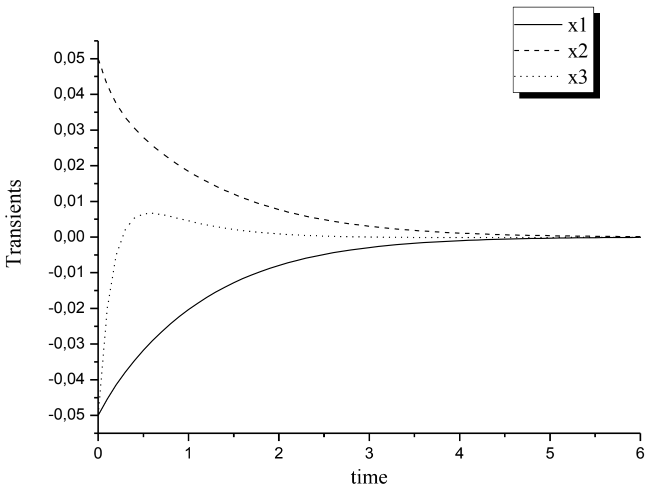

Рис. 1. Переходные процессы в замкнутой системе с регулятором Θ

4.2. ОПТИМАЛЬНЫЙ СТАТИЧЕСКИЙ РЕГУЛЯТОР

Построим оптимальный статический регулятор по критерию (6) для системы электромагнитного подвеса, в котором матрицы целевого выхода имеют следующий вид:

$$(37)\quad C_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad D = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

Матрица C_1 выбрана таким образом, чтобы оптимизация осуществлялась по всем переменным состояния. Квадратичный функционал в таком случае принимает представленный выше вид (5).

В результате решения системы линейных матричных неравенств (33) вычислены матрица Y, а затем из (13) вычислена искомый статический регулятор $\Theta_{opt} = (-12{,}8874;\ \ -11{,}5693)$.

Графики переходных процессов в замкнутой системе с оптимальным статическим регулятором представлены на рис. 2. Вычисленное минимальное значение параметра γ равно 18,7.

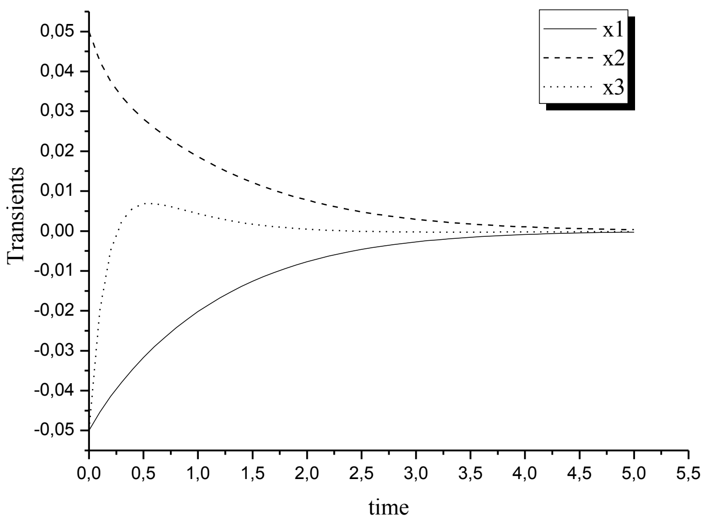

Рис. 2. Переходные процессы в замкнутой системе с регулятором Θ_{opt}

4. Заключение

В работе рассмотрен частный случай синтеза статических регуляторов по выходу, в котором изначально невыпуклая задача может быть преобразована к выпуклой. В результате решение такой задачи упрощается до решения двух линейных матричных неравенств. Условие реализации рассмотренного случая выражается в том, что размерность пространства измерений должна быть на единицу меньше размерности пространства состояния. На примере электромагнитного подвеса, в котором реализуется требуемое условие, вычислены стабилизирующий и оптимальный статические регуляторы по выходу в предположении измерения смещения ротора и возможности последующего вычисления скорости.

Автор благодарит профессора кафедры дифференциальных уравнений, математического и численного анализа Д.В. Баландина за консультацию, а также за ценные и полезные замечания.

Литература

1. БАЛАНДИН Д.В., КОГАН М.М. *Синтез законов управления на основе линейных матричных неравенств*. – М.: Физматлит, 2007. – 281 с.
2. БАЛАНДИН Д.В., КОГАН М.М. *Синтез регуляторов на основе решения линейных матричных неравенств и алгоритма поиска взаимно-обратных матриц* // Автоматика и телемеханика.- 2005.- №1.- С. 82–99.
3. БАЛАНДИН Д.В., БИРЮКОВ Р.С., КОГАН М.М., ФЕДЮКОВ А.А. *Оптимальная стабилизация тела в электромагнитном подвесе без изменения его положения* // Известия РАН. ТиСУ. – 2017. – № 3. – С. 12–24.
4. БАЛАНДИН Д.В., КОГАН М.М. *Линейно-квадратичные и γ-оптимальные законы управления по выходу* // Автоматика и телемеханика. – 2008. – №6. – С. 5–14.
5. ASTOLFI A., COLANERI P. *Static output feedback stabilization of linear and nonlinear systems* // 39th Conf. on Decision and Control, Sydney, Australia, 2000.

6. ASTOLFI A., COLANERI P. *An algebraic characterization for the static output feedback stabilization problem* // American Control Conference, Arlington, VA, 2001– P. 1408–1413.
7. BOYD S., EL GHAOUI L., FERON E., BALAKRISHNAN V. *Linear Matrix Inequalities in System and Control Theory*. – Philadelphia: SIAM, 1994.
8. CAO Y.-Y., LAM J., SUN Y.-X. *Static output feedback stabilization: an ILMI approach* // Automatica. – 1998. – Vol. 34. – P. 1641–1645.
9. EL GHAOUI L., OUSTRY F., AITRAMI M. *A cone complementarity linearization algorithm for static output-feedback and related problems* // IEEE Trans. on Automatic Control. – 1997. – Vol. 42. – P. 1171–1176.
10. GAHINET P., NEMIROVSKI A., LAUB A. J., CHILALI M. *The LMI Control Toolbox. For Use with Matlab*. *User's Guide*. – Natick, MA: The MathWorks, Inc., 1995.
11. GOH K.C. *Robust control synthesis via bilinear matrix inequalities*: Ph.D. thesis, University of Southern California, Los Angeles, CA, 1995.
12. HASSIBI A., HOW J., BOYD S. *A path following method for solving BMI problems in control* // Proc. of American Control Conference. – 1999. – Vol. 2. – P. 1385–1389.
13. HENRION D., LOEFBERG J., KOCVARA M., STINGL M. *Solving polynomial static output feedback problems with PENBMI* // Proc. joint IEEE Conf. Decision Control and Europ. Control Conf., Sevilla, Spain, 2005.
14. SADABADI M. S., PEAUCELLE D. *From static output feedback to structured robust static output feedback: A survey* // Annual Reviews in Control. – 2016. – Vol. 42. – P. 11–26.
15. SCHWEITZER G. *Magnetic bearings theory, design, and application to rotating machinery*. – Berlin: Springer, 2009.
16. ZHURAVLEV YU.N. *Active magnetic bearings. Theory, calculation, application.* – SPb.: Politechnica, 2003.

SYNTHESIS OF STATIC OUTPUT CONTROLLERS BASED ON THE SOLVING OF LINEAR MATRIX INEQUALITIES

Aleksey Mukhin, Lobachevsky State University, Nizhny Novgorod, postgraduate student (myhin-aleksey@yandex.ru).

Abstract: One of the most practically demanded methods of control in linear systems is control in the form of a static controller. To implement this control method, it is not necessary to measure all the phase variables of the system. With this method of control, the dimension of the closed system coincides with the dimension of the original object. The problem of static controller synthesis, in general, is reduced to the search for two mutually inverse matrices that satisfy a system of linear matrix inequalities. Such a problem is nonconvex and therefore cannot be solved using the apparatus of linear matrix inequalities. The solution of such a problem is reduced to finding two mutually inverse matrices that satisfy a system of linear matrix inequalities. The article considers a special case of the problem of synthesis of static controllers, which can be reduced to solving a system of linear matrix inequalities. The conditions for the implementation of such a case are shown. Two problems of the synthesis of static controllers are considered: the synthesis of the stabilizing controller and the synthesis of the optimal controller. The obtained results are applied to the stabilization of the electromagnetic suspension when the measured variable is the vertical displacement of the rotor. Graphs of transients are presented. A comparative analysis of the quality of transients in a closed system with calculated static controllers is performed.

Keywords: linear matrix inequalities, nonlinear matrix inequalities, Schur's lemma, static controller, electromagnetic suspension.

УДК 517.977
ББК 22.18
DOI: 10.25728/ubs.2021.92.2

Статья представлена к публикации членом редакционной коллегии Г.А. Угольницким.

Поступила в редакцию 11.04.2021.
Опубликована 31.07.2021.

ИЗУЧЕНИЕ ЭНЕРГЕТИЧЕСКОЙ УСТОЙЧИВОСТИ РЕГИОНОВ РОССИЙСКОЙ ФЕДЕРАЦИИ С ПРИМЕНЕНИЕМ МЕТОДОВ АНАЛИЗА ПАТТЕРНОВ

Мячин А. Л.[1]

(*Национальный исследовательский университет «Высшая школа экономики», Москва; ФГБУН Институт проблем управления им. В.А. Трапезникова РАН, Москва*)

Прокофьев В. Н.[2], **Степанов А. А.**[3]

(*Национальный исследовательский университет «Высшая школа экономики», Москва)*

Работа посвящена применению порядково-фиксированной и порядково-инвариантной паттерн-кластеризаций для исследования структурной схожести энергетического сектора в регионах Российской Федерации за пятилетний период. Методы анализа паттернов в работе обусловлены независимостью конечных результатов от разности в абсолютных значениях показателей и возможностью объединения регионов, близких по структуре (на основе рассматриваемых показателей). Выбор порядково-инвариантной паттерн-кластеризации обусловлен как эндогенным определением не только состава каждого паттерна, но их количества, так и отсутствием необходимости выбора первоначальной последовательности изучаемых показателей, к которой чувствительны некоторые методы анализа паттернов. Исследование проведено на основе таких показателей, как: снижение уровня потерь электроэнергии в сетях, электровооруженность труда работников, объемы произведенной электроэнергии, удельных расход топлива, стоимость услуг по технологическому присоединению, а также количество фактических подключений. В результате исследования изучены особенности энергетического развития отдельно взятых субъектов. Получены паттерны развития, отражающие схожесть внутренней структуры данных. Построены динамические траектории развития и выделены группы субъектов, придерживающиеся постоянной стратегии из года в год. Приведенные результаты могут быть учтены при разработке государственной политики в области энергоэффективности.

Ключевые слова: паттерн, анализ паттернов, энергетическая устойчивость, кластерный анализ.

---

[1] *Алексей Леонидович Мячин, к.т.н.* (amyachin@hse.ru).

[2] *Вадим Николаевич Прокофьев, к.ф.н.*(vprokofiev@hse.ru).

[3] *Александр Андреевич Степанов, студент* (aastepanov_3@edu.hse.ru).

1. Введение

Устойчивость энергетического развития российского государства определяется его способностью в полном объеме удовлетворять возникающие потребности в энергии и энергоносителях на внутреннем и внешнем рынке. Стабильное энергетическое развитие должно соотноситься с принципами энергосбережения и энергоэффективности, закрепленными в «Энергетической стратегии России на период до 2035 года» [21].

В настоящий момент энергетический сектор РФ характеризуется наличием достаточно серьезных проблем, ограничивающих дальнейший рост его эффективности. Одной из них является замедление спроса на энергию. На внутреннем рынке это вызвано снижающимися темпами роста экономической активности населения, на внешнем – трендом, связанным с усилением энергетической автономности. Снижение общего уровня энергоэффективности также обуславливается высокой степенью изношенности основных фондов и производственных мощностей. Это, в свою очередь, обнажает проблему серьезного технологического отставания от развитых стран и высокого уровня зависимости от импорта их оборудования [21]. Данные проблемы усугубляются неоднородностью социально-экономического и технико-технологического развития отдельно взятых субъектов страны, что создает дополнительные сложности для проведения эффективных реформ.

Для определения структурных различий и оценки устойчивости траекторий развития энергетического сектора предлагается использовать методы статического и динамического анализа паттернов. Результаты анализа могут быть использованы для формирования эффективной энергетической политики государства, учитывающей региональные энергетические особенности.

Различные методы анализа паттернов успешно применялись при изучении разных областей, как исключительно теоретических, так и практических. Эффективность их применения наглядно демонстрируется в работах по макроэкономике [23], банковскому сектору [22], финансовой устойчивости [3, 5] и изучению региональных особенностей в области науки и образования [1].

В основе методов анализа паттернов лежит простая идея: разбиение всех объектов исследуемого множества на непересекающиеся подмножества на основе системы показателей, значения которых могут меняться с течением времени. Разумеется, данная формулировка схожа с задачей кластерного анализа. Однако отличие состоит в принципе объединения исследуемых объектов в группы: при использовании методов анализа паттернов объединяются именно структурно близкие объекты (несмотря на возможные существенные различия в абсолютных значениях). Именно данная особенность преимущественно обуславливает использование методов анализа паттернов, поскольку некоторые субъекты РФ существенно отличаются друг от друга как по площади, так и по численности населения, что, разумеется, может находить свое отражение в серьезном разбросе значений отдельных показателей. Значительным преимуществом является также способность методов анализа паттернов эндогенно формировать группы объектов, характеризующиеся однородной структурой. Такой подход позволяет избежать неточностей, связанных с ошибочным определением начального числа паттернов.

Таким образом, использование данного подхода в исследовании позволит выявить основные структурные особенности энергетического сектора регионов РФ, а применение динамического анализа паттернов – оценить устойчивость их траекторий развития с течением времени.

2. Исследование энергетической отрасли

2.1. ЭНЕРГЕТИЧЕСКИЙ СЕКТОР: ОБЗОР

Стабильное функционирование энергетической отрасли становится основой для планового социально-экономического развития государства. К основным задачам отрасли относятся жизнеобеспечение и создание необходимых условий для поддержания производственного потенциала страны. Устойчивое энергетическое развитие является одним из главных факторов, определяющих перспективы роста ВРП отдельных субъектов РФ и, как следствие, ВВП всего государства.

Для поддержания конкурентоспособности экономики России была разработана программа [9], преследующая цели по реализации потенциала энергосбережения и повышению общего уровня энергоэффективности за счет рационального использования ресурсов и технико-технологической трансформации как генерирующих, так и потребляющих электроэнергию предприятий.

Под «энергоэффективностью» принято понимать экономически обоснованное и технически возможное качество использования энергоресурсов и энергии в текущих технологических условиях [10, с. 148].

Одной из проблем достижения устойчивого роста энергоэффективности является высокий уровень энергоемкости – величины потребляемой энергии, затрачиваемой на реализацию основных процессов, связанных с производством и потреблением экономических благ [11]. Постоянно растущий уровень потребления электроэнергии и соответствующий износ основных фондов кратно увеличивают энергоемкость произведённых товаров, выполненных работ и оказанных услуг, что ограничивает устойчивое энергетическое развитие.

Существует ряд исследований, концентрирующихся на изучении проблемы энергетической эффективности с помощью методов кластерного анализа [24, 26].

Например, в [26] авторы изучают модели потребления энергии различными предприятиями. Используя метод кластерного анализа *k*-means, специалисты разбивают все исследуемые объекты по разным группам таким образом, чтобы находящиеся в одном кластере характеризовались схожими моделями энергопотребления.

Изучение производственных циклов промышленных предприятий с использованием методов кластеризации позволило создать модель для оценки энергоэффективности и оптимизации процессов производства сложной и энергоемкой нефтехимической отрасли [24].

В [15] исследовался уровень энергоэффективности регионов РФ. В результате проведенного иерархического кластерного анализа по четырем показателям было выделено четыре группы объектов. Авторы показывают, что большинство регионов РФ принадлежат кластерам с высоким уровнем энергоемкости, что свидетельствует о низкой энергоэффективности экономики страны в целом.

Попытка разбиения регионов по группам с учетом их энергоэффективности была предпринята в [14] на основе кластеризация методом Варда. В результате анализа трех показателей все объекты были разбиты на три кластера, включающие в себя регионы с высокой, средней и низкой энергоемкостью соответственно.

Все перечисленные работы описывают процедуру разбиения объектов на основе их абсолютных значений показателей. Такой подход исключает возможность объединения в одну группу разномасштабных объектов, характеризующихся структурной схожестью. Кроме того, упомянутые исследования проводились в статике, т.е. за один конкретный год, что не позволяет делать выводы о тенденциях и закономерностях развития региональной энергетики, происходящих во времени.

