

ISSN 1998-0663

№4(26)—2013

<http://bijournal.hse.ru>

БИЗНЕС- ИНФОРМАТИКА

МЕЖДИСЦИПЛИНАРНЫЙ НАУЧНО-ПРАКТИЧЕСКИЙ ЖУРНАЛ НИУ ВШЭ

BUSINESS INFORMATICS

Учредитель:
Национальный
исследовательский университет
«Высшая школа экономики»

Редакционная коллегия

Абдульраб А.
Авдошин С.М.
Алескеров Ф.Т.
Бабкин Э.А.
Баранов А.П.
Беккер Й.
Белов В.В.
Грибов А.Ю.
Громов А.И.
Гурвич В.А.
Джейкобс Л.
Зандкуль К.
Ильин Н.И.
Калягин В.А.
Каменнова М.С.
Кузнецов С.О.
Мальцева С.В.
Миркин Б.Г.
Моттль В.В.
Пальчунов Д.Е.
Пардалос П.
Силантьев А.Ю.
Таратухин В.В.
Ульянов М.В.
Шалковский А.Г.

В ЭТОМ НОМЕРЕ:

**МАТЕМАТИКА
И БЕЗОПАСНОСТЬ**

**ФОРМАЛЬНЫЙ
АНАЛИЗ ПОНЯТИЙ**

**СЦЕНАРИЙ
ПО ПЕРСОНАЛУ**

**МОДЕЛЬ
ЖЕЛЕЗНОДОРОЖНЫХ
ПЕРЕГОНОВ**

В соответствии с решением президиума ВАК РФ журнал «Бизнес-информатика» с 19.02.2010 включен в Перечень ведущих рецензируемых научных журналов и изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученых степеней кандидата и доктора наук.

*Журнал зарегистрирован в «Роскомнадзоре».
Свидетельство ПИ № ФС 7752404 от 28 декабря 2012 г.*

БИЗНЕС- ИНФОРМАТИКА

№4(26)–2013

СОДЕРЖАНИЕ

Анализ данных и интеллектуальные системы

*Ф.Т. Алескеров, В.Ю. Белоусова, Л.Г. Егорова,
Б.Г. Миркин*

Анализ паттернов в статике и динамике,
часть 2: Примеры применения к анализу
социально-экономических процессов 3

*Т.А. Гаврилова, Д.В. Кудрявцев, И.А. Лещева,
Я.Ю. Павлов*

Об одном методе классификации
визуальных моделей 21

С.А. Глушенко

Применение системы MATLAB для оценки рисков
информационной безопасности организации 35

Г.А. Кухарев, Е.И. Каменская, Н.Л. Щеголева

Методы представления и сравнения семантически
разных классов изображений 43

С.Ю. Соловьев, Д.Е. Стельмашенко

Применение принципов экспертной классификации
для анализа формальных понятий 53

Математические методы и алгоритмы решения задач бизнес-информатики

А.И. Марон

Сценарный подход к планированию распределения
персонала НИИ с учетом рисков 58

Н.К. Хачатрян

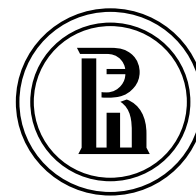
Динамическая модель организации грузоперевозок
при ограниченности емкостей перегонных путей 62

Программная инженерия

П.П. Олейник

Элементы среды разработки программных комплексов
на основе организации метамодели объектной системы 69

Annotations 77



БИЗНЕС-ИНФОРМАТИКА

№4(26)–2013

Междисциплинарный
научно-практический журнал
НИУ ВШЭ

Журнал рекомендован ВАК
для научных публикаций

Подписной индекс издания
в каталоге агентства
«Роспечать» – 72315

Учредитель:
Национальный
исследовательский университет
«Высшая школа экономики».
Выходит 4 раза в год.

Главный редактор
Голосов А.О.

Заместители главного редактора
Горбунов А.Р., Исаев Д.В.

Научный редактор
Лычкина Н.Н.

Технический редактор
Осипов В.И.

Дизайн обложки
Борисова С.Н.

Компьютерная верстка
Богданович О.А.

Администратор веб-сайта
Проценко Д.С.

Адрес редакции:
105187, г. Москва,
ул. Кирпичная, д. 33/5
Тел. +7 (495) 771-32-38
E-mail: bijournal@hse.ru

За точность приведенных сведений
и содержание данных,
не подлежащих открытой публикации,
несут ответственность авторы

При перепечатке ссылка на журнал
«Бизнес-информатика» обязательна

Тираж 500 экз.

Отпечатано в типографии НИУ ВШЭ
г. Москва, Кочновский проезд, 3.

© Национальный
исследовательский университет
«Высшая школа экономики»

АНАЛИЗ ПАТТЕРНОВ В СТАТИКЕ И ДИНАМИКЕ, ЧАСТЬ 2: ПРИМЕРЫ ПРИМЕНЕНИЯ К АНАЛИЗУ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ

Ф.Т. Алескеров,

доктор технических наук, руководитель Департамента математики факультета экономики Национального исследовательского университета «Высшая школа экономики», заведующий лабораторией Института проблем управления имени В.А. Трапезникова РАН

В.Ю. Белоусова,

кандидат экономических наук, заведующий отделом методологии бюджетного планирования Института статистических исследований и экономики знаний, доцент кафедры банковского дела Департамента финансов факультета экономики, с.н.с. Банковского института Национального исследовательского университета «Высшая школа экономики»

Л.Г. Егорова,

преподаватель Департамента математики факультета экономики Национального исследовательского университета «Высшая школа экономики»

Б.Г. Миркин,

доктор технических наук, профессор кафедры анализа данных и искусственного интеллекта отделения прикладной математики и информатики факультета бизнес-информатики Национального исследовательского университета «Высшая школа экономики»

E-mail: alesk@hse.ru, vbelousova@hse.ru, legorova@hse.ru, bmirkin@hse.ru

Адрес: г. Москва, ул. Мясницкая, 20

Приводится обзор работ по анализу паттернов в применении к макроэкономическому анализу стран в динамике, к анализу банковских систем Турции и России, электорального поведения в Великобритании и Финляндии, инновационного развития регионов России в долгосрочной перспективе.

Ключевые слова: паттерны данных, динамический анализ паттернов, кластерный анализ.

¹ Работа выполнена при финансовой поддержке Минобрнауки России по государственному контракту от 14.06.2012 г. № 07.514.11.4144 в рамках ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007-2013 годы».

Исследование осуществлено в рамках программы фундаментальных исследований НИУ ВШЭ в 2012 году. Работа авторов была также поддержана рядом лабораторий НИУ ВШЭ: Лаборатория анализа и выбора решений (Алескеров Ф.Т., Белоусова В.Ю., Егорова Л.Г., Миркин Б.Г.), Лаборатория алгоритмов и технологий анализа сетевых структур (Егорова Л.Г., Миркин Б.Г.). Мы благодарны за эту поддержку.

1. Введение

В данной работе термин «паттерн» используется для обозначения комбинаций параметров значимых признаков, характерных для определённых групп объектов, соответствующих отличающимся от других типам поведения. В предыдущей части мы сделали обзор литературы в каждом из трёх основных аспектов этого понятия: использование понятия паттерн в науке и технике, методы кластер-анализа, динамика многомерных объектов, а также предложили уточнение этого понятия в эквивалентных, но когнитивно-различающихся терминах параллельных координат, конъюнктивных описаний и геометрических «боксов». Для данных о динамике функционирования социально-экономических объектов предложено использовать типологию функционирования объектов по частоте смены паттерна. Этот подход в данной части статьи будет применён к анализу реальных данных на основе использования формализма параллельных координат. В частности, будут рассмотрены проблематика сопоставительного макроэкономического анализа, функционирование банковского сектора в Турции и России в современных условиях, паттерны поведения избирателей в условиях многопартийной системы и проблемы инновационного развития регионов РФ в долгосрочной перспективе. Следует иметь в виду, что многие из конкретных исследований, отражённых в статье, были выполнены ещё до того, как в полной мере сложилось понятие паттерна, развиваемое в данной работе.

2. Анализ паттернов в параллельных координатах

Как было определено в первой части данной статьи [16], паттерн — это такая совокупность объектов, которая характеризуется определенной комбинацией значений показателей объектов.

Для визуального отображения процесса анализа и поиска паттернов в n -мерном пространстве обычно используется система параллельных координат [12, 13], состоящая из n параллельных линий, вертикальных и равномерно распределённых. Каждая из этих параллельных осей отражает один из n выбранных для кластеризации признаков. Каждый объект X (точка в n -мерном пространстве признаков) представляется в виде ломаной с вершинами на параллельных осях (помеченных названиями показателей); положение вершины на i -й оси

соответствует i -й координате точки, т.е. значению i -го показателя. Таким образом, для M объектов будет построено M ломаных линий.

Сами числа, то есть конкретные значения параметров, характеризующие объект, не всегда важны для выделения паттернов. Гораздо более важной часто является информация о соотношении параметров. Чтобы проиллюстрировать это, рассмотрим возможные результаты голосования на выборах в трех округах по данным табл. 1 [9].

Таблица 1.
Гипотетические результаты выборов

	Партия 1	Партия 2	Партия 3
Округ 1	40	20	40
Округ 2	45	10	45
Округ 3	20	60	20

Объектами в этом случае являются округа, признаками — уровень поддержки трех политических партий, участвовавших в этих выборах. Эти данные изображены на рис. 1 в системе параллельных координат. Значения показателей были нормированы к 100% и представляют собой доли голосов, полученных партиями в ходе выборов.

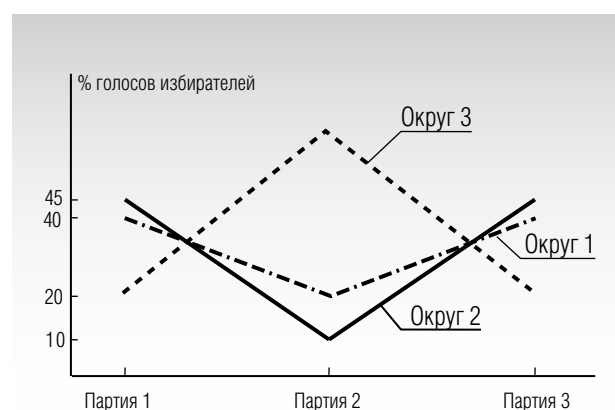


Рис. 1. Результаты гипотетических выборов

Очевидно, что первый и второй округа голосуют больше за первую и третью партии, в то время как третий округ поддерживает вторую партию. Эти результаты разделяют все округа на две группы схожих объектов: в первой группе будут округа 1 и 2, показывающие практически одинаковый результат голосования, а во второй группе — только округ 3.

Таблица 2.

Иллюстративная информация о банках

	Доля привлеченных депозитов в активах банка	Доля выданных кредитов в активах банка	Доля валютных операций в активах банка
Банк 1	40	20	40
Банк 2	45	10	45
Банк 3	20	60	20

Если целью исследования является, например, оценка деятельности банка (табл. 2), тогда можно предложить следующую интерпретацию результатов. Рассмотрим основные операции коммерческого банка: привлечение депозитов, предоставление кредитов и расчетно-кассовые операции, из которых наиболее востребован обмен валюты. Показателями, измеряющими активность банков в области оказания данных услуг, будут соответственно доли привлеченных депозитов, выданных кредитов и валютных операций в активах банка. В табл. 2 на пересечении строк и столбцов находятся значения параметров, характеризующих деятельность банка на определенный момент времени. Так, 20% активности первого банка приходится на кредитные и 40% - на валютные операции, 40% суммарных обязательств банк финансирует за счет привлечения депозитов.

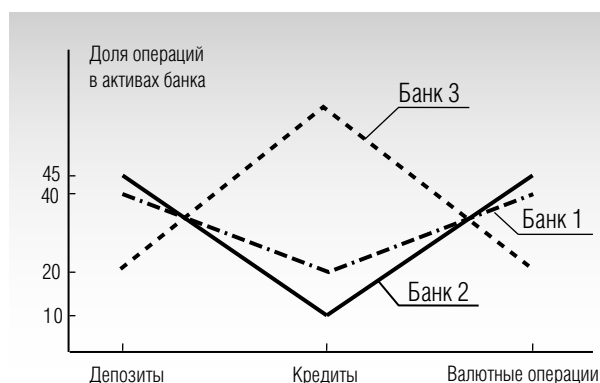
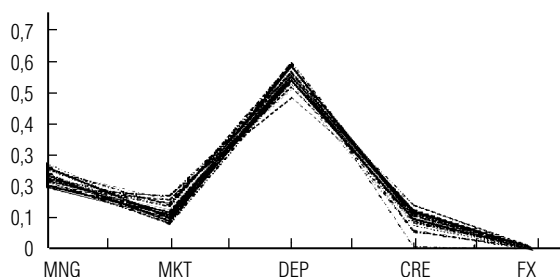


Рис. 2. Паттерны банков



Если изобразить эти данные на параллельных осях (рис. 2), то можно видеть, что первые два банка похожи, в то время как банк 3, очевидно, следует иной модели поведения [17]. Здесь показатели были также нормированы к 100% и представляют собой доли этих операций в активах банка.

Далее на выборке объектов, представленных в виде векторов в пространстве признаков или в виде ломаных в системе параллельных координат, проводится кластеризация и получается разбиение всех объектов выборки по кластерам-паттернам. Напомним, что объекты внутри одного кластера-паттерна должны быть схожими, а объекты разных паттернов — различными.

На рис. 3 представлен пример двух различных паттернов, полученных в [4, 6-7]. Объектами в данной задаче являются филиалы одного банка, в качестве признаков для оценки эффективности деятельности филиала были выбраны такие критерии, как менеджмент (MNG), маркетинг (MKT), активность банка на рынках депозитов (DEP), кредитов (CRE) и операций с валютой (FX).

Кластер 1 (слева) содержит 207 объектов, кластер 2 (справа) содержит 9 объектов. Видно, что объекты внутри каждого кластера обладают сходными оценками по всем признакам и, следовательно, относятся к одному классу. Паттерны кластеров между собой заметно разнятся и очевидно, что характер деятельности филиалов банка из первого паттерна отличается от характера деятельности филиалов из второго паттерна.

В классических терминах кластерного анализа задачу определения типов паттернов можно представить следующим образом.

Пусть X — множество объектов, Y — множество номеров (имён, меток) кластеров. Для формального описания понятия близости объектов задается функция расстояния (метрика) $\rho(x, x')$ между объектами $x, x' \in X$. Для рассматриваемой конечной выборки объектов $X^m = \{x_1, \dots, x_m\} \subset X$ требуется разбить ее на непересекающиеся подмножества, на-

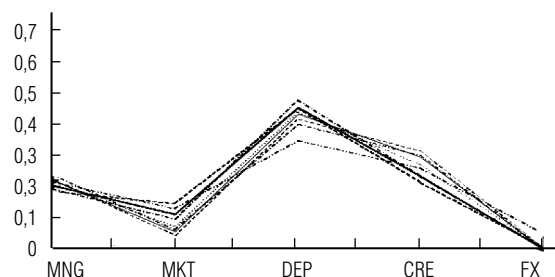


Рис. 3. Пример кластерных кривых различных паттернов

зываются *кластерами*, так, чтобы каждый кластер состоял из объектов, близких по метрике ρ , а объекты разных кластеров существенно отличались. При этом каждому объекту $x_i \in X^m$ приписывается номер кластера y_i .

Анализ паттернов — это такой кластер-анализ, при котором получаемые кластеры вполне адекватно описываются конъюнкциями некоторых значений или более точно — интервалов значений — некоторых признаков.

Как упоминалось выше, значения признаков в анализе паттернов не абсолютизируются — визуально паттерн определяется не столько положением объектов на осях, сколько углами наклона соответствующих линий. Чтобы выделить однородные группы среди всех объектов, необходимо оценивать объекты не по абсолютным значениям их признаков или параметров, а по углам наклона ломаных кривых, обозначающих объекты в системе параллельных координат. Такой алгоритм формирует кластеры, в которых система показателей для разных объектов имеет одну ту же структуру, хотя и отличается по абсолютному размеру. Например, если два изучаемых объекта А и В характеризуются векторами параметров (20; 50; 60) и (2; 5; 6), то при традиционной схеме кластерного анализа эти два объекта попадут в различные кластеры. Однако, если в кластеры объединяются объекты, похожие по структуре, то А и В попадут в один кластер [2-4, 6].

В частности, объекты выборки здесь могут характеризоваться не векторами значений, а кусочно-линейными функциями, характеризующими ломаные линии, проходящие через параллельные оси. Формально этот подход, впервые предложенный в [2], заключается в следующем.

Пусть X — множество объектов. Каждому объекту $x \in X$ поставим в соответствие n -мерный вектор $Z_x = \{z_1^x, \dots, z_n^x\}$, характеризующий объект x по n признакам, то есть его признаковое описание.

Построим по каждому вектору Z_x ломаную, которая проходит через точки $z_1^x, \dots, z_n^x$, то есть кусочно-линейную функцию

$$\begin{aligned} f^x(t) &= \{k_i^x t + a_i^x, \text{ если } i \leq t \leq i+1\}, \\ k_i^x i + a_i^x &= z_i^x, \\ k_i^x(i+1) + a_i^x &= z_{i+1}^x, i=1, \dots, n-1. \end{aligned} \quad (1)$$

Теперь каждому интервалу $[i, i+1]$ поставим в соответствие пару $V_i^x = (k_i^x, a_i^x)$, и каждой функции

f^x — вектор $V^x = \{V_1^x, \dots, V_n^x\}$. В качестве расстояния между точками V^{x_1} и V^{x_2} можно использовать евклидово расстояние, расстояние городских кварталов (манхэттенское расстояние) и др.

Далее одним из методов кластерного анализа проводим разбиение множества объектов $\{X_i\}$ так, чтобы ломаные внутри кластеров были похожи, а в разных кластерах находились «далеко» друг от друга в смысле выбранного расстояния. Для нахождения разбиения $\{X_i\}$, можно использовать различные методы кластерного анализа (см. часть 1 данного обзора).

Кроме того, анализ паттернов данных позволяет не только получать информацию о текущем состоянии исследуемых объектов, но также и отслеживать происходящие с течением времени изменения и иметь наглядное представление о динамике развития объектов с учетом всех признаков, характеризующих паттерны.

Если доступна информация о количественных характеристиках объектов, измеренная в последовательные моменты времени $t = 1, 2, \dots, k$, то задача выделения паттернов данных может быть выполнена для каждого момента времени. Тогда каждому объекту выборки $x_i \in X$ ставится в соответствие последовательность номеров кластеров (траектория) $\{y_i^1, y_i^2, \dots, y_i^k\}$, к которым данный объект x_i принадлежал в момент времени t . На основании полученных «траекторий развития» можно выделять динамические группы объектов, демонстрирующих одинаковое поведение и сходные характеристики в динамике.

Предположим, что в примере с результатами гипотетических выборов есть паттерны, построенные на итогах нескольких последовательных выборов. Каждый паттерн содержит линии, которые описывают определенный тип электоральной поддержки в округе, например, больше голосов отдается правым партиям и меньше — левым и националистам, или одинаковая поддержка правых и левых партий вместе с низким уровнем голосов за националистские партии. Если мы применим алгоритм кластерного анализа к результатам нескольких выборов, то получим паттерны для каждого голосования, и, возможно, типы электоральной поддержки, связанные с каким-то паттерном, будут повторяться, хотя этим паттерном в разные годы может описываться голосование в разных округах. Пусть в примере выборов из табл. 1 в момент времени T найдено два паттерна. Предположим, что на следующих выбо-

рах в момент времени $T + 1$ были другие результаты голосования (табл. 3).

Таблица 3.

**Гипотетические результаты выборов
в момент времени $T + 1$**

	Партия 1	Партия 2	Партия 3
Округ 1	35	30	35
Округ 2	38	24	38
Округ 3	40	20	40

Теперь уже второй и третий округ голосуют схожим образом, в то время как в округе 1 на выборах $T + 1$ -го года произошли значительные изменения, и теперь распределение голосов между партиями в этом округе уже не подходит к первому паттерну, как, впрочем, и ко второму. На этих выборах появился новый паттерн электоральной поддержки.

После получения всех возможных паттернов поддержки избирателей на выборах интересно проанализировать, какой паттерн был выбран каждым округом в начале анализируемого периода, как он менял свои паттерны на последующих выборах. На рис. 4 по горизонтальной оси отложен год выборов, на вертикальной — паттерны.

Легко увидеть, что округ 1 сменил паттерн с 1 на 3, округ 2 остался верен паттерну 1, а округ 3 поменял паттерн 2 на паттерн 1. Для целей анализа и планирования предвыборных стратегий очень важно понять, какие округа сохраняют свои предпочтения, а какие меняют их с течением времени и почему. Применение методологии динамического анализа паттернов к результатам голосований на парламентских выборах в Великобритании и к результатам финских муниципальных выборов описано в разделе 5 [1, 9, 20].

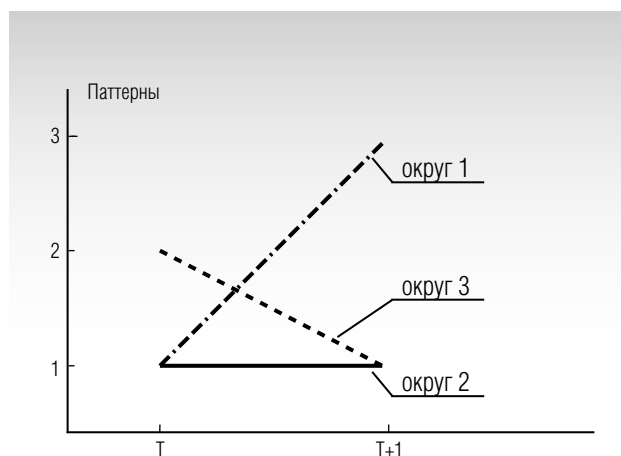


Рис. 4. Паттерны голосований

Динамический анализ подразумевает также классификацию объектов в соответствии с устойчивостью выбранных ими паттернов. Как правило, все объекты разбиваются на несколько групп: абсолютно устойчивые, устойчивые, неустойчивые и абсолютно неустойчивые.

К группе абсолютно траекторно устойчивых объектов относятся те объекты, которые придерживаются одного паттерна в течение всего периода наблюдений. Если объекты меняли свои паттерны небольшое число раз (число допустимых изменений задается исследователем в зависимости от конкретной задачи и длины выборки), то они называются устойчивыми. Абсолютно неустойчивыми называются объекты, меняющие паттерны в каждый момент времени. Все остальные объекты относятся к группе неустойчивых объектов.

В качестве иллюстрации приведем результаты исследования [17]. Было проанализировано, насколько 100 крупнейших российских банков по валюте баланса придерживаются бизнес-модели, выбранной ими после кризиса 2004 г., на временном интервале I кв. 2005 г. — II кв. 2007 г. Если на протяжении 10 отчетных периодов (квартал) банк избирал одну и ту же бизнес-модель, его в [17] отнесли к категории абсолютно устойчивых. «Траектория» абсолютно устойчивых банков может быть представлена как упорядоченный ряд паттернов во времени, например, 1-1-1-1-1-1-1-1-1-1. Оказалось, что абсолютно устойчивых банков всего 24 (более подробно см. далее).

3. Применение метода анализа паттернов в макроэкономическом анализе

Опишем применение метода анализа паттернов для поиска связи между такими экономическими показателями как инфляция (inflation), темп роста денежной массы (growth rate of money), реальный объем производства (real output) и реальный объем производства на душу населения (real output per capita) на данных 119 стран [2,3].

В качестве оценки для денежной массы использовались показатели M1 (наличные деньги в обращении, а также чеки и вклады до востребования) и M2 (M1 плюс срочные вклады). Для оценки уровня инфляции учитывался индекс потребительских цен (то есть стоимость корзины товаров по CPI), а для оценки производства — валовый внутренний продукт — GDP, отражающий рыночную стоимость всех конечных товаров и услуг, произведенных за

год на территории государства. В табл. 4 приведены результаты первичной статистической обработки данных, а именно средние значения коэффициента корреляции указанных показателей по нескольким группам стран.

Таблица 4.

**Коэффициенты корреляции
между темпами роста денежной массы,
индекса потребительских цен и ВВП**

	M1(M2)- CPI	M1(M2)- GDP	CPI- GDP	Число стран
Все страны	0.99 (0.99)	-0.08 (-0.06)	-0.16	119
Страны ОЭСР	0.96 (0.97)	0.56 (0.57)	0.48	23
Страны G-7	0.76 (0.77)	0.12 (0.23)	-0.31	7
Страны Латинской Америки	1.00 (0.99)	-0.21 (-0.20)	-0.25	30

На полной выборке из 119 стран (табл. 4) корреляция между темпом роста денежной массы и реальным объемом производства (в том числе и на душу населения), равно как и между инфляцией и реальным объемом производства, обнаружена не была. Проведенная кластеризация позволила выделить 6 групп стран, имеющих одинаковые паттерны соотношений между параметрами. Были рассчитаны коэффициенты корреляции отдельно для двух индикаторов денежной массы, для M1 (табл. 5) и M2 (табл. 6).

Таблица 5.

**Коэффициенты корреляции
между параметрами для M1**

	M1-CPI	M1-GDP per capita	CPI-GDP per capita	Число стран
All	0.99	-0.12	-0.18	119
Cluster 1	1.0	0.20	0.15	12
Cluster 2	0.99	0.36	0.37	25
Cluster 3	0.98	0.16	0.28	40
Cluster 4	0.92	0.09	0.37	19
Cluster 5	0.88	0.60	0.86	12
Cluster 6	-0.28	0.33	0.46	6

Интересно, что в один кластер попали страны со схожими экономическими и политическими условиями. Так, второй кластер содержит 23 развивающиеся страны, имеющие высокую инфляцию и нестабильную политическую ситуацию: Боливию, Чили, Эквадор, и т.д. Состав стран в кластерах меняется незначительно — как правило, одни и те же страны группируются в одни кластеры в обоих случаях.

Для всех кластеров и выборки в целом показано, что существует положительная зависимость между ростом денежной массы, находящейся в обращении, и инфляцией. Однако на всей выборке корреляция между темпом роста ВВП и темпом роста инфляции не наблюдается. В свою очередь, внутри двух из шести кластеров данная взаимосвязь обнаружена для расчетов с M1 и внутри пяти кластеров — для M2.

Таблица 6.

**Коэффициенты корреляции
для параметров с M2**

	M2 -CPI	M2-GDP per capita	CPI-GDP per capita	Число стран
All	0.99	-0.12	-0.18	119
Cluster 1	1.0	0.60	0.59	13
Cluster 2	1.0	-0.85	-0.87	7
Cluster 3	0.99	0.41	0.44	27
Cluster 4	0.98	0.33	0.44	33
Cluster 5	0.87	0.30	0.53	26
Cluster 6	0.74	0.65	0.91	10

Таким образом, эти результаты подтверждают гипотезу о том, что в каждом кластере страны используют различные механизмы разработки и реализации монетарной политики. Если данные паттерны не различать, то невозможно рассматривать дифференцированные механизмы экономического развития стран; виден лишь эффект общего характера — зависимость между эмиссией денежных знаков и инфляцией.

4. Применение метода анализа паттернов в банковской сфере

Применение метода паттернов для анализа структурных особенностей развития банковской системы проводилось в 1997-2004 гг. на данных Турции [5-8], а также для российских банков:

- ♦ в период восстановления экономики и финансового сектора после кризиса 1998 г. — с первого квартала 1999 года по третий квартал 2003 года [18];
- ♦ в период кризиса ликвидности 2004 г., а также кредитного бума и стабильного развития банковского сектора — с I квартала 1999 г. по II квартал 2007 г. [17];
- ♦ в преддверии финансового кризиса 2008 г. и его активной фазы — с апреля 2007 года по март 2009 года [15, 17].

Наиболее комплексный анализ бизнес-моделей российских банков в период с 2006-2009 с использованием метода паттернов представлен в [15].

Остановимся на описании некоторых проектов из этого цикла работ. В [21] для изучения развития российского банковского сектора, в частности, для определения структурных различий в банковской системе России, выявления тенденций и оценки состояния коммерческих банков, было исследовано 1018 кредитных организаций. Период анализа включал пять лет: с марта 1999 года по сентябрь 2003 года включительно.

Для использования предложенной методики в первую очередь потребовалось сформировать систему показателей, которая являлась бы отображением характеристик, описывающих различные стороны функционирования объекта. Например, для анализа функционирования банков в качестве основной концепции в [17] было предложено использовать модель CAMEL. Здесь деятельность банка характеризуется пятью его фундаментальными показателями: достаточностью капитала (C – Capital), качеством активов (A – Assets Quality), качеством менеджмента и системы управления (M – Management), прибыльностью (E – Earnings) и ликвидностью (L – Liquidity).

При разработке данной системы показателей банковской деятельности во внимание принимались следующие соображения:

- ✧ выбранные показатели должны отражать особенности формирования финансового рынка и деятельности банковских учреждений в России;
- ✧ так как банковская система России является достаточно неоднородной по размеру составляющих ее учреждений, то систему показателей целесообразно строить на основе относительных показателей. Последнее снимает проблему размерности;
- ✧ анализ структурных различий строится на основе тех количественных показателей деятельности банка, которые имеются в открытом доступе.

С учетом всех перечисленных факторов и на основании проведенного анализа динамики российского банковского сектора примером системы финансовых коэффициентов, отвечающих модели CAMEL, может служить следующая система показателей (см. табл. 7), используемая в [17].

Был проведен статистический анализ коэффициентов и получен 151 паттерн, которые определяли поведение банков в динамике. На основе анализа паттернов исследовалось поведение 10 крупных и

около 250 небольших, но стабильных по структурным характеристикам, банков. Было выделено 66 банков, принадлежащих к 16 паттернам, характеризующихся стабильными траекториями развития.

Таблица 7.

Список показателей по системе CAMEL [17]

Показатель	Коэффициент
Достаточность капитала (CAP)	Собственный капитал / Валюта баланса
Качество активов (GOS)	Вложения в госбумаги / Работающие активы
Качество менеджмента (MBK)	Полученные МБК / Суммарные обязательства
Качество менеджмента (CRED)	Кредиты коммерческим организациям / Работающие активы
Прибыльность (PR)	Фактическая прибыль / Работающие активы
Ликвидность (LIQ)	Ликвидные активы / Работающие активы

Полученные результаты позволили оценить степень однородности развития банковского сектора, его устойчивость, выявить тенденции развития, провести более глубокий анализ на уровне как группы финансовых институтов, так и отдельного банка с учетом реакции на изменение внешней среды, следования определенной стратегии развития, выявить конкурентные преимущества банков и проанализировать развитие рынка банковских услуг.

Кроме того, эндогенное выделение совокупностей банков, характеризующихся сходными неустойчивыми траекториями развития, позволило выявить информацию о слабых банках. Полученные результаты были полезны регулирующим органам для проведения превентивных мероприятий и отслеживания параметров развития банковской системы в целом и отдельных банков в частности.

В [15] представлен комплексный анализ бизнес-моделей ведущих, средних и малых российских коммерческих банков для наиболее актуального периода времени с учетом финансового кризиса 2007 г.: с 2006 г. по 2009 г. В его основе лежит группировка кредитных организаций по принципу однородности результатов их операционной и финансовой деятельности. При оценке бизнес-моделей учитывается структура активов и обязательств, уровень прибыльности и ликвидности банков. В этой задаче анализ был проведен для всех трех групп банков – крупные, средние и малые банки.

В качестве признаков для выделения паттернов были выбраны такие финансовые показатели, как достаточность капитала (С), качество активов (А), качество управления (М), прибыльность (Е), ликвидность (L), т.е. модель CAMEL плюс более тонкие характеристики: структура пассивов (F) и характеристика портфеля ценных бумаг (S). Например, на *рис. 5* представлены паттерны, полученные для банка «КИТ Финанс» в ходе анализа банковской системы России в период 2006–2009 гг. в [15]. Каждый паттерн, как оказалось, соответствует определенной бизнес-модели.

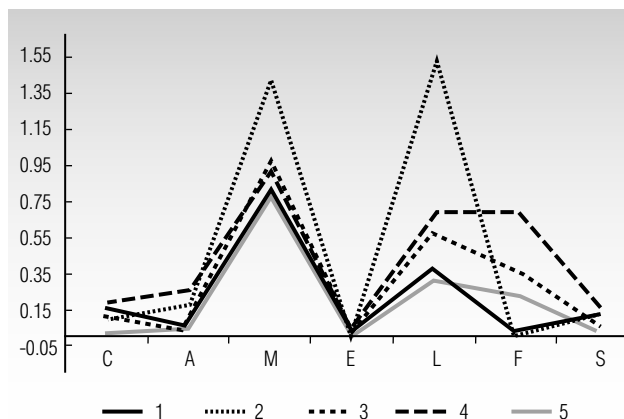


Рис. 5. Паттерны банка «КИТ Финанс»

Всего за исследуемый период (16 кварталов) данный банк придерживался 5 различных бизнес-моделей. Половину рассматриваемого временного интервала банк придерживался первой бизнес-модели, характеризующей достаточностью капитала на уровне 15%, просроченной задолженностью 7% и ликвидностью в пределах 40%. Убыточными бизнес-моделями стали для банка третья (которой банк придерживался на протяжении 4 кв. 2008 и 1 кв. 2009 г.) и пятая (2 кв. 2009 г.). Главной отличительной особенностью является уровень достаточности капитала (для третьей бизнес-модели – 10,7%, для пятой – 1,55%) и уровень ликвидности, который в два раза выше, чем у третьей модели поведения.

Деятельность банка «КИТ Финанс» была связана с инвестиционными операциями, которые, как показал мировой опыт, характеризуются высокой неопределенностью и часто приводят к финансовым потерям и списаниям с баланса убытков. В период кризиса, когда доверие инвесторов к инвестиционным активам резко падает, банк с подобной структурой операций теряет большин-

ство источников доходов, его финансовая устойчивость резко снижается, операции с ценными бумагами нужно либо приостановить, либо рассчитывать на еще большие убытки. В такой ситуации кредитной организации тяжело решить проблемы с убытками, и убыточность на протяжении нескольких периодов вполне объяснима, а на уровне крупных системно значимых банков – опасна, т.к. банкротство любого такого банка может привести к цепной реакции во всей банковской системе.

Применяя алгоритм кластерного анализа для данных, охватывающих нескольких периодов, мы получаем набор паттернов для каждого момента времени. Вполне возможно, что структура деятельности банков, описываемая различными паттернами, будет повторяться во времени, хотя на разных интервалах состав групп банков может меняться.

Таким образом, для каждого банка конструируются «траектории», которые показывают, какую бизнес-модель в определенный момент времени выбирает банк. Банки, имеющие одинаковые «траектории», составляют динамическую группу. Динамический анализ подразумевает также классификацию банков в соответствии с устойчивостью выбранных ими паттернов. Напомним, что при такой классификации все объекты разбиваются на несколько групп: абсолютно устойчивые, устойчивые, полуустойчивые, неустойчивые и абсолютно неустойчивые. К группе абсолютно устойчивых банков относятся те объекты, которые придерживаются одного паттерна в течение всего периода наблюдений. Если банки меняли свои бизнес-модели 1–2 раза за 10 отчетных периодов, то они называются полуустойчивыми. Абсолютно неустойчивыми называются банки, меняющие паттерны в каждый момент времени. Все остальные банков относятся к группе неустойчивых банков.

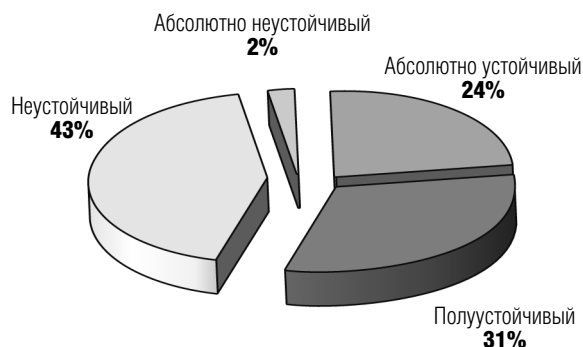


Рис. 6. Распределение крупнейших банков по типам устойчивости

Например, в [17] абсолютно устойчивым считается банк, который одной и той же бизнес-модели на протяжении 10 отчетных дат (кварталов) (см. распределение банков по типам устойчивости на рис. 6).

Если в течение 10 отчетных периодов банк произвел 1–2 смены бизнес-моделей, его можно отнести к категории полуустойчивых. «Траектория» полуустойчивых банков может иметь следующий вид: 1-2-1-1-1-1-1-1-1-1. Доля полуустойчивых банков в общем объеме анализируемой выборки составляет 31%. Таким образом, чуть более половины отнесены к устойчивым банкам данных двух категорий.

Банк является абсолютно неустойчивым, если в течение 10 отчетных дат сменил 9 бизнес-моделей, например, 65-37-28-68-38-28-39-23-29-22. Доля таких в банковском секторе РФ невелика – 2%. Остальные банки (43%) отнесены к категории неустойчивых банков.