Метод анализа паттернов помогает сформировать структурно схожие объединения, успешно справляясь с проблемой абсолютных значений показателей. В свою очередь, проведение динамического анализа паттернов предполагает построение траекторий развития каждого изучаемого объекта и оценку степени изменчивости этих траекторий во времени. Результаты анализа дают основания делать выводы об уровне устойчивости развития отдельно взятого объекта.

2.2. ФОРМИРОВАНИЕ СИСТЕМЫ БАЗОВЫХ ПОКАЗАТЕЛЕЙ

Для проведения процедуры анализа паттернов необходимо сформировать систему базовых показателей, которая будет удовлетворять определенному набору требований. Прежде всего необходимо удостовериться, что выбранные показатели четко определены и измеримы. Достижение объективных и точных результатов анализа возможно только при использовании достоверных данных, имеющих минимальную погрешность измерения. Стоит отметить, что необходимо использовать полную информацию об объекте исследования. Такой подход исключает любые пропуски данных.

Система базовых показателей формировалась на основе индикаторов, значимость которых подтверждается релевантными статьями, описывающими основные характеристики энергетической отрасли: докладом «Электроэнергетика России: Основные показатели функционирования и тенденции развития» [25] и государственной программой «Развитие энергетики» [9]. Для проведения исследования использовалась информация, предоставленная Федеральной службой государственной статистики. Агрегированные данные, характеризующие состояние энергетического сектора в каждом субъекте РФ, удовлетворяют всем перечисленным выше стандартам качества.

Уровень потерь электроэнергии в сетях является одним из наиболее важных индикаторов, отражающих эффективность работы электросетевых компаний. Эксперты утверждают, что производственные мощности электрогенерирующих компаний более чем на половину выработали свой срок службы, а степень износа электрических сетей в настоящий момент подходит к своей критической точке. Попытки увеличения объемов генерации электроэнергии в текущих условиях могут привести к драматическим последствиям [8]. Учитывая актуальность данной проблемы, было принято решение рассмотреть показатель уровня снижения потерь электроэнергии в сетях, рассчитывающийся как разность отношений величины потерь к общему объему отпущенной электроэнергии текущего и предыдущего годов.

Другим важным индикатором является стоимость услуг по технологическому присоединению к объектам электросетевого хозяйства для устройств, принимающих энергию свыше 150 кВт, рассчитанная в тысячах рублей. Кроме того, необходимо учитывать фактическое количество подключений таких объектов к электрическим сетям. Рост числа подключений энергопринимающих устройств свыше 150 кВт, сопровождающийся невысокой совокупной стоимостью присоединений, может свидетельствовать об эффективном процессе формирования производственной базы с минимальными технологическими затратами [12].

Объемы произведенной и потребленной электроэнергии – следующие два показателя, необходимые к рассмотрению. Их можно считать основополагающими индикаторами, характеризующими общее состояние энергетического сектора [7]. В настоящее время функционирование любого предприятия требует электроэнергии. Таким образом, генерирующие ее компании становятся неотъемлемой частью всего производственного процесса. В свою очередь, количество потребленной электроэнергии в условиях низкого уровня ее потерь в сетях во многом определяет возможные темпы роста производимых товаров и оказываемых услуг [19].

Еще одним немаловажным показателям является электровооруженность труда работников. Индикатор измеряется в киловатт-часах и рассчитывается как отношение совокупного объема электроэнергии, использованного на производстве, к фактически отработанным человеко-часам [6]. Повышение показателя электровооруженности труда может свидетельствовать о росте производительности и автоматизации труда, что обеспечивает устойчивое развитие социально-экономическое развитие и способствует достижению научно-технического прогресса [13].

Удельный расход условного топлива на отпущенную электроэнергию – последний показатель, определяющий степень технологического состояния генерирующих электричество предприятий. Чем ниже значения данного показателя, тем меньшие издержки несет электростанция. Эффективное использование ресурсов говорит о ее высокой конкурентоспособности [20].

Для проведения динамического анализа паттернов необходимо иметь количественные данные по всем показателям за исследуемый промежуток времени. В связи с этим было принято решение рассмотреть временной интервал равный пяти годам: с 2012 по 2016 г.

На этапе первичной статистической обработки данных необходимо удостовериться, что выбранные показатели не являются взаимосвязанными. Наличие взаимосвязи или ее отсутствие предлагается доказывать, основываясь на результатах корреляционного анализа. Показатели с высоким значением коэффициента корреляции должны быть исключены из базовой системы показателей. Использование двух взаимосвязанных показателей не даст существенной пользы при описании результатов из-за очевидного соотношения между ними.

Все изучаемые индикаторы представлены количественными шкалами, что позволяет использовать коэффициент корреляции Пирсона. Коэффициент принимает значения от –1 до 1. Чем ближе значение коэффициента к границам диапазона, тем сильнее линейная взаимосвязь между переменными.

Недостатком данного метода корреляционного анализа является его неустойчивость к выбросам – экстремальным значениям, существенно выделяющихся из общей выборки. В связи с чем предлагается исключить объекты, характеризующиеся аномальными значениями.

Ханты-Мансийский автономный округ – Югра и Тюменская область характеризуются крайне высокими объемами производства и потребления электроэнергии. Уровень снижения потерь электроэнергии в сетях принимает аномально низкие значения в Ингушетии и Дагестане. Рассматривая показатель электровооруженности труда работников невозможно не отметить экстремально высокие значения по республике Хакасия. Северная Осетия-Алания и Карачаево-Черкесская, Кабардино-Балкарская республики характеризуются низким удельным расходом топлива. Стоимость услуг по технологическому присоединению значительно отклоняется от общей выборки в таких субъектах, как Москва, Московская область и Санкт-Петербург. Наконец, Красноярский край, Орловская и Самарская области далеко оторвались в лидеры по общему числу подключений к объектам электросетевого хозяйства.

Выбросы способны формировать уникальные паттерны, в связи с чем все исключенные объекты будут рассмотрены в отдельной главе.

В результате проведенного корреляционного анализа за пятилетний промежуток времени была выявлена лишь одна пара линейно взаимосвязанных показателей. Для показателей, характеризующих объемы произведенной и потребленной электроэнергии, коэффициент корреляции Пирсона принимает значения выше 0,75. Исходя из этого предлагается исключить показатель «объем потребленной электроэнергии» из базовой системы показателей.

Для обеспечения однородности и сопоставимости показателей необходимо нормировать их значения. Процедура нормирования будет реализована по методу линейного масштабирования.

3. Анализ паттернов: методология, основанная на парном сравнении показателей

Понятие «паттерн» широко используется в различных областях знаний. В [18] под паттерном «понимается такая комбинация определённых с точностью до погрешности значений некоторого подмножества признаков, что объекты с этими значениями достаточно сильно отличаются от других объектов». В свою очередь, в [4, с. 9] паттерн определяют в качестве комбинации качественно похожих признаков.

Основная задача методов анализа паттернов заключается в выделении качественно схожих объектов и их объединении в уникальные группы. В этом плане поставленная задача весьма схожа с алгоритмами классификации и кластеризации данных. Однако в данном случае отсутствует тестовая выборка (как и количество предполагаемых групп, а также их типичные представители), как при классификации данных, а также результат не зависит от разности в абсолютных значениях показателей (как при использовании большинства методов кластерного анализа). К примеру, объекты с показателями (10, 5, 10) и (100, 50, 100) при использовании методов анализа паттернов должны быть объединены в единую группу, поскольку, несмотря на существенные отличия в абсолютных значениях, схожи по структуре данных.

Визуализация результатов анализа паттернов, как правило, осуществляется в системе параллельных координат, состоящей из вертикальных осей, число которых соответствует общему числу изучаемых показателей. На данных осях точками отмечаются определенные значения показателей для отдельно взятого объекта. Затем точки соединяются отрезками, образуя кусочно-линейные функции, характеризующие отдельные объекты [16].

Порядково-фиксированная паттерн-кластеризация – один из методов анализа паттернов, работа которого строится на парном сравнении показателей. Отличительная особенность порядково-фиксированной паттерн-кластеризации заключается в использовании только одной заранее заданной последовательности показателей. Кратко рассмотрим алгоритм работы данного метода.

Предположим, что имеется некоторое множество объектов X, которое содержит в себе k элементов. Каждый объект множества $x_i \in X$ характеризуется набором m показателей и может быть представлен в виде вектора $x_i = (x_{i1}, x_{i2}, \ldots, x_{ij})$, где x_{ij} – значение i-го объекта по показателю j. Кроме того, каждый объект множества X может быть описан последовательностью символов $r_i^1, r_i^2, \ldots, r_i^{m-1}$, которая будет учитывать характер парных отношений между соседними показателями. Данная последовательность определяется следующими формулами:

(1) $r_i^j = 1$ при $x_{ij} < x_{ij+1}$,

(2) $r_i^j = 0$ при $x_{ij} = x_{ij+1}$,

(3) $r_i^j = 2$ при $x_{ij} > x_{ij+1}$.

Процедура разбиения множества объектов X по отдельным группам предлагается осуществлять, основываясь на результатах оценки меры близости позиционных кодов, сформированных на основе последовательностей символов $r_i^1, r_i^2, \ldots, r_i^{m-1}$ каждого отдельно взятого объекта. В данном случае для оценки близости используется расстояние Хемминга. Объекты можно объединить в одну группу, если расстояние между ними равно нулю. Подробнее, данная методология описана в [16].

Другой метода анализ паттернов, основанный на парном сравнении показателей, называется порядково-инвариантная паттерн-кластеризация. Данный метод основан на формировании полного взвешенного ориентированного графа, ребра которого отражают значения парных сравнений показателей. Подобное представление позволяет учитывать все возможные парные сравнения (по формулам (1)–(3)) между показателями. Подробно методология описана в [16–17].

Программы по реализации обоих методов были написаны с помощью языка программирования Python.

4. Анализ паттернов энергетического развития регионов РФ в статике

Как сказано выше, для наглядности визуализация результатов, полученных в ходе применения методов анализа паттернов, осуществляется в системе параллельных координат. Необходимо представить объекты исследования в форме кусочно-линейных функций с вершинами, находящимися на вертикальных осях. На каждой оси, характеризующей изучаемые показатели, откладываются их количественные (в данном случае предварительно нормированные) значения, после чего соседние значения соединяются отрезками.

В результате реализации порядково-фиксированной и порядково-инвариантной паттерн-кластеризаций для всех объектов на пятилетнем промежутке времени удалось получить 16 уникальных паттернов. Каждый паттерн характеризуется индивидуальной структурой данных, свойственной только для включенных в него объектов. Паттерн состоит из множества объектов – регионов Российской Федерации, обладающих схожей структурой развития энергетического сектора в определенный период времени.

Рассмотрим несколько наиболее крупных групп объектов, а также два паттерна, описывающих выбросы. Первый паттерн является самым большим и включает в себя 162 объекта. Он характеризуется достаточно широким разбросом значений по первому показателю и низкими (менее 0,3) значениями электровооруженности труда и стоимости услуг по технологическому присоединению. Большинство объектов находятся на невысоком (менее 0,5) уровне по объемам произведенной электроэнергии и количеству фактических подключений. Что касается значений показателя удельного расхода топлива на отпущенную электроэнергии, то они концентрируются в диапазоне между 0,2 и 0,8.

Второй паттерн включает в себя 142 объекта. В данном случае можно обнаружить некоторые сходства с паттерном, представленном выше. Однако главное различие между первым и вторым паттерном развития можно заметить при рассмотрении показателя «объемы произведенной электроэнергии». В отличие от первого паттерна второе объединение объектов характеризуется крайне низкими (менее 0,2) значениями показателя.

Рис. 1. Кусочно-линейные функции, описывающие объекты паттерна №1

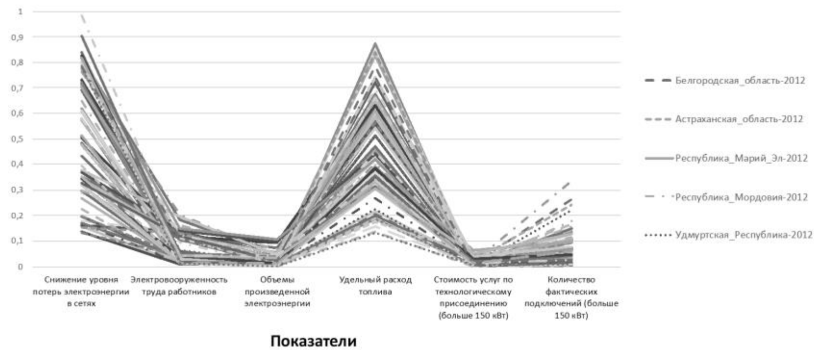

Рис. 2. Кусочно-линейные функции, описывающие объекты паттерна №2

Объекты, которые характеризуются крайне высокими или, наоборот, чрезвычайно низкими значениями по отдельным показателям, определяются как выбросы. Стоит отметить, что такие объекты часто образуют уникальные паттерны. Рассмотрим два субъекта РФ: Московскую область и город федерального значения Санкт-Петербург (рис. 3 и рис. 4 соответственно) и паттерны, к которым они относятся. Оба паттерна характеризуются крайне низкими (менее 0,1) значениями электровооруженности труда. Объемы произведенной электроэнергии и удельный расход топлива расположены в диапазоне от 0,1 до 0,3. Стоимость услуг по технологическому присоединению преимущественно демонстрирует растущую динамику для обоих паттернов.

Можно заметить, что рисунки схожи между собой, однако ключевое различие, которое сразу бросается в глаза, связано со значениями показателя числа фактических подключений к объектам электросетевого хозяйства. Московская область в четырёх из рассматриваемых пяти лет существенно превосходила Санкт-Петербург по данному показателю.

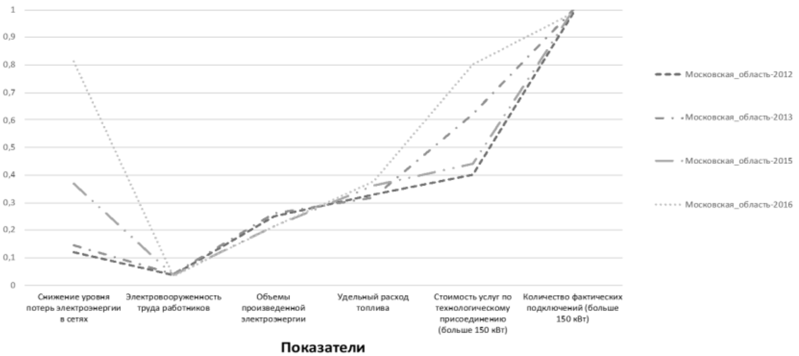

Рис. 3. Кусочно-линейные функции, описывающие объекты паттерна №3

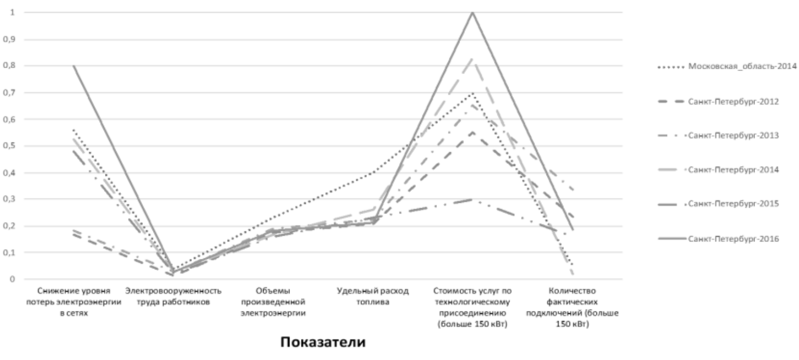

Рис. 4. Кусочно-линейные функции, описывающие объекты паттерна №4

5. Результаты динамического анализа паттернов

Динамический анализ паттернов основывается на построении «траекторий развития» для каждого рассматриваемого объекта. В данном случае нам необходимо понять, к какому паттерну энергетического развития относился регион в определенный момент времени.

Основная идея динамического анализа паттернов заключается в распределении объектов анализа по отдельным динамическим группам в соответствии с устойчивостью их принадлежности к одному и тому же паттерну. Принято выделять следующие категории устойчивости: абсолютно устойчивые, устойчивые, полуустойчивые, неустойчивые и абсолютно неустойчивые [2, с. 8].

В результате проведенного анализа удалось определить группу абсолютно устойчивых регионов. Она состоит из 30 субъектов РФ, которые ни разу не сменили паттерн своего энергетического развития за пятилетний промежуток времени.

Большинство регионов, которые попали в группу абсолютно устойчивых объектов, принадлежат первому и второму паттерну. Первый паттерн включает в себя 16 объектов, а второй – 11. Кроме того, абсолютно устойчивыми являются следующие регионы: г. Москва, г. Санкт-Петербург и республика Хакасия (рис. 5).