Распределение крупнейших банков по частоте смены паттернов представлено в табл. 8. Видно, что на те банки, которые сменили от 1 до 3 бизнес-модели, приходится 39% анализируемых банков. На банки, чья деятельность может быть описана числовыми значениями единственной бизнес-модели, – 24%. Таким образом, кумулятивный процент таких банков составляет 63%. Кумулятивный процент, равный 80%, достигается при включении банков, которые на протяжении последних 10 отчетных дат сменили четыре бизнес-модели.

Таблица 8.

Распределение крупнейших банков* по частоте смены бизнес-моделей

Частота смены бизнес-моделей	Кумулятивное количество банков	Кумулятивный процент
0	24	24
1	34	33
2	56	55
3	64	63
4	82	80
5	90	88
6	96	94
7	97	95
8	100	98
9	102	100

* Включая Сбербанк России и Внешторгбанк

Отметим, что устойчивость выбора банками бизнес-модели в течение рассматриваемого периода повышается: банки определяются со своей ры-

ночной нишей. Кредитные организации, которые аккумулируют практически 90% совокупных активов банковского сектора России на 01.01.2007 г., выбрали ориентацию на кредитование нефинансового сектора экономики, причем выполняют функцию финансового посредника с прибылью.

Таким образом, динамический анализ бизнес-моделей, выбираемых российскими банками, прояснил степень их устойчивости к условиям среды, позволил выявить тенденции их развития и определить, насколько банки ориентированы на кредитование нефинансового сектора экономики, т.е. выполняют ли они функцию финансового посредника.

Кроме этого, данный анализ показывает, насколько устойчива классификация банков по структуре операций и принимаемых ими основных финансовых рисков во времени. Таким образом, подтверждается гипотеза о том, что неоднородность российских банков необходимо учитывать в динамике, поскольку большинство банков из топ-100 с течением времени меняют свои бизнес-модели.

В [5–7] анализ паттернов данных был применен к задаче распределения персонала по филиалам турецкого банка Yapi Kredi Bank. Задача решалась в два этапа: на первом этапе все филиалы банка были разбиты на классы (паттерны) по микрорынкам, на которых они функционируют, на втором этапе осуществлялось ранжирование филиалов, описываемых одинаковыми паттернами, по эффективности их деятельности.

Распределение филиалов по паттернам было необходимо для их адекватной оценки с учетом разной направленности деятельности филиалов. Во-первых, деятельность филиала зависит от географического положения и характеристик локального рынка, на котором он находится – какие банковские продукты и услуги востребованы, какие конкуренты присутствуют на локальном рынке и т.п. Во-вторых, филиалы банка изначально неоднородны: есть головной офис, местные отделения, операционные офисы и т.п. Детальный анализ структуры деятельности каждого филиала банка слишком затратен как с финансовой, так и с временной точки зрения. Анализ паттернов позволяет разбить все множество разнородных отделений банка на классы, имеющие схожую структуру деятельности внутри каждого класса, и решать задачу оценки эффективности среди таких классов.

Объектами в данной задаче являются филиалы одного банка, в качестве признаков для оценки эф-

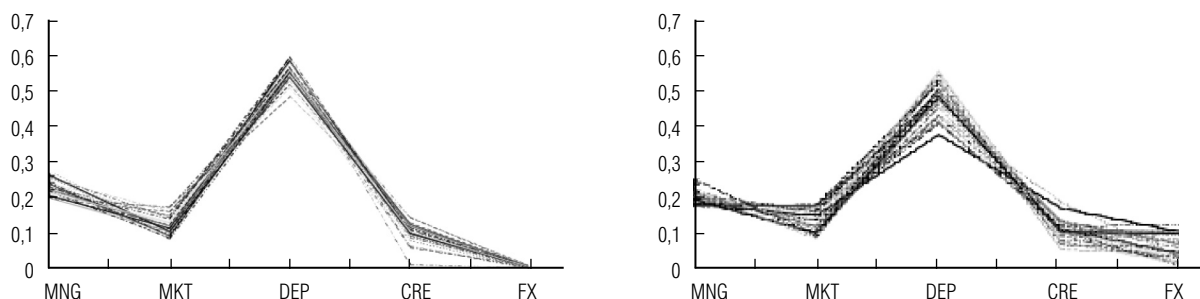


Рис. 7. Пример кластерных кривых различных паттернов

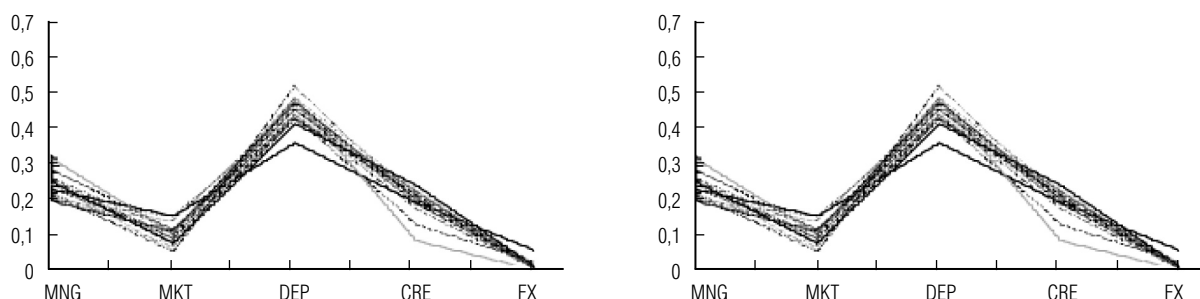


Рис. 8. Различные паттерны филиалов банка

эффективности деятельности филиала были выбраны такие критерии, как менеджмент (MNG), маркетинг (MKT), активность банка на рынках депозитов (DEP), кредитов (CRE) и операций с валютой (FX).

На рис. 7 и рис. 8 представлены примеры четырех различных паттернов, полученных в [7]. Паттерн 1 (рис. 7, слева) содержит 207 филиалов, главная активность которых сосредоточена вокруг депозитных операций. В паттерн 2 вошло 76 филиалов (рис. 7, справа), которые также делают упор на депозитах и операциях с валютой. В паттерне 3 расположено 20 филиалов, отличающихся от предшественников высоким значением параметра кредитования (рис. 8, слева). Четвертый паттерн, состоящий из 9 филиалов (рис. 8, справа), практически не характеризуется валютными операциями, но филиалы в нем более активны в области кредитования, чем филиалы первого и даже третьего паттерна.

В [4] аналогичное исследование было проведено для распределения персонала по отделениям банка внутри групп отделений, которые описываются одинаковыми паттернами. В качестве характеристик для выделения паттернов были выбраны показатели, характеризующие распределение трудозатрат на различные операции: трудозатраты на менеджмент и маркетинг, объем вкладов, объем кредитов и размер обменных операций. Расчет этих параметров см. в [10, 11, 14].

В [22] анализировалась эффективность конкурсного управления в Агентстве по страхованию вкладов — государственной корпорации, обеспечивающей эффективную процедуру банкротства российских кредитных организаций, в основном коммерческих банков. Предлагаемая методика для оценки эффективности деятельности управляющей компании при осуществлении процедур конкурсного производства (ликвидации) включала решение двух основных подзадач:

1. выявление однородных групп кредитных организаций (по паттернам, которых придерживаются банки);
2. построение регрессионных моделей возврата активов ликвидируемых банков.

Было продемонстрировано, что стандартные методы кластер-анализа в применении к задаче разделения всей совокупности кредитных организаций на однородные группы не дали адекватных результатов. С помощью этих методов не удалось разделить всю совокупность из 82 банков на небольшое число кластеров, содержащих сравнимые объекты, примерно равного объема. Как правило, при применении стандартных методов кластеризации, таких как метод k -средних, метод связанных компонент, а также иерархических процедур дальнего и ближнего соседа, при заданном количестве кластеров возникали одно-двухэлементные кластеры,

состоящие из банков, явно выделяющихся объёмами активов по сравнению с остальными. В связи с этим к первому классу задач — разработке процедуры, связанной с формированием и наполнением конкурсной массы¹, т.е. с выведением всех средств и активов ликвидируемой кредитной организации, — применялся метод анализа паттернов.

Основное отличие метода анализа паттернов от стандартных методов кластер-анализа состоит в том, что при его использовании предполагается независимость распределения по кластерам от совокупного объёма активов, т.е. крупный банк с большим объёмом активов на балансе и мелкий банк со сравнительно малым объёмом активов на балансе будут отнесены к одному кластеру при условии одинаковой структуры портфеля активов. Более формально процедура определения паттернов выглядит следующим образом: от пространства абсолютных значений показателей по каждому виду активов осуществляется переход к пространству пар из углов наклона и свободных коэффициентов линейных частей профильных кривых (при этом размерность пространства векторов меняется с n на $2n - 2$). Каждому банку ставится в соответствие вектор, состоящий из семи пар углов наклона и свободных коэффициентов. И в дальнейшем банки относятся к разным паттернам на основе сравнения этих векторов.

На следующих двух диаграммах приведены примеры двух похожих (Bank1 и Bank2, *рис. 9*) и двух различных паттернов банков (Bank3 и Bank4, *рис. 10*), полученных при анализе данных по незавершённым делам о ликвидации 82 российских кредитных организаций [22]. Среди показателей, характеризующих положение банка на момент начала процедуры ликвидации, были выделены следующие:

- ♦ касса и денежные средства на корреспондентском счёте в Банке России — «LA1»;
- ♦ обязательные резервы, перечисленные в Банк России, — «R1»;
- ♦ кредиты юридическим лицам — «CP1»;
- ♦ кредиты физическим лицам — «IP1»;
- ♦ вложения в ценные бумаги (векселя) «BILLS1»;
- ♦ основные средства — «BAS1»;
- ♦ прочие активы — «OTAS1».

После анализа всей совокупности, состоящей из 82 банков, оказалось, что фактически разделить все банки на несколько кластеров можно, учитывая в первую очередь значения показателей «кредиты юридическим лицам» и «объём вексельного

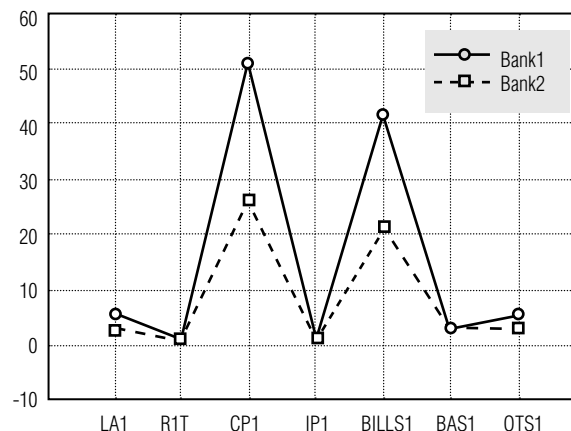


Рис. 9. Схожие паттерны для банков, подлежащих ликвидации

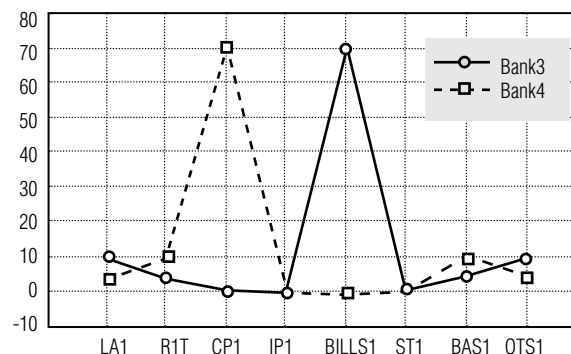


Рис. 10. Разные паттерны для банков, подлежащих ликвидации

портфеля». Сначала все банки были разделены на 6 групп, затем с целью укрупнения их количество было уменьшено до четырёх. Таким образом, каждый из 82 банков, проходящих процедуру ликвидации, был отнесён к одной из следующих групп (кластеров).

Кластер 1 включает банки, портфель активов которых практически полностью состоит из одних векселей. На *рис. 11* представлен паттерн 11 банков с пиком в ценных бумагах (значение показателя BILLS1). На этом и последующих рисунках отсутствуют числовые шкалы значений показателей, так как конкретные числа далее не имеют значения,

¹Под конкурсной массой понимаются основные и оборотные средства, нематериальные активы, дебиторская задолженность, кредиторская задолженность, прочие пассивы и активы, находящиеся на балансе должника. Всё имущество должника, имеющееся на момент открытия конкурсного производства и выявленное в ходе конкурсного производства, составляет конкурсную массу.

Кластер 1. Векселя.

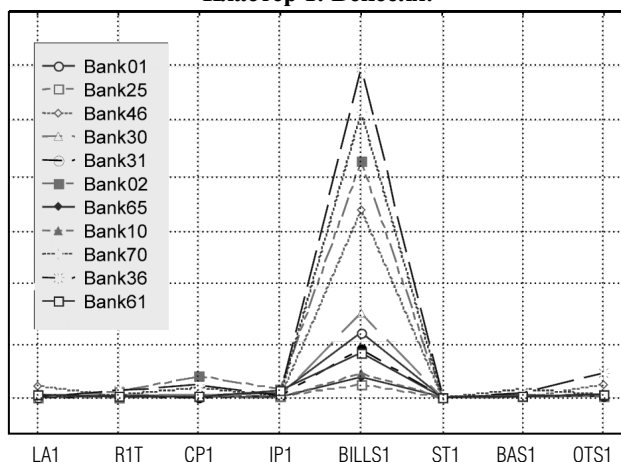


Рис. 11. Вексельные банки

важны лишь соотношения между параметрами и формы ломаных, соответствующих различным паттернам.

Кредитные организации из этого кластера, по мнению экспертов, относятся к группе так называемых «вексельных банков». Так как векселя — это активы, как правило, с достаточно низкой ликвидностью, то отнесение кредитной организации к данной группе банков может говорить о том, что перед банкротством в кредитной организации проводились операции по замене активов на менее ликвидные. А такое обстоятельство свидетельствует о возможно имевшем место «криминальном банкротстве», зачастую преднамеренном. Соответственно, эта группа банков характеризуется крайне малым процентом возврата активов.

Кластер 2 включает банки, большую часть порт-

Кластер 2. Кредиты юридическим лицам.

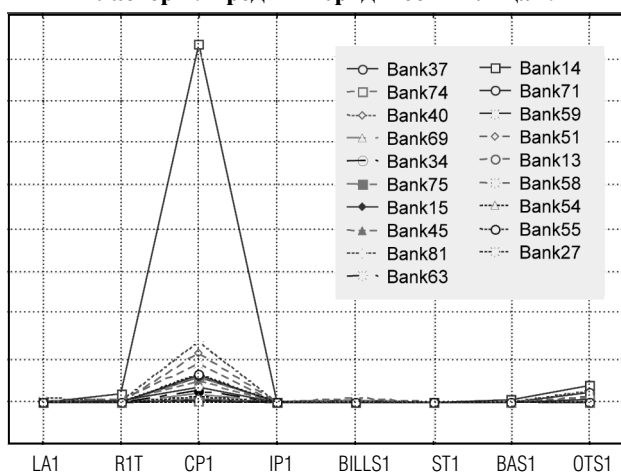


Рис. 12. Банки, активно кредитующие юридических лиц

феля активов которых занимают кредиты юридическим лицам. Диаграмма 19 банков с пиком в кредитах юридическим лицам (показатель CP1) показана на рис. 12.

Как видно из графика, на их балансе также присутствуют «другие активы» (OTAS1) и резервы в ЦБ, но в значительно меньшем объеме. И если наличие таких резервов — требование законодательства, то относительно большие объемы «других активов» у некоторых банков могут говорить о наличии, например, значительных основных фондов на балансе. Тем не менее, столь несбалансированный портфель активов, как правило, приводит к крайне малому проценту возврата активов. Банкротство такого банка может произойти из-за невозврата кредитов несколькими крупными клиентами.

Кластер 3 содержит банки, большая часть портфеля активов которых распределена между кредитами юридическим лицам и векселями. На рис. 13 приведены графики для 26 банков с двумя пиками в профилях: «кредиты юр. лицам» (CP1) и «векселя» (BILLS1), соответственно.

Кредитные организации из этой группы обладают сбалансированным портфелем активов. Банки этой группы имеют, по экспертным оценкам, средний процент возврата активов. Банкротство этих кредитных организаций, скорее всего, имеет рыночный характер, т.е. оно происходит, как правило, из-за неправильных (непреднамеренных) действий менеджмента банка или неблагоприятной ситуации на рынке для конкретного банка.

В кластер 4 вошли 26 банков, не вошедших в другие кластеры, а именно те банки, в портфеле активов которых присутствуют в примерно равных

Кластер 3. Кредиты юридическим лицам и Векселя.

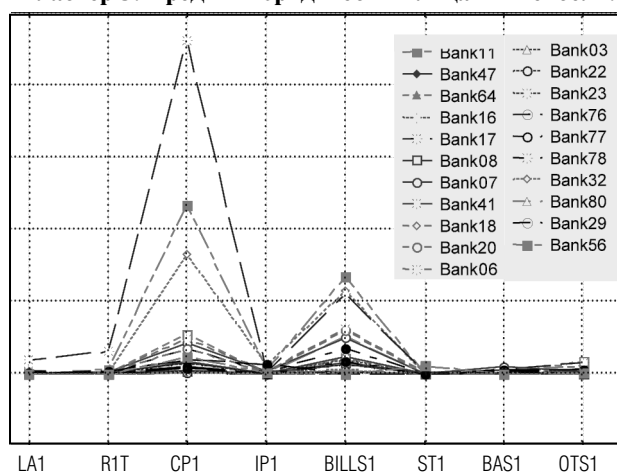


Рис. 13. Банки, имеющие в портфеле корпоративные кредиты и векселя

Кластер 4. Малые банки.

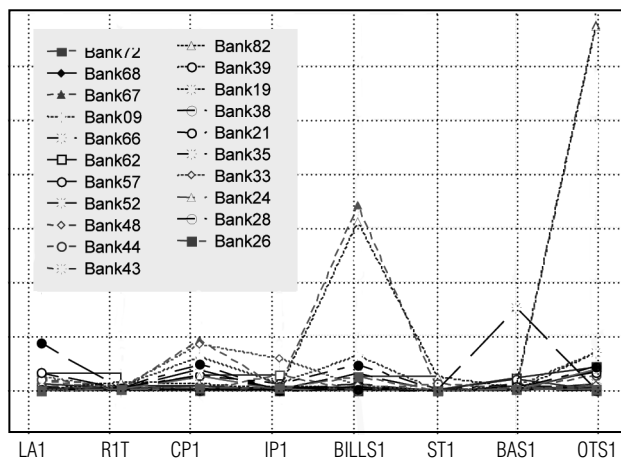


Рис. 14. «Мусорные» банки

долях разные типы активов и нет значимого преобладания кредитов юридическим лицам или векселей (рис. 14).

Можно заметить, что большинство кредитных организаций этого кластера имеют небольшие объёмы по всем видам активов, за исключением банков под номерами 82, 67 и 35. Это означает, что здесь собраны, в основном, мелкие банки, имеющие различную структуру портфеля активов. Об этом также говорит тот факт, что все 26 банков из этого кластера имеют минимальные резервы в ЦБ. 24 из 26 банков эксперты отнесли к классу так называемых «мусорных». Поскольку банкротство мелких кредитных организаций могло происходить по различным причинам, то и процент возврата активов может сильно отличаться для банков этого кластера. В то же время из-за небольшого совокупного объёма активов стоит ожидать достаточно высокого процента возврата. Стоит также отметить следующий факт – банки 67 и 82 можно было отнести в отдельный кластер, что обычно и происходило при применении стандартных процедур кластер-анализа.

Анализ финансового состояния ликвидируемых кредитных организаций с использованием метода паттернов показал то, что в подавляющем большинстве банков присутствовали признаки преднамеренного банкротства. Практически во всех банках были выявлены совершенные в преддверии банкротства сделки, наносящие существенный ущерб имущественным интересам банков и их кредиторов. Подчеркнем, что все кредитные организации были разбиты на однородные группы (кластеры) для того, чтобы строить модели возврата активов ликвидируемых банков уже внутри этих групп (банков со

схожими паттернами поведения). Было показано, что такая модель более адекватна, нежели модели, в которых не учитывается подобное разбиение.

5. Применение метода анализа паттернов в политологии

Метод анализа паттернов применялся для анализа результатов выборов в британских парламентских и финских муниципальных выборах [1, 9, 20]. Исследовались результаты парламентских выборов 1992, 1997 и 1999 гг. в округах Великобритании (отдельно по Англии, Шотландии, Уэльсу и Северной Ирландии) и семь муниципальных выборов в Финляндии в 1976-2000 гг. Был проведен как анализ паттернов поддержки политических партий по округам в статике, так и динамический анализ для оценки изменений такой поддержки с течением времени.

Объектами исследования были округа, в которых проводились выборы, а в качестве признаков для кластеризации были выбраны итоги голосования за различные партии.

Эмпирические данные о результатах выборов показывают широкий разброс итогов голосований по различным округам и в различные годы. Выделение и изучение типологий поведения населения округов при голосовании является очень важной задачей анализа политической конкуренции партий, особенно динамики изменений этой конкуренции и ее отражения в итогах голосований. В обоих случаях – и для Великобритании, и для Финляндии, и для парламентских, и для муниципальных выборов – оказалось, что распределение поддержки партий в среднем может быть описано несколькими похожими типами поведения.

Для выборов в Великобритании из 440 избирательных округов 199 округов показали стабильность предпочтений избирателей между партиями в течение исследуемого периода. Всего было выделено 13 паттернов поддержки партий (табл. 9), восемь из которых (паттерны 1-6,8,9) были найдены на данных о выборах 1992 г., три (7,10,12) и два (11,13) паттерна добавились по итогам изучения голосований 1997 и 2001 гг. При этом первые 6 паттернов поведения являются наиболее распространенными – более 90% округов имеет именно такие паттерны голосования. В выборах 1992 г. наиболее распространенным был паттерн 2 и, действительно, эти выборы выиграли консерваторы. А в выборах 1997 и 2001 гг. чаще всех остальных

Таблица 9.

Паттерны поддержки партий избирателями в Великобритании

Номер паттерна	Консервативная партия	Лейбористская партия	Либеральные демократы	Другие партии
1	23,18	60,23	12,49	4,1
2	48,82	32,88	15,76	2,53
3	38,08	47,61	11,8	2,52
4	50,97	19,05	26,78	3,19
5	41,42	12,98	42,42	3,17
6	14,22	46,29	36,40	3,07
7	31,40	38,75	24,77	5,08
8	35,24	28,08	33,74	3,0
9	12,72	36,25	48,50	2,50
10	0	65,30	0	34,70
11	38,0	12,60	15,60	33,80
12	37,5	0	0	62,5
13	19,1	22,1	0	58,8

наблюдался паттерн 1, и в этих парламентских выборах победу одержала Лейбористская партия во главе с Тони Блэром.

Построение «траекторий» изменения паттернов позволяет отследить динамику изменений паттернов поддержки партий (табл. 10). В таблице отражены самые распространенные траектории динамики голосования в Великобритании — результаты голосований по 440 округам (83% от общего числа).

Таблица 10.

Распределение английских округов по динамическим группам

Динамическая группа	Частота	% округов
x-x-x	99	37.62
x-x-y	21	13.97
y-x-x	251	47.45
x-y-x	20	3.78
x-y-z	38	7.18

Около 38% округов оказались стабильными в своих предпочтениях. Из них 14% округов оказались приверженцами лейбористов («1-1-1»), почти 10% поддерживали консерваторов («2-2-2»). Некоторые округа стабильно выбирали при голосовании паттерны 4 и 5.

Для групп, менявших паттерны, наиболее распространённой оказалась траектория вида «y-x-x», причем среди траекторий такого вида траектория «3-1-1» встречается чаще всего. Также популярными оказались траектории вида

«2-1-1» и «2-3-3». Поскольку и первый, и третий паттерн наибольшую поддержку дают как раз Лейбористской партии, это и предредило победу лейбористов на выборах 1997 года.

Финские выборы были проанализированы на данных о результатах 7 муниципальных выборов с 1976 г. по 2000 г., причем учитывались 6 главных политических партий: Социально-демократическая партия (the Social Democratic Party, сокращенно SDP), Центристская партия (the Center Party, KESK), Партия национальной коалиции (the National Coalition Party, KOK), Левый Альянс (the Left Alliance, VAS), Шведская партия (the Swedish People's Party, RKP) и партия Зеленых (the Green League, VIHR).

Всего было получено 87 паттернов электоральной поддержки партий, причем на подавляющем большинстве округов реализовалось 46 из них. Остальные паттерны являются редкими — каждый реализуется менее чем в 5-10 случаях. Десять паттернов, встречающихся более 100 раз в наблюдениях, приведены в табл. 11.

Из табл. 11 видно, что в наиболее распространенном паттерне партия KESK обладает абсолютным большинством голосов (53%), значимую поддержку имеет партия VAS (22%) и далеко позади партии SDP и KOK с 10% и 8% поддержки населения соответственно. Этот паттерн оказался характерным для сельских округов, особенно в северной и восточной части страны.

Во втором по распространенности паттерне данных также доминирует KESK (41%) с догоняющи-

Таблица 11.

**Паттерны поддержки политических партий
в Финляндии на данных 7 муниципальных выборов**

Номер паттерна	Количество наблюдений	Средняя поддержка партий внутри кластера, %						
		SDP	KESK	KOK	VAS	RKP	VIHR	Other
1	386	9.55	53.33	8.42	21.84	0	0.48	6.37
2	190	28.26	41.1	14.96	2.29	0	0.91	12.47
3	190	34.32	7.59	27.68	17.43	0.92	1.88	10.2
4	189	13.25	53.82	14.98	0.85	0	0.34	16.76
5	180	18.09	44.08	8.89	10.66	0	0.25	18.01
6	134	13.16	46.91	19.93	10.97	0	0.2	8.83
7	127	13.88	45.56	22.85	15.41	0	0.38	3.93
8	119	10.25	49.49	9.56	16.45	0	0.35	13.91
9	101	17.92	1.52	1.65	1.97	71.31	1.07	4.55
10	96	29.19	37.69	18.35	11.08	0	0.73	2.94

ми ее партиями SDP (29%) и KOK (15%). Округа, попавшие в третий распространенный паттерн, отдают предпочтение SDP (34%), а также партиям KOK (28%) и VAS (17%), KESK же имеет там очень низкую поддержку (8%). Такое распределение голосов характерно для городских муниципалитетов.

Очень сильная позиция партии KESK отражена в том факте, что в 8 из 10 самых распространенных паттернов она является победителем по абсолютному большинству голосов. В двух других паттернах победителями стали партии SDP и RKP. Как было упомянуто выше, первый из них относится к городским поселениям, а второй паттерн голосования типичен для множества мелких муниципалитетов на побережье или островах, где преобладает шведское население.

При анализе динамических групп было выявлено, что почти треть абсолютно стабильных групп (не менявших свои паттерны поведения в течение всей истории голосований) придерживается первого паттерна, обеспечивающего KESK положение доминирующей партии. Всего таких округов 19, они расположены большей частью в регионе Лапландии и Северной Остроботнии. Многие другие абсолютно стабильные паттерны также характеризуются доминирующей позицией KESK. Например, 6 муниципалитетов на протяжении всех 7 исследованных выборов голосовали согласно паттерну 4, то есть демонстрировали высокую поддержку KESK, затем KOK и SDP, и совсем немного VAS. Оказалось, что все такие муниципалитеты

расположены в сельской части юга Финляндии. Также в числе абсолютно стабильных муниципалитетов оказались муниципалитеты, выбравшие паттерн 9 с абсолютным большинством у партии RKP. Ими и в самом деле оказались муниципалитеты побережья с преобладанием шведского населения.

Так, анализ паттернов данных и динамический анализ паттернов помогает упростить анализ данных большой размерности, представить результаты в наглядной форме, отразить структуризацию политической жизни страны, скрывающуюся за огромным массивом разрозненной информации.

6. Применение метода анализа паттернов в оценке инновационного развития регионов

Успешное развитие регионов в долгосрочном плане неразрывно связано не только с экономическими факторами, но и напрямую определяется степенью развития в регионе образовательной и научной деятельности научных организаций. Любые инновации опираются на солидный фундамент научной деятельности, которая невозможна без соответствующих кадров — не только научных сотрудников, но и студентов всех ступеней образования, а также квалифицированных рабочих и служащих. Все эти факторы сложным образом взаимосвязаны. Для анализа структуры этих связей был использован метод анализа паттернов [19].

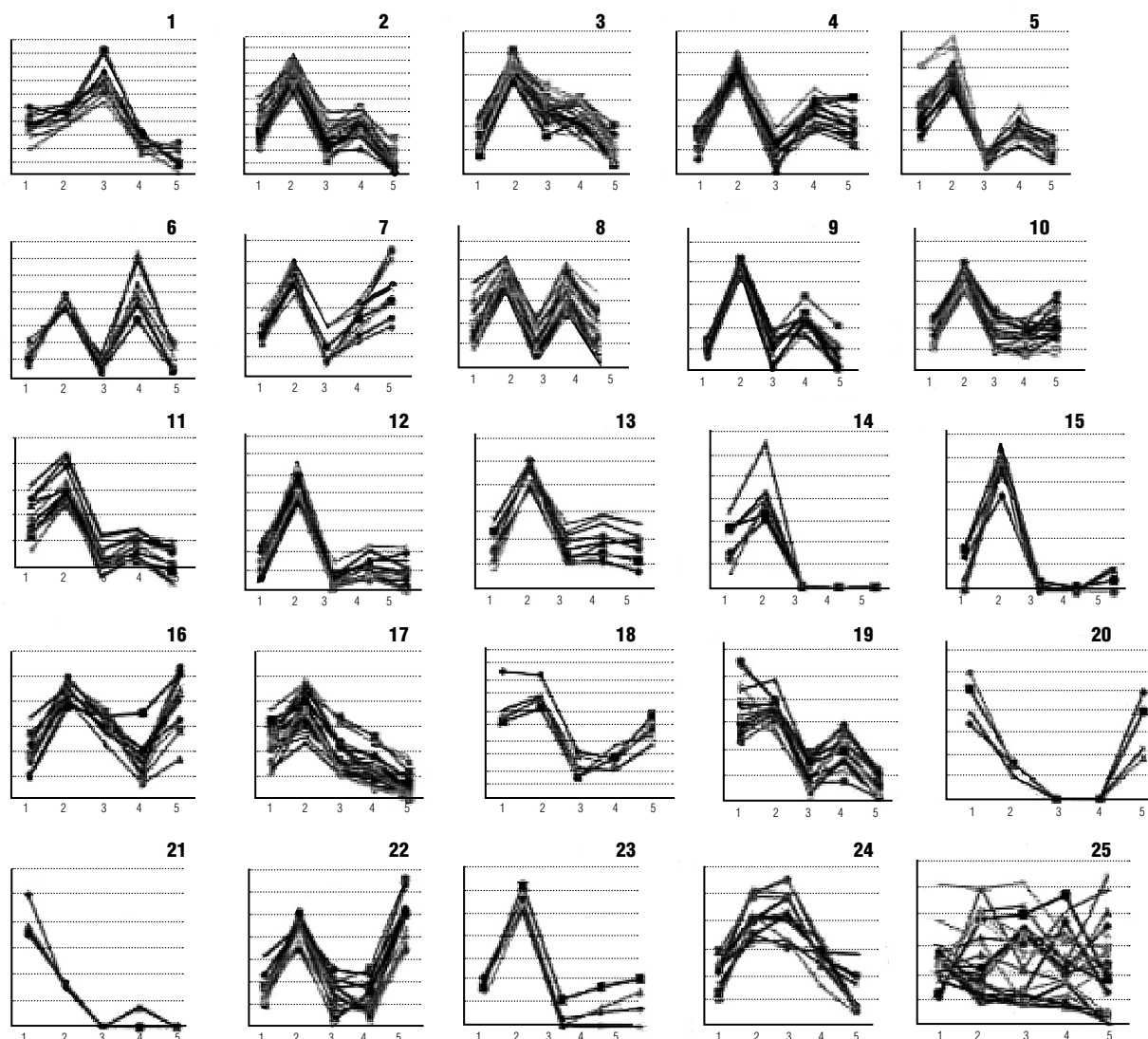


Рис. 15. Паттерны инновационного развития регионов

В этой задаче объектами анализа были регионы, цель заключалась в построении системы индикаторов, адекватно отражающих положение региона в части развития в нем науки, инновационной активности организаций и образования, и анализ полученных кластеров. Базовая система индикаторов создавалась на основе рейтинга инновационного развития регионов, разработанного коллективом ИСИЭЗ НИУ ВШЭ [23]. Были выделены следующие блоки показателей

1. Социально-экономические условия;
2. Образовательный потенциал;
3. Результативность исследований и разработок;
4. Потенциал инновационной деятельности;
5. Результативность инновационной деятельности.

Все показатели в базовой системе были нормированы в шкале от 0 до 1 — они либо приведены к

численности экономически активного населения, либо представляют собой процентное соотношение. Авторы намеренно отказались от абсолютных показателей, имея целью исследовать внутреннюю структуру взаимной зависимости науки, образования и инноваций.

Логично считать, что регионы, имеющие приблизительно одинаковый паттерн в базовой системе показателей, схожи в выбранной модели развития науки, образования и инноваций. Стоит отметить, что не ставилась задача построения рейтинга или «выставления оценок» регионам, интересовало именно положение дел как оно есть, исследование скрытых взаимосвязей между наукой, образованием и инновационной деятельностью в разрезе регионов, нахождение «близких» по структуре и соотношениям между указанными областями регионов.

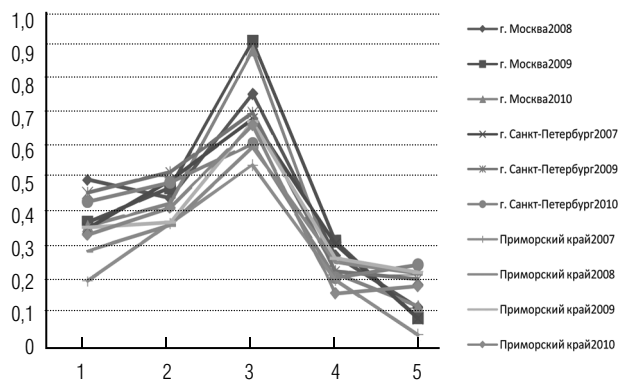


Рис. 16. Первый паттерн инновационного развития

Всего было получено 24 паттерна инновационно-го развития регионов (рис. 15), где под номером 25 объединены все уникальные объекты, не попавшие в 24 паттерна. Далее приведено описание первых 3 паттернов.

Первый паттерн (рис. 16) характерен для 11 объектов, представляющих город Москву и Приморский край за все 4 года с 2007 по 2010 гг., а также город Санкт-Петербург в 2007 г., 2009 г., 2010 г. Этот паттерн характеризуется умеренными (0,3-0,5) значениями первых двух показателей (Блок 1 «Социально-экономические условия» и Блок 2 «Образовательный потенциал»), чрезвычайно высоким (0,5-0,9) показателем Блока 3 «Результативность исследований и разработок», а также очень низкими (менее 0,2) показателями Блоков 4 и 5, отвечающими за потенциал и результативность инновационной деятельности.

Второму паттерну (рис. 17) соответствуют 19 объектов. Основные характеристики этого паттерна: низкие (0,1-0,3) показатели блоков 1, 3, 4, 5 (социально-экономические условия, результативность науки и потенциал и результативность инновационной деятельности) и средние (0,4-0,5) показатели образовательного потенциала регионов.

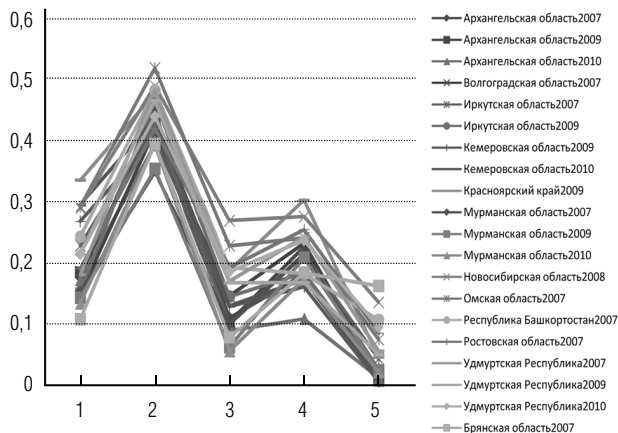


Рис. 17. Второй паттерн инновационного развития

Третий паттерн (рис. 18) объединяет 14 объектов. Он отличается от второго паттерна только двумя показателями: «результативность исследований и разработок» в нём ниже, чем во втором, а «результативность инновационной деятельности» наоборот, выше.

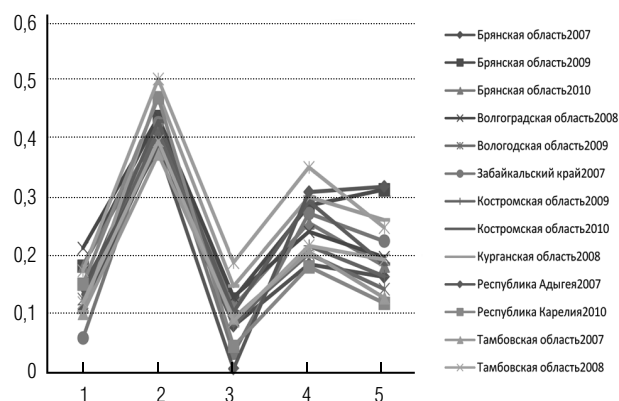


Рис. 18. Третий паттерн инновационного развития

7. Заключение

В работе рассмотрено понятие паттерна как устойчивой комбинации существенных переменных, связанной с характерной моделью поведения социально-экономических объектов. Предложено выделять типы устойчивости функционирования объектов в зависимости от частоты смены паттерна. Такая типологизация является полезным инструментом управления объектами. Кроме того, она позволяет выделить группы риска их стратегического развития. Сделаны обзоры литературы по использованию понятия паттерн в науке и культуре, методам факторного анализа, анализа многомерных систем, а также работ по анализу конкретных социально-экономических данных, выполненных под руководством Ф.Т. Алескерова, с использованием понятия паттерн. Эти примеры наглядно демонстрируют перспективность использования понятия паттерна в анализе данных, так как оно является эмпирическим аналогом логического понятия «тип», широко используемого на первых этапах изучения сложных феноменов и процессов.

В связи с этим актуальной задачей нам представляется формирование таких математических моделей и методов, которые бы выявляли паттерны в данных непосредственно, без предварительного выявления кластеров. Необходимость проведения соответствующих теоретических, алгоритмических и экспериментальных разработок давно назрела. ■

Литература

1. Aleskerov F. Patterns of party competition in British and Russian general and Finnish municipal elections // Proceedings of the 4-th International Conference on Operational Research. – Moscow, September 2004. – P. 11-13.
2. Aleskerov F., Alper C.E. Inflation, money, and output growths: Some observations // Bogazici University Research Paper. – 1996. – #SBE 96-06.
3. Aleskerov F., Alper C.E. A clustering approach to some monetary facts: a long-run analysis of cross-country data // The Japanese Economic Review. – 2000. – Vol. 51, No. 4. – P. 555-567.
4. Aleskerov F., Ersel H., Yolalan R. Personnel allocation among bank branches using a two-stage multicriterial approach // European Journal of Operational Research. – 2003. – Vol. 148/1. – P. 116-125.
5. Aleskerov F., Ersel H., Yolalan R. Clustering Turkish commercial banks according to structural similarities // Yapi Kredi Discussion Paper Series. – 1997. – No. 97-02. – Istanbul, Turkey.
6. Aleskerov F., Ersel H., Yolalan R. Multicriterial methods for evaluating the bank branches performance // Yapi Kredi Discussion Paper Series. – 1997. – No. 97-01. – Istanbul, Turkey.
7. Aleskerov F., Ersel H., Yolalan R. Multicriterial ranking approach for evaluating bank branch performance // International Journal of Information Technology and Decision Making. – 2004. – Vol. 3, No. 2. – P. 321-335.
8. Aleskerov F., Ersel H., Mercan M. Structural dissimilarity in Turkish banks 1988-1999 // Bogazici Journal Review of Social, Economic and Administrative Studies. – 2001. – Vol. 15, No. 1. – P. 57-69.
9. Aleskerov F., Nurmi H. A method for finding patterns of party support and electoral change: An analysis of British general and Finnish municipal elections // Mathematical and Computer Modelling. – 2008. – P. 1225-1253.
10. Caruth D.L. Surveying bank use of work measurement. – Burroughs Clearing House, 1970.
11. Caruth D.L. Work measurement in commercial banks. – Istanbul: Yapi Kredi Bank Publication. – 1985.
12. Few S. Multivariate Analysis Using Parallel Coordinates, 2006. Источник в Интернет: http://www.perceptualedge.com/articles/b-eye/parallel_coordinates.pdf.
13. Inselberg A. Parallel coordinates: Visual multidimensional geometry and its applications. – Springer, 2009.
14. Maynard H.B. (Ed.) Industrial engineering handbook. – NY: Mc-Graw Hill, 1971.
15. Алексашин П.Г., Алескеров Ф.Т., Белоусова В.Ю., Попова Е.С., Солодков В.М. Динамический анализ бизнес-моделей российских банков в период 2006-2009 гг. // Препринт WP7/2012/03. – М.: Издательский дом НИУ ВШЭ, 2012.
16. Алескеров Ф.Т., Белоусова В.Ю., Егорова Л.В., Миркин Б.Г. Анализ паттернов в статике и динамике, часть 1: Обзор литературы и уточнение понятия // Бизнес-информатика. – 2013. – № 3 (25). – С. 20.
17. Алескеров Ф.Т., Белоусова В.Ю., Кнурова А.А., Солодков В.М. Стереотипы поведения российских коммерческих банков в период финансового кризиса // Проблемы развития экономики и общества: в 3 кн. / Отв. ред. Е. Г. Ясин. – М.: Издательский дом ГУ-ВШЭ. – 2011. – Кн. 1. – С. 583-593.
18. Алескеров Ф.Т., Белоусова В.Ю., Сердюк М.Ю., Солодков В.М. Стереотипы поведения российских банков // Банковское дело. – 2008. – № 7. – С. 44-50.
19. Алескеров Ф.Т., Гохберг Л.М., Егорова Л.Г., Мячин А.Л., Сагиева Г.С. Анализ данных науки, образования и инновационной деятельности с использованием методов анализа паттернов // Препринт WP7/2012/07, Нац. исслед. ун-т «Высшая школа экономики». – М.: Издательский дом НИУ ВШЭ, 2012.
20. Алескеров Ф.Т., Нурми Х. Паттерны конкуренции партий в британских парламентских и финских муниципальных выборах // Препринт ГУ-ВШЭ, WP7/2003/07. – М., 2003.
21. 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.
22. Алескеров Ф.Т., Шерман И.В., Энтов Р.М. Анализ эффективности конкурсного управления при банкротстве банков // Банковское дело. – 2008. – № 12. – С. 70-76.
23. Гохберг Л.М. (ред.) Рейтинг инновационного развития субъектов Российской Федерации: аналитический доклад. – М.: НИУ ВШЭ, 2012.

ОБ ОДНОМ МЕТОДЕ КЛАССИФИКАЦИИ ВИЗУАЛЬНЫХ МОДЕЛЕЙ

Т.А. Гаврилова,

доктор технических наук, профессор, заведующий кафедрой
информационных технологий в менеджменте, Высшая Школа Менеджмента,
Санкт-Петербургский государственный университет

Д.В. Кудрявцев,

кандидат технических наук, руководитель направления управления знаниями
Бизнес Инжиниринг Групп (ООО «КВФ «БИГ-СПб»), доцент
Санкт-Петербургского государственного политехнического университета

И.А. Лещева,

старший преподаватель Высшей Школы Менеджмента,
Санкт-Петербургский государственный университет

Я.Ю. Павлов,

младший научный сотрудник Высшей Школы Менеджмента,
Санкт-Петербургский государственный университет

Адрес: Санкт-Петербург, Волховский пер., 3.

E-mail: gavriloa@gsom.pu.ru, dmitry.ku@gmail.com, leshcheva@gsom.pu.ru,
yaroslav.pavlov@gmail.com

Важность визуальных представлений для лучшего усвоения материала в системах традиционного и дистанционного обучения очевидна. Такие графические модели облегчают понимание и коммуникации «преподаватель-студент». Хотя количество визуальных языков моделирования велико, выбор среди них зачастую случаен и необоснован. В статье описаны результаты исследования по созданию новой таксономии для выбора визуального языка в зависимости от типа представляемых знаний (аспектов предметной области). В основу предлагаемого подхода положена семантическая классификация визуальных языков и типов знаний. Одним из наиболее значимых результатов применения предложенного подхода являются практические рекомендации преподавателям по выбору подходящей визуальной нотации при преподавании общественно-политических и социально-экономических дисциплин. Также предложенный обзор и классификация могут быть полезны практикующим бизнес-аналитикам для наглядного структурирования информации при моделировании компаний.

Ключевые слова: бизнес-моделирование, визуализация, таксономия, знания, бизнес-образование.

Введение

Использование графики и визуализации в преподавании является традиционным и очевидным дидактическим приёмом. Некоторые науки принципиально основываются на визуальных представлениях (геометрия, география, частично физика и химия). Активно используются графические представления в современных науках об информации [1–4]. Однако в гуманитарных науках они используются пока достаточно редко. Исключением является менеджмент, где последние десятилетия активно внедряются языки моделирования бизнес-процессов, табличные представления и различные диаграммы. В менеджменте как междисциплинарной науке разные виды визуализации позволяют рассмотреть предметную область с различных сторон и точек зрения, используя на первый взгляд противоречивые, но, на самом деле, взаимодополняющие парадигмы.

Неконтролируемый рост информации вызвал появление новых направлений в науках по изучению процессов восприятия и обработки информации, связанных с визуализацией (инфографика, когнитивная графика, виртуальная реальность, и т.д.) [5]. Основная цель новых видов графической подачи информации — сделать сложные информационные массивы более простыми для восприятия.

Визуализация знаний играет важную роль при получении и передаче знаний [6–8]. Визуальные языки хорошо дополняют технологии онтологического (или концептуального) моделирования, делая содержимое онтологий понятным и наглядным не только для экспертов и аналитиков, но также и для новичков. Всё чаще визуальные языки моделирования используются в менеджменте [9], управлении знаниями [10], научных исследованиях [11], учебном процессе [12].

Как отмечает ряд профессионалов в области дистанционного образования, визуализация играет значительную роль для улучшения усвоения материала [13, 14]. Визуализация материала не только упрощает его восприятие, но и повышает вовлеченность учеников, чего сложно достигнуть при отсутствии прямого контакта с преподавателем.

На сегодняшний день существует достаточно большое количество языков моделирования. Однако палитра используемых на данный момент на практике языков моделирования ограничена, а выбор конкретного языка случаен и зачастую ошибочен. Пожалуй, в первую очередь это связано с от-

сутствием обобщающих принципов использования языков и их классификации. Существующая литература, а также программные продукты, предназначенные для создания визуальных моделей, предлагают неоднородные и достаточно запутанные классификации языков. Как исключение, можно упомянуть работы Бабкина и соавторов, в которых они проводят сравнительный анализ языков бизнес-моделирования [15, 16]. Однако набор рассматриваемых языков в данных работах достаточно ограничен.

Данная ситуация ограничивает распространение языков моделирования вне узкого круга специалистов (таких как, например, бизнес-аналитики) и делает задачу выбора конкретного языка сложной для широкого круга работников, в частности, преподавателей.

Ранее мы уже отметили, что менеджмент во многом является исключением из ряда гуманитарных наук, так как в нем используется широкий спектр визуальных моделей, многие из которых привнесены из технических и естественнонаучных дисциплин.

Целью данного исследования был анализ современных визуальных языков, и **формирование новой классификации** визуальных моделей, основанной на таксономии языков моделирования. Такая классификация **впервые устанавливает взаимосвязь между типом знаний, которые нужно представить, и видом модели.**

При анализе любой сложной предметной области требуется взгляд с разных точек зрения — необходимо представить ЧТО-, КАК-, ЗАЧЕМ-, ПОЧЕМУ-, КТО-, ГДЕ- и КОГДА-знания [17]. В данной работе мы ставим задачу создания новой классификации современных языков моделирования, которая поможет делать обоснованный выбор языка для наиболее адекватного представления знаний в бизнесе. Данная задача актуальна как для преподавателей, так и для широкого круга профессионалов из областей моделирования предприятия (enterprise modeling), аналитиков бизнес-архитектуры (business architecture), дистанционного обучения (distance learning), а также для широкого круга специалистов бизнес-информатики.

1. Классификация и метод выбора визуальных языков моделирования

Обзор современных языков моделирования и первые попытки их классификации показывают,

что на сегодняшний день не существует общепринятой в научном сообществе таксономии, так как для более чем ста языков моделирования предложено не менее 15 видов классификаций [9, 18]. Существует ряд профессиональных программных инструментов, например, Visio или SmartDraw, каждый из которых предлагает несколько десятков моделей. Классификации визуальных языков в таких инструментах недостаточно проработаны — используют различные основания деления на одном уровне и содержимое классов часто пересекается.

Мы предлагаем новую классификацию, основанную на неформальном описании типов знаний [19]. Такой подход предоставляет возможность подобрать визуальный язык моделирования для конкретной управленческой задачи. Для неформального описания типов знаний предлагается использовать вопросы проверки компетентности (competency questions) [20].

Рис. 1 иллюстрирует новую классификацию вопросов по типу знаний. Список вопросов может быть существенно расширен. Рис. 2 уточняет какие понятия и типы связей между ними относятся к тому или иному типу знаний. Классификация на рис. 3 представляет расширение классификации из работы [19] и сопоставляет наиболее распространённые в бизнес-приложениях визуализации

с предложенной типологией знаний. Данная классификация была получена путём семантического анализа метамodelей языков моделирования с использованием вопросов проверки компетентности и шаблонов проектирования онтологий [21].

Для успешного выбора метода визуализации необходимо ответить на несколько вопросов [7]:

- ♦ Какой тип знаний необходимо визуализировать?
- ♦ Какова цель визуализации знаний и для кого необходимо визуализировать знаний?
- ♦ В каких условиях это будет использоваться: место размещения, канал передачи, формат?

Данная статья устанавливает связь между ответом на первый вопрос и методом визуализации. Любой преподаватель или специалист (не профессиональный бизнес аналитик) ищет визуальный язык моделирования, обладая неформализованным пониманием своих потребностей и желая доходчиво отобразить свои знания. В таком случае можно предложить практический рецепт:

А). Из рис. 1 выбрать, какой вопрос наиболее точно отражает потребность и определить тип знаний. Уточнить выбор типа знаний можно с помощью рис. 2, оценив какие взаимосвязи необходимо отразить.

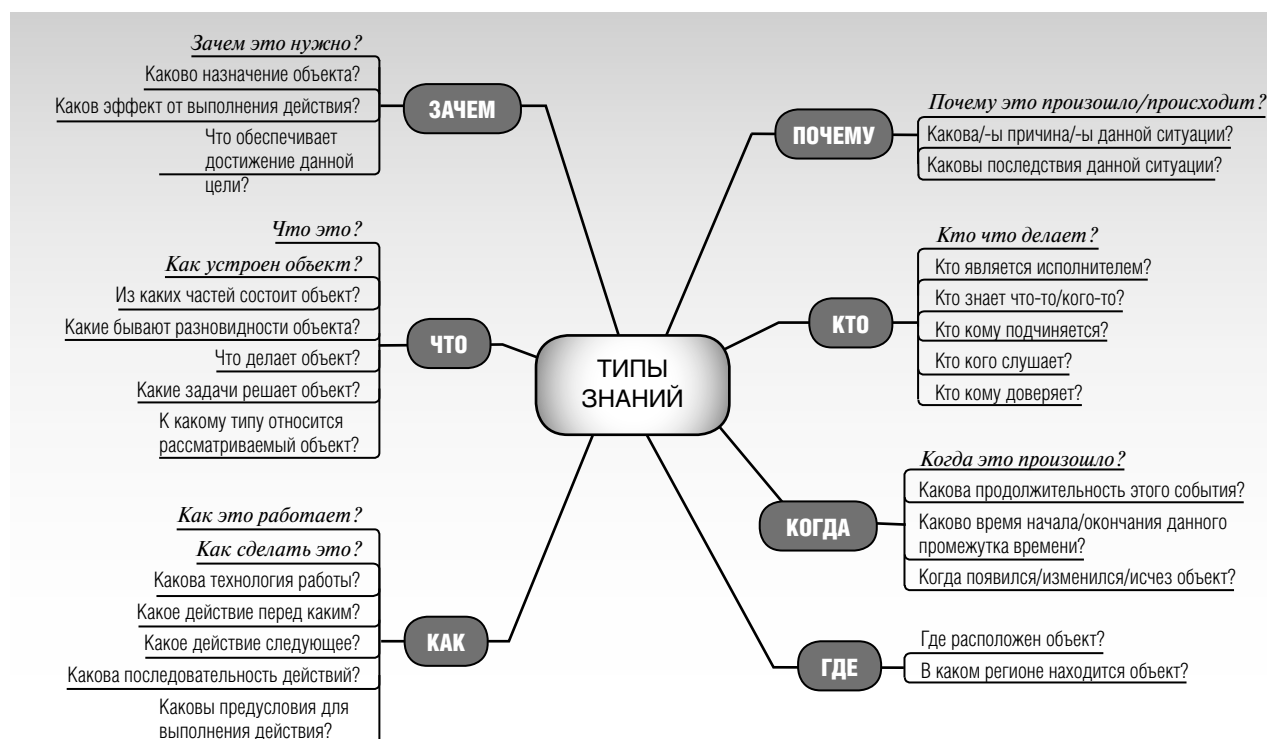


Рис. 1. Описание типов знаний с помощью вопросов проверки компетентности

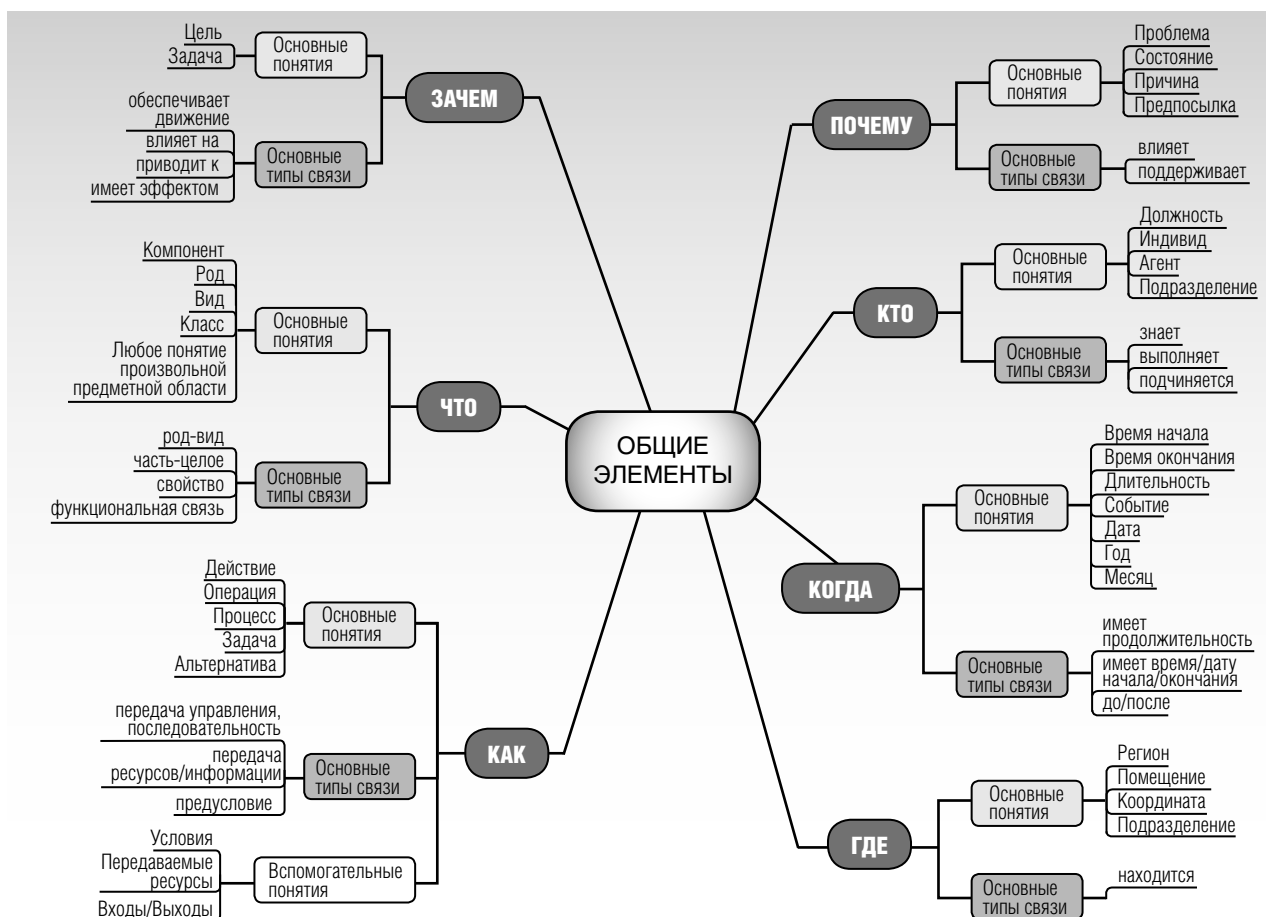


Рис. 2. Общие элементы для языков моделирования и типов знаний

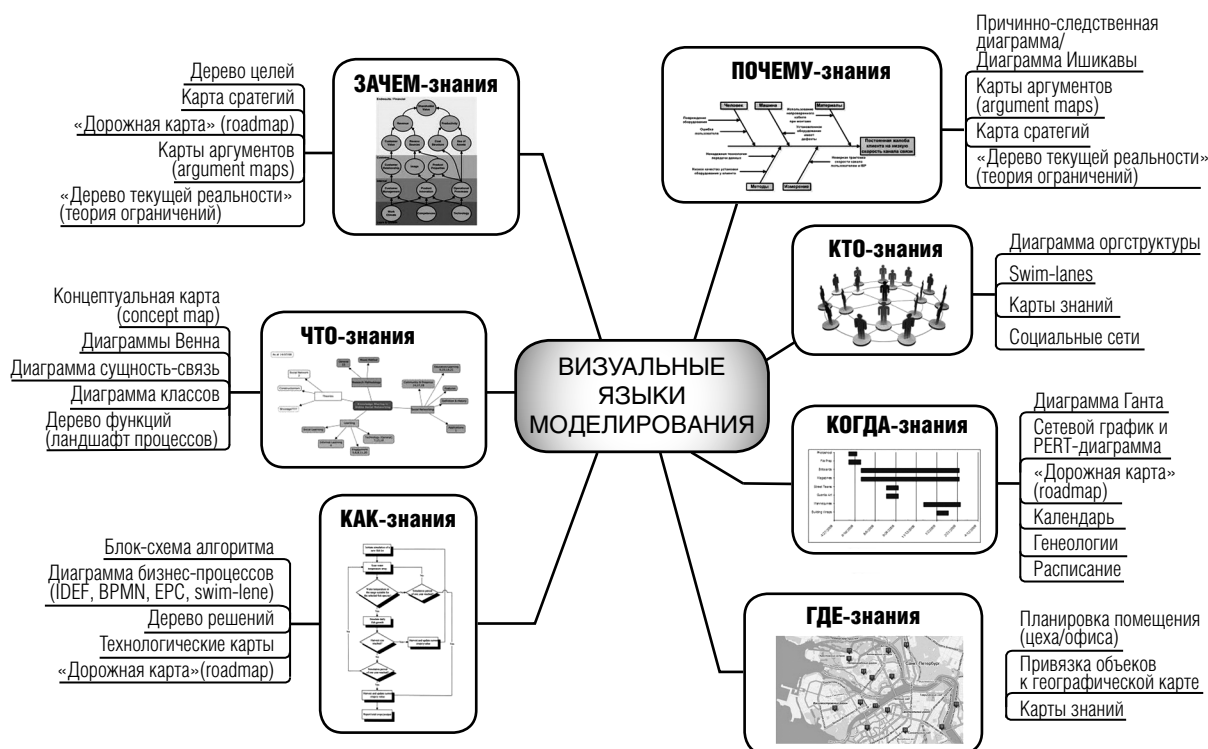


Рис. 3. Распределение визуальных языков моделирования по типам знаний

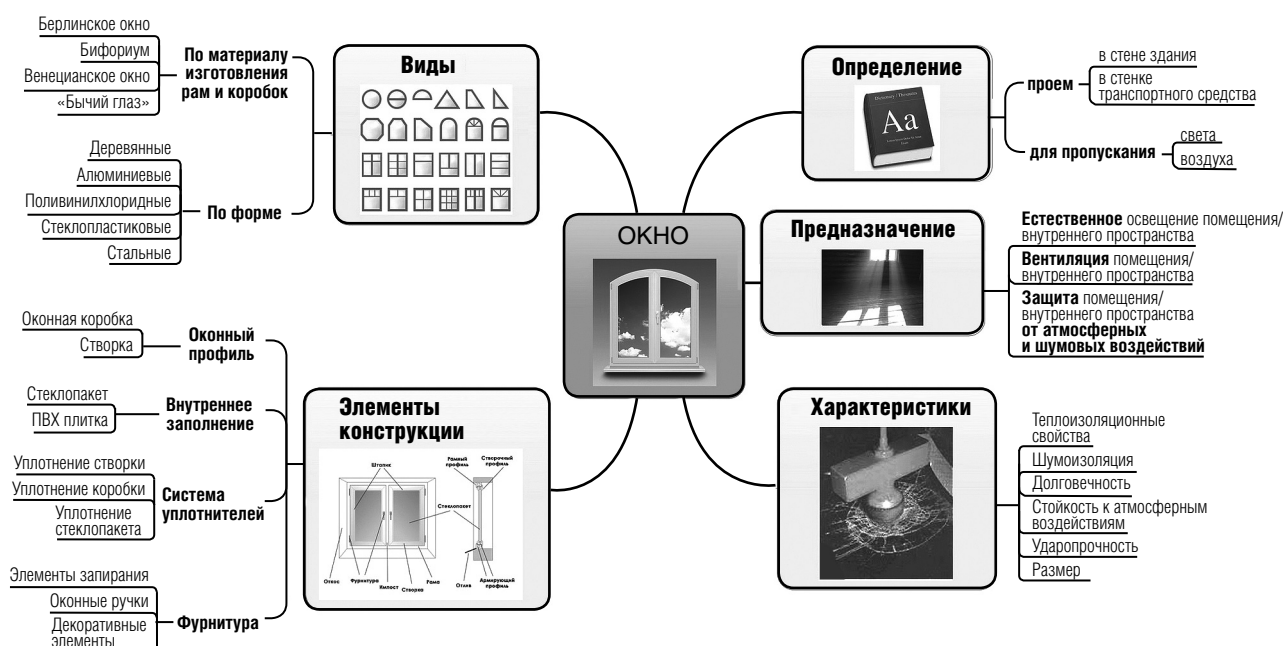


Рис. 4. Пример интеллект-карты

Б). Определить, какие визуальные языки можно использовать для представления требуемого типа знаний по рис. 3.

Далее следует анализ наиболее распространённых в бизнес-образовании визуальных дидактических моделей согласно предложенной классификации.

2. Обзор основных визуальных дидактических моделей

ЧТО-знания

В обучении ЧТО-знания являются первоочередными. Классический способ преподавания предполагает, что сначала даётся определение изучаемого понятия, затем изучаются его свойства и лишь затем — взаимодействие с другими объектами или понятиями. Именно поэтому для визуализации ЧТО-знаний существует множество языков. Рассмотрим основные из них.

Интеллект-карты (mind maps) — один из наиболее простых способов отображения понятийных структур. Хотя автором идеи интеллект-карт считается Тони Бьюзен [22], в России похожий метод использовался в педагогике уже в 20-е годы [23]. Интеллект-карты — это иерархические диаграммы, используемые для представления идей, проектов, заданий, которые связаны с центральным ключевым понятием и организованы радиально вокруг него. Рис. 14 иллюстрируют понятие интеллект-карты (и-карты).

И-карты применяются, с одной стороны, с целью генерирования, структурирования и классификации идей, с другой стороны, с целью облегчения процесса обучения [24], процесса принятия решения [25 — 27], а также при составлении различных документов [8].

Их можно использовать как для объяснения, так и для проверки усвоенного материала. В этом случае структура и-карты служит критерием понимания изучаемого предмета. Среди и-карт, созданных студентами, встречаются диаграммы, демонстрирующие частичное (или полное) непонимание дисциплины, что является обратной связью для преподавателя.

Благодаря упрощённому изображению идей в графическом виде и-карты используют для активизации «мозгового штурма» [28]. В образовательном процессе данная модель может быть полезна, например, при выполнении групповых проектов, анализе кейсов и т.д. [12].

На рис. 4 представлен пример и-карты для понятия «Окно» (предположим, это продукция компании), описывающий виды окон, их элементы и характеристики.

Поскольку и-карты позволяют представлять любые древовидные структуры, они могут использоваться не только для работы с ЧТО-знанием, но и для знаний других типов.

Если интеллект-карты демонстрируют понятия и древовидную структуру определённых фрагментов

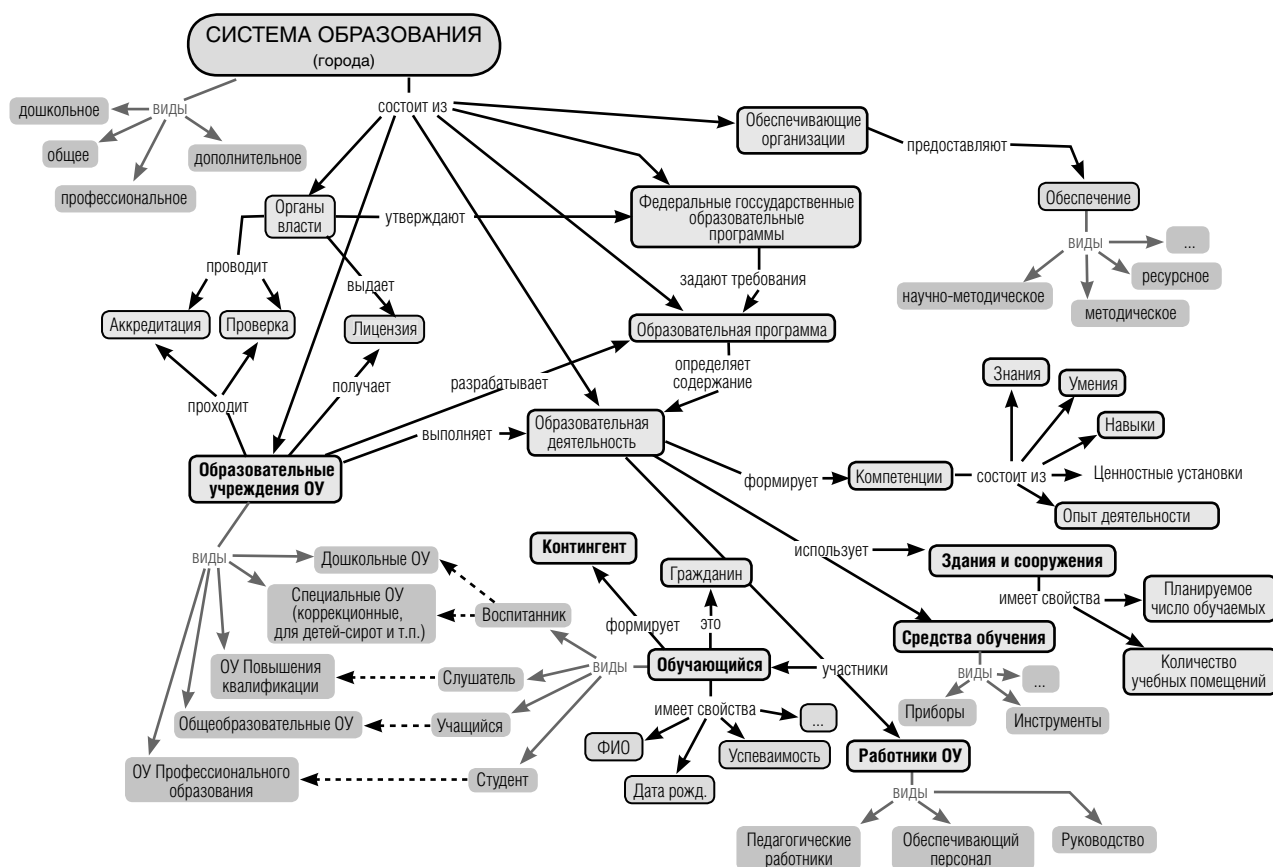


Рис. 5. Пример концептуальной карты

знаний, то **концептуальные карты** (concept maps) позволяют глубже раскрыть рассматриваемый вопрос, благодаря поддержке сетевых структур и включению в модель отношений между понятиями.

Концептуальная карта представляется в виде графа, узлы которого отображают понятия (объекты или концепты), а направленные поименованные дуги, соединяющие эти узлы, — отношения (связи). Могут использоваться связи различного типа, например, «является», «имеет свойство» и т. п. На рис. 5 представлена концептуальная карта для системы образования города, разработанная в рамках Системного проекта г. Москвы [29].

Впервые концептуальные карты были предложены Новаком в начале 70-х гг. прошлого века при изучении детского мышления. Это исследование использовало идеи Д. Асубеля [30] о формировании понятийного мышления.

В процессе создания концептуальной карты преподаватель анализирует структуру отношений предметной области, что помогает ему самому глубже понять её природу. Визуальные спецификации в форме концептуальных карт широко при-

меняются в обучающих системах (e-learning) и в традиционном преподавании [31], а также при моделировании корпоративных баз знаний [1].

Поскольку концептуальные карты, как и интеллект-карты, не накладывают ограничений на семантику элементов диаграммы, то они также используются не только для работы с ЧТО-знанием, но и для знаний других типов.

ПОЧЕМУ-знания

Причины, которые привели к текущей ситуации, относятся к ПОЧЕМУ-знаниям. Самый распространённый способ их визуализации — **причинно-следственные диаграммы Ишикавы (fishbone)** [32]. В области контроля качества диаграмма Ишикавы — это диаграмма, которая показывает отношение между показателем качества и воздействующими на него факторами. В частности, рис. 6 заимствован из стандарта по качеству Северной железной дороги — филиала ОАО «РЖД» — «Методика восьми шагов решения проблем качества технических средств железнодорожного транспорта».

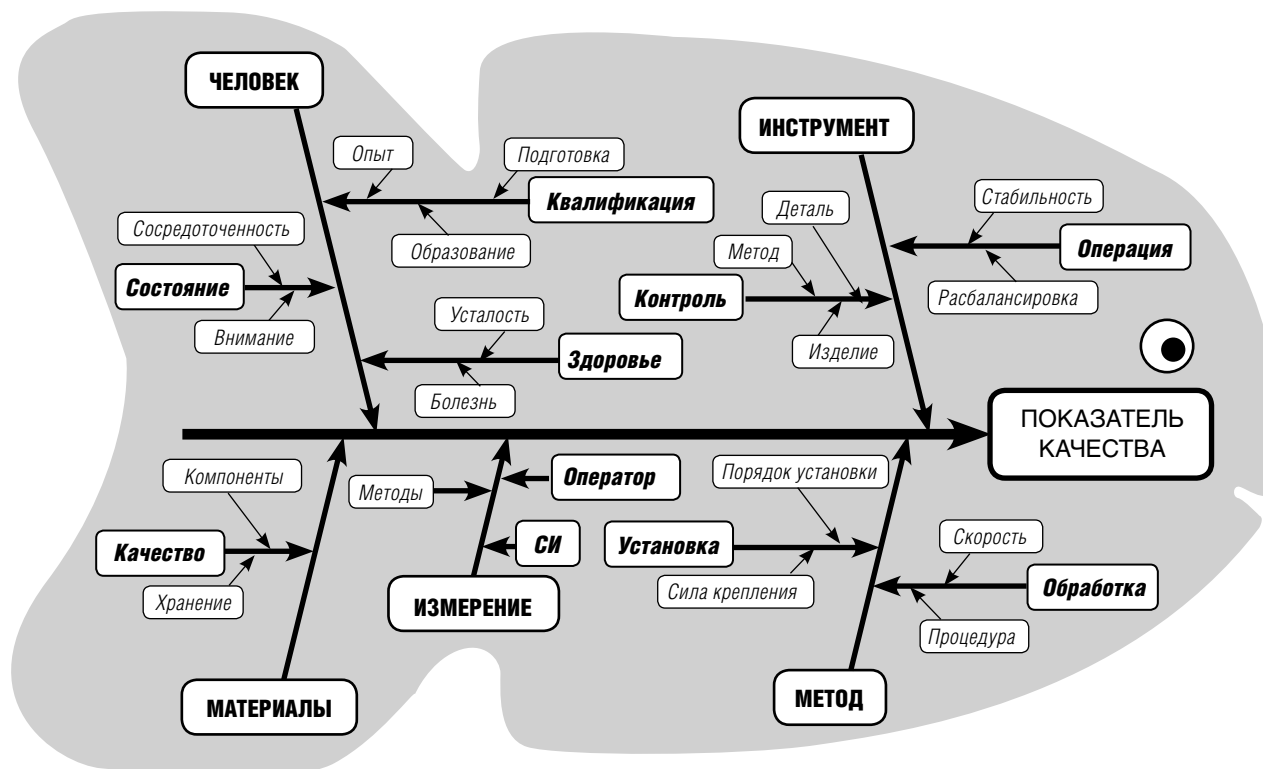


Рис. 6. Пример диаграммы Ишикавы «Причины хорошего или плохого качества результата»

Кроме того, данный тип диаграммы может служить для визуального описания истории компании (продукта, технологии и т.п.), разбивая её ход на определённые этапы и отмечая некоторые ключевые события.