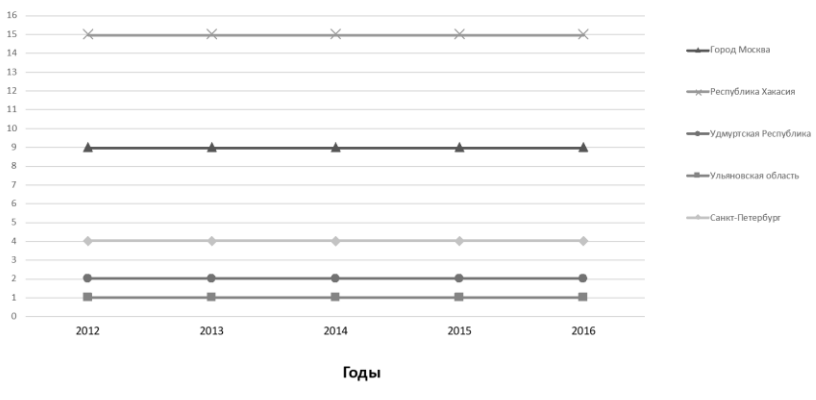

Рис. 5. Группы «абсолютно устойчивых» регионов РФ

Группа абсолютно неустойчивых объектов состоит из пяти субъектов РФ. Данные регионы меняли принадлежность паттерну каждый год в течение пяти лет: Архангельская обл., Иркутская обл., Краснодарский край, Красноярский край и Ханты-Мансийский автономный округ – Югра.

Результаты динамического анализа паттернов отражают тенденции развития энергетического сектора в Российской Федерации. Методы анализа паттернов не определяют причины, которые влияют на смену отдельными субъектами РФ паттернов своего энергетического развития, но описывают фактическое положение дел в исследуемой отрасли. Траектории развития отдельных субъектов РФ, полученные в ходе анализа, могут сигнализировать о наличии каких-либо проблем или, наоборот, улучшений в состоянии энергетического сектора субъектов РФ.

Республика Хакасия относится к группе абсолютно устойчивых объектов. На всем пятилетнем промежутке данный субъект РФ принадлежит паттерну №15 (рис. 6).

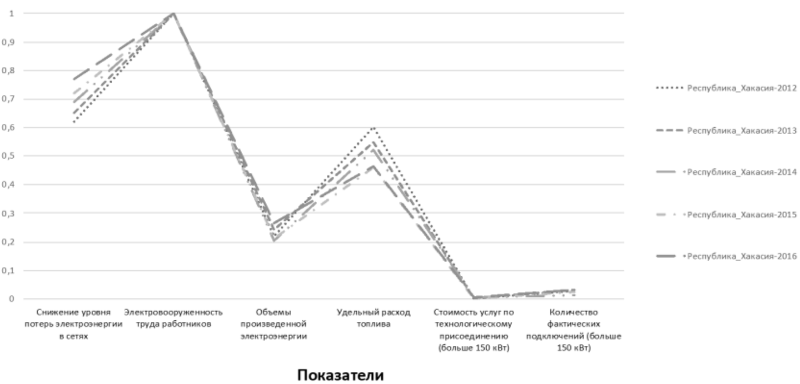

Рис. 6. Кусочно-линейные функции, описывающие объекты паттерна №15

Паттерн характеризуется достаточно высокими значениями первого показателя (более 0,6) и самыми высокими значениями электровооруженности труда среди всех регионов страны. Объекты находятся на крайне низком (менее 0,1) уровне по количеству фактических подключений и стоимости услуг по технологическому присоединению. Объемы произведенной электроэнергии концентрируются в диапазоне между 0,2 и 0,3. Что касается значений показателя удельного расхода топлива на отпущенную электроэнергии, то они находятся на среднем уровне (от 0,4 до 0,6).

Стоит отметить, что республика Хакасия демонстрирует стабильное энергетическое развитие на рассматриваемом промежутке времени. В течение пяти лет регион характеризовался одними из самых высоких темпов снижения уровня потерь электроэнергии в сетях. Кроме того, наблюдаются улучшения в значениях показателя удельного расхода топлива. С каждым годом динамика данного показателя улучшалась, что свидетельствует о повышении энергоэффективности в регионе.

Устойчивость региона может рассматриваться с двух точек зрения: стабильное развитие или, напротив, деградация. Это зависит от того, какому паттерну принадлежит объект в каждый период времени. Так, если в регионе проводится эффективная энергетическая политика, значения ключевых показателей будут улучшаться. В свою очередь, следование неправильной политике в области энергетического развития найдет свое отражение в ухудшающейся или неменяющейся динамике значений показателей энергоэффективности и принадлежности объектов соответствующим паттернам.

6. Заключение

Использование методов анализа паттернов позволило выявить основные структурные особенности функционирования энергетического сектора в регионах Российской Федерации в статике и динамике.

В результате исследования изучены характеристики энергетического развития субъектов Российской Федерации за пятилетний промежуток времени. Для разбиения регионов по отдельным группам, характеризующимся уникальной качественной структурой данных, использовались порядково-фиксированная и порядково-инвариантная паттерн-кластеризации. Результатом статического анализа паттернов стало выделение 16 уникальных объединений объектов, качественно отличающихся друг от друга. Для оценки устойчивости энергетического развития регионов была проведена процедура динамического анализа паттернов. Проанализировав полученные траектории развития, удалось обнаружить типовые динамические группы.

Разбиение регионов на группы объектов, характеризующиеся схожей структурой энергетического развития, и определение траекторий развития энергетического сектора субъектов РФ с помощью предложенного метода анализа паттернов может способствовать выработке более эффективной государственной политики в области энергоэффективности, которая будет учитывать особенности функционирования энергетического сектора в различающихся по масштабу и социально-экономическому развитию регионах.

Литература

1. АЛЕСКЕРОВ Ф.Т. и др. *Анализ данных науки, образования и инновационной деятельности с использованием методов анализа паттернов* // Высшая школа экономики. – Серия WP7. – 2012. – №7. – 62 с.
2. АЛЕСКЕРОВ Ф.Т. и др. *Анализ паттернов в статике и динамике, часть 2: Примеры применения к анализу социально-экономических процессов* // Бизнес-информатика. – 2013. – №4(26). – С. 3–20.
3. АЛЕСКЕРОВ Ф.Т., ШЕРМАН И.В., ЭНТОВ Р.М. *Анализ эффективности конкурсного управления при банкротстве банков* // Банковское дело. – 2008. – №12. – С. 70–76.
4. АЛЕСКЕРОВ Ф.Т., БЕЛОУСОВА В.Ю., ЕГОРОВА Л.Г., МИРКИН Б.Г. *Анализ паттернов в статике и динамике, часть 1: Обзор литературы и уточнение понятия* // Бизнес-информатика. – 2013. – №3(25). – С. 3–18.
5. АЛЕСКЕРОВ Ф.Т., СОЛОДКОВ В.М., ЧЕЛНОКОВА Д.С. *Динамический анализ паттернов поведения коммерческих банков России* // Экономический журнал Высшей школы экономики. – 2006. – №1. – С. 48–61.

6. БАКЛАНОВ Г.И., АДАМОВ В.Е., УСТИНОВ А.Н. *Статистика промышленности. Учебник*. / Под ред. В.Е. Адамова. – М.: Финансы и статистика, 1982. – 439 с.
7. БЕЛОВА Т.Д. *Методика оценки энергоэффективности региональной экономики* // Региональная экономика: теория и практика. – 2016. – №3. – С. 82–91.
8. ГОРЮНОВ В.Н., ДЕД А.В., ЖИЛЕНКО Е.П., ЛАВРИКОВ Ю.П., СМИРНОВ П. С. *Анализ сведений о потерях электрической энергии в филиалах ПАО «МРСК Сибири» за период с 2010 по 2017 год* // ОНВ. – 2018. – №6. – С. 25–30.
9. *Государственная программа российской федерации «Энергосбережение и повышение энергетической эффективности на период до 2020 года»* // Министерство энергетики Российской Федерации.
10. ЕФРЕМОВ В.В., МАРКМАН Г.З. *«Энергосбережение» и «энергоэффективность»: уточнение понятий, система сбалансированных показателей энергоэффективности* // Известия Томского политехнического университета. Инжиниринг георесурсов. – 2007. – Т. 311, №4. – С. 146–148.
11. КОЗЛОВА Е.И., НОВАК М.А., ЧЕРНИКОВА О.А. *Развитие электроэнергетики как фактор повышения конкурентоспособности отечественной экономики на мировой арене* // Инновационная экономика: перспективы развития и совершенствования. – 2018. – №3. – С. 81–87.
12. КОСЯКОВ С.В., САДЫКОВ А.М. *Метод зонирования территории по стоимости технологического присоединения к электрическим сетям* // Вестник Ивановского государственного энергетического университета. – 2013. – №5. – С. 1–5.
13. МАЗУРОВА О.В. *Электровооруженность промышленности как фактор качественного скачка в росте производительности труда* // Промышленная энергетика. – 2017. – №5. – С. 2–8.
14. МАРЧЕНКО Е.М., БЕЛОВА Т.Д. *Кластеризация регионов с учетом показателей энергоэффективности* // Региональная экономика: теория и практика. – 2016. – №1. – С. 51–60.

15. МИНИНА Е.А., БАКУМЕНКО Л.П. *Анализ уровня энергоэффективности в регионах российской федерации* // Редакционная коллегия. – 2017. – С. 92.
16. МЯЧИН А.Л. *Анализ паттернов в системе параллельных координат на базе парного сравнения показателей* // Автоматика и телемеханика. – 2019. – №1. – С. 138–152.
17. МЯЧИН А.Л. *Анализ паттернов: порядково–инвариантная паттерн–кластеризация* // Управление большими системами. – 2016. – №61. – С. 41–59.
18. МЯЧИН А.Л. *Определение центроидов для повышения точности порядково-инвариантной паттерн-кластеризации* // Управление большими системами. – 2019. – №78. – С. 6–22.
19. ХАБИБРАХМАНОВ Р.Р. *Моделирование влияния энергетического фактора на динамику изменения валового внутреннего продукта России* // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. – 2012. – №82. – С. 1–12.
20. ЧУЧУЕВА И.А. *Вычислительные методы определения удельных расходов условного топлива ТЭЦ на отпущенную электрическую и тепловую энергию в режиме комбинированной выработки* // Машиностроение и компьютерные технологии. – 2016. – №2. – С. 135–165.
21. *Энергетическая стратегия России на период до 2035 года* // Аналитический центр при Правительстве Российской Федерации.
22. ALESKEROV F., ERSEL H., YOLALAN R. *Clustering Turkish commercial banks according to structural similarities* // Yapi Kredi Discussion Paper Series. – 1997. – P. 97–102.
23. ALESKEROV F.T., ALPER C.E. *Inflation, Money, and Output Growths: Some Observations* // Bogazici Univer. Res. Paper. 1996. – No. SBE 96-06. – P. 3–13.
24. GENG Z., ZENG R., HAN Y., ZHONG Y., FU H. *Energy efficiency evaluation and energy saving based on DEA integrated affinity propagation clustering: Case study of complex petrochemical industries* // Energy. – 2019. – Vol. 179. – P. 863–875.
25. https://ipcrem.hse.ru/news/147589655.html (дата обращения: 01.01.2021).

26. LIU G., YANG J., HAO Y., ZHANG Y. *Big data-informed energy efficiency assessment of China industry sectors based on K-means clustering* // Journal of Cleaner Production. – 2018. – Vol. 183. – P. 304–314.

THE STUDY OF ENERGY SUSTAINABILITY OF REGIONS OF RUSSIAN FEDERATION BASED ON PATTERN ANALYSIS

Alexey Myachin, National Research University Higher School of Economics, Moscow, Cand.Sc., associate professor, Institute of Control Sciences of RAS, Moscow, senior research fellow (amyachin@hse.ru).
Vadim Prokofiev, National Research University Higher School of Economics, Moscow, Cand.Sc., associate professor (vprokofiev@hse.ru).
Alexander Stepanov, National Research University Higher School of Economics, Moscow, student (aastepanov_3@edu.hse.ru).

Abstract: The study is devoted to the application of ordinal-fixed and ordinal-invariant pattern clustering to study the structure of the energy sector in regions of Russian Federation over a five-year period. Methods for pattern analysis in the work are due to the independence of the final results from the difference in the absolute values of indicators and the possibility of combining regions that are similar in structure (based on the indicators under consideration). The choice of ordinal-invariant pattern clustering is due to both the endogenous determination of not only the composition of each pattern, but their number, and the lack of the need to choose the initial sequence of the studied indicators, to which some methods of pattern analysis are sensitive. The study was carried out on the basis of such indicators as: a decrease in the level of electricity losses in the networks, the electric power supply of workers, the volume of electricity produced, specific fuel consumption, the cost of services for technological connection, as well as the number of actual connections. As a result of the study, the features of the energy development of individual subjects were studied. The development patterns were obtained, reflecting the similarity of the internal data structure. Dynamic trajectories of development have been built and groups of subjects have been identified that adhere to a constant strategy from year to year. The above results can be taken into account in the development of state policy in the field of energy efficiency.

Keywords: pattern, pattern analysis, energy sustainability, cluster analysis.

УДК 021.8 + 025.1
ББК 78.34
DOI: 10.25728/ubs.2021.92.3

Статья представлена к публикации членом редакционной коллегии Я.И. Квинто.

*Поступила в редакцию 14.04.2021.
Опубликована 31.07.2021.*

ПОНЯТИЙНЫЙ АНАЛИЗ И ПОНЯТИЙНОЕ МОДЕЛИРОВАНИЕ

Выхованец В. С.[1]
(*ФГБУН Институт проблем управления им. В.А. Трапезникова РАН, Москва*)

Описываются понятийные модели, которые строятся на первичных ментальных абстракциях идентификации, обобщении и ассоциации. В процессе понятийного анализа строится понятийная модель предметной области, состоящая из понятийной структуры и содержания входящих в нее понятий. Понятийная структура определяет каждое понятие как результат обобщения или ассоциации других понятий. Содержание понятий задается перечислимыми или разрешимыми множествами сущностей предметной области. Основным отличием понятийной модели от других моделей знаний является отказ от описания ассоциации понятий в виде взаимосвязи. В понятийной модели ассоциации являются такими же понятиями, как и обобщения, что позволяет образовывать из ассоциаций новые понятия. Все это делает понятийную модель семантически инвариантной, т.е. не зависящий в своей интерпретации от знаний из предметной области. Еще одним отличием понятийных моделей является многоаспектное выражение понятий. Для обоснования предложенного подхода разработана формальная теория понятий. Семантическая часть теории (язык понятий) обосновывает методологию понятийного анализа, а синтаксическая ее часть (исчисление понятий) – технологию понятийного моделирования. Доказано, что язык понятий разрешим, полон и непротиворечив, в то время как исчисление понятий разрешимо и непротиворечиво на счетных моделях, а полно только на конечных моделях. Использование понятийного анализа и понятийного моделирования позволяет улучшить выразительность и наглядность представления знаний, повысить эффективность и надежность их обработки.

Ключевые слова: ментальные абстракции, понятие, концепт, понятийный анализ, понятийная модель, язык понятий, исчисление понятий, база знаний, вывод на знаниях, интеллектуальная система.

1. Введение

Нерешенные проблемы в области представления и обработки знаний затрудняют эффективную реализацию современных баз знаний и интеллектуальных информационных систем на их основе. В частности, извлечение знаний до сих пор не является

---

[1] *Валерий Святославович Выхованец, д.т.н., доцент* (valery@vykhovanets.ru).

полностью решенной задачей [16], формы представления знаний слабо адаптированы к психическим и психологическим особенностям человека [5, с. 31], обработка знаний занимает длительное время и подвержена возникновению скрытых противоречий [12].

Причина таких проблем заключается в том, что при моделировании человеческого мышления до сих пор используются традиции, заимствованные из логики, где не учитываются такие свойства человеческого мышления, как объективность (открытость к новому знанию, замена устаревшего знания новым), субъективность (перестройка мышления для решения конкретных прикладных задач), рефлексивность (самоорганизация познавательной деятельности), продуктивность (постоянная генерация нового знания) [1].

Некоторые из существующих проблем удалось избежать с помощью понятийного анализа и понятийного моделирования. Модели называются понятийными, чтобы отличать их от концептуальных моделей. В концептуальных моделях задаются понятия и различные взаимосвязи (отношения) между ними.

Например, в таких концептуальных моделях, как логические, продукционные, дескрипционные, грамматические, фреймовые, объектно-ориентированные, семантические сети и т.д. [11, с. 31], отношения между сущностями предметной области определяются как предикаты, которые интерпретируются как первичные семантические категории предметных знаний и не используются для образования других понятий. А если такие понятия все же требуется образовать, то выразительные возможности логического языка концептуальной модели необходимо повысить до логики второго порядка [10, с. 375]. Последнее порождает неразрешимые проблемы – в теоретическом плане, и непреодолимые трудности – в прикладном.

Наиболее близкой к понятийной модели является ER-модель, где абстракциями концептуального проектирования являются классификация, агрегация и обобщение [4, с. 41], которые используются для классификации примитивных объектов и построения из них сложных классов. Для задания взаимосвязи классов используется абстракция агрегации, которая не порождает ни объекта, ни класса, а рассматривается как семантическая категория, посредством которой производится интерпретация модели.

В понятийной модели отношения между понятиями являются обычными понятиями [15], понятийная модель строится путем указания абстракций, использованных для образования понятий, а также перечисления сущностей предметной области, им принадлежащих. Каждое понятие описываются в одном или нескольких аспектах, что позволяет учесть многовариантность и изменчивость предметных знаний.

2. Понятия и концепты

Понятие – это вид мысли, которая соотносится с определенным набором уникальных представлений (сущностей) внутреннего или внешнего мира человека (предметной области).

Понятие (англ. notion) – это простая идея, мнение, представление или понимание чего-либо; понятийный (англ. notional) – это гипотетический, воображаемый [14, с. 607]. Понятие соотносится с представлениями из предметной области – сущностями (англ. entities). В свою очередь концепт (англ. concept) – это общее понятие, абстрактная идея [14, с. 171]. Таким образом, в отличие от концепта, понятие не объективно, а субъективно по определению.

Концепт состоит из экземпляров (англ. instances), каждый экземпляр характеризуется совокупностью свойств (англ. properties), а свойства могут быть определенными (англ. defining) и неопределенными (англ. nondefining) [11, с. 38].

Понятия формируются (определяются) в процессе абстрагирования путем выполнения ментальных операций над сущностями. Каждое понятие имеет имя, которое обозначает некоторый набор сущностей из предметной области.

Существует три вида понятий и порождающих их ментальных абстракций (отвлечений): понятия-знаки (абстракция идентификации, отождествления), понятия-обобщения (абстракция обобщения) и понятия-ассоциации (абстракция ассоциации).

2.1. ИДЕНТИФИКАЦИЯ

Понятия-знаки – это результат мысленного отбора уникальных представлений из предметной области и присвоения им имен. Понятия-знаки формируются для фиксации определенного состояния органов чувств или элементарных представлений. При формировании понятия-знака сущность мысленно заменяется знаком (именем), в результате чего между сущностями и знаками устанавливается взаимно однозначное соответствие. Противоположной абстракцией для идентификации является абстракция интерпретации, при которой понятие-знак сопоставляется с некоторой уникальной сущностью из предметной области.

Пример 1. Примерами понятий-знаков могут служить понятия, обозначенные такими знаками (словами) как «Зеленый», «Длинный», «Первый», «Редко», «Много» и т.п. ♦

2.2. ОБОБЩЕНИЕ

Понятие-обобщение образуется при объединении сущностей обобщаемых понятий (объединение множеств сущностей). Абстракция обобщения используется также для образования частного вида понятия-обобщения – понятия-типа, представляющего собой обобщение (типизацию) понятий-знаков. Обобщение (типизация) имеет обратную абстракцию, называемую специализацией (конкретизацией).

Пример 2. Примером понятия-обобщения является понятие «Дерево», которое объединяет деревья различных пород: «Береза», «Бук», «Пихта», «Сосна», «Тополь» и т.д. В свою очередь «Береза» является специализацией понятия-обобщения «Дерево». ♦

Понятия-обобщения объединяют понятия, схожие в какомлибо смысле, в то время как понятия-типы используются для объединения однородных сущностей.

Пример 3. Понятие-тип «Цвет» может быть образовано путем типизации таких понятий-знаков как «Красный», «Оранжевый» и т.д., а понятие-знак «Красный» является конкретизацией понятия-типа «Цвет». ♦

2.3. АССОЦИАЦИЯ

Понятие-ассоциация образуется при соединении сущностей ассоциируемых понятий, когда каждая сущность понятия-ассоциации включает в себя по одной сущности ассоциируемых понятий (подмножество декартова произведения множеств сущностей). Не все комбинации сущностей ассоциированных понятий могут быть сущностями понятия-ассоциации.

Частным случаем абстракции ассоциации является абстракция агрегации, при которой образовываемому понятию-агрегату принадлежат все комбинации сущностей агрегируемых понятий (полное декартово произведение множеств сущностей). Ассоциация (агрегация) имеет обратную абстракцию, называемую индивидуализацией (декомпозицией).

Пример 4. Примером понятия-ассоциации является понятие «Погода», которое соединяет такие понятия как «Место», «Дата», «Температура», «Влажность», «Ветер», «Облачность» и т.п. Сущностью понятия «Погода» может быть «г. Москва, 15 марта 2021 г., 8°C, 45 %, 3 м/с, Облачно, …». ♦

3. Понятийный анализ

Понятийный анализ состоит в выявлении понятий и их абстракций, которые необходимы и достаточны для описания заданной предметной области и решаемых в ней задач.

Абстракции идентификации, ассоциации и обобщения, используемые при понятийном анализе, рассматриваются как мыслительные операции, необходимые и достаточные для выделения и превращения в понятия имеющихся представлений из описываемой предметной области.

Основная цель понятийного анализа – многовариантная структуризация предметной области в виде ее понятийной структуры.

3.1. ПРОБЛЕМНЫЕ ОБЛАСТИ

Проблемы однозначного и непротиворечивого описания концептов, связанные с изменчивостью набора свойств у экземпляров, привели к появлению альтернативных теорий концептуального моделирования [11, с. 39]. Часть таких теорий базируется на вероятностном подходе, при котором набор свойств расширяется и их появление у экземпляров концепта имеет вероятностную природу. Другие теории разделяют свойства экземпляров концепта на типичные, присущие всем экземплярам концепта, и атипичные, которые могут появляться лишь у некоторых экземпляров.

При понятийном моделировании проблемы, возникшие при описании концептов, решаются на основе выделения в предметной области проблемных областей [1]. Поскольку понятие предполагает субъективное отражение предметной области, то необходимо предусмотреть возможность нескольких альтернативных описаний одного и того же понятия, в каждой проблемной области по-своему. Это позволит в последующем объективировать понятие в его коллективной (концептуальной) форме.

Проблемная область – это предметная область, рассматриваемая в некотором узком смысле (аспекте), с точки зрения некоторой активной проблемы, всякий раз по иному структурирующей анализируемую предметную область. Одно и то же понятие в разных аспектах (в разных проблемных областях) может иметь разные определения и описания, а обобщение одноименных понятий из различных проблемных областей может рассматриваться как концепт предметной области.

Пример 5. Концепт «Противоположность» является обобщением понятий «Противоположность» в аспектах «Свет», «Красота» и т.д., которые, в свою очередь, являются понятиями-ассоциациями, соединяющими понятия, сущности которых противоположны в каком-то смысле (ассоциация по контрасту). Для понятия «Противоположность» в аспекте «Свет» это сущности «Яркий, Неяркий», «Яркий, Тусклый», «Тусклый, Сильный», «Видимый, Невидимый», и т.д. Для понятия «Противоположности» в аспекте «Красота» это сущности «Красивый, Некрасивый», «Красивый, Уродливый», «Прекрасный, Некрасивый» и т.д. ♦

3.2. ПОНЯТИЙНАЯ СТРУКТУРА

Результатом понятийного анализа является понятийная структура предметной области. В понятийной структуре перечисляются понятия и их аспекты, а также способы их абстрагирования.

Понятия, использованные для образования других понятий, называются атрибутами, а само понятие рассматривается как ассоциация своих атрибутов.

К общим атрибутам понятий относится понятие-тип «Имя» (имя понятия), понятие-тип «Аспект» (имя проблемной области), а также понятие-тип «Абстракция» (способ образования понятия), состоящий из понятий-знаков «Знак», «Тип», «Обобщение» и «Ассоциация». Частные атрибуты понятия или просто атрибуты – это понятия, необходимые для его образования (абстрагирования).

Понятийная структура состоит из схем понятий. Схема понятия – это упорядоченный набор атрибутов, первые три элемента которого – общие атрибуты, остальные – частные атрибуты понятия.

Пример 6. На рис. 1 показан фрагмент понятийной диаграммы некоторой предметной области. Диаграмма состоит из геометрических фигур, внутри которых указаны имена понятий и их аспекты. Окружности используются для задания понятий-знаков, овалы – понятий-типов, ромбы – понятий-ассоциаций, а прямоугольники – понятий-обобщений. Стрелки на диаграмме соединяют понятий-атрибуты с образовываемыми ими понятиями.

Из диаграммы могут быть получены схемы понятий, входящие в ее понятийную структуру:

– «Разряд»: «Типизация», «1», «2», «3»;
– «Рабочий»: «Ассоциация», «Профессия», «Разряд», «ФИО»;
– «Инженер»: «Ассоциация», «Специальность», «ФИО»;
– «Руководитель»: «Ассоциация», «Инженер», «Бригада»;
– «Сотрудник»: «Обобщение», «Инженер», «Рабочий»;
– «Участник»: «Ассоциация», «Сотрудник», «Бригада»;

где имя понятия отделено от других атрибутов двоеточием. Понятия-типы «Специальность», «ФИО», «Профессия», «Бригада» на диаграмме не определены, как и не указаны аспекты. ♦

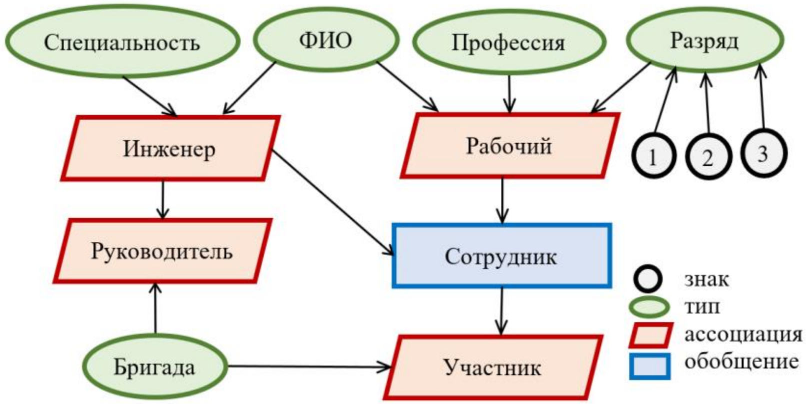

Рис. 1. Фрагмент понятийной диаграммы

Следует заметить, что по определению абстракций на понятийной диаграмме:

– к каждому понятию-обобщению и понятию-типу может идти только одна стрелка от каждого обобщаемого (типизируемого) понятия;

– к каждому понятию-ассоциации может идти одна и более стрелок от каждого ассоциируемого понятия.

Стрелка, идущая от обобщаемого (типизируемого) понятия и понятию-обобщению (понятию-типу) может интерпретироваться как взаимосвязь типа «член – группа» (англ. IsMemeberOf), а от ассоциируемого понятия к понятию-ассоциации – как взаимосвязь типа «часть – целое» (англ. sPartOf).

4. Понятийное моделирование

Понятийная модель состоит из понятийной структуры и описаний сущностей ее понятий, выполненных с использованием средств перечисления и разрешения множеств.

4.1. ПОНЯТИЙНАЯ МОДЕЛЬ

В современных информационных системах понятийная модель предметной области может быть реализована с использованием системы управления базами данных реляционного типа.

В качестве понятий-знаков используются значения простых типов данных, а сами простые типы данных рассматриваются как встроенные понятия-типы, например, бит «Bit», целое число «Integer», число с плавающей запятой «Float», дата и время «DateTime», символ «Char», строка «Varchar», двоичные данные «Binary» и т.д. [6, с. 51].

Специфические для предметной области понятия-типы создаются путем ограничения простых типов данных и реализуются в виде таблиц базы данных, в которых перечисляются сущности понятия-типа, или в виде формул, задающих разрешающие процедуры для множества сущностей понятия-типа [6, с. 268].

Понятия-ассоциации реализуются в виде таблиц базы данных (англ. tables), соединяющих в одной записи сущности ассоциированных понятий, а понятия-обобщения представляются как запросы (англ. queries) или представления (виртуальные таблицы, англ. views), объединяющие записи из таблиц обобщаемых понятий [6, с. 180, 188, 228].

Понятия-ассоциации по своей реализации близки к типам отношений (англ. relation types), а понятия-обобщения – к типам сущностей (англ. entity types) [11, с. 41, 59].

Помимо декларативных понятия могут иметь процедурные описания и реализовываться, например, в виде процедур (функций) на языке базы данных. Процедурные средства позволяют генерировать таблицы понятий, а также преобразовывать сущности-аргументы в сущности-результаты [6, с. 335].

Для отслеживания и обработки событий, возникающих в понятийной модели (создание, изменение и удаление сущностей), может использоваться механизм триггеров [6, с. 226].

Пример 7. Фрагмент диаграммы базы данных с понятийной моделью представлена на рис. 2, где показаны таблицы трех понятий: «Notion», «Attribute» и «Icon», а также ключевые поля, хранящие первичные ключи записей (англ. primary keys), и связи между таблицами, реализованные в виде полей с внешними ключами (англ. foreign keys) [6, с. 163].

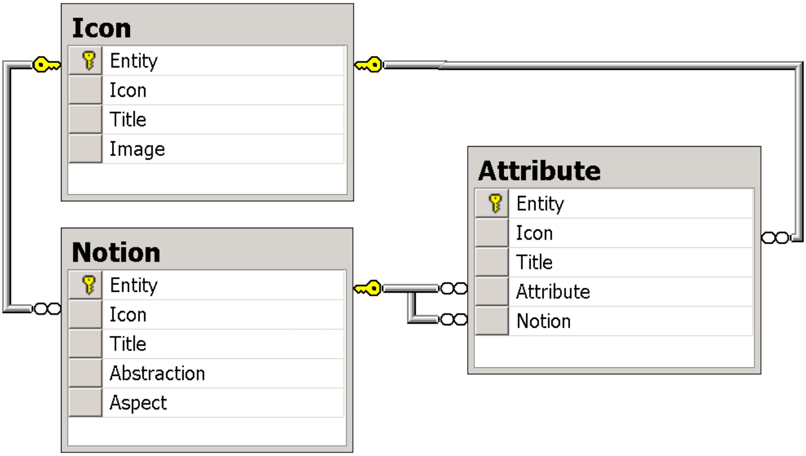

Рис. 2. Фрагмент диаграммы базы данных

Каждая запись в таблицах понятийной модели представляет собой сущность, имеющую следующие общие поля (общие атрибуты): Entity – уникальный идентификатор сущности, Tittle – наименование сущности (декоративное имя), Icon – иконка сущности (уникальный идентификатор изображения в таблице Icon). Поля Tittle и Icon используются для отображения сущностей в интерфейсе пользователя.

Таблица Notion имеет частные атрибуты Abstraction и Aspect, которые представляются полями встроенных типов данных с сущностями соответствующих понятий-типов: в поле Abstraction записывается один из символов, обозначающих абстракцию понятия: идентификация – *C*, типизация – *T*, ассоциация – *A* и обобщение – *G*, а в поля Aspect – строка имени аспекта (рис. 3).

Таблица Icon имеет частный атрибут Image, который является встроенным типом двоичных данных и содержит изображение иконки (рис. 4).

Таблица Attribute перечисляет частные атрибуты понятий и имеет два частных атрибута Attribute и Notion. Каждая запись в таблице Attribute описывает соответствующие поля в таблицах понятий, в том числе и поля самой таблицы Attribute (рис. 5). ♦

Entity	Icon	Title	Abstraction	Aspect
61000000060	63000000054	Attribute	A	Common
61000000061	63000000020	Notion	A	Common
61000000062	63000000018	Model	A	Common
61000000063	63000000019	Icon	A	Common
61000000064	63000000021	Abstraction	A	Common
61000000065	63000000034	Aspect	A	Common

Рис. 3. Фрагмент таблицы понятия Notion

Entity	Icon	Title	Image
63000000018	63000000018	Model	0x424D3605000000...
63000000019	63000000019	Icon	0x424D3605000000...
63000000020	63000000020	Notion	0x424D3605000000...
63000000021	63000000021	Abstract	0x424D3605000000...
63000000022	63000000022	Abstraction	0x424D3605000000...
63000000023	63000000023	Tag	0x424D3605000000...

Рис. 4. Фрагмент таблицы понятия Icon

Entity	Icon	Title	Attribute	Notion
6000000003	63000000020	Notion	61000000061	61000000060
6000000004	63000000023	Attribute	61000000061	61000000060
6000000001	63000000021	Abstract	61000000064	61000000061
6000000002	63000000034	Aspect	61000000065	61000000061
6000000019	63000000019	Image	61000000045	61000000063

Рис. 5. Фрагмент таблицы понятия Attribute

Для уникальной идентификации сущностей понятийной модели ключевые поля состоят из двух частей: первая часть – это номер записи сущности в таблице понятия, а вторая часть – это номер записи понятия как сущности в таблице понятий.

Пример 8. Пусть уникальным идентификатором иконки является число 63000000020. Из анализа идентификатора следует, что иконка как сущность имеет номер 20 в таблице иконок Icon (рис. 4), а иконка как понятие «Icon» имеет номер 63 в таблице понятий Notion (рис. 3). ♦

Следует обратить внимание на то, что в понятийной модели понятие само является понятием и описывается теми же средствами, что и другие понятия.