Аналогичные вопросы приводят к построению **дерева текущей реальности**, разработанного в рамках теории ограничений Э.М. Голдратта [33]. Цель дерева — выявить основную причину фактора, который препятствует достижению цели. При по-



Рис. 7. Пример карты аргументов

строении дерева текущей реальности обычно начинают с наблюдения нежелательных эффектов (undesirable effects, UDE). Далее, в обратном порядке сопоставляются причины и следствия, пока не обнаруживаются первопричины нежелательных эффектов, определённых в начале.

Кроме того, можно выделить также и **причинно-следственную цепь** (causal chain), которая помимо ответа на вопрос «почему это произошло», позволяет описать предполагаемые последствия. Если ранее описанные модели направлены в первую очередь на анализ текущей ситуации, то причинно-следственная цепь удобна для анализа редких событий (rare events) [34]. В процессе обучения данная модель может эффективно использоваться как раз для анализа редких событий, таких как экономические кризисы, внедрение инноваций и т.п.

Ещё одним способом объяснения является **карта аргументов**, заимствованная из неформальной логики (см. рис. 7).

КАК-знания

Рассмотрим языки, позволяющие визуализировать ответы на вопросы «Как достичь цели?» или «Как решить проблему?».

Простым способом является графическое описание алгоритма решения задачи в виде **блок-схемы**. Это же представление можно использовать для визуализации процессов (рис. 8).

Для моделирования бизнес-процессов разработано множество нотаций (например, IDEF, BPMN, EPC), заимствующих идею блок-схем, но их обсуждение выходит за рамки данной статьи.

Иногда при моделировании процесса бывает полезно разбить состояния деятельности на группы, каждая из которых представляет собой элемент организации или определённый ресурс. Визуально каждая группа отделяется от соседних чертой (см. рис. 9), это похоже на плавательные дорожки в

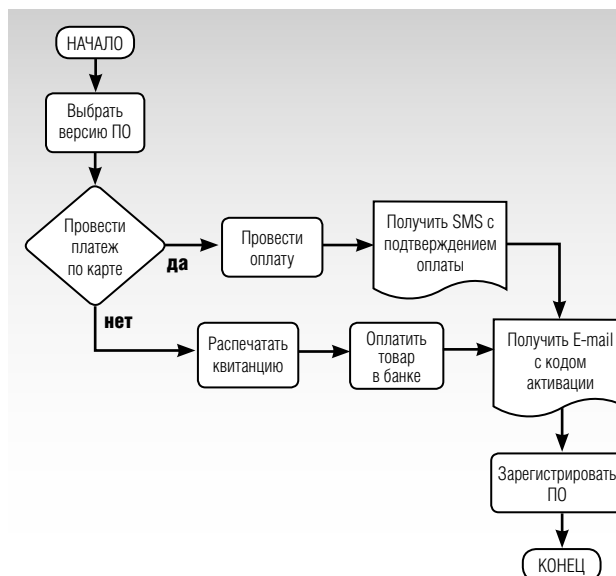


Рис. 8. Схема процесса «Приобретение ПО через web-сайт»

бассейне, поэтому такие диаграммы называют **swim lanes** (плавательные дорожки).

Swimlane является визуальным элементом, используемых в процессе технологических схем, которые описывают, что или кто работает на определённой части процесса.

Теория ограничений, которая уже упоминалась в этом обзоре, даёт ответ на вопрос «как» в виде **дерева перехода** и **плана перехода**, иллюстрирующих каким образом можно осуществить переход из текущего состояния в желаемое, описанное деревом будущей реальности.

Дерево целей — это структурированная, построенная по иерархическому принципу совокупность целей, в которой выделены генеральная цель («вершина дерева»); подчинённые ей подцели последующих уровней («ветви дерева») [36]. Концепция «дерева целей» впервые была предложена Ч. Чёрчменом и Р. Акоффом в 1957 году [37]. Важность подцелей по отношению друг к другу показывают коэффициенты значимости, которые отражают их

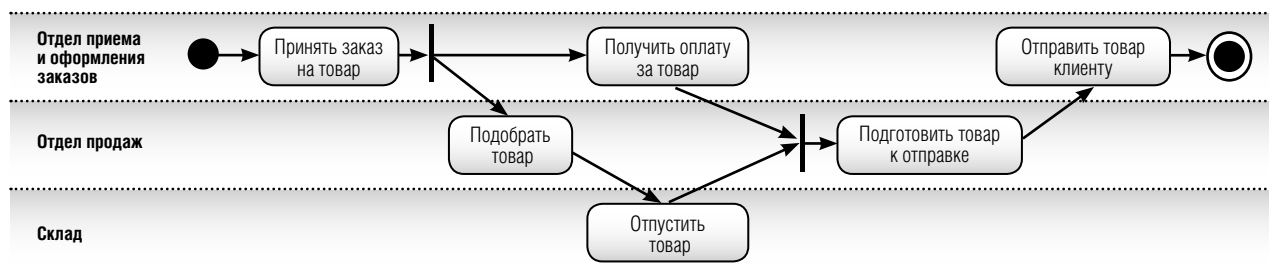


Рис. 9. Схема бизнес-процесса, на которой операции разнесены по отделам [35]

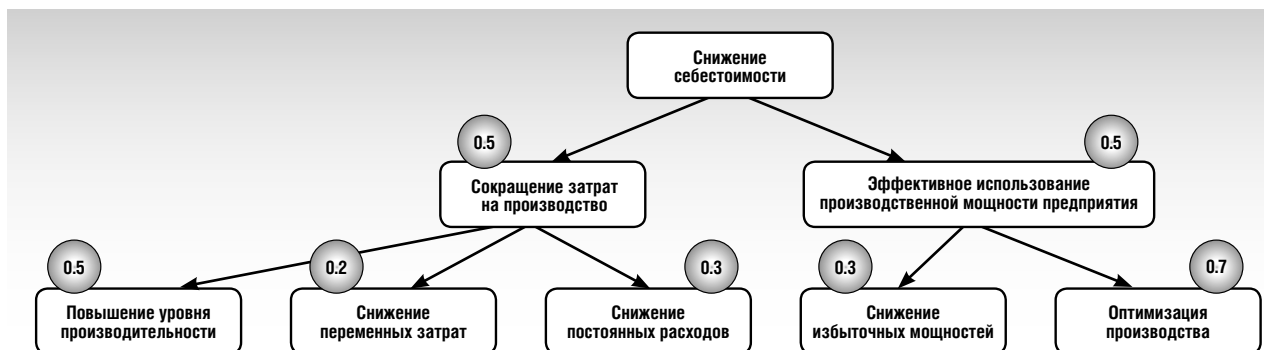


Рис. 10. Дерево целей: снижение себестоимости

вклад в достижение конечной цели (рис. 10).

Дерево решений рассматривает не единственный способ решения некоторой задачи, а помогает осуществить выбор из совокупности альтернативных способов (рис. 11).

ЗАЧЕМ-знания

В результате стратегического анализа получают ЗАЧЕМ-знания о предметной области. Для визуализации ЗАЧЕМ-знаний могут применяться методы визуализации КАК-знаний. В частности, дерево целей (рис. 10) будет давать ответ не столько на вопрос, как достичь основной цели, сколько показывать, зачем необходимы «подчинённые цели». В образовательном процессе эта модель полезна для описания программы курса, давая чёткое осознание необходимости частей материала для освоения всего курса.

Дерево решений (рис. 11) также помогает ответить на вопрос «зачем». Давая возможность сравнить между собой альтернативные решения проблемы и их результаты, дерево показывает, зачем делать именно так.

Карты стратегий [38] — инструмент стратегического управления, он позволяет устанавливать причинно-следственные связи между целями и структурировать цели по группам («перспективам»).

Кроме того, из теории ограничений можно позаимствовать диаграмму «дерево будущей реальности», показывающую иерархию прогнозируемых состояний системы и причинно-следственные связи между ними. Эта модель демонстрирует не только целевое состояние, но и пути его достижения. По сути, эта диаграмма также может использоваться и для представления КАК- и ЗАЧЕМ-знаний.

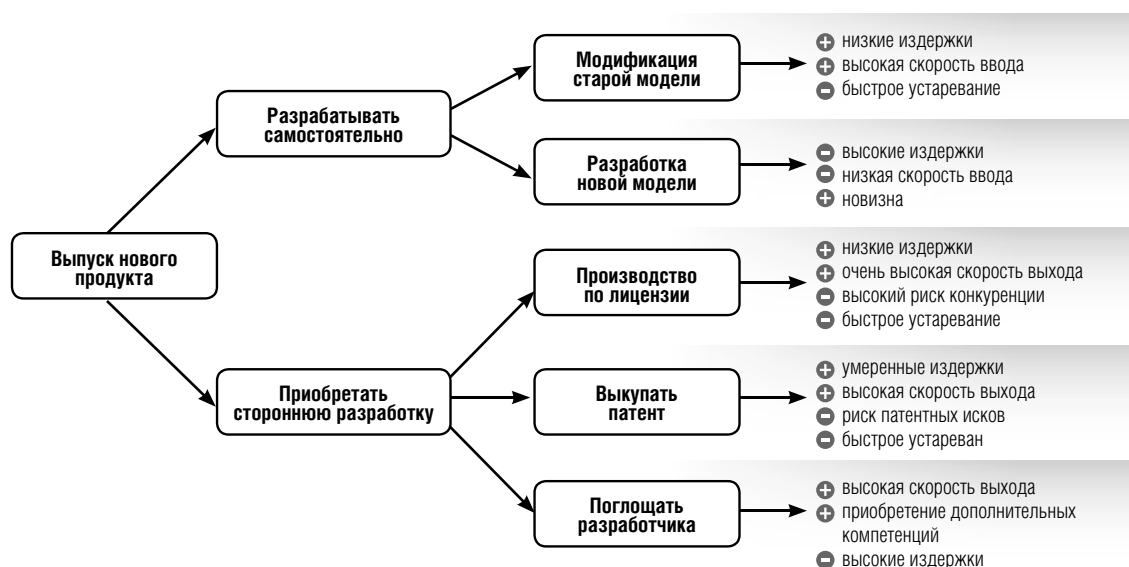


Рис. 11. Пример дерева решений «Выпуск нового продукта»

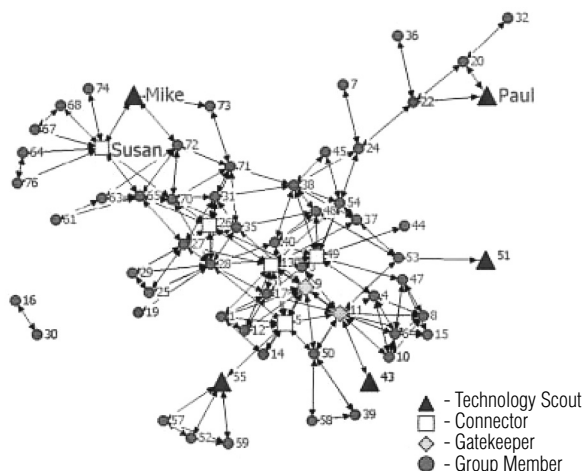


Рис. 12. Визуализация связей типичной организации [40]

КТО-знания

Знания о том, кто за что отвечает, обычно отображают в виде организационной диаграммы или в матрице ответственностей.

Также, на вопрос «Кто обладает знаниями в организации?» помогают ответить «плавательные дорожки» (см. рис. 9) и карты знаний. Карта знаний представляет собой изображение местоположения знаний, необходимых для работы. Она связывает три типа знаний — ЧТО-, ГДЕ- и КТО-в единую модель. Карта знаний показывает возможность интеграции разьединённых знаний (скрытых и явных), находящихся в базах знаний, в головах экспертов, внутри компании и даже за её пределами (у клиентов и партнёров). Это также инструмент обнаружения необходимых ресурсов и услуг [39].

Интересный способ визуализации КТО-знаний предлагает направление современной компьютерной социологии SNA (Social Network Analysis — анализ социальных сетей), которое занимается анализом возникающих в ходе коммуникации связей (сетей). Цель визуализации — выявление групп, а также тех, кто является связующим звеном между группами; тех, кто собирает группу; и тех, кто плохо вписывается в коммуникационную сеть (рис. 12).

КОГДА-знания

Традиционно информация о времени проведения мероприятий узнается из различного рода календарей и расписаний. Но они не дают возможности показать взаимосвязь большого количества событий. Для этого служат сетевые графики (или PERT-диаграммы).

Сетевой график — это граф, который служит для отображения последовательности работ, их связей и длительностей. Работы могут обозначаться как дугами (тогда вершины графа — это состояния проекта), так и узлами (тогда дуги — это связи между работами). Но сетевой график не привязан к календарю и длительности работ никак не влияют на длины дуг. То есть сетевой график даёт ответ на вопрос, в каком порядке должны осуществляться действия, но не очерчивает временных рамок их реализации.

Решить обе вышеназванные задачи можно с помощью диаграммы Ганта, объединившей в себе календарь и сетевой график. Пример учебной диаграммы Ганта с отмеченным критическим путём показан на рис. 13. В образовательном процессе эти

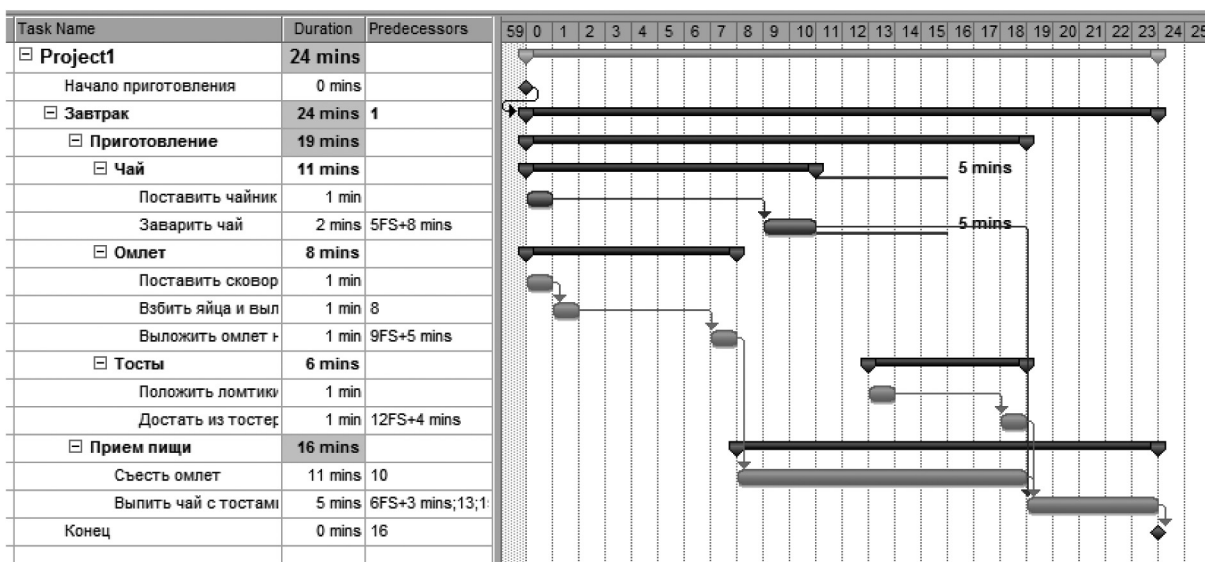


Рис. 13. Пример расширенной диаграммы Ганта

диаграммы могут применяться для планирования курса, групповых проектов студентами и для решения других схожих задач.

Кроме того, для визуализации КОГДА-знаний можно использовать диаграмму Ишикавы. Некоторые варианты использования этих диаграмм позволяют комбинировать ПОЧЕМУ-знания и КОГДА-знания. Такой подход особенно полезен при изучении истории организаций, развития событий и т.д.

ГДЕ-знания

Первые карты появились задолго до возникновения письменности. Поэтому неудивительно, что способы визуализации местоположения объектов столь широко развиты. Визуализация ГДЕ-знаний также активно используется в гуманитарных науках: исторические карты, карты из экономической географии. Особенно актуален этот тип знаний в современных ГИС (гео-информационных системах).

3. Ограничения и область применения подхода

Использование визуальных языков моделирования, выбираемых с помощью предложенного метода, имеет свои сложности и ограничения. Одним из проблемных вопросов в использовании графических языков является семантическая интероперабельность — различные пользователи могут по-разному интерпретировать значения символов на диаграмме, различные языки моделирования могут иметь элементы с похожими названиями, но с разным смыслом, а также, наоборот, при схожести смысла понятий могут различаться по формулировкам. Для устранения данной проблемы используются унифицированные метамодели [41] или онтологии [42, 43]. На практике в данном случае полезны непротиворечивые демонстрационные примеры диаграмм, выполненных в соответствии с требуемой нотацией. Чем их больше, тем лучше. Пользователи понимают значения элементов визуального языка через примеры (экстенциональные определения понятий).

Кроме того, для решения управленческих задач обычно нужна не отдельная диаграмма, а комплекс диаграмм, охватывающих разные типы знаний. Например, при оптимизации деятельности с помощью теории ограничений используются 5 взаимосвязанных типов диаграмм [44], а при проектировании

информационных систем — несколько моделей из нотации UML [45] или инструментария Archimate [46].

Наконец, не всегда диаграммы — это оптимальный способ представления знаний. В статьях [47, 48] было произведено сравнение матричных (табличных) представлений и наиболее популярных диаграмм вида «узел-связь». Гоньем М. (Ghoniem M., 2005) с соавторами [47] показал, что матрицы превосходят диаграммы в случае больших сильно-связанных графов в контексте нескольких задач чтения. Такие выводы подтверждаются экспериментами Келлера и его коллег [48], в результате которых было выявлено, что диаграммы «узел-связь» нагляднее для небольших структур, однако они становятся сложными для моделей больших систем.

Несмотря на вышесказанное, диаграммы являются полезным средством для представления знаний, однако нуждаются в поддерживающих механизмах. В частности, методы визуального моделирования необходимо сочетать с онтологическим инжинирингом и метамоделированием [49] для обеспечения семантической интероперабельности, с инженерией методов (method engineering) [50] и отраслевыми лучшими практиками для интеграции диаграмм со способами решения задач в различных областях, а также с неграфическими методами представления информации в ситуациях, когда последние более эффективны [51, 52].

4. Заключение

Все больше современных менеджеров и специалистов понимают важность проблемы структурирования и наглядного представления знаний. Визуализация позволяет наглядно показать изучаемые процессы, явления и понятия в тех случаях, когда непосредственное восприятие затруднено. Графические представления как метод компактной организации информации могут использоваться как мощное орудие мышления применительно ко всем областям интеллектуальной деятельности, включая моделирование предприятий, сложных организационных структур, а также процессы бизнес-обучения.

В статье предложена новая классификация языков визуального моделирования, которая позволяет выбрать тип диаграммы в зависимости от типа знаний, которые необходимо представить. В статье также представлено описание типов знаний с помощью вопросов проверки компетентности,

а также выявлены общие понятия и отношения, характерные для каждого типа знаний и соответствующих языков. Данные описания вопросов проверки компетентности и общих элементов, с одной стороны, обосновывают распределение визуальных языков по типам, а с другой — помогают понять, в визуализации какого типа знаний есть потребность. Все вышеуказанные классифи-

кации и описания интегрированы в метод выбора языка визуального моделирования. Кроме того, в статье выявлены ограничения визуальных средств и предложены подходы по их устранению — интеграция с онтологическим инжинирингом, инженерией методов и отраслевыми лучшими практиками, а также с неграфическими средствами моделирования. ■

Литература

1. Загорулько Ю.А., Загорулько Г.Б. Онтологический подход к разработке системы поддержки принятия решений на нефтегазодобывающем предприятии // Вестник Новосибирского государственного университета, серия «Информационные технологии». — 2012. — Т. 10. — № 1. — С. 121-128.
2. Романовский Д.Ю., Кознов Д.В. Язык DRL для проектирования и разработки документации семейства программных продуктов // Вестник Санкт-Петербургского государственного университета, Серия «Информатика». — 2007. — № 4. — С. 110-122.
3. Рыбина Г.В. Инструментальные средства построения динамических интегрированных экспертных систем: развитие комплекса АТТЕХНОЛОГИЯ // Искусственный интеллект и принятие решений. — 2010. — №1. — С. 41-48.
4. Тельнов Ю.Ф. Интеллектуальные информационные системы в экономике: учеб.пособие: 3 изд., расширенное и доработ. — М.: СИНТЕГ, 2002.
5. Рапуто А.Г. Дескриптивное моделирование образного мышления при репрезентации дидактических объектов // Сборник «Ученые записки». Вып. 34. — М.: ИИО РАО, 2011. — С. 114-116.
6. Кознов Д.В. Методика обучения программной инженерии на основе карт памяти // Системное программирование. — 2008. — Вып. 3. — С. 121-140.
7. Eppler M., Burkhard R. Visual representations in knowledge management: framework and cases. // Journal of Knowledge Management. — 2007. — Vol. 11. — №4. — P. 112-122.
8. Margulies N., Valenza C. Visual Thinking Tools for Mapping Your Ideas. — Bethel, CT, USA: Crown House, 2005.
9. Lengler R., Eppler M. Towards a Periodic Table of Visualization Methods for Management. // Proceedings of the Conference on Graphics and Visualization in Engineering, 2007. — P. 1-6.
10. Кудрявцев Д.В. Системы управления знаниями и применение онтологий: учеб. пособие. — СПб.: Изд-во Политехнического ун-та., 2010.
11. Хорошевский В.Ф. Пространства знаний в сети Интернет и Semantic Web (Часть 3) // Искусственный интеллект и принятие решений. — 2012. — № 1. — С. 3-38.
12. Гаврилова Т.А., Лещева И.А., Страхович Э.В. Об использовании визуальных концептуальных моделей в преподавании // Вестник Санкт-Петербургского государственного университета, серия «Менеджмент». — 2011. — № 4. — С. 125-151.
13. Vallano E. How data visualization promotes eLearning through effective story telling // MindJet Blog. 16.02.12. URL: <http://blog.mindjet.com/2012/02/how-data-visualization-promotes-elearning-through-effective-story-telling/> (дата обращения: 21.03.13).
14. Singh S. Why Visualization Matters in Elearning Courses? // CommLab India blog. 08.06.11. URL: <http://blog.commlabindia.com/elearning/visualization-in-elearning> (дата обращения: 19.03.13).
15. Бабкин Э.А., Князькин В.П., Шиткова М.С. Разработка метода проведения сравнительного анализа языков бизнес-моделирования // Бизнес-информатика. — 2010. — №3. — С. 41-46.
16. Бабкин Э.А., Князькин В.П., Шиткова М.С. Сравнительный анализ языковых средств, применяемых в методологии бизнес моделирования // Бизнес-информатика. — 2011. — №2. — С. 31-42.
17. Гаврилова Т.А., Яшин А.М. Некоторые аспекты методологии автоматизированного проектирования баз знаний // Научные труды СПбГТУ. — СПб.: СПбГТУ, 1995.
18. Lohse G.L., Biolsi K., Walker N., Rueter H.H. A classification of visual representations // Communications of the ACM. — 1994. — Vol. 37. — № 12. — P. 36-49.

19. Гаврилова Т.А., Гулякина Н.А. Визуальные методы работы со знаниями: попытка обзора // Искусственный интеллект и принятие решений. — 2008. — № 1. — С. 15-21.
20. Grüninger M., Fox M. Methodology for the design and evaluation of ontologies // Proceedings of IJCAI 1995 Workshop on Basic Ontological Issues in Knowledge Sharing / Ed. D. Skuce. — Montreal: AAAI Press, 1995. — P. 6.1–6.10.
21. Kudryavtsev D., Gavrilova T. Diagrammatic knowledge modeling for managers: ontology-based approach // Proceedings of International Conference on Knowledge engineering and Ontology Development, 2011. — P. 386-389.
22. Бьюзен Т., Бьюзен Б. Супермышление / пер. с англ. Е.А. Самсонов. Минск: Попурри, 2003.
23. Глаголева Е.А. Трудности усвоения школьных навыков. — М.: Педагогика, 1926.
24. Chei C. C. The effect of concept mapping on students' learning achievements and interests // Innovations in Education and Teaching International. — 2008. — Vol. 45. — № 4. — P. 375387.
25. Роэм Д. Визуальное мышление. Решение проблем и продажа идей при помощи картинок на салфетке / пер. с англ. О.И. Медведь. — М.: Эксмо, 2009.
26. Hyerle D. Visual Tools for Transforming Information Into Knowledge. — Thousand Oaks, CA: Corwin Press, 2009.
27. Jeffery A.B., Maes J.D., Bratton-Jeffery M.F. Improving team decision-making performance with collaborative modeling // Team Performance Management. — 2005. — Vol. 11. — № 1/2. — P. 4050.
28. Lim B.C., Klein K.J. Team mental models and team performance: A field study of the effects of team mental model similarity and accuracy // Journal of Organizational Behavior. — 2006. — Vol. 27. — № 4. — P. 403-418.
29. Костырко А., Кудрявцев Д., Григорьев Л., Кислова В., Жулин А., Синятулина Л., Ермаков Р. Моделирование комплексов городского хозяйства для системного развития ИКТ города // Сборник трудов конференции «Инженерия знаний и технологии семантического веба — 2012». — СПб.: СПбГУ ИТМО, 2012. — С. 81-88.
30. Ausubel D.P. Educational psychology: A cognitive view. — New York: Holt, Rinehart and Winston, 1968.
31. Kinchin I.M., Hay D.B., Adams A. How a qualitative approach to concept map analysis can be used to aid learning by illustrating patterns of conceptual development // Educational Research. — 2000. — Vol. 42. — № 1. — P. 43-57.
32. Ishikawa K. Guide to Quality Control. — Tokyo: Asian Productivity Organization, 1976.
33. Новак С. Бизнес-инструменты для производственного предприятия: от основ до высшего пилотажа / пер. с англ. Л.Ю. Смоляк; науч. ред. О.Ч. Шмигельский. — Минск: Гревцов Паблишер, 2008.
34. Christianson M.K., Farkas M.T., Sutcliffe K.M., Weick K.E. Learning Through Rare Events: Significant Interruptions at the Baltimore & Ohio Railroad Museum // Organization Science. — 2009. — Vol. 20. — № 5. — P. 846-860.
35. Леоненков А. Самоучитель UML. — СПб.: БХВ-Петербург, 2001.
36. Райзберг Б.А., Лозовский Л.Ш., Стародубцева Е.Б. Современный экономический словарь. — 5-е изд., перераб. и доп. — М.: Инфра-М, 2007.
37. Churchman C.W., Ackoff L.E. Introduction to Operations Research. — New York: Wiley, 1957.
38. Каплан Р.С., Нортон Д.П. Сбалансированная система показателей: от стратегии к действию / пер. с англ. М. Павлова. — 2-е изд., испр. и доп. — М.: Олимп-Бизнес, 2003.
39. Мариничева М.К. Управление знаниями на 100%: Путеводитель для практиков. — М.: Альпина Бизнес Букс, 2008.
40. Whelan E. It's who you know not what you know: a social network analysis approach to talent management // European Journal of International Management. — 2011. — Vol. 5. — № 5. — P. 484500.
41. A Meta-Meta-Model for Seven Business Process Modeling Languages // Proceedings of 2013 IEEE International Conference on Business Informatics, 2013. — P. 216-221.
42. Grüninger M., Atefi K., Fox M.S. Ontologies to support process integration in enterprise engineering // Computational & Mathematical Organization Theory. — 2000. — Vol. 6. — № 4. — P. 381394.
43. Andersson B., Bergholtz M., Edirisuriya A., Ilayperuma T., Johannesson P., Gordijn J., Gregoire B., Schmitt M., Dubois E., Abels S., Hahn A., Wangler B., Weigand H. Towards a reference ontology for business models // Proceedings of 25th International Conference on Conceptual Modeling Conceptual Modeling (ER 2006). —

- Berlin / Heidelberg: Springer, 2006. — P. 482-496.
44. Детмер У. Теория ограничений Голдратта: Системный подход к непрерывному совершенствованию / пер. с англ. У.В. Саламатова. — 2-е изд. — М.: Альпина Бизнес Букс, 2008.
 45. Буч Г., Рамбо Д., Джекобсон А. Язык UML. Руководство пользователя / пер. с англ. А.А. Слинкин. — 2-е изд. М.: ДМК Пресс; СПб.: Питер, 2004.
 46. ArchiMate 2.0 Specification / M. Iacob, H. Jonkers, M. Lankhorst, E. Proper, D.A.C. Quartel // The Open Group Publications Catalog, 2009–2012. URL: <http://pubs.opengroup.org/architecture/archimate2-doc/> (дата обращения: 10.02.2013)
 47. Ghoniem M., Fekete J., Castagliola P. On the readability of graphs using node-link and matrixbased representations: a controlled experiment and statistical analysis // Information Visualization. — 2005. — Vol. 4. — № 2. — P. 114-135.
 48. Keller R., Eckert C.M., Clarkson P.J. Matrices or node-link diagrams: which visual re-presentation is better for visualising connectivity models? // Information Visualization. — 2006. — Vol. 5. — P. 6276.
 49. Кудрявцев Д.В. Онтологии и организационное моделирование: согласование терминологии // Труды X Российской научно-практической конференции «Реинжиниринг бизнес-процессов на основе современных информационных технологий. Системы управления знаниями». — М.: МЭСИ, 2007. — С. 156-160.
 50. Henderson-Sellers B., Ralyté J. Situational Method Engineering: State-of-the-Art Review // Journal of Universal Computer Science. — 2010. — Vol. 16. — № 3. — P. 424478.
 51. Григорьев Л.Ю., Кудрявцев Д.В. Организационное проектирование на основе онтологий: методология и система ОРГМастер // Научно-технические ведомости СПбГПУ, серия «Информатика. Телекоммуникации. Управление». — 2012. — №1. — С. 21-28.
 52. Grigoriev L., Kudryavtsev D. Non-diagrammatic method and multi-representation tool for integrated enterprise architecture and business process engineering // Proceedings of 2013 IEEE International Conference on Business Informatics, 2013. — P. 216-221.

ЖДЕТ СВОИХ ПОЛЬЗОВАТЕЛЕЙ НОВЫЙ УНИКАЛЬНЫЙ В МИРОВОЙ ПРАКТИКЕ

Универсум связей между русскими словами CrossLexica

- ◆ К **сверхбольшому компьютерному словарю**, по объему уже не воспроизводимому в печати, обращается либо человек — **интерактивно**, либо внешний софт — **программно**.
- ◆ Словарь предельно **политематичен**: от математики и философии до политики, спорта, строительства, кулинарии, автомобилей, моды и бранной лексики, допущенной Госдумой.
- ◆ Включает информацию как **языкового**, так и **энциклопедического** характера.
- ◆ Беспрецедентен **по объему: 295 тыс. элементов** словника и **8.3 млн. связей** между ними: **4.7 млн.** связей в словосочетаниях, **2.8 млн.** смысловых связей и **0.8 млн.** связей внешнего сходства слов. **47%** элементов словника сами состоят из нескольких слов.
- ◆ Объединяет в себе словари **словосочетаний, синонимов, антонимов, паронимов, словоизменения, тезаурусных иерархий, глагольно-именного управления** и др.
- ◆ Рассчитан на **любой класс пользователей** — на ученого, преподавателя, инженера, журналиста, бизнесмена, военного, студента, школьника, пенсионера, домохозяйку.
- ◆ **Обучает** грамотному русскому языку как **русскоговорящего**, так и **иностранца**.
- ◆ Обеспечивает **доступ иностранцу** еще и через **английские** переводы словника.
- ◆ Быстро выбирает **WEB-запрос** из наличных миллионов, посылая его в **Google / Яндекс**.
- ◆ Отвечает на **запросы с клавиатуры** и те, что возникают при **навигации** по словнику.
- ◆ Размещается в ОП любого компьютера, запускается под **Windows XP** и **Windows 7**.
- ◆ Не повторяя существующие толковые словари, способен поднять как интерактивную, так и автоматическую обработку русских текстов на уровень, ныне не достижимый.

На эксклюзивную продажу выставляется пакет:

текущая версия, исходная текстовая база, грамматические таблицы, утилиты морфоклассификации слов и сборки рабочей версии, коды интерфейса с человеком и компьютером, технология правки и пополнения базы.

Автор и владелец: **доктор наук профессор ИГОРЬ АЛЕКСЕЕВИЧ БОЛЬШАКОВ.**

Контакт: **iabolshakov@gmail.com**

Подробнее: см. статью в номере 3(25) за 2013 г.

ПРИМЕНЕНИЕ СИСТЕМЫ MATLAB ДЛЯ ОЦЕНКИ РИСКОВ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ ОРГАНИЗАЦИИ

С.А. Глушенко,

ассистент кафедры «Информационные системы и прикладная информатика»
ФГБОУ ВПО «Ростовский государственный экономический университет (РИНХ)»

E-mail: www.555.sergey@mail.ru

Адрес: г. Ростов-на-Дону, ул. Большая Садовая, 69

В статье обосновывается целесообразность применения нечеткой логики для оценки риска информационной безопасности организации и предлагается нечеткая продукционная модель (НПМ). Проводится реализация процесса нечеткого моделирования базы правил посредством применения специализированного пакета Fuzzy Logic Toolbox программного средства MATLAB. Выполнение нечеткого вывода реализуется на основе алгоритма Мамдани (Mamdani).

Ключевые слова: риск, нечеткое множество, терм-множество, нечеткая продукционная модель, лингвистическая переменная, база правил, функция принадлежности.

1. Введение

Процесс внедрения информационных технологий и средств вычислительной техники в производство и управление современных организаций является эффективным инструментом повышения производительности труда. Однако информационная инфраструктура организаций часто приобретает неструктурированный характер и влечет за собой неконтролируемый рост уязвимостей и риск информационной безопасности (ИБ) организации в целом.

Информационная безопасность организации – защищенность информации и поддерживающей инфраструктуры от случайных или преднамеренных воздействий естественного или искусственно-

го характера, которые могут нанести неприемлемый ущерб [1].

В Российской Федерации внешнее регулирование деятельности по обеспечению ИБ осуществляется различными государственными органами, такими как:

- ✧ Комитет Государственной думы по безопасности;
- ✧ Совет безопасности России;
- ✧ Федеральная служба по техническому и экспортному контролю (ФСТЭК России);
- ✧ Федеральная служба безопасности Российской Федерации (ФСБ России);
- ✧ Федеральная служба охраны Российской Федерации (ФСО России);

❖ Федеральная служба по надзору в сфере связи, информационных технологий и массовых коммуникаций (Роскомнадзор).

На уровне предприятия для контроля ИБ и осуществления мер по защите информации могут быть организованы:

- ◆ Служба экономической безопасности;
- ◆ Служба информационной безопасности;
- ◆ Служба безопасности персонала.

Деятельность данных служб основана на методических документах государственных органов России (Доктрина информационной безопасности РФ [2], руководящие документы и приказы ФСТЭК (Гостехкомиссии России) и ФСБ) и стандартах ИБ (Международные стандарты, государственные (национальные) стандарты РФ [3], рекомендации по стандартизации, методические указания).

В целях обеспечения информационной безопасности организации строится Система обеспечения информационной безопасности (СОИБ). СОИБ предприятия представляет собой совокупность мер организационного и программно-технического уровня, направленных на защиту информационных ресурсов предприятия от потенциальных угроз. Меры защиты организационного уровня реализуются путем проведения соответствующих мероприятий, предусмотренных документированной политикой информационной безопасности. Меры защиты программно-технического уровня реализуются при помощи соответствующих средств и методов защиты информации [3].

Экономический эффект от внедрения СОИБ проявляется в виде снижения величины возможного материального, репутационного и иных видов ущерба, наносимого предприятию, за счет использования мер, направленных на формирование и поддержание режима ИБ.

Определить перечень необходимых мер защиты информации, выбрать стратегию развития информационной структуры организации и поддерживать на должном уровне безопасность организации, возможно только по результатам аудита уязвимостей предприятия и анализа рисков.

В данном случае риск рассматривается как фактор, сущность или элемент, представляющий опасность для ИБ организации, величина которой не определена [6].

Наиболее распространенными методиками оценки риска являются методы, изложенные в специ-

альных рекомендациях 800-30 Национального Института Стандартов и Технологий США (NIST) [7] и разработанный Службой Безопасности Великобритании метод CRAMM [6]. Они охватывают широкий круг вопросов, связанных со стратегией управления рисками и являются хорошей основой для разработки собственной системы управления рисками.

Стоит отметить, что в отечественной нормативной базе ИБ отсутствует ГОСТ по управлению рисками. Российские стандарты ГОСТ ИСО/МЭК 17799 [3] и ГОСТ ИСО/МЭК 27001 [4] являются техническим переводом первой и второй частей соответственно британского стандарта BS 7799 из трех. Согласно российскому стандарту ГОСТ 27001 введенному в 2007 г., система управления ИБ трактуется как часть общей системы управления предприятием и предназначается для создания, внедрения, эксплуатации, мониторинга, анализа, сопровождения и совершенствования ИБ, а ГОСТ 17799 описывает примеры по среде и системам ИБ, но не отражает вопросов руководства по оценке и управлению рисками. Таким образом, внедрение стандартов ГОСТ 17799 и ГОСТ 27001 в практику подразумевает наличие в организации как минимум двух документов: политики ИБ и методологии оценки рисков ИБ, однако российская нормативная база не содержит описаний формы и содержания для последнего документа [5].