4.2. БАЗА ЗНАНИЙ

Понятийная база знаний – это база данных, содержащая понятийную модель, а также механизм вывода, позволяющий обрабатывать предметные знания.

Обработка знаний подразумевает наличие форм представления знаний и методов манипулирования ими с целью имитации рассуждений человека. Обычно для представления знаний используются факты (суждения), а для обработки знаний – правила вывода, позволяющие на основе имеющихся фактов делать умозаключения и получать новые утверждения (факты) об имеющихся или вновь выводимых фактах [5, с. 49].

4.2.1. ФАКТЫ

В понятийной базе знаний фактами являются истинные высказывания с логическим связками И ($\wedge$), ИЛИ ($\vee$), НЕ ($\neg$), круглыми скобками и двумя видами элементарных выражений: одноместными предикатами принадлежности сущности X понятию N вида $N(X)$, а также отношениями вида $X[N] \div Y$, где $X[N]$ – функтор, возвращающий сущность понятия-атрибута N сущности X; $\div$ – знак отношения, допустимого между сущностями $X[N]$ и Y (равно =, больше >, и т.п.). Функтор $X[N]$ возвращает специальную сущность ε (пустое понятие), если сущность X не имеет атрибута N.

Универсальное отношение равенства сущностей определяется рекурсивно: две сущности равны тогда и только тогда, когда значения (сущности) атрибутов этих сущностей равны. Другие отношения могут быть унаследованы от встроенных типов данных, используемых в понятийной модели.

Пример 9. Для понятийной модели, диаграмма которой показана на рис. 1, могут быть выражены следующие факты:

– Разряд(2) – 2 есть сущность понятия-типа «Разряд»;

– Рабочий(X) – X есть сущность понятия-ассоциации «Рабочий» или, если кратко, X – Рабочий;

– X[Разряд] = 3 – Рабочий X имеет Разряд, равный 3. ♦

4.2.2. ПРАВИЛА ВЫВОДА

В понятийной модели схемы понятий задают правила вывода, а понятийная структура рассматривается как формальная теория предметной области, которая сохраняет истинность всех выводимых в ней следствий.

Из схемы понятия-обобщения следует правило вывода, согласно которому сущность X принадлежит понятию N тогда и только тогда, когда эта сущность принадлежит хотя бы одному понятию-атрибуту понятия N.

Из схемы понятия-ассоциации следует правило вывода, согласно которому если сущность X принадлежит понятию N, то атрибуты сущности X являются атрибутами понятия N.

Пример 10. Понятийная структура, диаграмма которой показана на рис. 1, порождает следующие правила вывода (но не все):

– $\forall X$ Сотрудник(X) $\leftrightarrow$ Инженер(X) $\vee$ Рабочий(X);

– $\forall X$ Участник(X) $\rightarrow$ Сотрудник$\{X\}$ $\wedge$ Бригада$\{X\}$,

где $\forall$ – квантор всеобщности; $\rightarrow$ ($\leftrightarrow$) – импликация (двухсторонняя импликация); $N\{X\}$ – предикат наличия у сущности X атрибута N, $N\{X\} \leftrightarrow \neg\, X[N] = \varepsilon$; ε – пустая сущность. ♦

При выводе на знаниях требуется установить или опровергнуть некоторый факт путем поиска сущностей, удовлетворяющих заданным условиям. В этом случае правила, следующие из схемы понятия-обобщения, задают понятия, подлежащие обработке, а правила, следующие из схемы понятия-ассоциации, позволяют установить наличие атрибутов у сущностей этого понятия и извлечь значения атрибутов, если это необходимо.

4.2.3. ЗАПРОСЫ

Для поиска в базе знаний сущностей, удовлетворяющих заданным условиями, используются запросы.

Пример 11. Рассмотрим базу знаний с понятийной структурой, изображенной на рис. 1. Из рисунка видно, что в понятийной модели определены следующие предикаты и функторы:

– Инженер(X), X[Специальность], X[ФИО];
– Рабочий(Y), Y[ФИО], Y[Профессия], Y[Разряд];
– Руководитель(Z), Z[Инженер], Z[Бригада];

– Сотрудник(V), V[Инженер], V[Рабочий];
– Участник(W), W[Сотрудник], W[Бригада],

где «Специальность», «ФИО», «Профессия», «Разряд» и «Бригада» являются понятиями-типами, заданными перечислением принадлежащих им сущностей, как, например, это сделано для понятия-типа «Разряд».

Следует обратить внимание на функторы V[Инженер] и V[Рабочий] понятия-обобщения «Сотрудник». Если V – сущность понятия «Инженер», то V[Инженер] = V, в противном случае V[Инженер] = ε. Аналогично ведет себя и функтор V[Рабочий].

Используя имена сущностей и их переменные, а также перечисленные выше предикаты и функторы, можно формулировать запросы к базе знаний, например, такой: «Кто из участников бригады В не является рабочим с третьим разрядом?». На языке формальной логики этот запрос будет иметь выражение

$$\text{Участник}(X) \wedge X[\text{Бригада}] = \text{В} \wedge \forall Y\, (X[\text{Сотрудник}] = Y \wedge$$
$$\wedge\ \text{Рабочий}(Y) \wedge \neg\, Y[\text{Разряд}] = 3),$$

в котором X является свободной переменной, а Y – связанной. ♦

4.2.4. ПРОЦЕДУРНЫЕ ЗНАНИЯ

Декларативные знания в понятийной базе знаний представляются понятиями и принадлежащими им сущностями. Для представления процедурных знаний предусмотрены процедуры и функции, которые, как и все в понятийных моделях, являются сущностями соответствующих понятий.

Без ограничения общности операциями, через которые выражаются процедурные знания, являются операции создания, изменения и удаления сущностей. Для повышения уровня процедурного языка до универсальных языков программирования предусматривается операция присваивания, а также условные операторы и операторы циклов. Условия у этих операторов выражаются в виде фактов, выполнимость которых проверяется. В свою очередь результат выполнения запроса может быть присвоен внутренней переменной для дальнейшей обработки извлеченных знаний.

Рассмотрим процедурный язык понятийной базы знаний (англ. knowledge language). Встроенными типами данных этого языка являются байты Byte, символы Char, натуральные числа Number, целые числа Integer, числа с плавающей запятой Float со встроенными операциями в стиле языка C [8].

В язык добавлен встроенный тип данных DateTime, а также строковый String и двоичный Binary типы с двуместными операциями конкатенации `` (двойной обратный апостроф, англ. grave accent) и присваивания с конкатенацией ``=. Сущностями последних двух типов являются последовательности символов и байтов конечной длины, например: ″строка″, 0x12DE3A. Сущности понятия-типа «DateTime» – это дата и/или время, заключенные в апострофы, например: ′29.05.2021 14:32:56′, ′14:32:56′.

Для представления процедурных знаний в язык введен встроенный тип данных Program, сущности которого – последовательности команд некоторой виртуальной машины, генерируемые лямбда-выражениями в стиле языка Java (*X*) –> *Y*, где *X* – аргументы (могут отсутствовать), а *Y* – оператор тела функции (процедуры).

Единственной встроенной понятийной константой является пустое понятие ′ ′ (двойной апостроф) с двуместными операциями объединения с пустым понятием ?? и присваивания с объединением с пустым понятием ??= в стиле языка C#.

Логическая часть языка стандартная:

– ==, != – отношения равенства и неравенства;
– >, <, <=, >= – отношения порядка;
– !, &&, || – логические связки НЕ, И, ИЛИ.

Для распараллеливания вычисления используются операции последовательного и параллельного вычисления:

– *X*, *Y*, …, *Z* – последовательное вычисление *X*, *Y*, …, *Z*;
– *X* :: *Y* :: … :: *Z* – параллельное вычисление *X*, *Y*, …, *Z* .

Для улучшения выразительности языка и повышения компактности программ расширен состав тернарных операций:

– *X* ? *Y* : *Z* – условное вычисление (если *X*, то *Y*, иначе *Z*);
– *X* ~ *Y* : *Z* – цикл (пока *X*, вычислять *Y*, потом *Z*);
– *X* ! *Y* : *Z* – блокировка *X* пока вычисляется *Y*, потом *Z*;
– *X* ` *Y* : *Z* – поиск в *X* вхождений *Y* и замена *Z*.

Понятийные операции языка:

– операции абстракций:

$[N1 : X1, \ \ldots, \ ND : XD]$ – создание понятия-ассоциации с атрибутами Xj и именами Nj ($j = 1, 2, \ldots, D$),

$\{N1 : X1, \ \ldots, \ ND : XD\}$ – создание понятия-обобщения с атрибутами Xj и именами Nj ($j = 1, 2, \ldots, D$);

– операции интенсионала:

$X[]$ – число атрибутов понятия X,

$X[Y]$ – доступ к атрибуту Y сущности X,

$[X]Y$ – доступ к атрибуту Y понятия X,

$[]X$ – понятие сущности X;

– операции экстенсионала:

$X\{\}$ – число сущностей понятия X,

$X\{Y\}$ – доступ к сущности Y понятия X,

$\{X\}Y$ – создание сущности X понятия Y,

$\{\}X$ – удаление сущности X;

– процедурные операции:

$X.Y$ – функция (процедура) Y понятия X,

$X()$ – выполнение процедуры X,

$X(Y)$ – выполнение функции X с аргументами Y,

$(X)Y$ – запрос Y со свободной переменной X,

$()X$ – обновление сущности X;

– операции отношений:

X =^ Y – понятие X включено в понятие Y (операция in),

X =? Y – понятие X есть понятие Y (операция is),

X =: Y – понятие X как понятие Y (операция as),

X =` Y – понятие X похоже на понятие Y (операция like);

– операции присваивания:

$X = Y$ – присваивание сущности X сущности Y,

X ||= Y – присваивание с объединением X и Y,

X &&= Y – присваивание с пересечением X и Y,

X ^^= Y – присваивание с исключением Y из X,

X ::= Y – присваивание с клонированием Y в X.

Операторы языка:

– операция присваивания и точка с запятой;

– составной оператор – операторы в фигурных скобках;

– goto – оператор перехода;

– switch – оператор выбора;
– if-then-else – условный оператор;
– while, do, for – операторы циклов;
– try-catch, throw – операторы исключений;
– begin, commit, rollback – операторы транзакций [6, с. 753].

Особенностью пространства имен языка является интерпретация всех переменных как понятий. Любая переменная – это массив сущностей. Если в массиве один элемент, то переменная является сущностью, но одновременно и понятием с одной принадлежащей ему сущностью.

Пример 12. Рассмотрим пример процедуры, которая повышает разряды рабочим – участникам бригады В (см. пример 11).

```
// Определение функции повышения разряда
Повышение = (Разряд) –> Разряд < 3 ? 1 : 0;
// Создание индексной переменной и ее инициализация
Индекс = 0;
// Открытие транзакции с именем Разряды
begin Разряды;
// Запрос сущностей локального понятия Рабочие
Рабочие = (Х)(Участник(Х) != ′ ′ && Х[Бригада] == В
    && Рабочий(Х[Сотрудник]) != ′ ′);
// Цикл по сущностям понятия Рабочие
while (Индекс < Рабочие{})
{
    // Создание локальной сущности Рабочий
    Рабочий = Рабочие{Индекс};
    // Увеличение атрибута Разряд сущности Рабочий
    Рабочий[Разряд] += Повышение(Рабочий[Разряд]);
    // Приращение индекса сущности понятия Рабочие
    Индекс = Индекс + 1;
}
```

// Обновление измененных сущностей

()Рабочие;

// Успешное завершение транзакции

commit Разряды;♦

Следует обратить внимание на то, что в программах операции присваивания создают только локальные переменные. Однако связь с базой знаний сохраняется для всех извлеченных из нее сущностей. Для внесения изменений в базу знаний после изменения сущностей используется оператор обновления.

Операция доступа к атрибуту позволяет извлекать и изменять не только частные атрибуты сущности *X* по их номерам или именам, но и общие атрибуты по их именам, например, наименование сущности *X*[Title], ее иконку *X*[Icon] (см. пример 7).

Однако изменение наименований и иконок никак не сказывается на обработке знаний, так как операции над сущностями осуществляются с использованием уникальных идентификаторов Entity (см. пример 8). Если для обеспечения целостности обработки знаний требуется зафиксировать наименования или иконки сущностей, то применяется механизм транзакций.

Функции на процедурном языке могут использоваться для вычисления виртуальных атрибутов понятий по значениям других атрибутов.

Так, для встроенного понятия-типа «DateTime» предусматриваются такие виртуальные атрибуты как Date, Year, Month, Day, Time, Hour, Minute, Second, Millisecond, а для понятия-типа String (Binary) – виртуальный атрибут длины Length.

Помимо виртуальных атрибутов могут определяться функции и процедуры понятий в стиле методов классов объектно-ориентированных языков программирования. Например, для понятия-типа «DateTime» таким методом является *X*.Diff(″w″, *Y*) – разница дат *X* и *Y* в неделях, а для понятий-типов String и Binary – индексаторы, возвращающие соответственно символ и байт с индексом *Y*: *X*.Char(*Y*) и *X*.Byte(*Y*).

Для задания статических функций может использоваться специальное понятие-ассоциация «Function», например, со следующими атрибутами:

Argument1: Notion, …, Argument16: Notion;
Arguments: Number;
Code: Program;
Return: Notion;
Text: String,

согласно которому сущность понятия «Function» состоит из аргументов с именами Argument1, …, Argument16, их числа Arguments, исполняемого кода Code, возвращаемого понятия Return и текста функции на процедурном языке Text. Примером такой функции может служить функция Function.Now(), которая возвращает текущую дату и время.

5. Формальная теория

Предыдущая часть статьи была посвящена разъяснению основных понятий и механизмов понятийного анализа и понятийного моделирования, а также описанию реализации базы знаний, в основе которой лежит понятийная модель. Рассмотрим теперь формальную сторону понятийного анализа и понятийного моделирования – теорию понятий.

Следует заметить, что иных более общих выразительных средств для формализации предметной области, кроме как ее описания в виде понятийной модели, не существует. Это объясняется тем, что понятийное мышление является предельно общим и единственным механизмом рационального познания окружающего мира, а понятийное моделирование – субъективной и, в конечном счете, объективной формой фиксации результатов понятийного мышления.

Очевидно, что предметная область является семантическим объектом теории понятий, а понятийная модель – ее синтаксическим объектом. Теория понятий будет полной, если каждая предметная область имеет свою понятийную модель, и каждая понятийная модель имеет свою предметную область, возможно, воображаемую.

По этой причине будем различать семантические и синтаксические понятийные модели, а также их теории. Семантические понятийные модели опишем формальной теорией, называемой языком понятий, а синтаксические понятийные модели формализуем в рамках аксиоматической теории, называемой исчислением понятий.

5.1. ОБЩИЕ ОПРЕДЕЛЕНИЯ

Теория понятий наследуется от теории множеств [7], которая, в свою очередь, включает в себя логику первого порядка [10, с. 41].

Определение 1. Алфавит Ω – это конечное упорядоченное множество без повторения элементов, состоящее из цифр и букв. ♦

Определение 2. Имя (строка) – это конечное упорядоченное множество с повторением элементов, принадлежащих алфавиту Ω. ♦

Определение 3. Пусть $A1$, $A2$, B, Dc и Dac – имена, тогда:

– $\{B, A2, A1, Dc, Dac\}$ – неупорядоченное множество без повторения элементов;

– $\langle B, A2, A1, A1, Dc, Dc, Dac\rangle$ – неупорядоченное множество с повторением элементов;

– $(B, A2, A1, Dc, A1, Dac, Dc)$ – упорядоченное множество с повторением элементов;

– $[A1, A1, A2, B, Dac, Dc, Dc]$ – упорядоченное множество с повторением и перечислением элементов в лексикографическом порядке, задаваемом алфавитом Ω;

– $Efg : \{B, A2, A1, Dc, Dac\}$ – именованное множество, где Efg – имя множества;

– $\{X \mid Y\}$ – множество, заданное порождающей процедурой, где X – конструктор элементов, а Y – условие включения элемента в множество. ♦

Определение 4. Пусть A – конечное множество имен проблемных областей, $A : \{Aj \mid j = 1, 2, ...\}$, где имя Aj называется аспектом предметной области или просто аспектом (англ. aspect). ♦

Определение 5. Пусть B – множество видов понятий (абстракций, используемых для их образования), $B : \{C, T, A, G\}$, где C обозначает понятие-знак (абстракция идентификации, англ. identification, comprehension), T – понятие-тип (абстракция типизации, англ. typification), A – понятие-ассоциацию (абстракция ассоциации, англ. association) и G – понятие-обобщение (абстракция обобщения, англ. generalization). ♦

Определение 6. Формальное понятие или просто понятие – это множество $N : (b, I, E)$, у которого:

– N – имя понятия, $N = a \bullet c$;
– a – имя аспекта, $a = Aspect(N) \in A$;
– c – имя концепта, $c = Concept(N)$;
– b – абстракция, $b = Abs(N) \in B$;
– I – интенсионал, $I = Int(N) = [Ij \mid j = 1, 2, ..., D]$;
– Ij – атрибут, $Ij = Atr(N, j), j = 1, 2, ..., D$;
– D – степень, $D = Degree(N)$;
– E – экстенсионал, $E = Ext(N) = \{Ek \mid k = 1, 2, ..., P\}$;
– Ek – сущность понятия, $Ek = Ent(N, k)$, $k = 1, 2, ..., P$;
– Ekj – сущность атрибута Ij сущности Ek, $Ekj = Ek[Ij]$;
– P – объем, $P = Scope(N)$;
– S – схема, $S = Schema(N) = (b, I)$,

где знак = обозначает равенство чисел, строк или множеств соответственно; • – разделитель аспекта и концепта; $\in$ – отношение принадлежности элемента множеству. ♦

Определение 7. Понятие $N1 : (B1, I1, E1)$ и понятие $N2 : (B2, I2, E2)$ равны тогда и только тогда, когда $B1 = B2$, $I1 = I2$, $E1 = E2$. ♦

Из определения 7 следует, что проверку понятий на равенство следует производить без учета имен понятий (с точностью до имен понятий).

Определение 8. Пусть $M : \{Nj : (Bj, Ij, Ej) \mid j = 1, 2, ..., P\}$ – множество понятий предметной области R, тогда M – понятийная модель предметной области R мощности P, а $S : \{Schema(Nj) \mid j = 1, 2, ..., P\}$ – ее понятийная структура. ♦

5.2. СЕМАНТИЧЕСКАЯ ТЕОРИЯ ПОНЯТИЙ

В семантической теории понятий изучаются понятия и операции над ними. Семантическая теория строится как формальный язык, предназначенный для первичного описания предметных областей в процессе понятийного анализа. Результатом понятийного анализа является именование понятий и выражение способов их образования путем задания операций над понятиями.

5.2.1. ОПЕРАЦИИ НАД ПОНЯТИЯМИ

Определение 9. Свободная переменная N обозначает понятие-знак $N : (C, (N), \{N\})$, которое имеет интенсионал и экстенсионал, определяемые именем N. ♦

Определение 10. Обобщение понятий $N : \{N1, ..., ND\}$ – это понятие $N : (G, I, E) = \oplus(N, N1, ..., ND)$ такое, что $I = [N1, ..., ND]$, $E = Ext(N1) \cup ... \cup Ext(ND)$, где D – степень понятия-обобщения, $D > 1$; $\oplus$ – операция обобщения; $\cup$ – объединение множеств. ♦

Определение 11. Ассоциация понятий $N : \langle N1, \dots, ND\rangle$ с селектором σ – это понятие $N : (A, I, E) = \otimes(N, N1, \dots, ND, \sigma)$ такое, что $I = [N1, \dots, ND]$, $E \subseteq Ext(N1) \times ... \times Ext(ND)$, где D – степень понятия-ассоциации, $D > 1$; $\otimes$ – операция ассоциации, $\subseteq$ – отношение нестрогого включения множеств, $\times$ – декартово произведение множеств, σ – имя множества E. ♦

Операция обобщения не зависит от порядка операндов и их повторения, так как операнды задаются неупорядоченным множеством без повторения элементов, а операция ассоциации зависит от повторения операндов, так как операнды образуют неупорядоченное множество с повторением элементов (см. определение 3).

Интенсионалы понятий являются множествами с повторением элементов, которые перечислены в лексикографическом порядке. Последнее необходимо для повышения эффективности сравнения понятий при практической реализации понятийной модели (см. определение 7).

Использование имени экстенсионала (селектора) необходимо потому, что могут существовать понятия, заданные на одном и том же множестве атрибутов, но с разными экстенсионалами. В этом случае селектор позволяет различать (сравнивать) такие понятия и рассматривается как некоторое характеристическое условие включения сущностей из декартова произведения атрибутов в экстенсионал понятия.

5.2.2. ФОРМУЛЫ

Определение 12. Формула на языке понятий задается следующими рекурсивными правилами:

– свободная переменная N (имя) является формулой N;

– если $F1$, ..., FD – формулы, а N – имя, то $\oplus(N, F1, ..., FD)$ – формула;

– если $F1$, ..., FD – формулы, N – имя и σ – селектор (имя), то $\otimes(N, F1, ..., FD, \sigma)$ – формула. ♦

Пример 13. Часть понятий понятийной диаграммы на рис. 1 выражаются следующими формулами:

– элементарные формулы – это 1, 2, 3;

– формула понятия «Разряд» – $\oplus$(Разряд, 1, 2, 3);

– формула понятия «Инженер» – $\otimes$(Инженер, $F1$, $F2$, Инженеры), где $F1$ и $F2$ – формулы, выражающие понятия «Специальность» и «ФИО», а Инженеры – селектор, который именует (характеризует) экстенсионал понятия «Инженер»;

– формула понятия «Сотрудник» – $\oplus$(Сотрудник, $F3$, $F4$), где $F3$ и $F4$ – формулы, выражающие понятия «Инженер» и «Рабочий». ♦

5.2.3. ОТНОШЕНИЕ ЭКВИВАЛЕНТНОСТИ

Определение 13. Две формулы $F1$ и $F2$ эквивалентны, если они выражают одно и то же понятие. ♦

Из определений 10–13 непосредственно следует то, что в семантической теории понятий справедливы следующие эквивалентности:

(1) $\oplus(N, \ldots, Fj, \ldots, Fj, \ldots) \Leftrightarrow \oplus(N, \ldots, Fj, \ldots)$;

(2) $\oplus(N, \ldots, Fj, \ldots, Fk, \ldots) \Leftrightarrow \oplus(N, \ldots, Fk, \ldots, Fj, \ldots)$;

(3) $\otimes(N, \ldots, Fj, \ldots, Fk, \ldots, \sigma) \Leftrightarrow \otimes(N, \ldots, Fk, \ldots, Fj, \ldots, \sigma)$,

где N – имя, Fj и Fk – формулы, σ – селектор, $\Leftrightarrow$ – отношение эквивалентности формул.

Определение 14. Формула находится в канонической форме, если она преобразована с помощью эквивалентностей вида (1) – удаление лишних подформул, и эквивалентностей вида (2) и (3) – перестановки подформул в лексикографическом порядке. ♦

***Утверждение* 1**. *Понятия равны тогда и только тогда, когда их канонические формы равны с точностью до имен понятий.*

Доказательство. Пусть понятия $N1$ и $N2$ равны (определение 7). Тогда они могут быть выражены эквивалентными формулами (определение 12), канонические формы которых равны с точностью до имен понятий (определения 7 и 14).

Верно и обратное. Пусть $F1$ и $F2$ – канонические формы, равные с точностью до имен понятий. Эти формы являются эквивалентными формулами (определения 7 и 14). Следовательно, выражаемые ими понятия равны (определение 13). ♦

Утверждение 1 позволяет выявить и удалить из понятийной модели равные понятия. Следует обратить внимание на игнорирование имен понятий при их сравнении, кроме имен понятий-знаков – первичных понятий предметной области. Последнее означает, что понятия сравниваются структурно, т.е. по порядку их образования из других понятий и с учетом используемых для этого абстракций.

Определение 15. Понятийная модель называется приведенной, если все ее формулы находятся в канонической форме и в модели нет формул равных понятий. ♦

Очевидно, модель в приведенной форме эквивалентна исходной понятийной модели.

Пример 14. Пусть понятийная модель задана следующими формулами:

$\oplus(A, C, \otimes(B, D, C, \sigma), C)$;

$\oplus(V, \otimes(Y, X, D, \delta), C)$;

$\oplus(X, C, \otimes(Y, D, C, \sigma))$,

где свободные переменные обозначают первичные (не раскрываемые) понятия предметной области.

Используя определения 13 и 14 находим канонические формы этих формул:

$\oplus(A, \otimes(B, C, D, \sigma), C)$;

$\oplus(V, \otimes(Y, D, X, \delta), C)$;

$\oplus(X, \otimes(Y, C, D, \sigma), C)$.

Из утверждения 1 следует, что понятия A и X равны. Оставляем формулу понятия A и удаляем формулу понятия X, а все вхождения имени X заменяем именем A. В итоге имеем приведенную модель:

$\oplus(A, \otimes(B, C, D, \sigma), C)$;

$\oplus(V, \otimes(Y, D, A, \delta), C)$. ♦

Заметим, что свободные переменные уникальны и выражают по определению 9 различные понятия.

5.2.4. СЕМАНТИЧЕСКАЯ РАЗРЕШИМОСТЬ

Важным свойством любой формальной теории является ее разрешимость (англ. decidability), заключающаяся в возможности определить, принадлежит ли теории некоторая формула или нет.

Для доказательства разрешимости семантической теории понятий построим формальную грамматику, порождающую формулы языка понятий (см. определение 12).

Определение 16. Формальная грамматика языка понятий – это множество $G : (V, W, P, F)$, у которого:

– $V : (Nj, \sigma j, \oplus, \otimes, (,), ; \mid j = 1, 2, ...)$ – терминальный алфавит грамматики;

– $W : (F, L)$ – нетерминальный алфавит, где нетерминальный символ F означает «Первая формула списка», а нетерминальный символ L – «Последняя формула списка»;

– $P : \{F \Rightarrow Nj,\ F \Rightarrow \oplus(Nj; F; L),\ F \Rightarrow \otimes(Nj; F; L; \sigma j),\ L \Rightarrow F;L,$ $L \Rightarrow F\}$ – продукции грамматики;

– F – начальный нетерминальный символ, $F \in W$;

где Nj – понятийные переменные, σj – селекторы, ; – разделитель формул в списке (вместо запятой), $\Rightarrow$ – разделитель правых и левых частей продукций, а также знак грамматического вывода. ♦

Язык, порожденный грамматикой из определения 16, является бесконечным контекстно-свободным формальным языком [9, с. 11].

Пример 15. Выведем формулы из примера 13 в пустом аспекте, когда разделитель • не используется для построения имен понятий.

Из начального символа F и продукции $F \Rightarrow Nj$ получаем терминальные понятия:

$F \Rightarrow Nj \Rightarrow 1$, $F \Rightarrow Nj \Rightarrow 2$, $F \Rightarrow Nj \Rightarrow 3$.

Далее выведем формулу ⊕(Разряд, 1, 2, 3):

$$F \Rightarrow \oplus(Nj; F; L) \Rightarrow \oplus(Nj; F; F; L) \Rightarrow \oplus(Nj; F; F; F) \Rightarrow$$

$$\Rightarrow \oplus(\text{Разряд}; F; F; F) \Rightarrow \oplus(\text{Разряд}; 1; F; F) \Rightarrow$$

$$\Rightarrow \oplus(\text{Разряд}; 1; 2; F) \Rightarrow \oplus(\text{Разряд}; 1; 2; 3).$$

Аналогично выводятся другие формулы. ♦

Для грамматики из определения 16 существует эффективная процедура, с помощью которой можно проверить, принадлежит ли формула языку, порождаемому этой грамматикой, или нет [9, с. 21]. Однако выводимость формулы языка понятий оказывается недостаточной для проверки принадлежности этой формулы семантической теории.

Каждое формальное понятие, выраженное на языке семантической теории, должно иметь интерпретацию в предметной области. При понятийном мышлении недопустимы понятия, которые прямо или косвенно определяются через самих себя, а также не имеющие принадлежащих им сущностей. Следовательно, понятия с пустым экстенсионалом и понятия, выраженные через самих себя, не имеют интерпретации в любой предметной области. Для семантической теории это означает, что в формулах языка понятий не должно быть понятийных циклов, а также не должно быть формул пустых понятий (понятий с пустым экстенсионалом).

Определение 17. Формальная теория понятий разрешима, если существует эффективная процедура, распознающая понятийные модели с понятийными циклами и пустыми понятиями. ♦

Утверждение 2. *Язык понятий разрешим*.

Доказательство. Пусть M – семантическая понятийная модель, $M : \{Fj \mid j = 1, 2, ..., P\}$, где Fj – формула, выражающая понятие с именем Nj, P – мощность модели.

Рассмотрим распознавание формул модели на пустоту выражаемых ими понятий. Из определения 9 следует, что понятия-знаки не пусты. Образование понятий-обобщений также не приводит к определению пустых понятий, если обобщаемые понятия не пусты (определение 10). Однако при образовании понятий-ассоциаций возможно появление пустых понятий, если именованный экстенсионал образуемого понятий пуст (определение 11). Однако появление таких понятий в процессе понятийного анализа считается невозможным, так как понятия с пустым экстенсионалом не могут мыслится. Следовательно, формулы языка понятий не выражают пустых понятий.

Для нахождения понятийных циклов в модели применим следующую процедуру. В начале удаляются все формулы из M, которые выражают равные понятия (см. утверждение 1).

Для каждого оставшегося понятия Nj, выраженного формулой Fj, ищется вхождение этого понятия в формулу Fj. Если вхождение найдено, то понятийный цикл обнаружен и процедура завершается. В противном случае понятие Nj ищется в других формулах Fk, имена понятий которых Nk встречаются в формуле Fj. Если все формулы просмотрены и имя понятия Nj не обнаружено, то понятийные циклы с понятием Nj отсутствуют.

Если все понятия просмотрены и понятийные циклы не обнаружены, то процедура завершается. Число понятий P конечно, как и конечна длина формул модели. Следовательно, найдена эффективная процедура определения циклов в понятийных моделях.

В итоге разрешимость языка понятий доказана. ♦

Интерпретируемость понятийный модели может быть также проверена по понятийной диаграмме предметной области (см. рис. 1). В этом случае на диаграмме, рассматриваемой как ориентированный граф, не должно быть циклов.

Следует обратить внимание на то, что разрешимость языка понятий в теории понятий – это то же самое, что регулярность множеств в теории множеств (см. аксиому основания [7, с. 17]).

5.2.5. НЕПРОТИВОРЕЧИВОСТЬ И ПОЛНОТА

Другими важными свойством любой формальной теории является ее непротиворечивость (англ. consistent) и полнота (англ. completeness). Для доказательства полноты и непротиворечивости синтаксической теории понятий нам потребуется утверждение о полноте и непротиворечивости семантической теории.

Противоречием (англ. contradiction) в теории понятий будем называть ситуацию, при которой одна и та же сущность мыслится (описывается) как принадлежащая, так и не принадлежащая одному и тому же понятию в одно и то же время. Следует заметить, что логическое противоречие является частным случаем понятийного. В логике противоречием называется ситуация, при которой некоторое высказывание принадлежит как множеству истинных высказываний, так и множеству ложных.

Утверждение 3. *Язык понятий непротиворечив.*

Доказательство. Язык понятий непротиворечив по своему построению, так как нет никаких выразительных средств для выражения противоречий, нет ни одного способа указать непринадлежность какой-либо сущности какому-либо понятию, нет операции отрицания. Это служит основанием для утверждения о непротиворечивости языка понятий. ♦

Вывод о семантической полноте любой теории можно сделать только относительно класса моделей, для описания которых теория предназначена. Семантическая теория понятий полна, если любая предметная область может быть описана в этой теории, т.е. имеет описывающую ее формулу на языке понятий.

Утверждение 4. *Язык понятий является полной теорией.*

Доказательство. Пусть имеется предметная область с принадлежащими ей множеством первичных сущностей – понятий-знаков $C = \{c1, c2, \ldots\}$. Будем считать множество C разрешимым, так как оно как-то выделено в предметной области и предъявлено в процессе понятийного анализа.

Как доказано в утверждении 1 и показано на примере 14, выводимые в семантической теории понятия сравнимы с точностью до имен понятий-знаков. Используя процедуру генерации типов из теории типов [2] определим множество сущностей предметной области E на основе первичных понятий-знаков C так:

– если $\varphi \in C$, то φ – сущность (выделение);
– если φ – сущность, то $\{\varphi\}$ – сущность (группировка);
– если φ и ψ – сущности, то $\varphi\bullet\psi$ – сущность (соединение).

Множество E является счетным и содержит все мыслимые сущности предметной области, т.к. ему принадлежат все строки конечной длины $\{c1, c2, \ldots, c1\bullet c1, c1\bullet c2, \ldots, c2\bullet c1, c2\bullet c2, \ldots, c1\bullet c1\bullet c1, c1\bullet c1\bullet c2, \ldots, c1\bullet c2\bullet c1, \ldots\}$ со всеми вариантами правильно расставленных скобок, по-разному структурирующих исходную строку. Так как множество C разрешимо, то разрешимым будет и множество E.

Для конструирования всех мыслимых понятий образуем булеан $B(E)$ – множество всех подмножеств множеств E. Существование булеана следует из аксиомы степени теории множеств [7, с. 17].