Методики управления рисками NIST и CRAMM для целей анализа предлагают использовать для каждого вида риска вероятность его появления и ущерб негативных последствий от риска. В качестве интегральной оценки риска используется произведение величины вероятности и ущерба. В [6] отмечается, что реально оценить вероятность риска достаточно сложно. В большинстве случаев ИТ-менеджеры и эксперты, основываясь на собственном опыте и/или имеющихся данных других организаций, проводят оценку в виде словесных формулировок, которые затем связывают с числовыми данными.

Такой «жесткий» механизм получения оценок риска существенно ограничивает возможности всей методики в целом. Например, вероятность риска оценивается как «низкая», «средняя» или «высокая» и этим определениям ставятся в соответствии следующие числовые значения вероятности риска: 17%, 50% и 84%. При этом «низкое» значение риска определяет интервал от 1% до 33% со средним значением 17%, «среднее» — интервал от 34% до 67% со

средним 50% и «высокое» — интервал от 68% до 99% со средним 84%.

Применяемый подход к оценке вероятности риска имеет ряд недостатков. Во-первых, оценки вероятности риска формируются в отдельных точках (17%, 50% и 84%), что не всегда может быть приемлемым. Во-вторых, уверенность эксперта в предлагаемой оценке может быть различной, что не отражается в процедуре оценивания вероятности риска.

Для устранения вышеприведенных недостатков предлагается использовать нечеткие модели [8], применение которых эффективно в следующих случаях:

- имеется недостаточность или неопределенность знаний об исследуемой системе или процессе;
- получение требуемой информации сопряжено с различными трудностями или вообще невозможно;
- основная часть информации получена на основе экспертных данных или эмпирических описаний процессов; параметры и входные данные не являются точными и корректно представленными.

Применение лингвистического подхода, во время оценки состояния системы обеспечения информационной безопасности организации, является общеизвестным. Оценка компонентов СОИБ проводится терминами «средний уровень программно-аппаратной защиты», «высокий уровень организационной защиты», «низкая рыночная ценность информационного ресурса» и др., т.е. ИТ-менеджерам сложно придать им точную (объективную) количественную оценку и описать с помощью математического языка. При таком подходе целесообразно рассматривать характеристики системы с точки зрения теории нечетких множеств, как лингвистические переменные. Использование методов получения оценок рисков на основе нечеткой логики позволяют использовать как количественные характеристики, которым объективно свойственна неопределенность, так и качественные, субъективные оценки экспертов, выраженные нечеткими понятиями, а также формализовать нечеткие описания с помощью нечетких чисел, множеств, лингвистических переменных и нечетких свидетельств.

Недостатками данного подхода являются субъективность в выборе функций принадлежности и формировании правил нечеткого ввода, а также необходимость специального программного обеспечения и специалистов, умеющих с ним работать.

2. Нечеткая продукционная модель оценки риска

Для моделирования риска информационной безопасности организации, нечеткие модели целесообразно представлять в виде нечетких сетей, элементы и совокупности элементов которых реализуют различные компоненты нечетких моделей и этапы нечеткого вывода.

Нечеткая продукционная модель (НПМ) может быть представлена следующим образом [9]:

$$(i): Q; P; A \Rightarrow B; S; F; N,$$

- Q — сфера применения нечеткой продукции;
- P — условие активизации ядра нечеткой продукции;
- A — условие ядра (антецедент);
- B — заключение ядра (консеквент);
- S — метод определения количественного значения степени истинности заключения ядра;
- F — коэффициент уверенности нечеткой продукции;
- N — постусловие продукционного правила.

Нечеткое причинно-следственное отношение между антецедентом и консеквентом задается в виде нечеткой продукции:

ЕСЛИ x есть A , ТО y есть B

- X — область определения антецедента;
- A — нечеткое множество, определенное на X ;
- $\mu_A(x) \in [0,1]$ — функция принадлежности нечетного множества A ;
- Y — область определения консеквента;
- B — нечеткое множество, определенное на Y ;
- $\mu_B(y) \in [0,1]$ — функция принадлежности нечетного множества B .

Если известна функция принадлежности нечетного множества A — $\mu_A(x)$, тогда для нечеткого множества B функция принадлежности определяется по правилу композиции:

$$\mu_B(y) = \sup_{x \in X} \{T(\mu_A(x), \mu_R(x, y))\} \quad (1)$$

$\sup$ — операция определения верхней границы множества элементов;

T — операция Т-нормы.

При моделировании риска информационной безопасности организации, в качестве правила вычисления нечеткой импликации, применяется классическая нечеткая импликация Л. Заде:

$$\mu_R(x, y) = \max \{ \min [\mu_A(x), \mu_B(y)], [1 - \mu_A(x)] \} \quad (2)$$

Основными способами нечеткого вывода заключений в НПМ являются прямой и обратный вывод. Прямой вывод основывается на правиле вывода «нечеткий модус поненс» (*fuzzy modus ponens*).

Во время построения нечеткой продукционной модели оценки рисков ИБ организации необходимо сформировать полное пространство предположений $X = \{x_i\}, i = \overline{1, n}$ – факторов, являющихся источниками риска, и полное пространство заключений $Y = \{y_j\}, j = \overline{1, m}$ – показателей риска различных областей информационной безопасности организации.

В процессе анализа факторов риска выявлены показатели, которые могут быть источниками риска ИБ организации (табл. 1). При задании лингвистических переменных, характеризующих факторы риска, могут использоваться следующие термножества, определяющие уровни факторов [10]:

$T2 = \{\text{Низкий (H)}, \text{Высокий (B)}\};$

$T3 = \{\text{Низкий (H)}, \text{Средний (C)}, \text{Высокий (B)}\};$

$T4 = \{\text{Очень Низкий (ОчН)}, \text{Низкий (H)}, \text{Средний (C)}, \text{Высокий (B)}\};$

$T5 = \{\text{Очень Низкий (ОчН)}, \text{Низкий (H)}, \text{Средний (C)}, \text{Высокий (B)}, \text{Очень Высокий (ОчВ)}\}.$

Таблица 1.

Факторы риска ИБ организации (фрагмент)

Обозначение	Наименование ЛП	Вид термножества
x_1	Программно-аппаратный уровень защиты	$T3$. H – удовлетворительная, для обеспечения начально уровня защиты; C – достаточна, для базовой информационной защиты; B – полностью соответствует уровню конфиденциальности информации
x_2	Уровень организационной защиты	$T3$. H – слабое планирование и отсутствие мониторинга уязвимостей; C – планирование и мониторинг уязвимостей проводятся нерегулярно; B – своевременное планирование и мониторинг уязвимостей
x_3	Уровень правовой защиты	$T3$. H – обрывочная и неполная документация; C – документация имеется, но недостаточно детальная; B – документация полная и синхронизированная

В процессе анализа риска выявлены показатели, которые могут характеризовать риски ИБ организации (табл. 2). При задании лингвистических переменных, характеризующих показатели риска, используется следующее термножество, определяющие показатели риска:

$T1 = \{\text{Низкая очевидность риска (НОР)}, \text{Средняя очевидность риска (СОР)}, \text{Высокая очевидность риска (ВОР)}\};$

$T2 = \{\text{Очень низкая очевидность риска (ОНОР)}, \text{Низкая очевидность риска (НОР)}, \text{Средняя очевидность риска (СОР)}, \text{Высокая очевидность риска (ВОР)}, \text{Очень высокая очевидность риска (ОВОР)}\}.$

Таблица 2.

Показатели риска ИБ организации (фрагмент)

Обозначение	Наименование ЛП	Примечание
y_1	Риск снижения эффективности защиты	Характеризует потенциальную возможность снижения/увеличения эффективности защиты, по отношению к требуемой эффективности для конкретного предприятия.
y_2	Риск возникновения потенциальных угроз	Характеризует возможности возникновения потенциальных угроз для предприятия
y_3	Риск материального ущерба	Характеризует возможность возникновения материального ущерба для предприятия при нарушениях параметров информационной безопасности предприятия

Взаимосвязь между факторами (антецедентом) и показателями риска (консеквентом) представляет собой бинарное нечеткое отношение на декартовом произведении соответствующих нечетких множеств. Нечеткое причинно-следственное отношение между антецедентом и консеквентом задается в виде нечеткой продукции [9]. Продукционные правила приведены в табл. 3 (фрагмент).

3. Применение пакета Fuzzy Logic Toolbox для построения НПМ

Реализация процесса нечеткого моделирования базы правил проводится посредством применения специализированного пакета *Fuzzy Logic Toolbox* программного средства *MATLAB* [11]. Выполнение нечеткого вывода реализуется на основе алгоритма Мамдани (*Mamdani*).

Шаг 1. Фазификация — введение нечеткости. На этом шаге необходимо задать функции принадлежности для термножеств входных и выходных лингвистических переменных:

$PA3$ в модели соответствует лингвистической переменной «Программно-аппаратный уровень защиты» — x_1 ;

Таблица 3.

Нечеткие продукционные правила модели
(фрагмент)

Обозначение	Антецедент	Консеквент
База правил П1		
П1.1	$(x_1=H \wedge x_2=H \wedge x_3=H) \vee (x_1=C \wedge x_2=H \wedge x_3=H) \vee (x_1=H \wedge x_2=C \wedge x_3=H)$	$y_1 = \text{Очень ВОР}$
П1.2	$(x_1=B \wedge x_2=H \wedge x_3=H) \vee (x_1=C \wedge x_2=C \wedge x_3=H) \vee (x_1=H \wedge x_2=B \wedge x_3=H) \vee (x_1=C \wedge x_2=B \wedge x_3=H) \vee (x_1=H \wedge x_2=H \wedge x_3=C) \vee (x_1=H \wedge x_2=C \wedge x_3=C) \vee (x_1=H \wedge x_2=B \wedge x_3=C) \vee (x_1=H \wedge x_2=H \wedge x_3=B)$	$y_1 = \text{ВОР}$
П1.3	$(x_1=B \wedge x_2=C \wedge x_3=H) \vee (x_1=B \wedge x_2=B \wedge x_3=H) \vee (x_1=C \wedge x_2=H \wedge x_3=C) \vee (x_1=B \wedge x_2=H \wedge x_3=C) \vee (x_1=C \wedge x_2=C \wedge x_3=C) \vee (x_1=C \wedge x_2=B \wedge x_3=C) \vee (x_1=H \wedge x_2=C \wedge x_3=B) \vee (x_1=H \wedge x_2=C \wedge x_3=B) \vee (x_1=C \wedge x_2=C \wedge x_3=B) \vee (x_1=H \wedge x_2=B \wedge x_3=B)$	$y_1 = \text{СОР}$
П1.4	$(x_1=B \wedge x_2=C \wedge x_3=C) \vee (x_1=B \wedge x_2=B \wedge x_3=C) \vee (x_1=B \wedge x_2=H \wedge x_3=B) \vee (x_1=B \wedge x_2=C \wedge x_3=B) \vee (x_1=C \wedge x_2=B \wedge x_3=B)$	$y_1 = \text{НОР}$
П1.5	$x_1=B \wedge x_2=B \wedge x_3=B$	$y_1 = \text{Очень НОР}$

ОргЗ в модели соответствует лингвистической переменной «Уровень организационной защиты» — x_2 ;

ПравЗ в модели соответствует лингвистической переменной «Уровень правовой защиты» — x_3 ;

РискЗ в модели соответствует лингвистической переменной «Риск снижения эффективности защиты» — y_1 .

Для входной переменной *ПАЗ* терм-множество состоит из трех термов $T=\{\text{Низкий (H)}, \text{Средний (C)}, \text{Высокий (B)}\}$, которые характеризуют низкий, средний и высокий уровень программно-аппаратной защиты организации. Функции принадлежности для входной переменной *ПАЗ* являются трапециевидными. В общем случае трапециевидная функция принадлежности имеет следующий вид:

$$\mu_T(x, a, b, c, d) = \begin{cases} 0, x \leq a \\ \frac{x-a}{b-a}, a \leq x \leq b \\ \frac{d-x}{d-c}, c \leq x \leq d \\ 0, d \leq x \end{cases}, \quad (3)$$

где a, b, c, d — числовые параметры, характеризующие нижнее основание трапеции, (a, d) и верхнее (b, c), причем должно выполняться условие $a \leq b \leq c \leq d$.

С учетом (3) функции принадлежности нечетких терм-множеств лингвистической переменной «Программно-аппаратный уровень защиты» будут иметь следующий вид:

$$\mu_{\Delta}^H(x; 0; 0; 0; 15; 0; 45), \mu_{\Delta}^C(x; 0; 1; 0; 4; 0; 6; 0; 9), \\ \mu_{\Delta}^B(x; 0; 55; 0; 85; 1; 0; 1; 0).$$

На *рис. 1* приведены графики функций принадлежности терм-множеств лингвистической переменной *ПАЗ* — «Программно-аппаратный уровень защиты».

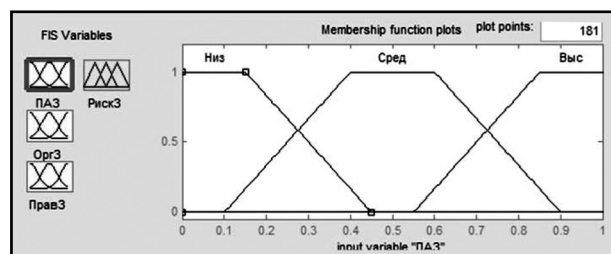


Рис. 1. Функции принадлежности для входной переменной ПАЗ

Для входной переменной *ОргЗ* терм-множество состоит из трех термов $T=\{H, C, B\}$, функции принадлежности которых являются треугольными. В общем случае треугольная функция принадлежности имеет следующий вид:

$$\mu_{\Delta}(x, a, b, c) = \begin{cases} 0, x \leq a \\ \frac{x-a}{b-a}, a \leq x \leq b \\ \frac{c-x}{c-b}, b \leq x \leq c \\ 0, c \leq x \end{cases}, \quad (4)$$

где a, b, c — числовые параметры, характеризующие основание треугольника (a, c) и его вершину (b), причем должно выполняться условие $a \leq b \leq c$.

С учетом (4) функции принадлежности нечетких терм-множеств лингвистической переменной «Уровень организационной защиты» будут иметь следующий вид:

$$\mu_{\Delta}^H(x; 0; 0; 0; 0; 4), \mu_{\Delta}^C(x; 0; 1; 0; 5; 0; 9), \\ \mu_{\Delta}^B(x; 0; 6; 1; 0; 1; 0).$$

На *рис. 2* приведены графики функций принадлежности терм-множеств лингвистической переменной *ОргЗ* — «Уровень организационной защиты».

Для входной переменной *ПравЗ* терм-множество состоит из трех термов $T=\{H, C, B\}$, функции принадлежности которых являются трапециевидными.

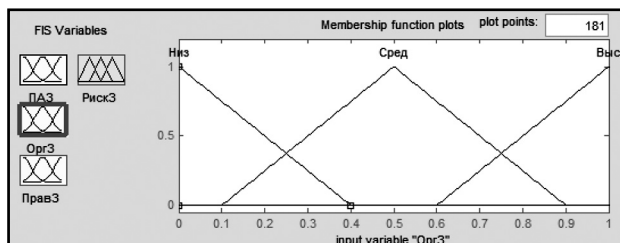


Рис. 2. Функции принадлежности для входной переменной Opr3

С учетом (3) функции принадлежности нечетких терм-множеств лингвистической переменной «Уровень правовой защиты» будут иметь следующий вид (рис. 3):

$$\mu_{\Delta}^H(x; 0; 0; 0,15; 0,45), \mu_{\Delta}^C(x; 0,05; 0,45; 0,55; 0,95), \\ \mu_{\Delta}^B(x; 0,55; 0,85; 1,0; 1,0).$$

На рис. 3 приведены графики функций принадлежности терм-множеств лингвистической переменной Прав3 – «Уровень правовой защиты».

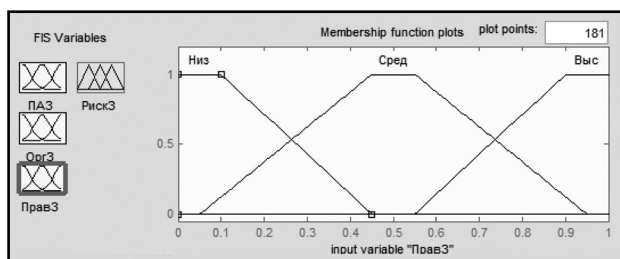


Рис. 3. Функции принадлежности для входной переменной Прав3

Для выходной переменной Риск3 (лингвистическая переменная «Риск снижения эффективности защиты») терм-множество состоит из пяти термов: $T=\{\text{Очень низкая очевидность риска (ОНОР)}; \text{Низкая очевидность риска (НОР)}; \text{Средняя очевидность риска (СОР)}; \text{Высокая очевидность риска (ВОР)}; \text{Очень высокая очевидность риска (ОВОР)}\}$. Функции принадлежности лингвистических переменных являются трапециевидными.

С учетом (3) функции принадлежности нечетких терм-множеств лингвистической переменной «Риск снижения эффективности защиты» будут иметь следующий вид:

$$\mu_{\Delta}^{\text{ОНОР}}(x; 0; 0; 0,075; 0,22), \mu_{\Delta}^{\text{НОР}}(x; 0,02; 0,2; 0,3; 0,48), \\ \mu_{\Delta}^{\text{СОР}}(x; 0,28; 0,45; 0,55; 0,72), \mu_{\Delta}^{\text{ВОР}}(x; 0,52; 0,7; 0,8; 0,98), \\ \mu_{\Delta}^{\text{ОВОР}}(x; 0,78; 0,925; 1,0).$$

На рис. 4 приведены графики функций принадлежности терм-множеств лингвистической переменной Риск3 – «Риск снижения эффективности защиты».

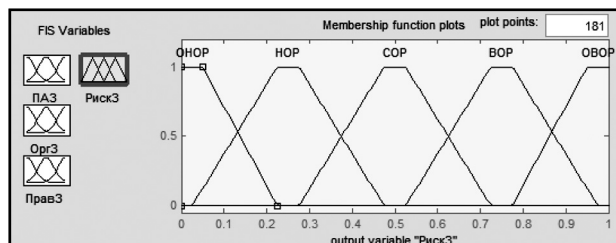


Рис. 4. Функции принадлежности для выходной переменной Риск3

Шаг 2. Задание нечетких правил. В алгоритме Мамдани база правил должна задаваться в виде структуры с тремя входами и одним выходом (рис. 5).

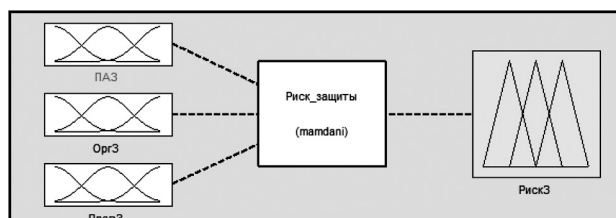


Рис. 5. Структура нечеткой модели для базы правил П1

Правила модели формируются на основе общих закономерностей поведения исследуемой системы и позволяют «вложить» в механизм вывода логическую модель прикладного уровня.

В алгоритме Мамдани для агрегирования степени истинности предпосылок используем T -норму и $\min$ -конъюнкция:

$$T(\mu_A(x), \mu_B(x)) = \min(\mu_A(x), \mu_B(x)) \quad (5)$$

Определение степени истинности заключений по каждому правилу (импликация) основано на операции $\min$ -активизации:

$$\mu_R(x, y) = \min\{\mu_A(x), \mu_B(y)\}. \quad (6)$$

Шаг 3. Аккумуляция заключения по всем правилам проведено с применением операции $\max$ -дизъюнкции. При дефазификации использован метод центра тяжести для дискретного множества значений функций принадлежности:

$$y' = \frac{\sum_{r=1}^{Y_{\max}} y_r \mu_{B'}(y_r)}{\sum_{r=1}^{Y_{\max}} \mu_{B'}(y_r)}, \quad (7)$$

где $Y_{\max}$ — число элементов уг в дискретизированной для вычисления «центра тяжести» области Y .

Реализуя систему нечеткого вывода на этапе дефазификации, получим оценку приоритета риска. Приоритезация является основной целью анализа рисков и основополагающим фактором в процессе принятия решений по управлению рисками ИБ организации.

4. Использование модели оценки рисков ИБ организации

Предположим, что на основе предварительного обследования получены некоторые оценки уровня программно-аппаратной, организационной и правовой защиты организации, которые введем в окно механизма вывода графического интерфейса *Fuzzy Logic Toolbox*.

При значении классификатора $ПАЗ = 0,7$ значение лингвистической переменной x_1 — «Программно-аппаратный уровень защиты» соответствует терму H — «достаточно, для базовой информационной защиты» с уровнем уверенности $\mu^H x_1 = 0,65\mu$. При значении классификатора $ОргЗ = 0,75$ значение лингвистической переменной x_2 — «Уровень организационной защиты» соответствует терму C — «планирование и мониторинг уязвимостей проводятся нерегулярно» с уровнем уверенности $\mu^C x_2 = 0,7\mu$. При значении классификатора $ПравЗ = 0,65$ значение лингвистической переменной x_3 — «Уровень правовой защиты» соответствует терму C — «документация имеется, но недостаточно детальная» с уровнем уверенности $\mu^C x_3 = 1,0\mu$.

По заданным исходным условиям активизируются правила 14 и 15. Результирующее значение классификатора выходной переменной $РискЗ$ соответствует значению 0,386, что определяет значение лингвистической переменной риска проекта y_1 — «Риск снижения эффективности защиты» равное НОР — «Низкая очевидность риска» с уровнем уверенности $\mu^{НОР} y_1 = 0,5\mu$.

Графический интерфейс *Fuzzy Logic Toolbox* позволяет получить график зависимости выходной величины от любой из входных переменных.

На *рис. 6* представлен график «кривой вывода» зависимости выходной переменной $РискЗ$ — «Риск снижения эффективности защиты» от входной

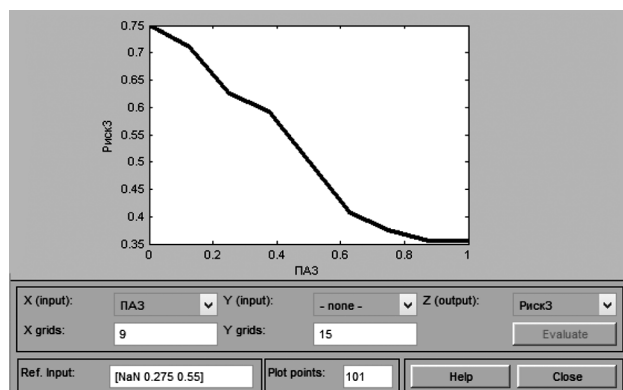


Рис. 6. Зависимость переменной РискЗ от ПАЗ

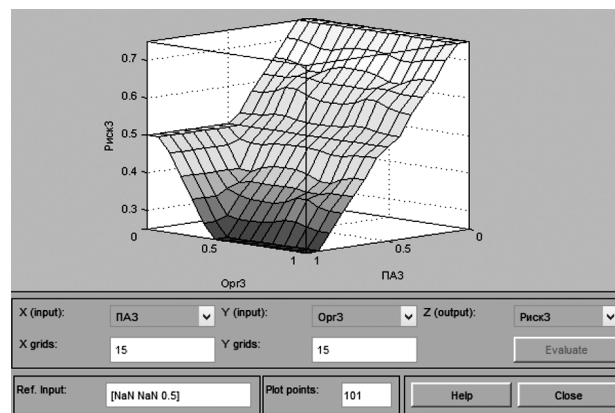


Рис. 7. Поверхность системы нечеткой модели относительно входных переменных ПАЗ и ОргЗ для базы правил П1

переменной $ПАЗ$ — «Программно-аппаратный уровень защиты» при фиксированном значении двух других входных переменных для базы правил П1 нечеткой модели.

График показывает обратную зависимость величины риска снижения эффективности защиты организации от уровня программно-аппаратной защиты, а также что значение классификатора выходной переменной $РискЗ$ не опускается ниже 0,35, что определяет значение лингвистической переменной риска информационной безопасности организации y_1 равное НОР — «Низкая очевидность риска» с уровнем уверенности $\mu^{НОР} y_1 = 0,57\mu$, при значении классификатора $ПАЗ = 0,85$, при котором значение лингвистической переменной x_1 соответствует терму B — «полностью соответствует уровню конфиденциальности информации» с уровнем уверенности $\mu^B x_1 = 0,95\mu$ и фиксированных значениях классификатора $ОргЗ = 0,275$, при котором значение лингвистической переменной x_2 соответствует терму C — «планирование и мониторинг уязвимостей проводятся нерегулярно» с уровнем уверенности $\mu^C x_2 = 0,5\mu$, и классификатора $ПравЗ = 0,55$, при котором значение лингвистической переменной x_3 соответствует терму C — «документация имеется, но недостаточно детальная» с уровнем уверенности $\mu^C x_3 = 1,0\mu$.

На *рис. 7* приведена поверхность зависимости выходной лингвистической переменной от двух входных с фиксированным значением третьей переменной для базы правил П1 нечеткой модели.

Графический вид зависимостей выходной лингвистической переменной $РискЗ$ — «Риск снижения эффективности защиты» от входных значений переменных $ПАЗ$ — «Программно-аппаратный уровень

защиты» и *Org3* – «Уровень организационной защиты» показывает закономерный рост величины риска снижения эффективности защиты организации при уменьшении уровня программно-аппаратной защиты и уровня организационной защиты.

Интерфейс *Fuzzy Logic Toolbox* обеспечивается возможность контролировать «качество» механизма вывода. Таким образом, гладкие и монотонные графики зависимостей приведенных «поверхностей вывода» и «кривых вывода» свидетельствуют о хорошем «качестве» механизма вывода и о достаточности и непротиворечивости используемых правил вывода.

5. Заключение

Разработанная нечеткая продукционная модель позволяет существенно расширить возможности существующих методик, снять ограничения на число учитываемых входных переменных и интегрировать как качественные, так и количественные подходы к оценке рисков.

В НПМ определены 7 входных лингвистических переменных, характеризующих факторы риска, 4 выходных лингвистических переменных, характе-

ризующих риски различных областей ИБ. Модель содержит 4 базы правил и позволяет проводить лингвистический анализ рисков, которые несут потенциальные угрозы и ущерб организации. Используемый в методике механизм получения оценок риска на основе нечеткой логики позволяет получить численное значение риска, лингвистическое описание степени риска, а также уровень уверенности эксперта в возникновении рискового события, которые позволяют ИТ-менеджерам выявить приоритеты рисков (очень высокий, высокий, средний, низкий, очень низкий) и выработать план мероприятий по снижению влияния наиболее опасных угроз на информационную безопасность организации.

Основная сложность механизма получения оценок риска на основе нечеткой логики состоит в построении модели для проведения лингвистического анализа рисков СОИБ, однако, данный механизм является эффективным инструментом, когда другие подходы к оценке риска неприменимы. Он обладает широкими возможностями и позволяет адаптировать его к имеющимся на предприятии моделям управления рисками, а также модифицировать с учетом реальных условий политики информационной безопасности организации. ■

Литература

1. Рекомендации по стандартизации «Информационные технологии. Основные термины и определения в области технической защиты информации» (Р 50.1.053-2005).
2. Доктрина информационной безопасности Российской Федерации (утверждена Президентом РФ В.В.Путиным 09.09.2000 г., № Пр-1895). – Российская газета, № 187, 28.09.2000 г.
3. Национальный стандарт РФ «Информационная технология. Практические правила управления информационной безопасностью» (ГОСТ Р ИСО/МЭК 17799—2005).
4. Национальный стандарт РФ «Методы и средства обеспечения безопасности. Системы менеджмента информационной безопасности» (ГОСТ Р ИСО/МЭК 27001-2006).
5. Марков А., Цирлов В. Управление рисками – нормативный вакуум информационной безопасности // Открытые системы. СУБД: Журнал для профессионалов в области информационных технологий. – 2007. – №8. – С. 63-67.
6. Симонов С.В. Анализ рисков, управление рисками. – Jet Info, 2003, №2.
7. Risk Management Guide for Information Technology Systems. – NIST, Special Publication 800-30.
8. Борисов В.В., Круглов В.В., Федюлов А.С. Нечеткие модели и сети. – М.: Горячая линия-Телеком, 2007. .
9. Заде Л.А. Понятие лингвистической переменной и его применение к принятию приближенных решений. – М.: Мир, 1976.
10. Долженко А.И. Модель анализа риска потребительского качества проектов экономических информационных систем // Вестник Северо-Кавказского государственного технического университета. – 2009. – №1 (18). – С.129-134.
11. Леоненков А.В. Нечеткое моделирование в среде MATLAB и fuzzyTECH. – СПб.: БХВ-Петербург, 2005.

МЕТОДЫ ПРЕДСТАВЛЕНИЯ И СРАВНЕНИЯ СЕМАНТИЧЕСКИ РАЗНЫХ КЛАССОВ ИЗОБРАЖЕНИЙ

Г.А. Кухарев,

доктор технических наук, профессор кафедры математического обеспечения ЭВМ Санкт-Петербургского государственного электротехнического университета, профессор факультета информатики Западно-поморского технологического университета, Щецин, Польша (West Pomeranian University of Technology, Szczecin, Poland)

Е.И. Каменская,

кандидат технических наук, инженер программного обеспечения компании Google, Цюрих, Швейцария

Н.Л. Щеголева,

кандидат технических наук, доцент кафедры математического обеспечения ЭВМ Санкт-Петербургского государственного электротехнического университета

E-mail: : kuga41@gmail.ru, ekamenskaya@google.com, stil_hope@mail.ru

Адрес: г. Санкт-Петербург, улица Профессора Попова, 5

В статье обсуждаются методы представления и сравнения семантически не связанных изображений с оценкой их меры подобия в исходном пространстве признаков, а также в пространстве канонических переменных (ПКП). Последнее достигается в рамках алгоритма 2D CCA/2D KLT для реализации проекции этих изображений в ПКП.

Ключевые слова: сравнение семантически не связанных изображений, канонический корреляционный анализ, фазовая корреляция, индекс подобия.

1. Введение

В последнее десятилетие активно развивается область информатики, связанная с биометрическими технологиями распознавания и понимания людей [1 – 3]. В рамках этих технологий решаются как относительно простые задачи, например, поиск людей в больших базах по фотопор-

третам, идентификация и аутентификация людей, так и более сложные задачи – понимание намерений людей в процессе распознавания и интерпретации их эмоционального и психофизического состояния, а также задачи, охватываемые, проблемой человеко-компьютерного взаимодействия (human-computer interaction) или проблемой интерфейса мозг-компьютер (brain-computer interface).

Особенность этих задач состоит в наблюдении за человеком с разных, внешне не связанных между собой, сторон и использование информации этих наблюдений для интерпретации или понимания одних наблюдений через другие. Именно поэтому, в последнее время интересы специалистов по компьютерным технологиям и распознаванию образов обратились к математическим методам, позволяющим трансформировать два (не связанных между собой) набора данных из исходного пространства признаков (ИПП) в новое общее подпространство признаков, в котором исходные наборы данных сильно коррелируют. Эти методы строятся на основе проекций на латентные структуры (Partial Least Squares, Projection Latent Structures – PLS) и канонического корреляционного анализа (Canonical Correlation Analysis – CCA).

Метод CCA был представлен Хоттелингом в 30-х годах прошлого столетия и описывал соотношение между двумя **одномерными** рядами данных [7] с прогнозом его применения к многомерным данным. Однако сложность и объем вычислений в CCA привела к тому, что потребовалось семьдесят лет ожиданий на первое практическое применение CCA к наборам цифровых изображений. При использовании CCA в приложении к обработке цифровых изображений возникли две проблемы. Первая связана с многомерностью исходных данных – изображений, как объектов 2D, с третьим измерением, определяющим яркость и четвертым измерением, определяющим цветовые компоненты. Вторая связана с проблемой малой выборки (Small Sample Size) [20], когда размерность ИПП значительно превосходит число изображений в исходной выборке.

Поэтому публикации, связанные с широким применением методов PLS и CCA в задачах обработки изображений вообще и в задачах биометрии, в частности, появились только в последние 5 – 7 лет, когда сформировались новые подходы к реализации PLS и CCA на основе двумерных методов проекции.

Однако и здесь оказалось все не так просто. Можно с сожалением отметить, например, что работа [10], в которой впервые была представлена революционная идея реализации двумерного анализа главных компонент и двумерного преобразования Карунена-Лоэва в приложении к обработке изображений, осталась незамеченной большинством исследователей. Более того, исказилась история развития этого метода: ошибочно считается, что первый раз этот метод был представлен в работе [13], хотя речь в ней идет о проекции только по

одному направлению (а именно по строкам изображения), что не соответствует названию «**Two-dimensional PCA**».

Подчеркнем еще раз, что история метода двумерных проекций в приложении к обработке изображений началась с работы [10], продолжена в работах [1, 3, 11, 12], а для задач, решаемых на основе CCA и PLS, развита и показана в работах [14, 15]. Полная история развития методов представления и обработки изображений в подпространствах, определяемых на основе двумерных проекций в собственных базисах, получаемых в рамках PCA, LDA, PLS и CCA, показана в работе [15].

В работах [14, 15] рассмотрено решение нескольких прикладных задач с использованием бенчмарковых баз изображений лиц ORL, Face94, FERET, EQUINOX. Здесь были показаны: специфика обработки и взаимного распознавания изображений разных категорий (например, инфракрасных, термо- и видео- изображений), представляющих одно и то же лицо человека; особенности представления и обработки гендерно смешанных лиц (мужского и женского), состоящих в одном классе и методы их взаимного распознавания. При этом были рассмотрены случаи, когда все изображения в составах X и Y относились к одному и тому же глобальному классу – классу «лица людей». Возможно, это повлияло на полученные высокие результаты корреляции в подпространствах.

Цель настоящей статьи – представление методов сравнения семантически не связанных изображений с оценкой меры их подобия в исходном пространстве признаков, а также представление алгоритма 2D CCA/2D KLT для реализации проекции этих изображений в ПКП и оценка меры их подобия в ПКП.

2. Структура вычислений в CCA/KLT

В рамках CCA информация об объекте (форма, модель или иная группа признаков) переводится из ИПП в ПКП, и все дальнейшие операции осуществляются в этом пространстве. Рассмотрим структуру вычислений в CCA.

Пусть нам заданы два набора исходных данных X и Y , состоящие из K векторов размером $DIM \times 1$ каждый так, что:

$$X = [X^{(1)} X^{(2)} \dots X^{(K)}] \text{ и } Y = [Y^{(1)} Y^{(2)} \dots Y^{(K)}]. \quad (1)$$

Целью CCA является нахождение двух матриц проекции, трансформирующих исходные

данные в пространство канонических переменных $X^{(k)} \rightarrow U^{(k)}$ и $Y^{(k)} \rightarrow V^{(k)}$ так, чтобы выполнялось условие: $\|U^{(k)} - V^{(k)}\| \rightarrow \min$. Исходные данные $X^{(k)}$ и $Y^{(k)}$ могут быть не связаны между собой корреляцией, в то время как канонические переменные $U^{(k)}$ и $V^{(k)}$, полученные в результате ССА, связываются между собой устойчивой корреляцией, максимум которой достигается при решении двух совместных задач на собственные значения [7]:

$$\begin{cases} (C_{xx}^{-1}C_{xy}C_{yy}^{-1}C_{yx})W_x = W_x\Lambda_x \\ (C_{yy}^{-1}C_{yx}C_{xx}^{-1}C_{xy})W_y = W_y\Lambda_y, \end{cases} \quad (2)$$

где: $C_{xx}, C_{yy}, C_{xy}, C_{yx}$ — матрицы ковариации порядка DIM, причем

$C_{xx} = XX^T, C_{yy} = YY^T, C_{xy} = XY^T, C_{yx} = C_{xy}^T$; Λ_x, Λ_y — диагональные матрицы собственных значений; W_x и W_y — матрицы собственных векторов (матрицы проекций).

Если представить результаты ССА следующим образом:

$$U = [U^{(1)}U^{(2)} \dots U^{(K)}] \text{ и } V = [V^{(1)}V^{(2)} \dots V^{(K)}], \quad (3)$$

то условие $\|U^{(k)} - V^{(k)}\| \rightarrow \min$ для $\forall k$, заменяется условием $\|U - V\| \rightarrow \min$, которое достигается при максимизации отношения:

$$\rho^2 = \frac{w_x^T C_{xy} w_y}{\sqrt{(w_x^T C_{xx} w_x)(w_y^T C_{yy} w_y)}} = \max, \quad (4)$$

что связано с поиском векторов w_x и w_y , входящих в матрицы W_x и W_y .

Обозначение ССА/KLT подчеркивает тот факт, что реализация собственно канонического корреляционного анализа (ССА) является отдельным этапом, на котором определяются матрицы проекции W_x и W_y (или собственный базис). Заметим, что первый этап может выполнять самостоятельную роль — анализ исходных данных.