Каждый элемент $B(E)$ является объединением конечного числа элементов из множества E. Так как разрешающая процедура для элементов множества E существует, то существует и разрешающая процедура для элементов множества $B(E)$, так как объединение конечного числа множеств разрешимо.

Таким образом, удалось показать, что с помощью выделения (идентификации), соединения и группировки (ассоциации), а также объединения (обобщения) сущностей предметной области могут быть конструктивно построены все мыслимые понятия. Осталось только выбрать те понятия, которые интерпретируются в предметной области, и включить их в понятийную модель (абстракция ограничения, игнорирования, англ. restriction).

Очевидно, верно и обратное. Для любого множества понятий M, выведенных в семантической теории из C, в $B(E)$ содержатся прототипы всех понятий из M. ♦

Аналогично доказательству утверждения 4 выполнено обоснование полноты логики высказываний через функциональную полноту логических операций [10, с. 18], т.е. путем доказательства возможности выразить любое сложное высказывание из простых с помощью логических связок.

Доказательство утверждения 4 также близко технологии концептуализации предметных областей [3, с. 47], где основными шагами концептуального проектирования являются:

1) выделение базисных концептов X (идентификация);
2) построение булеана $B(X)$ (синтез обобщений);
3) построение декартиана $B(X \times B(X))$ (синтез ассоциаций);
4) выделение из булеана и декартиана множества концептов X, имеющих интерпретацию в предметной области (ограничение).
5) если концептов X недостаточно, то переход на шаг 2.

В итоге имеем, что предметом теории понятий являются понятийные модели, выводимые в этой теории, а ее объектом – предметные области, рассматриваемые как совокупности ментальных представлений и иерархии понятий, абстрагированных от них.

5.3. СИНТАКСИЧЕСКАЯ ТЕОРИЯ ПОНЯТИЙ

Семантическая теория понятий ответственна за первичное описание предметной области в виде ее понятийной структуры, выраженной на языке понятий.

Синтаксическая теория понятий, также называемая исчислением понятий, используется для вывода понятийных моделей, включающих в себя не только понятийную структуру предметной области, но и перечисление сущностей из экстенсионалов понятий.

Исчисление определяется алфавитом, аксиомами и правилами вывода.

5.3.1. АЛФАВИТ

Алфавит исчисления понятий включает:

– имена понятий (аспектов, селекторов), состоящие из букв и цифр (см. определение 2);

– разделитель имен • (точка), отношение именования : (двоеточие), операции над понятиями $\oplus$ (обобщение) и $\otimes$ (ассоциация);

– символы логического вывода, логические связки и кванторы: $\models$, $\vdash$, $\neg$, $\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$, $\forall$, $\exists$ [10, с. 447];

– предикаты, операции и отношения над множествами: $\varnothing$, $\in$, $\notin$, $-$, $\cup$, $\cap$, $\times$, $\subset$, $\supset$, $\subseteq$, $\supseteq$, $<$, $>$, $=$, $\neq$ [7, с. 49];

– скобки (круглые, квадратные, угловые, фигурные): (,), [,], $\langle$, $\rangle$, {, } (см. определение 3).

5.3.2. АКСИОМЫ

Аксиомы исчисления понятий – это аксиомы логики первого порядка [10, с. 62], аксиомы теории множеств [7, с. 15], а также следующие четыре аксиомы.

Аксиома 1 (ограничение). Существует пустое понятие $\varepsilon : (C, \varnothing, \varnothing)$, не принадлежащее интенсионалу и экстенсионалу любого понятия:

$$\exists \varepsilon : (C, \varnothing, \varnothing)\ \forall N\ (\varepsilon \notin Int(N) \wedge \varepsilon \notin Ext(N)).\ \blacklozenge$$

Аксиома 2 (идентификация). Существует понятие-знак $N : (C, (N), \{N\})$ такое, что его интенсионал и экстенсионал состоит только из этого понятия:

$$\exists N : (C, (N), \{N\}).\ \blacklozenge$$

Аксиома 3 (обобщение). Для любых $D > 1$ понятий $N : \{N1, \ldots, ND\}$ существует понятие-обобщение $\oplus(N, N1, \ldots, ND)$:

$$\forall N : \{N1, \ldots, ND\}$$

$$\exists N : (G, [N1, \ldots, ND], Ext(N1) \cup \ldots \cup Ext(ND)).\ \blacklozenge$$

Аксиома 4 (ассоциация). Для любых $D > 1$ понятий $N : \langle N1, \ldots, ND\rangle$ и селектора σ существует понятие-ассоциация $\otimes(N, N1, \ldots, ND, \sigma)$:

$$\forall N : \langle N1, \ldots, ND\rangle$$

$$\exists N : (A, [N1, \ldots, ND],$$

$$\{Ej : (E1, \ldots, ED) \mid E1 \in Ext(N1) \wedge \ldots \wedge ED \in Ext(ND) \wedge \sigma(Ej)\}),$$

где $\sigma(Ej)$ – логическая формула принадлежности сущности Ej экстенсионалу понятия-ассоциации. ♦

Пустое понятие ε используется для обозначения сущностей, отсутствующих в модели. Очевидно, такое понятие не может использоваться при выводе других понятий, так как обозначаемые им сущности исключаются из рассмотрения. По своей сути пустое понятие реализует абстракцию ограничения.

Как и в семантической теории, операции $\oplus$ и $\otimes$ обозначают обобщение и ассоциацию понятий, а селектор σ – это логическая формула со свободным вхождением одной переменной, которая используется для разрешения сущностей из экстенсионалов понятий-ассоциаций.

Возможность построения понятий-обобщений с помощью аксиомы 3 следует из аксиомы объемности теории множеств, а построение понятий-ассоциаций с помощью аксиомы 4 – из аксиомы выделения [7, с. 16].

5.3.3. ПРАВИЛА ВЫВОДА

Синтаксическая модель предметной области состоит из понятий (определение 8), каждое из которых должно быть выведено с помощью аксиом исчисления понятий. Исчисление понятий, как и теория множеств, наследует правила вывода из логики первого порядка: правило отделения (modus ponens) и правило обобщения [10, с. 62].

Определение 18. Вывод понятийной модели $M : \{N1, \ldots, NL\}$ – это упорядоченное множество понятий ($N1$, ..., Nj, Nk, ..., NL) таких, что предыдущие понятия $N1$, ..., Nj (базовая модель) используются для вывода следующих понятий Nk, ..., NL (наследуемой модели) с помощью аксиом исчисления понятий: $\vDash (N1, \ldots, NL)$ – вывод модели M; $\{N1, ..., Nj\} \vdash (N1, ..., NL)$ – вывод наследуемой модели M из базовой модели $\{N1, ..., Nj\}$, где $\vDash$ ($\vdash$) – символ абсолютной (относительной) выводимости. ♦

Пример 16. Выведем фрагмент понятийной модели из примера 13:

– $\vDash 1 : (C, [1], \{1\})$, $\vDash 2 : (C, [2], \{2\})$, $\vDash 3 : (C, [3], \{3\})$ (аксиома 2);

– $1, 2, 3 \vdash$ Разряд : (G, [1, 2, 3], $\{1, 2, 3\}$) (аксиома 3);

– ФИО, Специальность $\vdash$ Инженер : (A, [Специальность, ФИО], $\{X$ | Инженеры(X)$\}$) (аксиома 4).

В последнем выводе селектор «Инженеры» характеризует экстенсионал понятия «Инженер» путем разрешения соединения сущностей понятий «Специальность» и «ФИО», например, так:

Инженеры(X) $\leftrightarrow$ X[Специальность] = Металлург $\wedge$
$\wedge$ X[ФИО] = Иванов Иван Иванович $\vee$
$\vee$ X[Специальность] = Программист $\vee$
$\vee$ X[ФИО] = Петров Петр Петрович,

где X – свободная переменная, обозначающая сущность понятия «Инженер».

Селектор «Инженеры» включает в экстенсионал понятия «Инженер» сущности (Металлург, Иванов Иван Иванович), (Программист, *) и (*, Петров Петр Петрович), где * – произвольные вхождения сущностей «ФИО» и «Специальность». ♦

Заметим, что правила вывода, полученные из понятийной модели, являются правилами вывода в предметной области и используются для построения предметной формальной теории (см. пример 10). Здесь же используются правила вывода исчисления понятий, с помощью которых строится сама понятийная модель.

5.3.4. ТИПЫ И КЛАССЫ

Теорема 1 (типизация). *Для множества понятий-знаков* $T : \{Nj \mid j = 1, \ldots, D\}$ *существует понятие-тип* $T : (G, I, E)$ *такое, что* $I = [N1, \ldots, ND]$, $E = \{N1, \ldots, ND\}$.

Доказательство. Существование понятий-знаков следует из аксиомы 2, а существование искомого понятия-типа – из аксиомы 3. ♦

Теорема 1 позволяет вывести все конечные понятия-типы, используемый на практике, как-то: число с ограниченной разрядностью представления, символ, строка и т.п.

Однако в исчислении понятий справедлива и теорема, в которой доказывается существование понятия-класса – абстрактного понятия-типа со счетным объемом.

Теорема 2 (классификация). *Для множества понятий-знаков* $C : \{Nj \mid j = 1, 2, \ldots\}$ *существует понятие-класс* $C : (G, I, E)$ *такое, что* $I = [N1, N2, N3, \ldots]$, $E = \{N1, N2, N3, \ldots\}$.

Доказательство. Существование счетного числа понятий-знаков следует из аксиомы 2 и аксиомы бесконечности теории множеств [7, с. 16].

Выведем семейство понятий-типов $\{Cj \mid j = 1, 2, \ldots\}$, используя математическую индукцию и аксиому 3, добавляя по одному понятию-знаку на каждом шаге:

$$\{\oplus(Cj, Ck, Nk) \mid C1{=}N1 \wedge j = k{+}1 \wedge k = 1, 2, \ldots\}.$$

В пределе (при индуктивном переходе) получаем существование искомого понятия-класса C счетного объема. ♦

5.3.5. АКТУАЛИЗАЦИЯ МОДЕЛЕЙ

Теорема 3 (насыщение и усечение). *Для любого понятия-ассоциации* $N : (A, I, E)$ *существует насыщенное (усеченное) понятие-ассоциация* $Nx : (A, I, Ex)$ *такое, что* $E \subseteq Ex$ ($E \supseteq Ex$).

Доказательство. Из аксиомы 4 следует, что понятие-ассоциация может быть выведено с тем же интенсионалом, но с расширенным (усеченным) экстенсионалом, поскольку экстенсионал – это любое подмножество декартова произведения атрибутов понятия. Для этого случая существует селектор $\sigma \vee \delta$ ($\sigma \wedge \gamma$), где δ (γ) – логическая формула, которая выбирает новые (исключает имеющиеся) сущности экстенсионала Ex. ♦

Из теоремы 3 следует, что сущность предметной области является полноправным понятием. Действительно, усекая экстенсионал понятия до состояния, когда в нем остается всего одна сущность Ex, имеем понятие-сущность $Nx : (A, I, \{Ex\})$.

Теорема 3 позволяет актуализировать понятийную модель с помощью таких операций, как удаление и добавление сущностей в экстенсионалы понятий. Кроме того, теорема 3 позволяет создавать новые понятия с помощью запросов, выбирающих сущности из экстенсионала одного понятия-ассоциации для включения их в экстенсионал подобного понятия.

5.3.6. РАВЕНСТВО СУЩНОСТЕЙ

Определение 19. Сущности X и Y равны, если и только если равны сущности их атрибутов, $X = Y \leftrightarrow \forall j\ (X[j] = Y[j])$. ♦

Для обеспечения выполнимости определения 19 и повышения эффективности реализации понятийных моделей:

– интенсионалы понятий отсортированы в лексикографическом порядке (определение 6);

– для отсутствующего атрибута у сущности функтор возвращает пустое понятие (определение 22).

Определение 20. Пусть X – сущность некоторого понятия. Тогда $Name(X)$ – функтор, возвращающий имя сущности X, определяемый рекурсивно:

– если X – сущность понятия-знака, понятия-типа или понятия-класса, то $Name(X) = X$;

– если X – сущность понятия-ассоциации, то $Name(X) = \{Name(X[1])\bullet\ldots\bullet Name(X[D])\}$, где $D = Degree(N)$ – степень понятия сущности;

– если X – сущность понятия-обобщения, то имя X равно имени сущности понятия-знака, понятия-типа или понятия-ассоциации, которому принадлежит X. ♦

Из определения 20 видно, что имя сущности задается атрибутами понятия-ассоциации или понятия-знака, поскольку интенсионал понятия-обобщения не описывает структуру понятия, а задает лишь множество обобщаемых понятий (аксиома 3).

Пример 17. Именем некоторой сущности понятия-ассоциации «Участник» (см. рис. 1) может быть {В•{Металлург•3•Иванов Иван Иванович}}, где В – имя сущности понятия-типа «Бригада», Металлург – имя сущности понятия-типа «Профессия», 3 – имя сущности понятия-типа «Разряд», Иванов Иван Иванович – имя сущности понятия-типа ФИО. В свою очередь {Металлург•3•Иванов Иван Иванович} – имя сущности понятия-ассоциации «Рабочий». Порядок перечисления имен – как в интенсионалах соответствующих понятий-ассоциаций. ♦

Из аксиомы 4 следует, что существуют понятия с одинаковым интенсионалом и различными экстенсионалами. Тогда у таких понятий могут быть равные сущности.

Теорема 4 (равенство сущностей). *Сущности равны тогда и только тогда, когда они имеют одинаковые имена:*

$$X1 = X2 \leftrightarrow Name(X1) = Name(X2).$$

Доказательство. Из определения 20 следует, что имена сущностей состоят из имен понятий-знаков, которые уникальны (аксиома 2) и могут быть пронумерованы, а благодаря использованию разделителя имен и фигурных скобок неоднозначностей в структурировании составных имен не возникает.

Тогда из определения 19, аксиом 3 и 4 следует, что имена сущностей можно сравнивать, а равные сущности имеют одинаковые имена. Верно и обратное: если сущности имеют одинаковые имена, то они равны. ♦

Теорема 4 позволяет реализовать эффективный поиск равных сущностей в базе знаний. Для такого поиска достаточно сравнить имена сущностей без доступа к их атрибутам. Более того, если пронумеровать понятия-знаки предметной области, то сравнение имен сводится к сравнению последовательностей натуральных чисел, разделенных точками и структурированных фигурными скобками.

Пример 18. Пусть понятия-знаки некоторой предметной области – это *Abc*, 12,3, 0x2359AD, 73, 15.05.2021. Пронумеруем их в лексикографическом порядке: 0x2359AD ~ 1, 12,3 ~ 2, 15.05.2021 ~ 3, 73 ~ 4, *Abc* ~ 5, где ~ – отношение соответствия. Пусть также образованы понятия с сущностями, имена которых и их числовые эквиваленты такие:

73 ~ 4,
{15.05.2021•0x2359AD} ~ {3•1},
{15.05.2021•0x2359AD•73} ~ {3•1•4},
{*Abc*•{12,3•15.05.2021}} ~ {5•{2•3}},
{*Abc*•0x2359AD•{12,3•73}} ~ {5•1•{2•4}}.

Тогда сущности {3•1} и {3•1} равны, сущность {3•1•4} больше сущности {3•1}, а сущности {5•{2•5}} и {5•1•{2•4}} несравнимы, так как сущность {2•3} не сравнима с сущностью 1. ♦

В прикладном плане отношение равенства сущностей проверяется сравнением числовых идентификаторов сущностей (см. пример 8) и, если идентификаторы не равны, то сравнением их имен.

5.3.7. СИГНАТУРА МОДЕЛИ

Селекторы, используемые в исчислении понятий, отличаются от селекторов языка понятий. В языке понятий селекторы – это имена подмножеств (определение 11), а в исчислении понятий – это логические формулы (аксиома 4).

Сигнатура понятийной модели – это множество предикатов, функторов и отношений, которые используются в логических формулах для выражения экстенсионалов, а также в запросах к базе знаний (см. пример 11).

Определение 21. Пусть N – понятие. Тогда $N(X)$ – предикат принадлежности сущности X экстенсионалу понятия N, $N(X) \leftrightarrow X \in Ext(N)$. ♦

Определение 22. Пусть X – сущность. Тогда $X[j]$ – функтор, возвращающий сущность (значение) j-го атрибута сущности X или пустое понятие ε, если у сущности X нет атрибута с номером (именем) j. ♦

Из практических соображений помимо функторов, извлекающих значения простых типов данных или идентификаторы сущностей атрибутов понятия, могут использоваться и другие функторы, общие для всех понятий, например: функторы наименования $Title(X)$, иконки $Icon(X)$ (см. пример 7), имени $Name(X)$, аспекта $Aspect(X)$, концепта $Concept(X)$, числа атрибутов $Degree(X)$ и т.д. (см. определение 6).