Второй этап — собственно проекция исходных данных в пространство канонических переменных (KLT подчеркивает реализацию проекции в собственном базисе, определенном матрицами W_x и W_y) выполняется следующим образом:

$$U = W_x^T X, \quad V = W_y^T Y. \quad (5)$$

В приложении к обработке изображений в биометрии входной информацией для ССА могут быть некоторые признаки исходных изображений (например, яркостные или спектральные в базисе Фурье или базисе собственного преобразования),

антропометрические точки на области лица или координаты (параметры) эластичной модели лица). В отличие от исходных признаков, канонические переменные U и V всегда относятся к категории спектральных признаков в собственном базисе, а число признаков d определяется выбором главных компонент на первом этапе (или числом выбранных базовых функций в матрицах W_x и W_y на втором этапе). В любом случае, размерность d всегда меньше размерности DIM (то есть $d \ll DIM$), что обычно связывают с редукцией размерности пространства признаков.

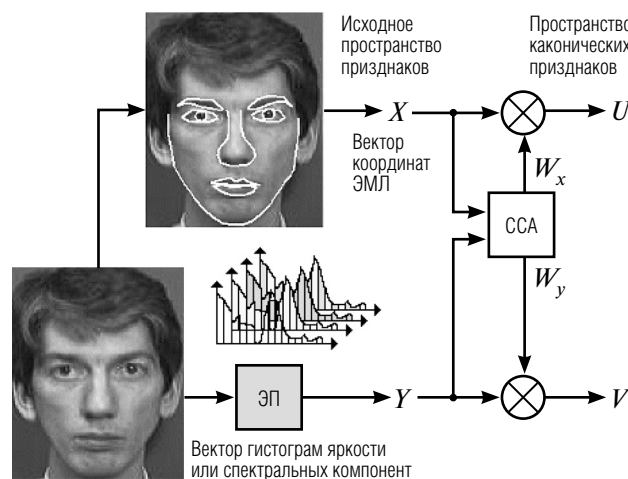


Рис. 1. Пример системы, основанной на ССА

На рис. 1 показан пример системы распознавания изображений лиц на основе ССА, где входными данными являются параметры ЭМЛ и гистограммы яркости изображений или параметры ЭМЛ и спектр изображения лица. Система реализует идею «multirepresentation» — представление одного и того же лица различными составами признаков, которые в общем случае не коррелируют между собой.

Эта система представляет собой типичный пример ранних идей применения ССА в приложении к обработке изображений (как двумерных данных), в которой отразился этап векторного представления исходных изображений (ИПП). При этом часто перед векторизацией исходных изображений, предварительно уменьшался их размер и только после этого выполнялся ССА или сначала выполнялась векторизация исходных изображений, затем редукция размерности ИПП с использованием PCA/KLT или LDA/KLT и, только потом, ССА. Аналогичный путь прошло и применение PLS, поэтому вместо блока ССА на рис. 1, можно было бы поставить блок PLS.

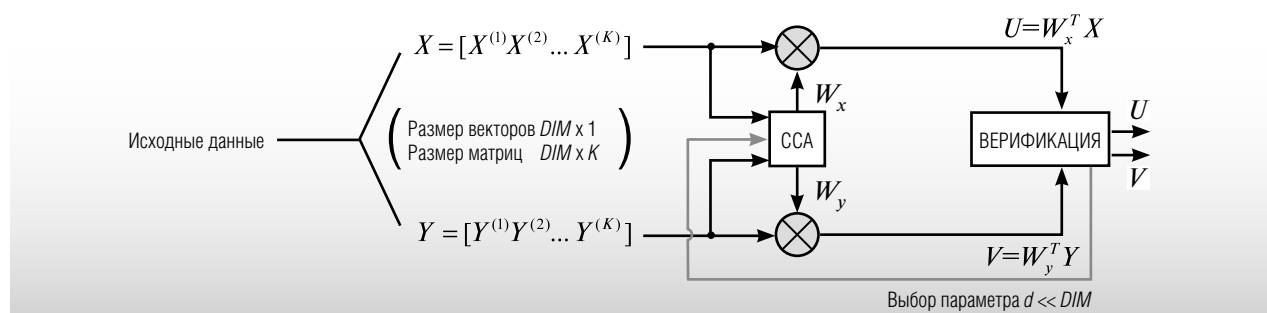


Рис. 2. Исходные данные и вычисления в CCA/KLT

Исходные данные, структуру вычислений в CCA, а также их результаты можно схематически представить так, как показано на рис. 2, где блоки « $\otimes$ » реализуют процедуру проекции исходных данных в новое пространство переменных.

Вычисления в CCA заканчиваются выбором параметра d редукции размерности пространства признаков (причем $d < DIM$), нахождением коэффициентов корреляции $\rho_k = \sqrt{\lambda_k} \forall k$ и проверкой качества взаимного соответствия переменных в новом пространстве признаков. Эти операции реализуются в блоке «Верификация».

Вычислительная сложность в CCA/KLT определяется размерностью DIM , а возможность устойчивого решения задач на собственные значения связана с условием:

$$DIM < K \quad (6)$$

Выполнение этого условия особенно важно в приложении к обработке изображений, поскольку размерность DIM связана с размерами $M \times N$ изображения (или общим числом его пикселей MN). Размерность исходного признакового пространства в этом случае составляет $DIM = MN$ и обычно $DIM \gg K$, что известно как проблема малой выборки. Именно поэтому, методология CCA в приложении к обработке изображений в биометрии и была основана на двух подходах, связанных, с одной стороны, с обеспечением выполнения условия (6), а с другой стороны, с возможностью использования базового метода [7], изначально ориентированного на обработку одномерных данных, представляющих каждый образ.

Представим метод реализации CCA, не требующий предварительной процедуры «DownSampling», и не требующий формального выполнения условия (6). Этот метод развивает идеи двумерной проекции [10].

3. Алгоритмы 2D CCA/2D KLT

Пусть заданы два набора исходных данных, состоящие из K изображений размером $M \times N$ каждое:

$$X = [X^{(1)} X^{(2)} \dots X^{(K)}] \text{ и } Y = [Y^{(1)} Y^{(2)} \dots Y^{(K)}], \quad (7)$$

где $X^{(k)}$ и $Y^{(k)}$ — пары изображений, при этом $MN \gg K$.

Определим средние образы:

$$\bar{X} = \frac{1}{K} \sum_{k=1}^K X^{(k)} \text{ и } \bar{Y} = \frac{1}{K} \sum_{k=1}^K Y^{(k)} \quad \forall k=1, 2, \dots, K. \quad (8)$$

Положим $\tilde{X}^{(k)} = (X^{(k)} - \bar{X})$ и $\tilde{Y}^{(k)} = (Y^{(k)} - \bar{Y})$, тогда вместо (8) получим наборы данных, центрированных относительно средних значений:

$$\tilde{X} = [\tilde{X}^{(1)} \tilde{X}^{(2)} \dots \tilde{X}^{(K)}]; \tilde{Y} = [\tilde{Y}^{(1)} \tilde{Y}^{(2)} \dots \tilde{Y}^{(K)}], \quad (9)$$

для которых вычислим по четыре матрицы ковариации:

$$C_{xx}^{(r)} = (\tilde{X}) \tilde{X}^T; C_{yy}^{(r)} = (\tilde{Y}) \tilde{Y}^T; C_{xy}^{(r)} = \tilde{X} (\tilde{Y})^T; C_{yx}^{(r)} = (C_{xy}^{(r)})^T; \quad (10a)$$

$$C_{xx}^{(c)} = (\tilde{X})^T \tilde{X}; C_{yy}^{(c)} = (\tilde{Y})^T \tilde{Y}; C_{xy}^{(c)} = (\tilde{X})^T \tilde{Y}; C_{yx}^{(c)} = (C_{xy}^{(c)})^T. \quad (10b)$$

В выражении (10) знаки « r » и « c » подчеркивают факт вычисления матриц ковариации относительно строк (*row*) и столбцов (*column*) исходных изображений.

Вычислим (с использованием матриц (10)), по две общие (total) матрицы рассеяния:

$$S^{(total1,r)} = [C_{xx}^{(r)}]^{-1} C_{xy}^{(r)} [C_{yy}^{(r)}]^{-1} C_{yx}^{(r)}, \quad (11a)$$

$$S^{(total2,r)} = [C_{yy}^{(r)}]^{-1} C_{yx}^{(r)} [C_{xx}^{(r)}]^{-1} C_{xy}^{(r)}.$$

$$S^{(total1,c)} = [C_{xx}^{(c)}]^{-1} C_{xy}^{(c)} [C_{yy}^{(c)}]^{-1} C_{yx}^{(c)} \quad (11b)$$

$$S^{(total2,c)} = [C_{yy}^{(c)}]^{-1} C_{yx}^{(c)} [C_{xx}^{(c)}]^{-1} C_{xy}^{(c)}$$

В рамках двумерных методов, целью 2D CCA является определение четырех матриц проекции W_{x1} , W_{x2} и W_{y1} , W_{y2} трансформирующих исходные

данные в пространство переменных $X^{(k)} \rightarrow U^{(k)}$ и $Y^{(k)} \rightarrow V^{(k)}$, что достигается решением четырех задач на собственные значения:

$$\begin{aligned} S^{(total1,r)} W_{x_1}^{(r)} &= \Lambda_{x_1}^{(r)} W_{x_1}^{(r)}; & S^{(total1,c)} W_{x_2}^{(c)} &= \Lambda_{x_2}^{(c)} W_{x_2}^{(c)}; \\ S^{(total2,r)} W_{y_1}^{(r)} &= \Lambda_{y_1}^{(r)} W_{y_1}^{(r)}; & S^{(total2,c)} W_{y_2}^{(c)} &= \Lambda_{y_2}^{(c)} W_{y_2}^{(c)}, \end{aligned} \quad (12)$$

где $\Lambda_{x_1}, \Lambda_{y_1}, \Lambda_{x_2}, \Lambda_{y_2}$ – диагональные матрицы собственных значений для данных по строкам (1) и столбцам (2) соответственно; W_{x_1}, W_{x_2} и W_{y_1}, W_{y_2} – матрицы проекции.

Трансформация исходных данных в новое пространство реализуется как двумерное преобразование Карунена-Лоэва в форме:

$$U^{(k)} = W_{x_1}^T \bar{X}^{(k)} W_{y_1} \text{ и } V^{(k)} = W_{x_2}^T \bar{Y}^{(k)} W_{y_2}, \forall k, \quad (13)$$

где $U^{(k)}, V^{(k)}$ – матрицы, представляющие исходные изображения в новом пространстве признаков.

Определим главные компоненты. Ими будут те из них, которые имеют наибольшие значения. При этом основная изменчивость матриц ковариации может быть «описана» всего $d \ll K$ наибольшими собственными значениями, т.е. если собственные значения упорядочены по убыванию, то при любых $d_1 \geq 1$ и $d_2 \geq 1$ всегда выполняются следующие условия:

$$\sum_{i=1}^{d_1} (\lambda_i^{(r)})^2 \gg \sum_{i=d_1+1}^M (\lambda_i^{(r)})^2 \text{ и } \sum_{i=1}^{d_2} (\lambda_i^{(c)})^2 \gg \sum_{i=d_2+1}^N (\lambda_i^{(c)})^2 \quad (14)$$

где $\lambda_i^{(r)}$ и $\lambda_i^{(c)}$ – собственные значения; $d_1 \ll M$; $d_2 \ll N$ и $d_1 \neq d_2$ в общем случае.

Нижняя граница параметра d может быть выбрана экспериментально на этапе верификации результата выполненного анализа по критерию (4).

Для редукции размерности процедуру (13) необходимо модифицировать так, чтобы в проекции участвовали только те собственные векторы, которые соответствуют d главным компонентам. Для этого, из матриц W_{x_1}, W_{y_1} выберем d строк, соответствующих d наибольшим собственным значениям, и на их основе сформируем матрицы редукции F_{x1} и F_{y1} . А из матриц W_{x_2}, W_{y_2} выберем d столбцов, соответствующих d наибольшим собственным значениям, и на их основе сформируем матрицы F_{x2} и F_{y2} .

Для каждого исходного изображения выполним «усеченное» двумерное преобразование Карунена-Лоэва, которое может быть представлено в следующей форме:

$$\hat{U}^{(k)} = F_{x1} \bar{X}^{(k)} F_{x2}, \hat{V}^{(k)} = F_{y1} \bar{Y}^{(k)} F_{y2} \forall k. \quad (15)$$

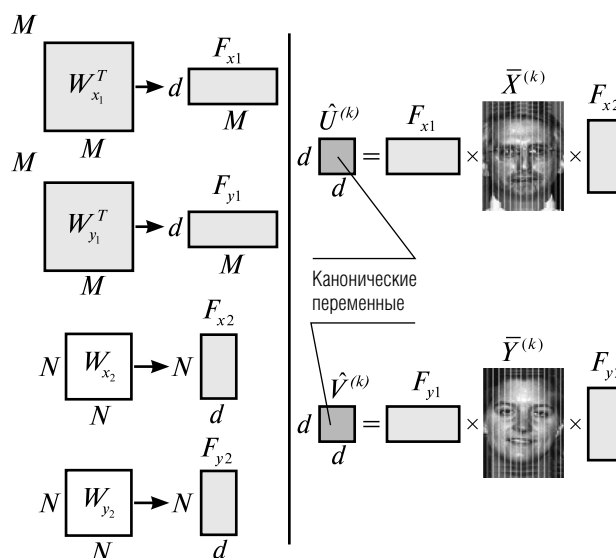


Рис. 3. Матрицы проекции и их использование

Здесь знак «^» определяет отличие результата от (13).

Размер матриц-результатов в (15) равен $d \times d$ или $d_1 \times d_2$. Результирующие матрицы (наборы) переменных определяемые как

$$\hat{U} = [\hat{U}^{(1)} \hat{U}^{(2)} \dots \hat{U}^{(K)}] \text{ и } \hat{V} = [\hat{V}^{(1)} \hat{V}^{(2)} \dots \hat{V}^{(K)}] \quad (16)$$

будут иметь размеры $d \times d \times K$ или $d_1 \times d_2 \times K$.

Реализация этапа формирования матриц проекции и выполнение проекции (15) отражена на рис. 3: формирование матриц F_{x1}, F_{x2}, F_{y1} и F_{y2} (слева) и схема вычисления переменных (справа). И, как видно, именно на этом этапе и реализуется двумерное преобразование Карунена-Лоэва (2D KLT – Two Dimensional Karhunen-Loeva Transform).

4. Характеристика метода 2D CCA/2D KLT

Метод 2D CCA/2D KLT основан на представлении исходных изображений как совокупности строк и столбцов и реализации CCA по этим совокупностям.

Метод не требует предварительного уменьшения размеров исходных изображений, не является итерационным и прямо реализуется по двум направлениям координат – а именно по строкам и столбцам исходного изображения, что определено в (13) и (15).

Максимальный порядок матриц ковариации в методе 2D CCA/2D KLT равен $DIM = \max\{M, N\}$,

что предопределяет практическую возможность решения задачи на собственные значения и стабильность этого решения даже для изображений очень больших размеров.

Кроме того, проблема SSS в этих методах также обходится стороной, поскольку вместо каждого изображения размером $M \times N$ фактически используется N изображений размером $M \times 1$ и M изображений размером $N \times 1$. При таком представлении исходных данных, всегда выполняется условие: $DIM = \max\{M, N\} < (KM)$ или $DIM = \max\{M, N\} < (KN)$, если $K \geq 2$.

5. Исходная база изображений для экспериментов

Рассмотрим случай, когда исходные изображения относятся к разным глобальным классам. Примером являются изображения базы «Люди и Собаки» («People and Dogs»), содержащей пары изображений — портреты хозяина и его собаки, подобранные так, чтобы показать, что между портретами существует некоторое подобие (см. рис. 4) [16].

При помощи данной базы изображений можно выяснить два принципиальных вопроса:

- 1) насколько наша визуальная оценка подобия этих портретов в исходной их форме, сравнима с формальными оценками подобия этих портретов;
- 2) как изменяется подобие этих портретов в ПКП.

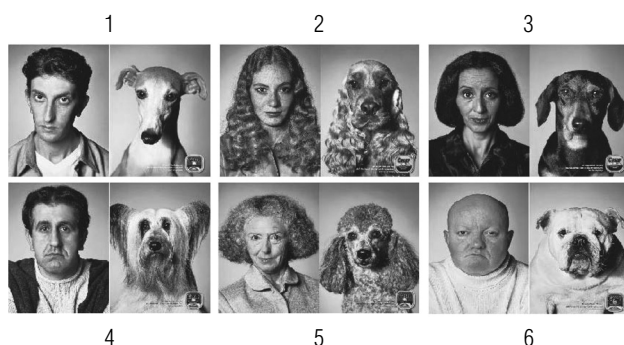


Рис. 4. Примеры пар изображений базы «Люди и собаки» [16]

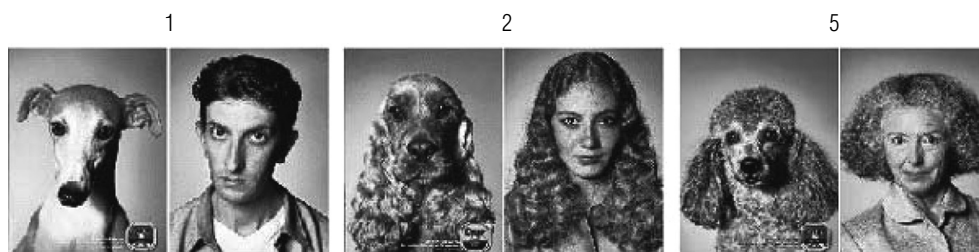


Рис. 5. Выбранные пары изображений

6. Сравнение в исходном пространстве признаков

Пары изображений на рис. 4, имеют некоторое внешнее сходство между хозяином и его собакой, которое определяется следующими факторами: близким по форме выражением лица хозяина и «морды» собаки; одинаковым ракурсом двух портретов; близкой по форме прической хозяина и экстерьером собаки; цветовой гаммой — одинаковый цвет волос хозяина и окраски шерсти собак, соответствующий цвет одежды.

Очевидно, что наиболее просто можно проверить отмеченное сходство по цветовой гамме. Формально его можно оценить, сравнивая соответствующие яркостные гистограммы исходных цветных изображений.

Покажем это для 1, 2 и 5 пар исходных изображений, поскольку они имеют наиболее выразительное сходство (см. рис. 5).

Вычислим для каждого изображения цветовую яркостную гистограмму [19] и фазовую корреляцию для каждой пары изображений [1] (см. рис. 6).

Сравнивая гистограммы изображений каждой пары можно «на глаз увидеть их сходство», которое формально оценено через фазовую корреляцию. Пик этой корреляции превышает значение 0.75, при относительно малых ее боковых значениях, что свидетельствует о высокой степени подобия между изображениями внутри каждой пары. Максимум корреляции (≈ 0.9) достигается для пары № 1, изображения которой имеют наиболее выразительное сходство.

Тем самым показано, что, по крайней мере, субъективная и формальная оценки фактора цветовой гаммы указывают на существование подобия в двух изображениях, семантически не связанных между собой.

На рис. 7 и 8 представлены изображения базы «Люди и Собаки» и результаты фазовой корреляции самих исходных изображений. Здесь: 1, 2 — портрет человека и собаки; 3 — автокорреляции для портрета человека; 4 — взаимная фазовая корреляция между портретом человека и собаки.

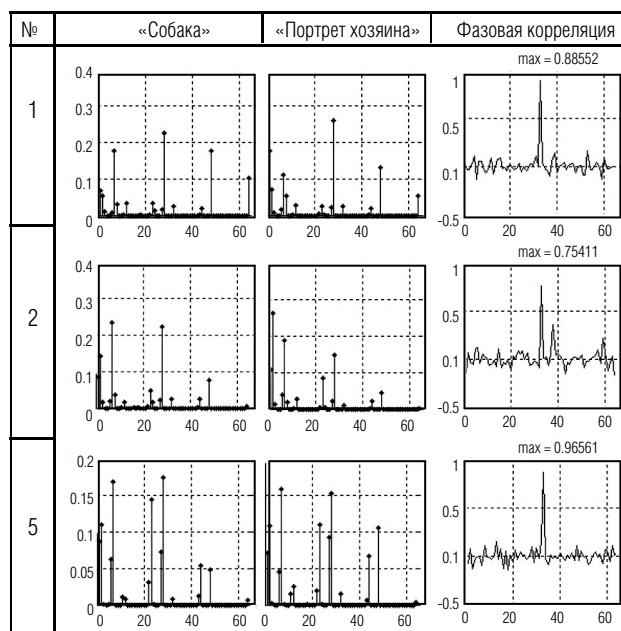


Рис. 6. Яркостные гистограммы исходных цветных изображений

Фазовая автокорреляция представлена ярко выраженным импульсом со значением равным 1. Взаимная корреляция при этом практически равна 0. Таким образом, между парами исходных изображений нет никакой корреляции (!), в то время как наша субъективная оценка и предыдущий эксперимент свидетельствует об обратном – нам кажется, что портреты подобны между собой! И трудно даже представить, какими формальными способами это можно подтвердить и оценить.

На нашу субъективную оценку влияют факторы цвета, форма и текстура изображений, поэтому сравним их по индексу структурного подобия (Structural SIMilarity Index – SSIM (I_{SSIM}) [17,18]), который позволяет оценить степень подобия (искажения) двух изображений как комбинацию трех факторов: яркостных изменений, изменений контраста и потери корреляции [18]:

$$I_{SSIM}(X, Y) = [L(X, Y)]^\alpha [C(X, Y)]^\beta [R(X, Y)]^\gamma, \quad (17)$$

где: X, Y – пара исходных изображений; $\alpha > 0, \beta > 0, \gamma > 0$ определяют значимость каждого фактора в (17).

Поскольку исходные изображения «в парах» по субъективным оценкам не значительно отличаются друг от друга, положим $\alpha = \beta = \gamma = 1$. Тогда изменения яркости L , контраста C и корреляции R можно выразить следующим образом:

$$L(X, Y) = (2\mu_x\mu_y + C1) / (\mu_x^2 + \mu_y^2 + C1), \quad (18)$$

$$C(X, Y) = (2\sigma_x\sigma_y + C2) / (\sigma_x^2 + \sigma_y^2 + C2), \quad (19)$$

$$R(X, Y) = (\sigma_{xy} + C3) / (\sigma_x\sigma_y + C3), \quad (20)$$

где μ_x и μ_y – среднее значение каждого изображения в паре; σ_x и σ_y – стандартное отклонение; σ_x^2 и σ_y^2 – дисперсии; σ_{xy} – ковариация изображений X и Y .

При обработке изображений может возникнуть ситуация, когда знаменатели в (18) – (20) становятся близки к нулю. Использование констант $C1, C2$ и $C3$ позволяет это учесть [18]. Для изображений шкалы GRAY с максимальным уровнем серого цвета имеем $Lmax = 255$, поэтому, если $C1 = (K1L)2$,

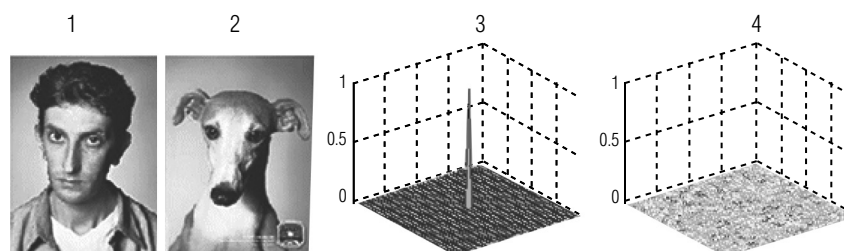


Рис. 7. Исходные изображения (пара 1)

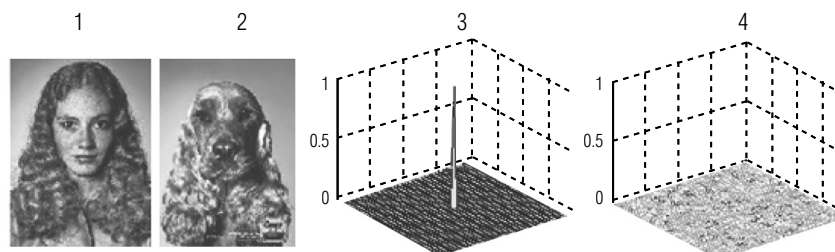


Рис. 8. Исходные изображения (пара 2)

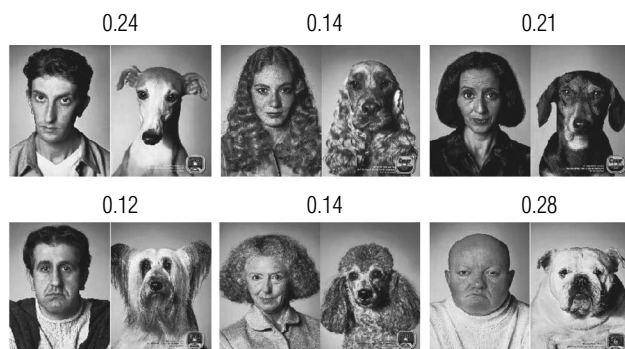


Рис. 9. Значения SSIM для исходных пар изображений

$C2 = K2L/2$, $C3 = C2/2$ [17] и $K1 = 0.01$ и $K2 = 0.03$, то значения констант $C1$, $C2$ и $C3$ могут быть легко вычислены.

Результат оценки SSIM для исходных пар изображений показаны на рис. 9.

Если положить $C1 = C2 = 0$, то (17) сводится к универсальному индексу качества (Universal Quality Index – UQI [17]), который является предшественником индекса SSIM. Тогда (20), при отсутствующем $C3$, представляет собой амплитудную линейную корреляцию двух изображений [1]. Из оценок на рис. 9 видно, что индекс SSIM также не выявил заметного сходства внутри пар изображений «хозяин-собака».

7. Представление семантически разных классов изображений в ПКП

Теперь сформулируем новую цель – проверка наличия корреляции между изображениями «хозяин-собака» в ПКП на основании двумерных проекций (13) и (15).

На рис. 10 приведены пары исходных данных из набора X (набор портретов хозяина) и Y (набор портретов собаки) для одного и того же класса и разных классов, а также коэффициенты взаимной фазовой

корреляции между ними. И, как видно, в ПКП фазовая корреляция «в своем классе» определена на уровне ≈ 0.6 , что существенно выше фазовой корреляции для изображений разных классов (здесь $\max \approx 0.15$) и существенно выше соответствующих значений SSIM.

Рассмотрим рис. 11, где представлены новые результаты анализа подобия образов в ПКП. Здесь в левом верхнем углу показан график зависимости $U = f(V)$, далее – две пары исходных изображений и, наконец, – канонические переменные U и V , соответствующие исходным изображениям.

Легко заметить, что три канонические переменные из четырех содержат одинаковые значения (яркостные элементы) и, следовательно, даже внешне похожи друг на друга.

В нижнем ряду рис. 11 слева направо показаны следующие виды фазовой корреляции:

- между всеми переменными U и V ;
- между переменными, соответствующими изображениям из одного класса;
- между переменными, соответствующими изображениям из разных классов.

Из представленных результатов видно, что в ПКП между переменными U и V появилась корреляция. При этом корреляция между переменными из одного класса значительно выше значения 0,5. Корреляцию в другом случае можно считать несущественной, а отношение этих корреляций составляет 4 раза.

Графическое представление связи между переменными в новом пространстве признаков для ССА, полученными по (15), показано на рис. 12.

Заметим, что точечные области представления канонических переменных (слева) – как функции U от V связаны практически линейной зависимостью (прямая линия, проходящая через середину точечной области). И, кроме того, с ростом значений V

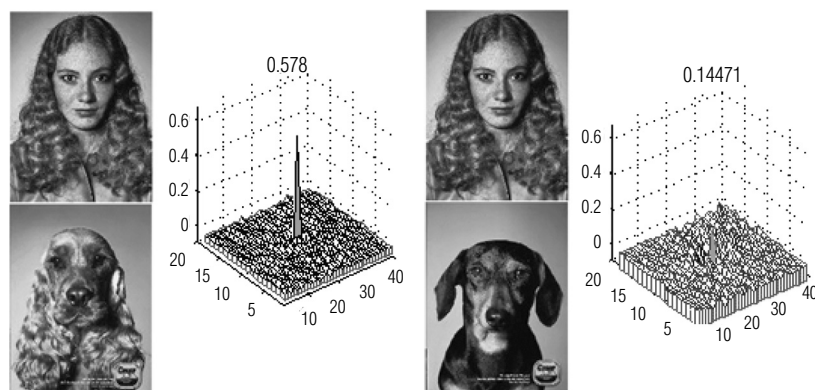


Рис. 10. Исходные данные, вид и коэффициенты фазовой корреляции

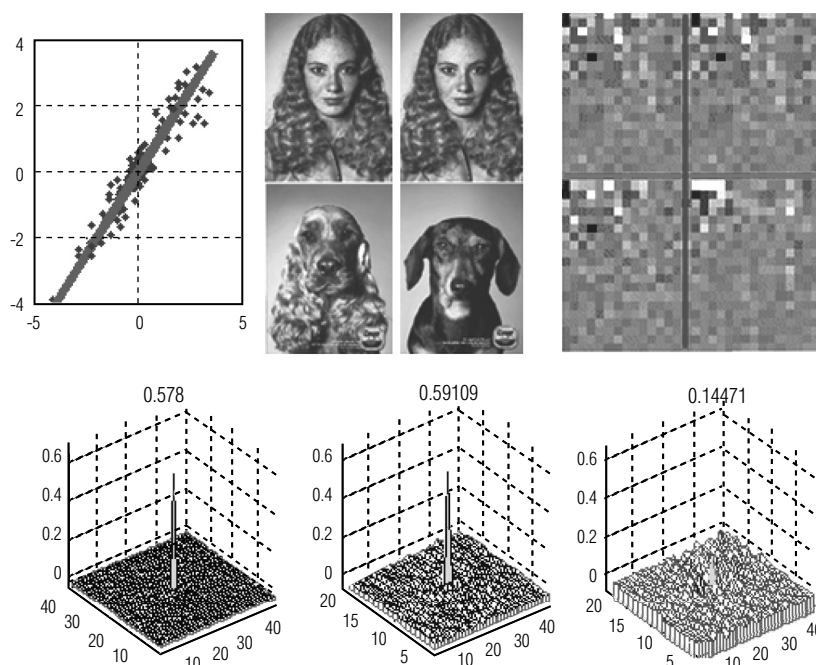


Рис. 11. Исходные данные и результаты их преобразований

растут и значения U , что соответствует критерию (4).

Другой вариант соотношений переменных U и V показан на рис. 12 справа — кривые $U^{(k)}$ и $V^{(k)}$ практически синхронно совпадают (коррелируют между собой) и находятся «в фазе».

Высокая корреляция переменных U и V в ПКП, их линейная зависимость друг от друга, «синхронизм в поведении» создают предпосылки для взаимного распознавания и модельного отображения одних переменных по другим. И эти задачи легко реализуется в ПКП, что было показано, например, в работах [13, 14].

Аналогичные результаты были получены и для других пар исходных образов, хотя мы исходили из того, что соответствующие пары исходных образов (хозяин-собака) принадлежали к разным глобальным (или семантически разным) классам и факти-

чески их подобие было доказано только на уровне цветовой гаммы.

Главный итог выполненного анализа состоит в следующем:

1) подтверждается факт установления корреляции между образами собаки и ее хозяина в ПКП, в то время как никакими другими способами это подтвердить не удалось;

2) используя п. 1, можно в рамках методов 2D ССА решать задачи индексирования одних изображений через другие (поиск, распознавание, модельное отображение одних образов в другие, реконструкция образов);

В целом же, показанный пример использования разработанных алгоритмов 2D ССА/2D KLT показал, что они могут быть широко использованы в задачах распознавания и классификации образов и сокращения избыточности их представления. Последнее связано с тем, что выбираемое значение параметра $d \ll MN$ — размерности ИПП.

8. Заключение

В статье обсуждались методы представления и сравнения семантически не связанных изображений с оценкой их меры подобия в исходном пространстве признаков, а также представление алгоритма 2D ССА/2D KLT для реализации проекции

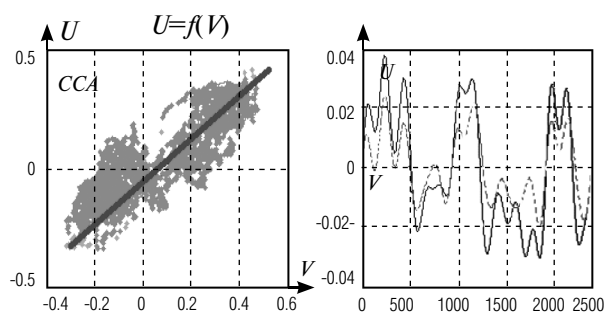


Рис. 12. Соотношение переменных в новом пространстве признаков

этих изображений в ПКП и оценка их меры подобия в ПКП.

Для сравнения изображений в исходном пространстве признаков использованы гистограммы цветных изображений и их фазовая корреляция, двумерная фазовая корреляция исходных изображений, а также индекс структурного подобия. Однако только частично удалось доказать подобие, соответствующее субъективным ощущениям для выбранных примеров.

При этом показано, что «неподобное в исходном пространстве признаков» может быть подобным в ПКП. Проекция в ПКП реализуется в рамках метода 2D ССА/2D KLT, специально предназначенного для обработки двух составов изображений и детально представленного в статье. И этот результат показал, что методы 2D ССА/2D KLT могут быть широко использованы в задачах поиска, распознавания и классификации образов, а также сокращения избыточности их представления независимо от их семантической взаимосвязи. ■

Литература

1. Kukharev G. Biometric Systems: Methods and Means of People Identification. — Sankt-Petersburg: Politehnika, 2001.
2. Encyclopedia of Biometrics / Li Stan Z. (editor). — Springer Science+Business Media, 2009.
3. Kukharev G., Kuzminski A. Biometric Technology: Face Recognition Methods. — Szczecin: WI PS, 2003.
4. Bie Tijl De, Nello C., Roman R. Eigenproblems in Pattern Recognition // Handbook of Geometric Computing: Applications in Pattern Recognition, Computer Vision, Neuralcomputing, and Robotics. — Heidelberg: Springer-Verlag, 2005, P. 1-39.
5. Pearson K. On lines and planes of closest fit to systems of points in space // Edinburgh and Dublin Philosophical Magazineand Journal of Sciences. — 1901. — V. 6, № 2. — P. 559-572.
6. Hotelling H. Analysis of complex variables into principal components // Journal of Educational Psychology. — 1933. — №24. — P. 417-441.
7. Hotelling H. Relations between two sets of variates // Biometryka. — 1936. — №28. — P. 321-377.
8. Sirovich L., Kirby M. Low-dimensional Procedure for the Characterization of Human Faces // Journal of the Optical Society of America, A - Optics, Image Science and Vision. — 1987. — V. 4, №3. — P. 519-524.
9. Turk M., Pentland A. Eigenfaces for Recognition // Journal of Cognitive Neuroscience. — 1991. — V. 3, №1. — P. 71-86.
10. Tsapatsoulis N., Alexopoulos V., Kollias S. A Vector Based Approximation of KLT and Its Application to Face Recognition. — Proc. IX European Signal Processing Conf. EUSIPCO'98, Greece. — 1998. — V. III. — P. 1581-1584.
11. Kukharev G., Forczmański P. Data Dimensionality Reduction for Face Recognition // Journal Machine Graphics & Vision. — 2004. — V.13 №1/2. — P. 99-121.
12. Kukharev G., Forczmanski P. Face Recognition by Means of Two-Dimensional Direct Linear Discriminant Analysis // PRIP'05, Minsk, Republic of Belarus. — 2005. — V. 2. — P. 63-67.
13. Yang J., Zhang D., Frangi A.F., Yang J-Y. Two-dimensional PCA: A new approach to appearance-based face representation and recognition // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2004. — №26(1). — P. 131-137.
14. Kukharev G., Kamenskaya E. Application of Two_Dimensional Canonical Correlation Analysis for Face Image Processing and Recognition // Pattern Recognition and Image Analysis. — 2010. — V. 20, №2. — P. 210-219.
15. Kukharev G., Tujaka A., Forczmański P. Face Recognition using Two-dimensional CCA and PLS // International Journal of Biometrics. — 2011. — №3. — P. 300-321.
16. www.popular-pics.com/Funny_People_And_Dog_Similarity_Pictures__1
17. Wang Z., Bovik A.C. A universal image quality index // IEEE Signal Process. Letters. — 2002. — V. 9, №3. — P. 81-84.
18. Dosselmann R., Yang X.D. A comprehensive assessment of the structural similarity index // SIViP. — 2011. — №5. — P. 81-91.
19. Smith John R., Chang Shih-Fu. Tools and Techniques for Color Image Retrieval // In IS&T/SPIE Symposium on Electronic Imaging: Science and Technology (EI'96) - Storage and Retrieval for Image and Video Databases IV. — 1996. — V. 2670, San Jose, CA, February 1996. — P. 426-437.
20. Мокеев В.В., Томилов С.В. О решении проблемы выборки малого размера при использовании линейного дискриминантного анализа в задачах распознавания лиц // Бизнес-информатика. — 2013. — №1(23). — С. 37-42.