При создании исторических баз знаний отражение времени в понятийной модели может быть реализовано добавлением общих функторов времени, возвращающих время появления $Descent(X)$ и исчезновения $Ascent(X)$ сущности X.

Определение 23. Отношения понятийной модели – это отношения равенства = и порядка <, заданные на множестве натуральных чисел. ♦

Основанием для включения в сигнатуру понятийной модели только двух отношений и их достаточности для разрешения экстенсионалов понятий непосредственно следует из теоремы 4, где сравнение сущностей сведено к сравнению последовательностей натуральных чисел – в теоретическом плане, и сравнению значений встроенных типов данных, которые кодируются теми же последовательностями натуральных чисел, – в прикладном плане.

Однако из-за наличия фигурных скобок сущности оказались упорядоченными только частично (см. пример 18). Но и частичная упорядоченность – достаточно сильное свойство, повышающее эффективность обработки знаний.

В итоге имеем, что сигнатура понятийном модели состоит из одноместных предикатов, одноместных функторов, а также отношений равенства и порядка, заданных на множестве натуральных чисел.

Известно, что одноместное исчислении предикатов (англ. pure monadic predicate calculus) имеет много полезных свойств, делающих его похожим на исчисление высказываний [10, с. 24].

Но наличие функторов и отношений в сигнатуре модели не позволяет сделать заключение о том, что в исчислении понятий логические формулы являются формулами одноместного исчисления предикатов.

Теорема 5 (одноместное исчисление предикатов). *Логические формулы, используемые в исчислении понятий для выражения экстенсионалов, являются формулами одноместного исчисления предикатов.*

Доказательство. Все вхождения подформул $\psi(X, Y)$ с функторами и отношениями в логическую формулу φ могут иметь следующий вид (см. пример 11):

$$X \div Y; \quad X[j] \div Y; \quad X \div Y[k]; \quad X[j] \div Y[k],$$

где X и Y – сущности понятий, $X[j]$ и $Y[k]$ – функторы, возвращающий значение j-го атрибута сущности X и k-го атрибута сущности Y; $\div$ – отношения между сущностями.

Так как формула φ используется для разрешения экстенсионала понятия, то не менее чем одна переменная из двух в $\psi(X, Y)$ является связанной. Без потери общности выберем в φ подформулу $QY\ \psi(X, Y)$, где Q – связывающий квантор для Y. Пусть P – свободный предикатный символ, не входящий в формулу φ, и ϕ – формула, полученная из φ заменой вхождений $QY\ \psi(X, Y)$ предикатом $P(X)$. Тогда формулы φ и $\phi \wedge \forall X\ P(X) \leftrightarrow QY\ \psi(X, Y)$ эквивалентны.

Проделав аналогичные подстановки для всех вхождений $\psi(X, Y)$ в φ и добавив предикаты $P(X)$ в сигнатуру модели, получим формулу ϕ, не содержащую вхождений функторов и отношений, а составленную только из одноместных предикатов. Что требовалось доказать. ♦

Следует заметить, что в одноместном исчислении предикатов нет ограничений на мощность сигнатуры модели. Однако сама сигнатура значительно ограничивает выразительность логического языка (англ. expressivity of logic).

Для концептуальных моделей также были предприняты попытки ограничить выразительность логического языка путем использования дескрипционных логик. Дескрипционные логики строятся на одноместных предикатах принадлежности экземпляра концепту и двуместных предикатах ролей (отношений между экземплярами).

Для сведения дескрипционной логики к одноместному исчислению предикатов использование ролей ограничено конструкциями $\forall R.C$ и $\exists R.C$, которые сводят применение двуместных предикатов для разрешения множеств экземпляров к одноместным, связывая вторую переменную роли R экземплярами концепта C [13, с. 50, 150]:

$\forall R.C = \{a | \forall b\ R(a, b) \to C(b)\}$;

$\exists R.C = \{a | \exists b\ R(a, b) \wedge C(b)\}$.

Это позволило добиться разрешимости формул и относительно невысокой вычислительной сложности решения основных задач представления и обработки знаний в дескрипционных логиках.

В понятийных моделях удалось еще более понизить выразительность логического языка. Отсутствие в языке двуместных предметных предикатов (ролей) позволяет надеяться на большую эффективность обработки знаний в формализме понятийных моделей по сравнению с моделями дескрипционных логик.

5.3.8. ПРИВЕДЕННАЯ МОДЕЛЬ

Приведенная понятийная модель не содержит равных понятий (определение 15).

Теорема 6 (приведенная модель). *Для любой понятийной модели существует эквивалентная ей приведенная понятийная модель.*

Доказательство. Для семантической теории существует процедура приведения понятийных моделей (см. пример 14). В отличие от семантической теории в исчислении понятий сравнению подлежат не уникальные имена селекторов, а логические формулы, их выражающие.

Логические формулы в исчислении понятий выражаются в одноместном исчислении предикатов (теорема 5), которое разрешимо [10, с. 221]. Отсюда следует и разрешимость проблемы эквивалентности логических формул с одноместными предикатами.

Действительно, доказательство разрешимости формул в одноместном исчислении предикатов сводится к проверке выполнимости формулы на конечном множестве точек области определения.

Применив аналогичную процедуру для двух формул, заданных на общем множестве предикатов, можно сделать вывод об эквивалентности формул, если они выполнимы на одном и том же подмножестве точек области определения и только на нем.

Следовательно, существует процедура приведения понятийных моделей и в синтаксической теории. ♦

Для уменьшения избыточности данных, хранимых в базе знаний, и устранения потенциальных неоднозначностей, связанных с именованием одного и того понятия разными именам, при понятийном моделировании используются только приведенные модели.

5.3.9. СИНТАКСИЧЕСКАЯ РАЗРЕШИМОСТЬ

Исчисление понятий разрешимо, если существует эффективная процедура распознавания формул исчисления, содержащих понятийные циклы и пустые понятия (определение 17).

Теорема 7 (синтаксическая разрешимость). *Исчисление понятий разрешимо.*

Доказательство. Формулы, порождаемые исчислением понятий, отличаются от формул языка понятий только селекторами, которые в исчислении понятий являются логическими формулами (аксиома 4), а на языке понятий – именами множеств (определение 11).

Отсюда следует, что в исчислении понятий наличие в понятийной модели циклов проверяется той же процедурой, которая используется и в семантической теории (утверждение 2).

Теперь осталось доказать разрешимость в исчислении понятий проблемы пустоты экстенсионала. Ранее показано, что язык понятий разрешим (утверждение 7). Следовательно, доказательство разрешимости исчисления понятий сводится к доказательству выполнимости логических формул (англ. satisfiability), используемых для выражения экстенсионалов.

Логические формулы, используемые в исчислении понятий, принадлежат одноместному исчислению предикатов (теорема 5), которое разрешимо на счетных моделях [10, с. 221]. Это позволяет проверить селектор на выполнимость процедурой, аналогичной процедуре из доказательства теоремы 6. В этом случае выполнимость селектора хотя бы в одной точке гарантирует непустоту соответствующего понятия.

В итоге разрешимость исчисления понятий доказана. ♦

5.3.10. НЕПРОТИВОРЕЧИВОСТЬ И ПОЛНОТА

Теорема 8 (синтаксическая непротиворечивость). *Исчисление понятий непротиворечиво на конечных и счетных моделях.*

Доказательство. Язык понятий непротиворечив (утверждение 3). Чтобы доказать непротиворечивость исчисления понятий, достаточно доказать непротиворечивость теории множеств, формулы которой на языке одноместного исчисления предикатов используется для выражения экстенсионалов понятий (теорема 5). Доказано, что теория натуральных чисел непротиворечива [7, с. 27]. Тогда и исчисление понятий непротиворечиво на конечных и счетных моделях. ♦

Теорема 9 (синтаксическая полнота). *Исчисление понятий является полным на конечных моделях.*

Доказательство. Семантическая теория полна (утверждение 4). Чтобы доказать полноту исчисления понятий относительно семантической теории, достаточно доказать полноту теории множеств, которая используется в исчислении понятий для выражения экстенсионалов. Полнота теория множеств доказана только для конечных моделей, а на счетных моделях теория множеств существенно неполна [7, с. 27]. Следовательно, на счетных моделях неполным будет и исчисление понятий. ♦

6. Заключение

Принципиальным отличием рассмотренного подхода к анализу и моделированию предметных областей, представлению и обработке знаний является использование помимо формальной логики еще одного семантического инварианта – формальной теории понятий. Формальная логика основана на трех логических связках (НЕ, И, ИЛИ), которые используются для построения формальных высказываний, а формальная теория понятий – на трех ментальных абстракциях (идентификация, ассоциация, обобщение), которые используются для построения формальных понятий.

Семантическими средствами формальной теории понятий, делающими ее инвариантной предметным областям, являются:

– именование сущностей предметной области при идентификации и образование первичных семантических категорий;

– объединение в каком-либо смысле сущностей предметной области при обобщении;

– соединение сущностей предметной области при ассоциации так, что смысл такого соединения заключается в создании составных (укрупненных) сущностей;

– перечисление сущностей предметной области при обобщении и ассоциации в смысле ограничения существования подобных сущностей.

Анализ и моделирование названы понятийными, чтобы отличить их от концептуального анализа и концептуального моделирования. Понятие отличается от концепта. Концепт – это абстрактное объективное понятие, а понятие – конкретный субъективный концепт. По этой причине существует множество формальных понятий, имеющих одно и то же имя, но разные структуру и содержание в различных аспектах (проблемных областях). В этом случае концепт может быть определен как обобщение одноименных понятий в различных аспектах.

Формальная теория понятий состоит из языка понятий (семантическая теория) и исчисления понятий (синтаксическая теория). Язык понятий используется для описания предметных областей в процессе понятийного анализа. Результатом понятийного анализа является понятийная структура, содержащая интенсионалы понятий (множества понятий, использованных для их абстрагирования), но без детального описания экстенсионалов понятий (множеств сущностей, принадлежащих понятиям). Семантическая теория понятий оказалась разрешимой, непротиворечивой и полной.

Исчисление понятий используется для понятийного моделирования. Результатом понятийного моделирования является понятийная модель, состоящая из понятийной структуры, у которой экстенсионалы понятий выражены в виде формулы одноместного исчисления предикатов с отношениями равенства и порядка на множестве натуральных чисел. Синтаксическая теория понятий оказалась разрешимой и непротиворечивой на счетных и полной на конечных моделях. Благодаря понижению выразительности логического языка, используемого в исчислении понятий, снижена вычислительная сложность задач представления и обработки знаний по сравнению дескрипционными логиками.

Формальная теория понятий является формальной теорией понятийных моделей, а сами понятные модели – формальными теориями предметных областей. Это позволяет рассматривать понятийную модель как форму представления знаний о предметной области, а также осуществлять вывод предметных знаний.

Основная трудность использования понятийных моделей видится в необходимости использования новой методологии анализа и моделирования предметных областей, а также новой технологии представления и обработки знаний. Однако высокая эффективность и выразительность реализуемой модели знаний, ее легкая актуализация и адаптация, не приводящая к возникновению скрытых противоречий, позволяет надеяться на широкое внедрение понятийного моделирования в практику создания и использования современных интеллектуальных информационных систем.

Следующим шагом в исследовании формальной теории понятий видится разработка эффективных алгоритмов выполнения запросов к понятийной базе знаний и получение оценок их вычислительной сложности. Однако уже сейчас можно сказать, что время выполнения запроса будет иметь хорошую полиномиальную оценку, так как поиск сущности по ее идентификатору в таблице понятия с числом записей n осуществляется за время с асимптотической оценкой n [6, с. 667], а в случае использования индекса – за время с асимптотической оценкой $\log n$ [6, с. 601], где log – логарифмическая функция.

Еще одним направлением развития понятийного анализа и понятийного моделирования видится разработка методов и средств обучения понятийных моделей на основе восприятия окружающего мира в виде поступающего потока неструктурированных данных. Понятийная модель позволяет структурировать этот поток, выделяя в нем сущности известных и неизвестных понятий. Найденные неизвестные сущности могут использоваться для изменения понятийной модели путем пополнения экстенсионалов существующих понятий или путем образования новых. К этому направлению относится и большой класс задач по распознаванию образов.

Литература

1. ВЫХОВАНЕЦ В.С. *О существенной неполноте формального метода* // International Interdisciplinary Conference «Philosophy, Mathematics, Linguistics: Aspects of Interaction». – СПб.: Международный математический институт им. Л. Эйлера, 2009. – С. 79-87.
2. ВЫХОВАНЕЦ В.С., КРЫЖАНОВСКАЯ А.В. *Совмещенные сети управления и данных* // Управление большими системами. – 2015. – Вып. 58. – С. 41–66.
3. НИКАНОРОВ С.П. *Концептуализация предметных областей*. – М.: Концепт, 2009. – 268 с.
4. BATINI C., CERI S., NAVATHE S.B. *Conceptual Database Design: An Entity-Relationship Approach*. – Redwood City, CA: The Benjamin/Cummings Publishing Company, 1992. – 496 p.
5. BRACHMAN R.J., LEVESQUE H.J. *Knowledge Representation and Reasoning*. – San Francisco, CA: Morgan Kaufmann, 2004. – 381 p.
6. ELMASRI R., NAVATHE S.B. *Fundamentals of Database Systems*. Hoboken. – NJ: Pearson, 2016. – 1273 p.
7. HAO W., NAUGHTON R. *Les Systemes Axiomatiques de la Theorie des Ensembles*. – Paris, France: Gauthier-Villars, 1953. – 54 p.
8. *ISO/IEC 9899:2018. – Information technology. – Programming language – C.* – URL: https://www.iso.org/standard/74528.html.

9. KALLMEYER L. *Parsing Beyond Context-Free Grammars*. – New York, NY: Springer, 2010. – 260 p.
10. MENDELSON E. *Introduction to Mathematical Logic*. – New York, NY: CRC Press, 2010. – 496 p.
11. OLIVÉ A. *Conceptual Modeling of Information Systems*. – New York, NY: Springer, 2007. – 471 p.
12. TALENS G., BOULANGER D., SÉGURAN M. *Domain Ontologies Evolutions to Solve Semantic Conflicts* // Ontologies-Based Databases and Information Systems / Ed. M. Collard. – Berlin, Heidelberg: Springer, 2007. – P. 51–67.
13. *The description Logic Handbook: Theory, Implementation and Applications* / Eds. F. Baader, D. Calvanese, D. L. McGuinness, et al. – New York: Cambridge University Press, 2003. – 573 p.
14. *The Oxford Dictionary of Current English* / Ed. D. Thompson. – New York, NY: Oxford University Press, 1993. – 1091 p.
15. VYKHOVANETS V.S. *Large-Scale Information Systems based on Conceptual Models* // Proc. of 19th Int. Conf. on Management of Large-Scale System Development. – Moscow, Russia: V.A. Trapeznikov Institute of Control Sciences, October 1-3, 2019. – Режим доступа: https://ieeexplore.ieee.org/document/8911106.
16. ZHAO Y., ZHANG C., CAO L. *Post-Mining of Association Rules: Techniques for Effective Knowledge Extraction*. – Hershey, New York: Information science reference, 2009. – P. 81–149.

THE NOTIONAL ANALISYS AND NOTIONAL MODELLING

Valeriy Vykhovanets, Institute of Control Sciences of RAS, Moscow, Dr. Sc., associative professor (valery@vykhovanets.ru).

Abstract: The article describes notional models that are based on primary mental abstractions: identification, generalization and association. In the process of the notional analysis, a notional model is constructed, which consists of a notional structure and contents of its notions. The notional structure defines each notion as the result of generalization or association of the other notions. The content of the notion is described by an enumerable or solvable set consisting of subject domain entities. The main difference between the notional model and the other knowledge models is the refusal to describe the association of the notions in the form of a *relationship. In the notional model, the association are the same notion as the generalization, which makes it possible to form the other notions from the associations. All this makes the notional model semantically invariant, i.e. independent in its interpretation from the knowledge of the subject domain. Another difference between notional and conceptual models is the multi-aspect expression of the notions. In order to prove the proposed approach to the representation and processing of knowledge, a formal theory of notions is given. The semantic part of the theory (notional language) proves the methodology of the notional analysis, and its syntactic part (notional calculus) proves the technology of notional modeling. It is proved that the notional language is decidable, complete and consistent, while the notional calculus is decidable and consistent on countable models, but complete only on finite models. The use of the notional analysis improves the expressiveness and clarity of knowledge representation, and the use of the notional models increases the efficiency and reliability of knowledge processing.*

Keywords: primary mental abstractions, notion, concept, notional analysis, notional model, language of notions, notion calculus, knowledge base, knowledge inference, intelligent system.

УДК 510.8+004.89
ББК 22.128+32.972
DOI: 10.25728/ubs.2021.92.4

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

Поступила в редакцию 12.05.2021.
Опубликована 31.07.2021.