ПРИМЕНЕНИЕ ПРИНЦИПОВ ЭКСПЕРТНОЙ КЛАССИФИКАЦИИ ДЛЯ АНАЛИЗА ФОРМАЛЬНЫХ ПОНЯТИЙ

С.Ю. Соловьев,

доктор физико-математических наук, профессор кафедры алгоритмических языков, факультет вычислительной математики и кибернетики Московского государственного университета им. М.В. Ломоносова

Д.Е. Стельмашенко,

инженер факультета вычислительной математики и кибернетики Московского государственного университета им. М.В. Ломоносова

E-mail: soloviev@glossary.ru, dashanikolaeva@gmail.com

Адрес: г. Москва, ГСП-1, Ленинские горы, МГУ имени М.В. Ломоносова, 1, стр. 52

В статье рассматривается вариант задачи конструирования решетки формальных понятий и обсуждаются особенности алгоритма ее решения. Для повышения эффективности алгоритма предлагается использовать принципы экспертной классификации.

Ключевые слова: анализ формальных понятий, метод экспертной классификации, принцип доминирования по характерности, шкала характерности.

1. Введение

В последнее время анализ формальных понятий (англ. Formal Concept Analysis; FCA) выступает своеобразным собирателем различных методов обработки нечисловых данных. В работе рассматривается вопрос о применении основных принципов экспертной классификации для одной задачи анализа формальных понятий. Систематическому изложению этих родственных направлений искусственного интеллекта посвящены монографии [1-4], здесь же приводится мини-

мально необходимый набор сведений из FCA (п. 2), а также адаптированные под конкретную задачу (п. 3) основы экспертной классификации (п. 4). Результатом такого «симбиоза» стали новые подходы к обработке формальных понятий (п. 5) и формулировки новых задач.

2. Некоторые определения FCA

Формальный контекст есть тройка множеств (G, M, I) , в которой элементы G называются объектами, элементы M — признаками, а I есть подмно-

жество из $G \times M$. Если $(g, m) \in I$, то говорят «объект g обладает признаком m » либо, что то же самое, «признак m встречается в описании объекта g ». При переходе от традиционных неколичественных моделей проблемных областей (первичных моделей) к формальным контекстам (вторичным моделям) каждый атрибут-свойство X с конечным числом значений $X_1, \dots, X_n$ кодируется набором $(0,1)$ -признаков $m_{x_1}, \dots, m_{x_n}$ по принципу «одно значение атрибута – один признак контекста».

Формальный контекст $K = (G, M, I)$ порождает так называемые формальные понятия. Пара непустых подмножеств (A, B) , где $A \subseteq G$ и $B \subseteq M$, называется формальным понятием, если (1) каждый объект из A обладает всеми признаками из B , и (2) никакая другая пара $(A \cup A', B \cup B')$ не удовлетворяет требованию (1). Пары множеств $(\emptyset, M)$ и $(G, \emptyset)$ являются формальными понятиями по определению. Если (A, B) – формальное понятие контекста K , то A называется объемом понятия, B – его содержанием. В дальнейшем изложении будем полагать, что множества G и M конечны, а все термины «контекст», «понятие» и «содержание» автоматически подразумевают спецификацию «формальный».

Множество $C(K)$ всех понятий контекста K образует полную решетку, с отношением частичного порядка

$$(A_1, B_1) \prec (A_2, B_2) \Leftrightarrow B_1 \supseteq B_2.$$

Соответственно, решеткой также является множество $S(K) = \{B | (A, B) \in C(K)\}$ с отношением частичного порядка $\supseteq$. Решетки $C(K)$ и $S(K)$ изоморфны; располагая одной из них, можно построить другую решетку. Рассмотрим подробнее те наборы признаков, которые не относятся к $S(K)$ и образуют множество $S^\circ(K) = 2^M \setminus S(K)$.

Подмножество признаков H контекста $K = (G, M, I)$ будем называть запретом, если $H \neq M$ и в описаниях объектов признаки из H вместе никогда не встречаются: $\forall g \in G \exists b \in H : (g, b) \notin I$. Требование $H \neq M$ объясняется лишь тем, что в содержаниях понятий – элементах множества $S(K)$ – все признаки запретов встречаются исключительно в понятии $(\emptyset, M)$. Множество всех запретов контекста K обозначим $P(K)$. Кроме того, обозначим $P^\circ(K)$ множество $S^\circ(K) \setminus P(K)$. Описанная классификация всевозможных наборов признаков представлена на рис. 1.

Эквивалентным представлением контекста $K = (G, M, I)$ является объектно-признаковая таблица $Tab(K)$, в которой именами столбцов служат признаки из M , кодами строк – объекты из G , а на

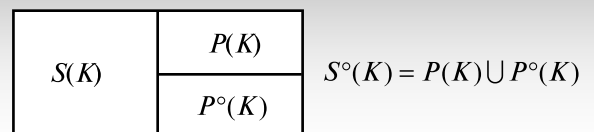


Рис. 1. Строение множества 2^M для контекста K

пересечении строки g и столбца m стоит $\times$, если $(g, m) \in I$.

Общее определение контекста допускает наличие в нем объектов-дубликатов, которым в $Tab(K)$ соответствуют строки, отличающиеся только кодами. Объекты-дубликаты не влияют на структуру решетки понятий, поэтому для многих приложений интерес представляют контексты без объектов-дубликатов, именуемые далее простыми контекстами. Не ограничивая общности, можно полагать, что в простом контексте (G, M, I) каждый объект g из G представляет собой подмножество признаков, а $I = \varphi(G) \equiv \{(g, m) | g \in G, m \in g\}$. При таком подходе для задания простого контекста достаточно перечислить описания его объектов.

3. Задача совмещения формальных контекстов

Очевидно, что любой нетривиальный контекст K способен порождать некоторые производные контексты посредством вычеркивания из $Tab(K)$ отдельных строк-объектов и/или столбцов-признаков. На содержательном уровне задача совмещения контекстов состоит в конструировании решетки $C(K)$ неизвестного контекста K по производным от него контекстам $K_1 = (G_1, M_1, I_1)$ и $K_2 = (G_2, M_2, I_2)$. Известные подходы к решению задачи совмещения различаются дополнительными предположениями и целевыми установками.

Для контекстов общего вида в случае $G_1 = G_2 = G$, $M_1 \cup M_2 = M$ и $M_1 \cap M_2 = \emptyset$ задача совмещения контекстов часто решается посредством так называемой «Nested Line» диаграммы [1]. Построение «Nested Line» диаграммы предполагает подстановку диаграммы Хассе решетки $C(K_1)$ в каждый узел соответствующей диаграммы решетки $C(K_2)$. Считается, что построенная таким образом диаграмма более наглядна, чем диаграмма Хассе решетки $C(K)$.

Для контекстов общего вида в случае $G_1 \cup G_2 = G$ и $M_1 \cup M_2 = M$ искомую решетку $C(K)$ предлагается [5] строить по контексту $K = (G_1 \cup G_2, M_1 \cup M_2, I_1 \cup I_2)$. Метод используется для пополнения баз знаний. Модификация метода [6] позволяет выявить в

объединенном множестве признаков дубликаты и модифицировать отношение $I_1 \cup I_2$.

Упомянутые подходы к решению задачи совмещения не претендуют на точное построение решетки $S(K)$. Иначе устроен метод @FC [7,8], о котором речь пойдет далее. Название метода происходит от первых букв ключевых слов его описания: Alternative design, Testing, Formal context, Concept. Метод @FC использует специальную терминологию простых контекстов.

Если $K = (G, M, \varphi(G))$ — простой контекст, заданный множеством своих объектов G (глобальный контекст), то проекцию G на фиксированное подмножество признаков M_T будем называть локальным контекстом. Формально, локальный контекст K_T есть

$$(G_T, M_T, \varphi(G_T)), \text{ где } G_T = \{g \cap M_T \mid g \in G\}.$$

На практике получение K_T может быть связано, например, со снаряжением целой экспедиции к местам обитания изучаемых объектов G , причем исследовательские возможности экспедиции ограничены набором признаков M_T .

Для глобального контекста K и пары локальных контекстов K_1, K_2 в случае $M_1 \cup M_2 = M$ метод @FC предлагает строить множество $S(K)$ в виде

$$V = \{B_1 \cup B_2 \mid B_1 \in S(K_1), B_2 \in S(K_2)\}$$

с последующей пошаговой классификацией элементов V на элементы $S(K)$ и $S^\circ(K)$. Доказано, что метод @FC позволяет построить множество $S(K)$ за конечное число шагов. На каждом шаге построения принимается решение о запросе нового локального контекста, и по итогам его специальной обработки для некоторых элементов V принимаются классификационные решения. Процесс формирования $S(K)$ заканчивается, когда классификационные решения приняты для всех элементов множества V .

Запрос нового локального контекста в виде множества его признаков M_T формируется автоматически или в диалоге с пользователем. При этом варианты запросов известны заранее, и необходимо «лишь» выбрать лучший из них.

В части обработки локальных контекстов K_T метод @FC использует правила классификации, которые иногда позволяют относить элементы V к классам $S(K)$ или $S^\circ(K)$. Приведем в качестве примера два классификационных правила:

(K1) Если $v \in S(K_T)$ и $\{v' \mid v \cup v' \in V\} = \{\emptyset\}$, то $v \in S(K)$.

(K2) Если $v \in P(K_T)$ и $v \cup v' \in V$ для некоторого v' , то $v \cup v' \in P(K)$.

С одной стороны, классификационные правила представляют собой строго доказанные утверждения [8], а с другой стороны — их можно использовать в качестве условных операторов алгоритма реализации метода @FC. Обработка очередного локального контекста заканчивается, когда все правила классификации оказываются неприменимыми.

При высокой стоимости локальных контекстов возникает потребность в технологиях их «глубокой переработки», позволяющих на каждом шаге классифицировать как можно больше элементов V . Аналогичная потребность исследовалась в 1980-х научным коллективом под руководством О.И. Ларищева. Результатом этой работы стал метод экспертной классификации, позволяющий при незначительных дополнительных, но проверяемых предположениях существенно увеличить количество классификационных решений в задаче совмещения формальных контекстов.

4. Принципы экспертной классификации

В терминах FCA главная идея метода экспертной классификации состоит в обоснованном переносе классификационных решений, полученных для одного объекта, на некоторые другие объекты. Базой знаний для обоснования и переноса служит отношение доминирования, которое, в свою очередь, строится из отношений характерности. Считается, что контекст $K = (G, M, I)$ представляет собой вторичную модель проблемной области, а в первичной модели используются атрибуты-свойства, принимающие конечное число значений. При этом контекст K наследует от первичной модели разбиение множества признаков $M = M_1 \cup M_2 \cup \dots \cup M_k$, в котором каждое подмножество M_i соответствует ровно одному атрибуту.

Теоретическую основу метода экспертной классификации составляет гипотеза, согласно которой «признаки каждого множества M_i различаются по степени характерности для объектов G ». Если для конкретной проблемной области гипотеза справедлива, то эксперт способен построить на элементах множества M_i асимметричное, антирефлексивное и транзитивное отношение характерности $<_i$; запись « $m_{i1} < m_{i2}$ » означает, что признак m_{i1} менее характерен для объектов G , чем признак m_{i2} . Экспериментально установлено, что фраза «быть характерным

для объектов G и фразы, производные от нее, понятны эксперту без разъяснений, и он почти всегда способен построить k штук отношений характерности — по одному для каждого множества M_i .

Строго говоря, в оригинальном описании экспертной классификации предполагается линейность всех отношений характерности. Это обстоятельство существенно для представления данных и для организации диалога с экспертом при извлечения знаний об отношениях характерности. В настоящей работе вместо линейного порядка предполагается наличие частичного порядка. Соответственно, изменяется процедура опроса эксперта, в основу которой полагается сравнение на характерность в различных парах признаков из M_i .

Совокупность отношений характерности $<_1, \dots, <_k$ порождает рефлексивное, антисимметричное и транзитивное отношение доминирования $\leq$ на всевозможных подмножествах признаков из M . По определению, пара подмножеств b_1 и b_2 связана отношением $b_1 \leq b_2$, если для любого $i = 1, \dots, k$ выполняется одно из двух: либо $b_1 \cap M_i = b_2 \cap M_i$, либо $b_1 \cap M_i = \{m_{i1}\}$, $b_2 \cap M_i = \{m_{i2}\}$ и $m_{i1} < m_{i2}$.

Если отношение доминирования удастся построить, то механизм переноса классификационных решений применительно к объектам контекста — элементам множества G — описывается двумя постулированными, но вполне разумными правилами:

- (П1) Если $g_1 \in G$, $g \in 2^M$ и $g_1 \leq g$, то $g \in G$;
- (П2) Если $g_1 \notin G$, $g \in 2^M$ и $g \leq g_1$, то $g \notin G$.

5. Метод доминирования для формальных контекстов

Метод @FC, в отличие от оригинальной версии экспертной классификации, оперирует не только описаниями объектов, но и другими подмножествами признаков. В связи с этим возникает необходимость корректного обобщения правил П1 и П2 на случай подмножеств из V , и главная роль в таком обобщении отводится формальной интерпретации

отношения характерности.

Простейшая, хотя и не единственная теоретико-множественная интерпретация отношений $<_i$ имеет вид:

$$m_{i1} <_i m_{i2} \Leftrightarrow \{g \in G \mid m_{i1} \in g\} \subseteq \{g \in G \mid m_{i2} \in g\}.$$

Доказано, что при таком подходе к отношениям характерности правила П1 и П2 превращаются в формальные утверждения, а кроме того, справедливы также утверждения

- (ПЗ) Если $v_1 \in S(K)$, $v \in V$ и $v_1 \leq v$, то $v \in S(K)$;
- (П4) Если $v_1 \in S^\circ(K)$, $v \in V$ и $v \leq v_1$, то $v \in S^\circ(K)$;
- (П5) Если $v_1 \in P(K)$, $v \in V$ и $v \leq v_1$, то $v \in P(K)$;
- (П6) Если $v_1 \in P^\circ(K)$, $v \in V$ и $v \leq v_1$, то $v \in P^\circ(K)$.

В доказательствах утверждений ПЗ–П6 существенно используются свойства объектов глобального контекста, однако в окончательных формулировках объекты не фигурируют, что позволяет использовать утверждения ПЗ–П6 в качестве дополнительных правил распространения классификационных решений метода @FC.

6. Заключение

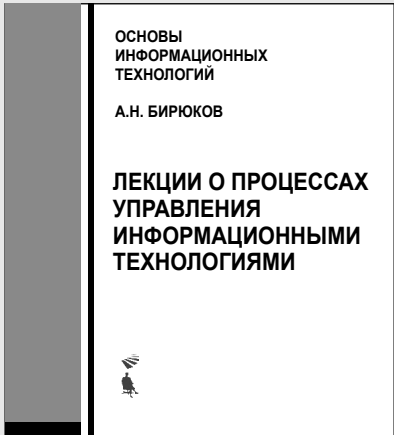
Представленная модификация метода совмещения @FC применима для глобальных контекстов, удовлетворяющих гипотезе доминирования в ее простейшей интерпретации. Проверка гипотезы по решетке содержаний $S(K)$ представляет собой самостоятельную теоретическую задачу, способную в перспективе породить новую версию алгоритма совмещения локальных контекстов. Кроме того, один из принципов экспертной классификации состоит в обязательном исследовании наборов признаков на границах классов. Адаптация этого принципа для задач FCA также представляется весьма перспективной.

Предложенный и намеченные алгоритмы могут использоваться при решении практических задач бизнес-информатики, сводимых к обработке формальных понятий. ■

Литература

1. Ganter B., Wille R. Formal Concept Analysis: Mathematical Foundations. — Springer, 1999.
2. Гуров С.И. Булевы алгебры, упорядоченные множества, решетки: Определения, свойства, примеры. — М.: ЛИБРОКОМ, 2013.
3. Ларичев О.И. Теория и методы принятия решений, а также Хроника событий в Волшебных Странах. — М.: Логос, 2000.

4. Ларичев О.И., Мечитов А.И., Мошкович Е.М., Фуремс Е.М. Выявление экспертных знаний. — М.: Наука, 1989.
5. Li G.Y., Liu S.P., Zhao Y. Formal Concept Analysis based Ontology Merging Method // Proceeding of 3rd IEEE International Conference on Computer Science and Information Technology. — 2010 — volume 8 — P. 279-282.
6. Bendaoud R., Napoli A., Toussaint Y. A proposal for an Interactive Ontology Design Process based on Formal Concept Analysis // Frontiers in Artificial Intelligence and Applications. — 2008 — volume 183: Formal Ontology in Information Systems — P. 311-323.
7. Стельмашенко Д.Е. Алгоритм восстановления глобального контекста // Сборник статей молодых ученых факультета ВМК МГУ. — 2012 — выпуск 9 — С.191-209.
8. Соловьев С.Ю., Стельмашенко Д.Е. Подходы к исследованию формальных контекстов // Информационные процессы. — 2011 — том 11 — № 2 — С. 277-290.



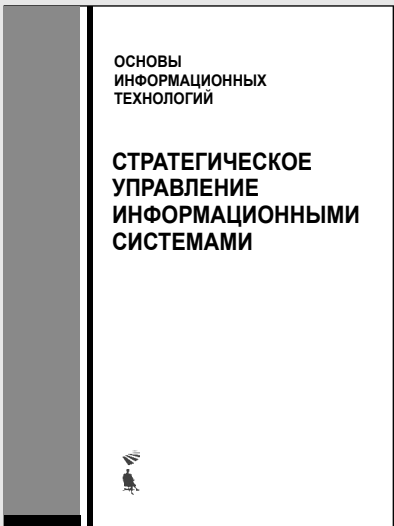
**ЛЕКЦИИ О ПРОЦЕССАХ УПРАВЛЕНИЯ
ИНФОРМАЦИОННЫМИ ТЕХНОЛОГИЯМИ**

Учебное пособие

А.Н. Бирюков

Москва: Интуит.РУ, БИНОМ.ЛТЗ, 2010.

Рассматриваются основные процессные модели и методики, связанные с управлением ИТ, появившиеся в последние годы. Основное внимание уделяется анализу их взаимосвязей и выявлению общих концепций и подходов. Изложение в большой степени базируется на оригинальных материалах, не переведенных на русский язык.



**СТРАТЕГИЧЕСКОЕ УПРАВЛЕНИЕ
ИНФОРМАЦИОННЫМИ СИСТЕМАМИ**

Учебник.

**Р.Б. Васильев, Г.Н. Калянов,
Г.А. Левочкина, О.В. Лукинова**

Под ред. Г.Н. Калянова

*Москва: Интернет-университет
Информационных технологий: Бином.
Лаборатория знаний, 2010.*

Учебник дает читателю представление о методологической базе и современных подходах и методах стратегического управления информационными системами, обеспечивающего целостный, процессно-ориентированный подход к принятию управленческих решений, направленных на повышение эффективности владения и развития информационных систем для достижения бизнес-целей предприятий и создания новых конкурентных преимуществ.

СЦЕНАРНЫЙ ПОДХОД К ПЛАНИРОВАНИЮ РАСПРЕДЕЛЕНИЯ ПЕРСОНАЛА НИИ С УЧЕТОМ РИСКОВ

А.И. Марон,

кандидат технических наук, доцент кафедры бизнес-аналитики

Национального исследовательского университета «Высшая школа экономики»

E-mail: amaron@hse.ru

Адрес: г. Москва, ул. Кирпичная, д. 33/5

Предложена соответствующая современным реалиям постановка задачи оптимального распределения сотрудников НИИ по видам работ и заказчикам. Разработан метод решения на основе теории принятия решений в условиях неопределённости.

Ключевые слова: распределение сотрудников, оптимизация, неопределённость, риск.

1. Введение

Многие ведущие отраслевые научно – исследовательские и проектно – конструкторские институты (НИИ) сегодня являются дочерними предприятиями монопольных отраслевых корпоративных структур, которые являются основными заказчиками работ. Их заказ в дальнейшем будем называть корпоративным. Корпоративный заказ должен быть выполнен. Если при этом НИИ располагает трудовыми ресурсами для выполнения работ по договорам с другими заказчиками, то он обязан это делать, поскольку является коммерческой структурой, которая должна приносить акционерам максимально возможную прибыль. Кроме того, зачастую корпоративного заказа не хватает даже для поддержания нормальной деятельности НИИ. Возникает вопрос, как правильно распределить имеющиеся трудовые ресурсы. При этом принципиально различными являются ситуации, когда планирова-

ние осуществляется в условиях ненасыщенного и конкурентного рынка. Кроме того, важнейшим фактором является то, утверждён ли корпоративный заказ на момент планирования или имеется неопределённость. При утверждённом корпоративном заказе и неконкурентном рынке задачу оптимального распределения сотрудников НИИ можно свести к задаче об оптимальной загрузке производственных мощностей [1; 2].

Существенно труднее разработать рациональный план в условиях конкурентного рынка. Действительно, корпоративный заказ включает в себя научно-исследовательские работы, опытно-конструкторские работы и промышленное внедрение результатов НИР и ОКР. Соответственно, сложившийся производственный коллектив института состоит из следующих трёх основных групп: научные сотрудники; конструкторы; специалисты по внедрению результатов НИР и ОКР – далее «специалисты». Для каждой группы известны ко-

личество сотрудников и среднее значение годовой заработной платы сотрудника с учётом всех начислений. Кроме корпоративного заказчика, договора на выполнение ОКР и внедрение результатов НИР и ОКР могут быть заключены с его дочерними и зависимыми обществами (ДЗО), а также и с другими сторонними потребителями. Ценовая политика корпоративного заказчика такова, что размер прибыли по договору устанавливается с определённым коэффициентом от затрат НИИ на оплату труда своих сотрудников. Прочие расходы оплачиваются строго по себестоимости. В числе прочего это относится к затратам на оборудование и комплектующие элементы, а также услугам сторонних организаций. Такой же политики строго придерживаются и ДЗО. Сторонние заказчики, как правило, являются давними партнёрами, хорошо осведомлёнными о принципах ценообразования в отрасли. Поэтому можно считать, что при выполнении всех видов работ для любого заказчика прибыль пропорциональна стоимости трудовых затрат сотрудников НИИ. Однако размер коэффициентов пропорциональности различен. Можно сказать, что он минимален для сходных видов работ, когда они выполняются для корпоративного заказчика.

Работы представляют собой набор проектов. За каждым проектом, который относится к определённому виду работ, должен быть закреплён, как минимум, один профильный сотрудник. Очевидно, для выполнения научно-исследовательских проектов необходимо участия научных сотрудников, опытно-конструкторские работы (ОКР) требуют конструкторов, а внедренческие проекты — соответствующих специалистов. На практике в большинстве случаев при выполнении НИР требуются не только научные сотрудники, но также конструкторы и специалисты по внедрению. Соответственно при выполнении ОКР необходимы не только конструкторы, но специалисты по внедрению, а также научные сотрудники (хотя бы для консультаций). Аналогично, при внедрении требуются не только специалисты, но также научные сотрудники и конструкторы (хотя бы для авторского надзора). Современные жёсткие условия деятельности НИИ требуют перехода к матричной организационной структуре для эффективного выполнения проектов. При этом вполне реальным является закрепление одного непрофильного сотрудника за несколькими проектами [3; 4].

В условиях конкурентного рынка независимо от того, утверждён корпоративный заказ или нет, воз-

никает неопределённость относительно объёмов работ, которые можно получить от ДЗО и сторонних потребителей. Вместе с тем, кампанию по заключению договоров с ДЗО и сторонними заказчиками надо начинать заблаговременно. Зачастую раньше, чем утверждён корпоративный заказ. Возникают риски:

- ♦ заключить слишком много таких договоров, не оставив достаточно ресурсов для выполнения корпоративного заказа;
- ♦ заключить мало договоров, зарезервировав избыточные трудовые ресурсы для выполнения корпоративного заказа.

Первый из этих рисков является критическим. Его надо избежать. Реализация второго риска приведёт к тому, что определённое число сотрудников не будет привлечено к выполнению договоров, приносящих доход в рассматриваемом периоде. Они могут быть заняты в осуществлении перспективных, или как их иногда называют «задельных» работ, которые финансируются за счёт средств самого НИИ. Естественно это снизит прибыль.

Поскольку объём договоров неизвестен, приходится составить несколько возможных сценариев. Обычно составляется по три сценария: пессимистический, умеренный и оптимистический; для корпоративного заказчика, ДЗО и сторонних потребителей. В результате получается 27 возможных вариантов. Однако, поскольку риск невыполнения корпоративного заказа является критическим, при сценарном планировании следует принять для корпоративного заказчика оптимистический сценарий, которому соответствует больший объём заказа, чем при двух других сценариях. Это позволит гарантированно выделить ресурсы, достаточные для того, чтобы избежать риска невыполнения установленного корпоративного заказа. Соответственно число вариантов, подлежащих рассмотрению, уменьшается до девяти. Каждому из них соответствует своё отличное от других закрепление сотрудников за корпоративным заказом как по группам, так и численности. Требуется сравнить эти стратегии предварительного резервирования сотрудников для гарантированного выполнения корпоративного заказа. Для решения этой задачи в статье предложено прибегнуть к комбинированному использованию метода линейного программирования и методов поддержки принятия в условиях неопределённости, не имеющей вероятностной оценки. Для того чтобы избежать излишне

сложных обозначений и сделать изложение метода более лёгким для восприятия, ограничимся случаем раннего планирования, когда разделение на три сценария: пессимистический, умеренный и оптимистический; осуществляется для коммерческого заказа в целом.

Перейдём к формальной постановке задачи и изложению предлагаемого метода решения.

2. Формальная постановка и метод решения задачи

Имеется: i — квалификационных групп сотрудников, j — видов работ, k — заказчиков ($i, j, k = 1, 2, 3$). Численность научных сотрудников (группа 1) составляет b_1 , конструкторов (группа 2) b_2 и специалистов (группа 3) b_3 . Ставки заработной платы сотрудников в группах 1, 2 и 3 составляют c_1, c_2, c_3 , соответственно. Однако затраты на оплату сотрудников, которые можно согласовать с заказчиком, зависят от того для какого заказчика сотрудник выполняет работу и по своей ли специализации. Другими словами, затраты, которые можно заложить в договор, на сотрудника из группы i , выполняющего работу вида j для заказчика k равна $s_{ijk} = c_i \cdot h_{ijk}$, где h_{ijk} заданный коэффициент. Обычно, если сотрудник направляется на работу в проекте, выполняемом для корпоративного заказчика или ДЗО ($k = 1, 2$), то затраты определяются исходя из базовой ставки. Также считается, что он получает по базовой ставке, когда выполняет работу своего профиля для стороннего заказчика. Если же он направляется на работу для стороннего заказчика не по своей специализации, то затраты согласовываются исходя из базовой ставки профильного сотрудника по данному виду работы, умноженной на некоторый, как правило, понижающий коэффициент. Если x_{ijk} — число сотрудников группы i , которые выполняют работу вида j для заказчика — k , то трудозатраты $S_{ijk} = x_{ijk} \cdot s_{ijk}$. Прибыль начисляется на трудозатраты. Заданы коэффициенты прибыли — m_{jk} , которые зависят от заказчика и вида работы.

Точные объёмы работ на момент планирования не установлены. Поэтому применяется сценарный подход. Для ДЗО и сторонних заказчиков рассматривается по три сценария, каждому из которых соответствуют свои объёмы работ. Сценарии носят названия: пессимистический, умеренный, оптимистический. Объём работ минимален при пессимистическом сценарии и максимален при оптимистическом сценарии. Присвоим им номера 1, 2, 3,

соответственно. Условимся использовать для обозначения номера сценария символ r ($r = 1, 2, 3$). В случае необходимости будем его применять в качестве верхнего индекса. Соответственно, возможные сценарии обозначим через Π^1, Π^2, Π^3 . Объёмы работ, которые надо выполнить для ДЗО, при сценарии r обозначим как V_{rj2} ($j, r = 1, 2, 3$). Объёмы работ, которые надо выполнить для сторонних заказчиков, при сценарии r , обозначим как V_{rj3} . Для корпоративного заказчика рассматривается только один сценарий — оптимистический, как было сказано выше, это позволяет гарантированно избежать невыполнения корпоративного заказа. Объёмы по видам работ, которые организация должна выполнить по корпоративному заказу V^1 составляют $V_{11}^1, V_{21}^1, V_{31}^1$.

Количество проектов вида j , выполняемых для заказчика k , при рассматриваемом сценарии r , обозначим через L_{jk}^r ($j, k, r = 1, 2, 3$; при $k = 1$ и $r = 1$). На каждый проект вида j необходимо назначить минимум a_j (a_j — целое большее единицы) сотрудников соответствующего профиля. Число непрофильных сотрудников, назначаемых на один проект, находится в пределах от d_{jmin} до d_{jmax} .

Для каждого сценария r можно найти распределение сотрудников, при котором прибыль W^r достигает максимального значения. Причём поиск оптимума можно выполнить с помощью метода линейного программирования, решив следующую задачу.

Требуется найти значения x_{ijk} , при которых достигается максимум целевой функции

$$W^r = W_1^r + W_2^r + W_3^r - \sum_{i=1}^3 \left[c_i \cdot (b_i - \sum_{j=1}^3 \sum_{k=1}^3 x_{ijk}) \right] \quad (1)$$

где $W_k^r = \sum_{i=1}^3 \sum_{j=1}^3 (m_{jk} \cdot S_{ijk})$; ($k = 1, 2, 3$)

и выполняются ограничения:

$$\sum_{j=1}^3 \sum_{k=1}^3 x_{ijk} \leq b_i \quad (i = 1, 2, 3), \quad (2)$$

$$\sum_{i=1}^3 (1 + m_{jk}) \cdot S_{ijk} = V_{jk}^r, \quad (3)$$

(здесь и ниже $j, k, r = 1, 2, 3$ и если $k = 1$, то $r = 1$),

$$x_{ijk} \geq a_j \cdot L_{jk}^r \quad (\text{если } i = j), \quad (4)$$

$$d_{jmin} \cdot L_{jk}^r \leq x_{ijk} \leq d_{jmax} \cdot L_{jk}^r \quad (\text{если } i \neq j). \quad (5)$$

Эту задачу линейного программирования при отказе от условия о целых значениях искомых переменных можно решить даже штатными средствами оптимизации, входящими в Excel.

Найдя решения для всех трёх сценариев, получим значения W^1 , W^2 , W^3 . Каждое из них определяет максимальные значения прибыли в том и только том случае, когда действительно будет реализован соответствующий сценарий. Значению W^1 соответствует распределение сотрудников X^1 . В него входит пул сотрудников X^1_1 , которые будут выполнять работы по корпоративному заказу. Аналогично для сценариев Π^2 и Π^3 будут найдены — X^2_1 и X^3_1 . Будем рассматривать X^1_1 , X^2_1 , X^3_1 как допустимые стратегии закрепления персонала за корпоративным заказчиком. Возникает вопрос: «Какая стратегия лучше»? Для ответа на него надо выбрать критерий оптимальности. В нашем случае вероятности реализации сценариев неизвестны. Поэтому целесообразно обратиться к критериям Вальда или Сэвиджа. Первый из них наилучшим образом подходит для НИИ, финансовое состояние которого не позволяет принимать существенные риски. Второй применим, когда собственных средств достаточно, для того, чтобы считать, что главное не упустить возможную прибыль. Остановимся на нахождении стратегии предварительного закрепления сотрудников за корпоративным заказчиком, оптимальной по критерию Вальда. Согласно этому критерию оптимальной является стратегии, при которой минимальный выигрыш максимален. В нашем случае под выигрышем понимается прибыль.

Составим матрицу прибыли, записав в строках допустимые стратегии X^1_1 , X^2_1 , X^3_1 , а в столбцах возможные сценарии Π^1 , Π^2 , Π^3 . В ячейку, стоящую на пересечении строки r , и столбца t надо записать прибыль, которая может быть получена, если за корпоративным заказчиком будет закреплён пул сотрудников X^r_1 , а реализуется сценарий Π^t . В ячейки диагонали матрицы прибыли, где $r = t$, сле-

дует записать найденные ранее значения W^1 , W^2 , W^3 . Остаётся заполнить ещё 6 ячеек. Найти соответствующие им значения прибыли можно следующим образом.

При $r \neq t$ прибыль складывается из двух слагаемых. Одно из них равно W^r_1 . Эта величина уже найдена (смотри формулу (1)). Для нахождения второго слагаемого необходимо сделать следующее.

2. Уменьшить численность сотрудников в каждой группе в соответствии с тем, сколько сотрудников каждой квалификационной группы вошло в пул X^r_1 .

3. Решить задачу оптимизации прибыли, используя приведенную выше постановку, но для новых значений численности персонала с учётом следующих изменений:

- ♦ из формулы (1) убрать значение W^r_1 ;
- ♦ ограничения (3) оставить только для $k = 2$ и 3 , заменив в них строгое равенство на неравенство «меньше или равно».

После того как таблица прибыли заполнена остаётся в каждой её строке найти минимальное значение. После чего, найти среди них максимальное значение. Ему соответствует стратегии закрепления персонала оптимальная по критерию Вальда.

3. Заключение

Таким образом, решена задача закрепления ресурсов для выполнения корпоративного заказа в условиях, когда он ещё не утверждён. Эту задачу руководству НИИ приходится решать для того, чтобы успеть провести договорную компанию с коммерческими заказчиками, не ставя под угрозу гарантированное выполнение работ, заказанных основным учредителем. ■

Литература

1. Мищенко А.В., Джамай Е.В. Динамическая задача определения оптимальной производственной программы // Менеджмент в России и за рубежом. — 2002, №2.
2. Вентцель Е. С. Исследование операций: задачи, принципы, методология. — М.: Наука, 1988.
3. Ципес Г.Л., Товб А.С. Проекты и управление проектами в современной компании. — М.: Олимп — Бизнес, 2009.
4. Баркалов С., Воропаев В.И., Секлетова Г.И. Математические основы управления проектами. — М.: Высшая школа, 2005.

ДИНАМИЧЕСКАЯ МОДЕЛЬ ОРГАНИЗАЦИИ ГРУЗОПЕРЕВОЗОК ПРИ ОГРАНИЧЕННОСТИ ЕМКОСТЕЙ ПЕРЕГОННЫХ ПУТЕЙ¹

Н.К. Хачатрян,

кандидат физико-математических наук, старший научный сотрудник
Центрального экономико-математического института Российской академии
наук, старший преподаватель кафедры бизнес-аналитики Национального
исследовательского университета «Высшая школа экономики»

E-mail: nerses@cemi.rssi.ru

Адрес: г. Москва, Нахимовский проспект, 47

В работе [1] была исследована модель, описывающая процесс организации грузоперевозок с заданной системой контроля. Изучены возможные режимы грузоперевозок, исследован вопрос устойчивости таких режимов, описана область их устойчивости. В данной работе изучается модифицированная модель при дополнительных ресурсных ограничениях.

Ключевые слова: дифференциальные уравнения, стационарный режим грузоперевозок, устойчивость.

1. Введение

Транспорту принадлежит особая роль в экономике страны, он связывает воедино все отрасли экономики, обеспечивая перемещение сырья, полуфабрикатов и готовой продукции. Транспортной системе присущи черты, свойственные любой другой производственной системе. Однако по сравнению с остальными отраслями транспорт обладает целым рядом специфических особенностей, порождаемых характером производственного процесса. В процессе своего функционирования транспортная система не создает нового материального продукта, ее продукцией является сам процесс пере-

мещения грузов и пассажиров. В отличие от продукции других отраслей транспортная продукция не взаимозаменяема: перевыполнение плана перевозок какого-либо груза между одними пунктами не может скомпенсировать невыполнение перевозок того же груза между другими пунктами. Эта продукция не существует отдельно от транспорта и не может производиться в запас, т.е. невыполнение перевозок в один период времени не может быть скомпенсировано перевыполнением их в другой период времени. Масштабы деятельности отрасли, рассредоточенность ее объектов, динамический характер производственного процесса обуславливают чрезвычайную сложность управления транспортной системой.

¹ Работа выполнена при поддержке Российского фонда фундаментальных исследований (проект № 12-01-00768)

Среди проблем, связанных с работой транспорта, центральное место занимают задачи планирования и организации грузоперевозок. Впервые методы нахождения оптимального плана перевозок в нашей стране были предложены в 30-ых годах прошлого столетия. В 1939 г. Л.В. Канторовичем [2] математически описана транспортная задача линейного программирования. Им же определен целый класс задач, близких к транспортной, предложен алгоритм для решения транспортной задачи, названный методом разрешающих множителей. В 1949 г. Л.В. Канторович и М.К. Гавурин опубликовали работу [3], в которой решалась транспортная задача с ограничениями на пропускные способности. Используя идеи общего метода Л.В.Канторовича, для решения задач линейного программирования был разработан метод потенциалов. Через год этот же метод был предложен американским математиком Дж. Данцигом [4]. В 1985 г. О.И. Авен, С.Е. Ловецкий и Г.Е. Моисеенко опубликовали работу [5], посвященную проблемам оптимального планирования и управления транспортными потоками на транспортных сетях. Были рассмотрены математические модели транспортных сетей и транспортных потоков. Другой важной задачей, связанной с работой транспорта, является организация грузоперевозок и их системы контроля. Такая задача рассмотрена в работах [6 – 8].

В работе [1] была построена и исследована модель организации грузоперевозок на протяженном участке пути с большим количеством промежуточных станций, через которые проходит грузопоток. Приведем основные положения указанной модели. Предполагается, что между двумя соседними станциями существует межстанционный перегонный путь, где временно может храниться часть грузов. Движение грузов происходит в одном направлении. На произвольную промежуточную станцию с номером i груз может поступать как с предыдущей станции с номером $(i - 1)$ так и перегонного пути, расположенного между ними. Аналогично, с произвольной промежуточной станцией с номером i груз может быть отправлен либо на следующую станцию с номером $(i + 1)$ либо на перегонный путь, расположенный между ними.

Обработка грузов происходит в узлах станций. Очевидно, что количество задействованных узлов обработки грузов при бесперебойной работе всей цепи перевозок ограничено. Максимальное количество таких узлов, обозначаемое через Δ , определяет пропускную способность станций.

Рассматривались четыре варианта модели. Пер-

вый вариант модели описывает транснациональные грузоперевозки и поэтому предполагалось, что число промежуточных станций бесконечно как в левую, так и в правую стороны. Во втором варианте модели рассматривался грузопоток, идущий от узловой станции, и поэтому предполагалось, что число промежуточных станций бесконечно лишь в правую сторону. Третий вариант модели описывал грузопоток между двумя узловыми станциями. Наконец, четвертый вариант модели описывал транспортные грузоперевозки по круговой цепочке станций. Для первых трех вариантов модели приводились постановка задачи и основные результаты. Детальное описание этих вариантов модели можно найти в работах [9 – 11]. Для четвертого варианта модели изучены возможные режимы организации грузоперевозок, проведено исследование на устойчивость стационарного режима грузоперевозок.

Недостатком этой модели является предположение о неограниченности емкостей перегонных путей. Данная работа посвящена исследованию четвертого варианта указанной модели в случае, когда емкости перегонных путей ограничены.

2. Формальное описание исходной модели и основные результаты

Организация подобных грузопотоков зависит от технологий по приему, обработке и отправлению грузов. Опишем эти технологии.

Первая технология основана на установленных нормативных правилах взаимодействия соседних станций. Для каждой промежуточной станции с номером i существуют правила взаимодействия с предыдущей $(i - 1)$ -ой станцией и последующей $(i + 1)$ -ой станцией. Согласно правилу взаимодействия с предыдущей станцией, станция с номером i , в зависимости от количества задействованных узлов на $(i - 1)$ -ой станции, должна увеличивать или уменьшать количество задействованных узлов со скоростью $\alpha(z_{i-1} - z_i)$ (т.е. принимать груз с предыдущей станции если количество задействованных узлов на $(i - 1)$ -ой станции больше чем на i -ой станции, или отправлять на перегонный путь если количество задействованных узлов на $(i - 1)$ -ой станции меньше чем на i -ой станции). Согласно правилу взаимодействия с последующей станцией, станция с номером i , в зависимости от количества задействованных узлов на $(i + 1)$ -ой станции, должна уменьшать или увеличивать количество задействованных узлов со скоростью $\alpha(z_i - z_{i+1})$ (т.е. отправлять на следующую

шую станцию если число задействованных узлов на i -ой станции больше чем на $(i + 1)$ -ой станции, или принимать с перегонного пути если число задействованных узлов на i -ой станции меньше чем на $(i + 1)$ -ой станции). Первая технология не учитывает условия ограниченности пропускной способности станций. Кроме того, она не позволяет использовать весь потенциал станций. В связи с этим, наряду с первой технологией, используется и вторая технология. Она позволяет, как увеличить число задействованных узлов (если оно не превышает Δ) так и уменьшать (если оно превышает Δ). При этом груз принимается с перегонного пути либо отправляется на перегонный путь. Из определения второй технологии следует, что функция $\varphi(\cdot)$, задающая скорость изменения числа задействованных узлов обработки в рамках данной технологии, обладает следующими свойствами: на полупрямой $(-\infty, 0]$ тождественно равна 0, на интервале $(0, x_{opt})$ является возрастающей, в точке x_{opt} принимает максимальное значение, на полупрямой $(x_{opt}, +\infty)$ является убывающей, в точке Δ принимает нулевое значение, а на полупрямой $(\Delta, +\infty)$ является линейной.

Приведем описание четвертого варианта модели (исследование грузопотока по круговой цепочке станций) и основной результат данного исследования. Этот вариант модели описывается следующей системой

$$\dot{x}_1(t) = \alpha(x_n - x_1) - \alpha(x_1 - x_2) + \varphi(x_1), \quad t \in [0, +\infty), \quad (1)$$

$$\begin{aligned} \dot{x}_i(t) &= \alpha(x_{i-1} - x_i) - \alpha(x_i - x_{i+1}) + \varphi(x_i), \\ i &= 2, \dots, n-1, \quad t \in [0, +\infty), \end{aligned} \quad (2)$$

$$\dot{x}_n(t) = \alpha(x_{n-1} - x_n) - \alpha(x_n - x_1) + \varphi(x_n), \quad t \in [0, +\infty), \quad (3)$$

$$x_i(t) = x_{i+1}(t + \tau), \quad i = 1, \dots, n-1, \quad t \in [0, +\infty) \quad (4)$$

$$x_n(t) = x_1(t + \tau), \quad t \in [0, +\infty). \quad (5)$$

Здесь $x_i(t)$ — число задействованных узлов на i -ой станции в момент времени t . Дифференциальные уравнения (1)–(3) определяют интенсивность движения грузопотока по круговой цепочке, состоящей из n станций, а соотношения (4) и (5) задают систему контроля. Она заключается в том, что объемы обрабатываемых грузов для любого планового интервала времени на всех станциях должны совпадать с определенным лагом времени, единым для всех станций. Константу τ , которая является сдвигом между моментами замеров и сравнения объемов грузов, будем называть характеристикой системы контроля. Решения системы дифференци-

альных уравнений (1)–(3), удовлетворяющие условиям (4)–(5), называются решениями типа бегущей волны. В работе [1] было показано, что система (1)–(3) с положительными координатами начального значения имеет единственное решение типа бегущей волны (удовлетворяющее условиям (4)–(5)), а именно, стационарное решение $(\Delta, \Delta, \dots, \Delta)$. Кроме того, всякое решение системы (1)–(3) с положительными координатами начального значения сходится к стационарному решению $(\Delta, \Delta, \dots, \Delta)$.

3. Модификация модели

Проведем нумерацию перегонных путей. Перегонный путь, расположенный между станциями с номерами i и $i + 1$ обозначим номером i . Количество задействованных узлов на i -ом перегонном пути в момент времени t обозначим через $y_i(t)$. Предположим, что количество узлов на всех перегонных путях одинаково и равно V . Напомним, что в исходной модели емкости перегонных путей считались неограниченными, т.е. число задействованных узлов на них могло быть произвольным.

При такой модификации модели изменятся уравнения (1)–(3), описывающие интенсивность движения грузопотока на станциях. Кроме того появятся дополнительные уравнения, описывающие интенсивность движения грузопотока на перегонных путях, решения которых должны удовлетворять условию ограниченности емкостей перегонных путей. Начнем с пересмотра уравнения (2). Для того чтобы переписать уравнения (2) заметим следующее.

Первое слагаемое в правой части уравнения (2) описывает взаимодействие i -ой станции с $(i - 1)$ -ой станцией. Согласно правилу взаимодействия с предыдущей станцией, если $x_{i-1} > x_i$, то грузопоток поступает с предыдущей станции, если же $x_{i-1} < x_i$ то грузопоток отправляется на перегонный путь, расположенный между i -ой и $(i + 1)$ -ой станциями (данный перегонный путь мы обозначили номером i). С учетом ограниченности емкостей перегонных путей, первое слагаемое в правой части уравнения (2) примет вид:

$$\begin{aligned} &\alpha(x_{i-1} - x_i) \text{sign}(x_{i-1} - x_i) + \\ &+ \alpha(x_{i-1} - x_i) \text{sign}(x_i - x_{i-1}) \text{sign}(V - y_i). \end{aligned} \quad (6)$$

Функция sign , участвующая в выражении (6) определяется следующим образом:

$$\text{sign}(t) = \begin{cases} 1, & t > 0 \\ 0, & t \leq 0. \end{cases}$$

Второе слагаемое в правой части уравнения (2)

описывает взаимодействие i -ой станции с $(i + 1)$ -ой станцией. Согласно правилу взаимодействия с последующей станцией, если $x_i > x_{i+1}$, то грузопоток отправляется на следующую станцию, если же $x_i < x_{i+1}$, то грузопоток поступает с перегонного пути, расположенного между $(i - 1)$ -ой и i -ой станциями (данный перегонный путь мы обозначили номером $(i - 1)$). С учетом ограниченности емкостей перегонных путей, второе слагаемое в правой части уравнения (2) примет вид:

$$\alpha(x_i - x_{i+1})\text{sign}(x_i - x_{i+1}) + \alpha(x_i - x_{i+1})\text{sign}(x_{i+1} - x_i)\text{sign}(y_{i-1}). \quad (7)$$

Третье слагаемое в правой части уравнения (2) представляет собой интенсивность движения грузопотока в рамках второй технологии и описывает взаимодействие i -ой станции с соседними перегонными путями (с номерами $(i - 1)$ и i). Напомним, что в рамках второй технологии, станция с номером i принимает грузопоток с $(i - 1)$ -ого перегонного пути с интенсивностью $\varphi(x_n)$ если $x_n < \Delta$ и отправляет на i -ый перегонный путь с интенсивностью $\varphi(x_n)$ если $x_n > \Delta$. С учетом ограниченности емкостей перегонных путей, третье слагаемое в правой части уравнения (2) примет вид:

$$\varphi(x_i)\text{sign}(\Delta - x_i)\text{sign}(y_{i-1}) + \varphi(x_i)\text{sign}(x_i - \Delta)\text{sign}(V - y_i). \quad (8)$$

Таким образом, согласно выражениям (6), (7) и (8) уравнение (2) преобразуется в следующее уравнение

$$\begin{aligned} \dot{x}_i = & \alpha(x_{i-1} - x_i)\text{sign}(x_{i-1} - x_i) + \\ & + \alpha(x_{i-1} - x_i)\text{sign}(x_i - x_{i-1})\text{sign}(V - y_i) - \\ & - [\alpha(x_i - x_{i+1})\text{sign}(x_i - x_{i+1}) + \\ & + \alpha(x_i - x_{i+1})\text{sign}(x_{i+1} - x_i)\text{sign}(y_{i-1})] + \\ & + \varphi(x_i)\text{sign}(\Delta - x_i)\text{sign}(y_{i-1}) + \\ & + \varphi(x_i)\text{sign}(x_i - \Delta)\text{sign}(V - y_i), \\ & i = 2, \dots, n-1, t \in [0, +\infty) \end{aligned}$$

Аналогичным образом преобразуются и уравнения описывающие интенсивность движения грузопотока на станциях с номерами 1 и n . Необходимо лишь учесть, что для станции с номером 1 предыдущей будет станция с номером n и, соответственно, для станции с номером n последующей будет станция с номером 1. Итак, дифференциальные уравнения (1) и (3) преобразуются следующим образом

$$\begin{aligned} \dot{x}_1 = & \alpha(x_n - x_1)\text{sign}(x_n - x_1) + \\ & + \alpha(x_n - x_1)\text{sign}(x_1 - x_n)\text{sign}(V - y_1) - \end{aligned}$$

$$\begin{aligned} & - [\alpha(x_1 - x_2)\text{sign}(x_1 - x_2) + \\ & + \alpha(x_1 - x_2)\text{sign}(x_2 - x_1)\text{sign}(y_n)] + \\ & + \varphi(x_1)\text{sign}(\Delta - x_1)\text{sign}(y_n) + \\ & + \varphi(x_1)\text{sign}(x_1 - \Delta)\text{sign}(V - y_1), \quad t \in [0, +\infty); \end{aligned}$$

$$\begin{aligned} \dot{x}_n = & \alpha(x_{n-1} - x_n)\text{sign}(x_{n-1} - x_n) + \\ & + \alpha(x_{n-1} - x_n)\text{sign}(x_n - x_{n-1})\text{sign}(V - y_n) - \\ & - [\alpha(x_n - x_1)\text{sign}(x_n - x_1) + \\ & + \alpha(x_n - x_1)\text{sign}(x_1 - x_n)\text{sign}(y_{n-1})] + \\ & + \varphi(x_n)\text{sign}(\Delta - x_n)\text{sign}(y_{n-1}) + \\ & + \varphi(x_n)\text{sign}(x_n - \Delta)\text{sign}(V - y_n), \quad t \in [0, +\infty). \end{aligned}$$

Теперь необходимо описать закон, по которому меняется число задействованных узлов на перегонных путях. На произвольный перегонный путь грузопоток поступает и с него отправляется как в рамках первой, так и второй технологий. Вспоминая описанные выше обе технологии несложно убедиться, что интенсивность движения грузопотока на перегонных путях будет описываться следующими дифференциальными уравнениями

$$\begin{aligned} \dot{y}_1 = & \alpha(x_1 - x_n)\text{sign}(x_1 - x_n) - \alpha(x_3 - x_2)\text{sign}(x_3 - x_2) + \\ & + \varphi(x_1)\text{sign}(x_1 - \Delta)\text{sign}(V - y_1) - \varphi(x_2)\text{sign}(\Delta - x_2)\text{sign}(y_1), \\ & t \in [0, +\infty); \end{aligned}$$

$$\begin{aligned} \dot{y}_i = & \alpha(x_i - x_{i-1})\text{sign}(x_i - x_{i-1}) - \alpha(x_{i+2} - x_{i+1})\text{sign}(x_{i+2} - x_{i+1}) + \\ & + \varphi(x_i)\text{sign}(x_i - \Delta)\text{sign}(V - y_i) - \\ & - \varphi(x_{i+1})\text{sign}(\Delta - x_{i+1})\text{sign}(y_i), \quad i = 2, \dots, n-1, \quad t \in [0, +\infty); \end{aligned}$$

$$\begin{aligned} \dot{y}_n = & \alpha(x_n - x_{n-1})\text{sign}(x_n - x_{n-1}) - \alpha(x_2 - x_1)\text{sign}(x_2 - x_1) + \\ & + \varphi(x_n)\text{sign}(x_n - \Delta)\text{sign}(V - y_n) - \\ & - \varphi(x_1)\text{sign}(\Delta - x_1)\text{sign}(y_n), \quad t \in [0, +\infty). \end{aligned}$$

Линейная часть правых частей этих уравнений описывает интенсивность движения грузопотока на перегонных путях в рамках первой технологии, а нелинейная часть — в рамках второй технологии.

Кроме этого, должны выполняться следующие неравенства

$$0 \leq y_i(t) \leq V, \quad i = 1, \dots, n, \quad t \in [0, +\infty).$$

Эти неравенства накладывает ограничение на число задействованных узлов на перегонных путях. Окончательно, наша модель, описывающая процесс грузоперевозок и их систему контроля, задается следующим образом:

$$\begin{aligned} \dot{x}_1 = & \alpha(x_n - x_1)\text{sign}(x_n - x_1) + \\ & + \alpha(x_n - x_1)\text{sign}(x_1 - x_n)\text{sign}(V - y_1) - \end{aligned}$$

$$\begin{aligned}
& -[\alpha(x_1 - x_2) \operatorname{sign}(x_1 - x_2) + \\
& + \alpha(x_1 - x_2) \operatorname{sign}(x_2 - x_1) \operatorname{sign}(y_n)] + \\
& + \varphi(x_1) \operatorname{sign}(\Delta - x_1) \operatorname{sign}(y_n) + \\
& + \varphi(x_1) \operatorname{sign}(x_1 - \Delta) \operatorname{sign}(V - y_1), \quad t \in [0, +\infty)
\end{aligned} \quad (9)$$

$$\begin{aligned}
\dot{x}_i = & \alpha(x_{i-1} - x_i) \operatorname{sign}(x_{i-1} - x_i) + \\
& + \alpha(x_{i-1} - x_i) \operatorname{sign}(x_i - x_{i-1}) \operatorname{sign}(V - y_i) - \\
& - [\alpha(x_i - x_{i+1}) \operatorname{sign}(x_i - x_{i+1}) + \\
& + \alpha(x_i - x_{i+1}) \operatorname{sign}(x_{i+1} - x_i) \operatorname{sign}(y_{i+1})] + \\
& + \varphi(x_i) \operatorname{sign}(\Delta - x_i) \operatorname{sign}(y_{i-1}) + \varphi(x_i) \operatorname{sign}(x_i - \Delta) \operatorname{sign}(V - y_i), \\
& i = 2, \dots, n-1, \quad t \in [0, +\infty)
\end{aligned} \quad (10)$$

$$\begin{aligned}
\dot{x}_n = & \alpha(x_{n-1} - x_n) \operatorname{sign}(x_{n-1} - x_n) + \\
& + \alpha(x_{n-1} - x_n) \operatorname{sign}(x_n - x_{n-1}) \operatorname{sign}(V - y_n) - \\
& - [\alpha(x_n - x_1) \operatorname{sign}(x_n - x_1) + \\
& + \alpha(x_n - x_1) \operatorname{sign}(x_1 - x_n) \operatorname{sign}(y_{n-1})] + \\
& + \varphi(x_n) \operatorname{sign}(\Delta - x_n) \operatorname{sign}(y_{n-1}) + \\
& + \varphi(x_n) \operatorname{sign}(x_n - \Delta) \operatorname{sign}(V - y_n), \quad t \in [0, +\infty)
\end{aligned} \quad (11)$$

$$\begin{aligned}
\dot{y}_1 = & \alpha(x_1 - x_n) \operatorname{sign}(x_1 - x_n) - \alpha(x_3 - x_2) \operatorname{sign}(x_3 - x_2) + \\
& + \varphi(x_1) \operatorname{sign}(x_1 - \Delta) \operatorname{sign}(V - y_1) - \\
& - \varphi(x_2) \operatorname{sign}(\Delta - x_2) \operatorname{sign}(y_1), \quad t \in [0, +\infty)
\end{aligned} \quad (12)$$

$$\begin{aligned}
y_i = & \alpha(x_i - x_{i-1}) \operatorname{sign}(x_i - x_{i-1}) - \alpha(x_{i+2} - x_{i+1}) \operatorname{sign}(x_{i+2} - x_{i+1}) + \\
& + \varphi(x_i) \operatorname{sign}(x_i - \Delta) \operatorname{sign}(V - y_i) - \\
& - \varphi(x_{i+1}) \operatorname{sign}(\Delta - x_{i+1}) \operatorname{sign}(y_i), \quad i = 2, \dots, n-1, \quad t \in [0, +\infty)
\end{aligned} \quad (13)$$

$$\begin{aligned}
\dot{y}_n = & \alpha(x_n - x_{n-1}) \operatorname{sign}(x_n - x_{n-1}) - \alpha(x_2 - x_1) \operatorname{sign}(x_2 - x_1) + \\
& + \varphi(x_n) \operatorname{sign}(x_n - \Delta) \operatorname{sign}(V - y_n) - \varphi(x_1) \operatorname{sign}(\Delta - x_1) \operatorname{sign}(y_n), \\
& t \in [0, +\infty)
\end{aligned} \quad (14)$$

$$0 \leq y_i(t) \leq V, \quad i = 1, \dots, n, \quad t \in [0, +\infty) \quad (15)$$

$$x_i(t) = x_{i+1}(t + \tau), \quad i = 1, \dots, n-1, \quad t \in [0, +\infty) \quad (16)$$

$$x_n(t) = x_1(t + \tau), \quad t \in [0, +\infty) \quad (17)$$

Система дифференциальных уравнений (9)-(14) имеет три типа стационарных решений. Опишем их.

1. $x_i = \Delta$, $y_i = c_i$, $0 \leq c_i \leq V$, $i = 1, \dots, n$;
2. $x_i = b$, $b < \Delta$, $y_i = 0$, $i = 1, \dots, n$;
3. $x_i = b$, $b > \Delta$, $y_i = V$, $i = 1, \dots, n$.

Очевидно, что эти решения удовлетворяют также и условиям (15)-(17). Аналитическое исследование других решений системы (9)-(17) (если они существуют) крайне затруднительно, т.к. правые части дифференциальных уравнений (9)-(14) являются

разрывными функциями. В связи с этим система (9)-(17) была исследована численно. Приведем результаты численного исследования.

4. Численная реализация системы (9) - (17)

Вначале были исследованы все решения системы (9)-(15) (т.е. не только решения типа бегущей волны, удовлетворяющие условиям (16)-(17)). В численных экспериментах значение n было равным 5. Как показывают численные эксперименты при любых положительных координатах начального значения решения системы (9)-(15) сходятся к одному из трех указанных типов стационарных решений. Например, на *рис. 1* и *рис. 2* решения системы (9)-(15) сходятся к стационарному решению первого типа.

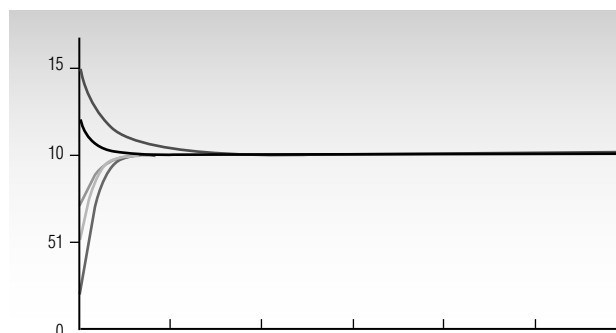


Рис. 1. Количество задействованных узлов на станциях

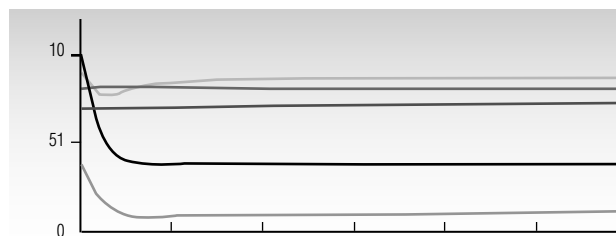


Рис. 2. Количество задействованных узлов на перегонных путях

Здесь $\Delta = 10$, $V = 20$,

$x_1(0) = 2$, $x_2(0) = 15$, $x_3(0) = 5$, $x_4(0) = 7$, $x_5(0) = 12$,
 $y_1(0) = 7$, $y_2(0) = 9$, $y_3(0) = 4$, $y_4(0) = 8$, $y_5(0) = 10$.

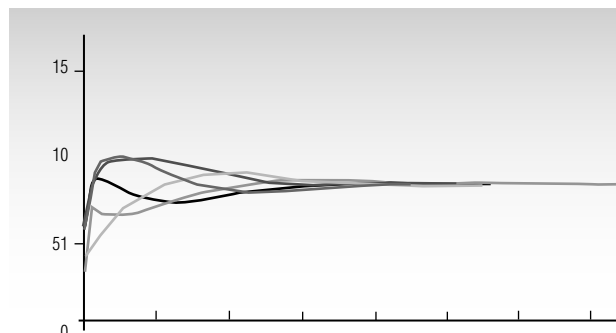


Рис. 3. Количество задействованных узлов на станциях

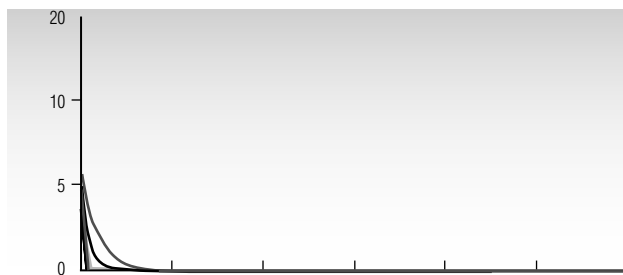


Рис. 4. Количество задействованных узлов на перегонных путях

На рис. 3 и рис. 4 решения системы (9)-(15) сходятся к стационарному решению второго типа.

Здесь $\Delta = 10$, $V = 20$,

$x_1(0) = 4$, $x_2(0) = 5$, $x_3(0) = 2$, $x_4(0) = 3$, $x_5(0) = 6$,
 $y_1(0) = 7$, $y_2(0) = 2$, $y_3(0) = 4$, $y_4(0) = 3$, $y_5(0) = 5$.

На рис. 5 и рис. 6 решения системы (9)-(15) сходятся к стационарному решению третьего типа.

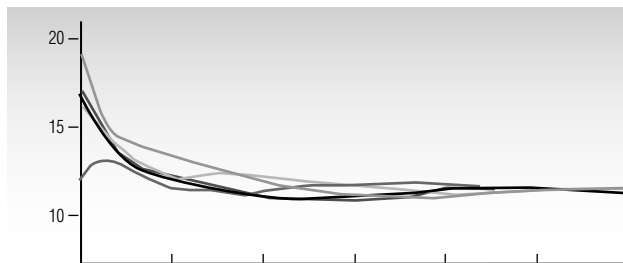


Рис. 5. Количество задействованных узлов на станциях

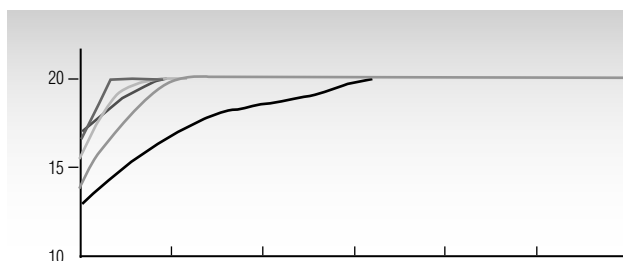


Рис. 6. Количество задействованных узлов на перегонных путях

Здесь $\Delta = 10$, $V = 20$,

$x_1(0) = 12$, $x_2(0) = 18$, $x_3(0) = 16$, $x_4(0) = 19$, $x_5(0) = 17$,
 $y_1(0) = 17$, $y_2(0) = 15$, $y_3(0) = 14$, $y_4(0) = 16$, $y_5(0) = 13$.

Поскольку произвольное решение системы (9)-(15) с положительными координатами начального значения сходится к стационарному решению одного из трех типов, то указанная система не имеет других решений, которые удовлетворяют условиям (16)-(17) бегущей волны, кроме стационарных. Таким образом, система (9)-(17) с положительными координатами начального значения имеет только стационарные решения.

5. Область устойчивости стационарных решений

Стационарные решения системы (9)-(17) исследованы на устойчивость в классе всех решений системы (9)-(15). Здесь следует отметить, что понятие устойчивости применяется только к компонентам стационарного решения, описывающим интенсивность грузопотока на станциях. Таким образом, стационарное решение системы (9)-(17) называется устойчивым, если компоненты произвольного решения системы (9)-(15), описывающие интенсивность грузопотока на станциях, т.е. $x_i(\cdot)$, $i = 1, \dots, n$ близкие к соответствующим компонентам стационарного решения в начальный момент времени сколь угодно близко к ним приближаются, начиная с некоторого момента времени. Как показали численные эксперименты, стационарные решения системы (9)-(17)

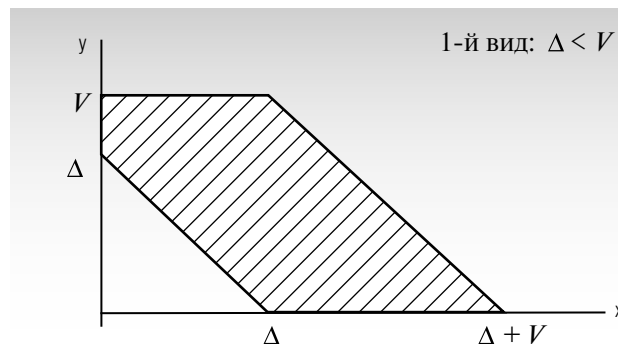


Рис. 7

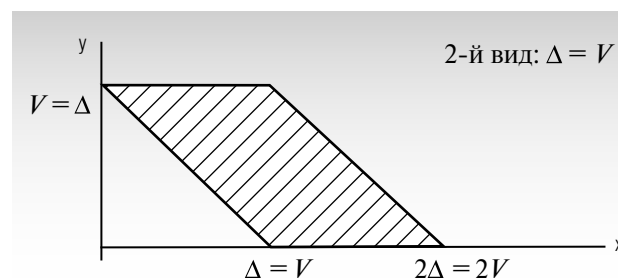


Рис. 8

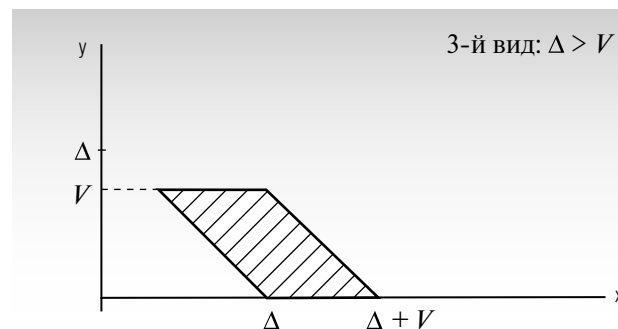


Рис. 9

второго и третьего типов являются неустойчивыми, а первого — устойчивыми. С помощью численных экспериментов определена область устойчивости. В зависимости от взаимного расположения на числовой прямой точек Δ и V получаются три вида областей устойчивости.

По оси x откладывается среднее значение числа задействованных узлов на станциях, а по оси y — среднее значение числа задействованных узлов на перегонных путях в начальный момент времени.

Вернемся к стационарным режимам. С экономической точки зрения наиболее оптимальным является первый режим. Во-первых, только он является устойчивым. Во-вторых, в этом режиме наиболее эффективно используются возможности станций и при этом можно организовать бесперебойную работу системы. Из рис. 7, 8 и 9 следует, что система переходит в стационарный режим первого типа, если сумма средних значений чисел задействованных узлов на станциях и чисел задействованных узлов на перегонных путях в начальный момент времени не меньше Δ и при этом не превышает значения $\Delta + V$. Если сумма средних значений чисел задействованных узлов на станциях и чисел задействованных узлов на перегонных путях в начальный момент времени меньше Δ , то система переходит в стационарный режим второго типа. Наконец, если сумма средних значений чисел задействованных узлов на станциях и чисел задействованных узлов на перегонных путях в начальный

момент времени больше $\Delta + V$, то система переходит в стационарный режим третьего типа.

6. Заключение

В данной статье изучена модель организации грузоперевозок с заданной системой контроля, учитывающая ограниченность емкостей перегонных путей. Такая модель описывается системой дифференциальных уравнений, удовлетворяющих дополнительным условиям, которые определяют систему контроля (задают решения типа бегущей волны) и накладывают ограничения на емкости перегонных путей. Указанная система имеет три типа стационарных решений. Аналитическое исследование множества всех решений сильно осложняется тем, что правые части дифференциальных уравнений являются разрывными функциями. В связи с этим данная модель была исследована численно. Численное исследование позволило показать, что описываемая модель не имеет других режимов грузоперевозок кроме стационарных, удовлетворяющих заданной системе контроля (т.е. решений типа бегущей волны). Кроме того были решены следующие задачи:

- стационарные режимы грузоперевозок исследованы на устойчивость в классе всех режимов (т.е. не только решений типа бегущей волны);
- для устойчивых режимов грузоперевозок определена область их устойчивости. ■

Литература

1. Бекларян Л.А., Хачатрян Н.К. Об одном классе динамических моделей грузоперевозок // Журнал вычислительной математики и математической физики. — 2013. — № 10.
2. Канторович Л.В. Математические методы организации и планирования производства. — Л.: Изд-во ЛГУ, 1939.
3. Канторович Л.В., Гавурин М.К. Применение математических методов в вопросах анализа грузопотоков // Проблемы повышения эффективности работы транспорта: Сб. научных статей. — М.: Изд-во АН СССР. — 1949. — С. 110-138.
4. Данциг Дж., Вольф Ф. Алгоритм разложения для задач линейного программирования // Математика: Сб. переводов. — 1964. — Т. 8, № 1. — С. 151-160.
5. Авен О.И., Ловецкий С.Е., Моисеенко Г.Е. Оптимизация транспортных потоков. — М.: Наука, 1985.
6. Галабурда В.Г. Совершенствование технологии перевозок и увеличение пропускной способности железных дорог. — М.: МИИТ, 1983.
7. Галабурда В.Г. Оптимальное планирование грузопотоков. — М.: Транспорт, 1985.
8. Козовский И.Г. Рационализация перевозок грузов на железных дорогах. — М.: Транспорт, 1977.
9. Beklaryan L.A., Khachatryan N.K. Traveling wave type solutions in dynamic transport models // Functional differential equations. — 2006. — V. 13, № 2. — P. 125-155.
10. Хачатрян Н.К. О решениях типа бегущей волны в одной транспортной модели // Автоматика и телемеханика. — 2003. — №3. — С. 137-149.
11. Хачатрян Н.К. Динамическая модель транснациональных грузоперевозок / Препринт #WP/2002/145. — М.: ЦЭМИ РАН, 2002.

ЭЛЕМЕНТЫ СРЕДЫ РАЗРАБОТКИ ПРОГРАММНЫХ КОМПЛЕКСОВ НА ОСНОВЕ ОРГАНИЗАЦИИ МЕТАМОДЕЛИ ОБЪЕКТНОЙ СИСТЕМЫ

П.П. Олейник,

*кандидат технических наук, доцент, системный архитектор
программного обеспечения, ОАО «Астон»*

E-mail: xsl@list.ru

Адрес: Ростовская обл., г. Шахты, пер. Бугроватый, 80

Рассмотрены основные элементы унифицированной среды быстрой разработки корпоративных информационных систем различных прикладных предметных областей. Описана структура уровня доступа к данным, методика проектирования приложений и реализованные проекты.

Ключевые слова: информационные системы (ИС), методы (паттерны) объектно-реляционного отображения, интегрированная среда разработки.

Введение

Неотъемлемой составляющей любого крупного предприятия или организации является информационная система. В широком смысле под информационной системой (ИС) понимают совокупность технического, программного и организационного обеспечения, а также персонала, предназначенная для того, чтобы своевременно обеспечивать надлежащих людей надлежащей информацией. Крупным подклассом ИС являются корпоративные информационные системы (КИС) под которыми понимают определенную совокупность методов и решений, используемых для создания единого информационного простран-

ства управления и обеспечения деятельности компании [1]. Разработка и внедрение единой КИС, автоматизирующей все ключевые бизнес-процессы организации, является неотъемлемым условием функционирования предприятия любой сферы деятельности. В специфических условиях ведения бизнеса постоянно возникают задачи доработки и внесения изменений в существующий функционал. Именно поэтому многие крупные организации предпочитают собственными силами разрабатывать программное обеспечение и внедрять у себя на производстве. При этом выбор среды разработки и языка программирования во многом определяется именно возможностью быстрого внесения измене-

ний в систему. Основой КИС выступает хранилище информации, т.е. база данных (БД). Автоматизация ключевых бизнес-процессов организации заключается в создании приложений БД. В настоящее время широкое распространение получили реляционные системы управления базами данных (РСУБД) [2-3], а для проектирования и реализации приложений используют объектно-ориентированные языки программирования (ООЯП) [4-6]. Из-за различий в методах обработки информации (реляционная модель построена на основе теории множеств и оперирует группами строк, а ОО-парадигма основана на принципах инкапсуляции и проектирования независимых классов, экземпляры которых могут иметь состояние и методы обработки) возникает так называемое объектно-реляционное несоответствие (Object-Relational Impedance Mismatch) [5] для преодоления последствий которого используют методы (паттерны/шаблоны) объектно-реляционного отображения (ОРО, Object-Relational Mapping (ORM) Patterns) [5-6]. Все имеющиеся паттерны ОРО можно разделить на две большие категории: к первой относятся шаблоны, предназначенные для представления иерархии классов конкретной предметной области в виде набора таблиц реляционной БД (РБД) [5-6, 15]; работы второй категории предоставляют возможность организации метамодели объектной системы в РСУБД, которая используется для различных целей приложения [6-14, 17]. Статьи последней категории рассматривают различные аспекты реализации популярной в настоящее время технологии, названной «архитектура, управляемая моделью (Model Driven Architecture, MDA)», позволяющей описать различные аспекты приложения и предполагающей создание среды разработки приложений.

В статье представлена архитектура унифицированной среды разработки КИС, которая создавалась под руководством автора в течение нескольких последних лет и апробирована в качестве целевой платформы для реализации приложений различных предметных областей. Рассматриваемая среда быстрой разработки получила название SharpArchitect RAD Studio.

1. Критерии оптимальности среды разработки

Создание системы начнем с формирования требований, критериев оптимальности (КО), которым должна удовлетворять реализация среды разработки

(КОср). Т.к. рассматриваемая среда предназначена для разработки корпоративных информационных систем, то созданные КИС также должны удовлетворять требованиям, предъявляемым к КИС [1] на основании которых сформулированы следующие КОср для SharpArchitect RAD Studio:

1. Предусмотреть возможность описания любых прикладных предметных областей. Заранее невозможно предугадать сферу применения системы, поэтому необходим общий способ описания прикладных предметных областей. Имеет смысл использовать структурные диаграммы (например, диаграммы классов) языка UML и на основе их интерпретации строить масштабируемые КИС [1, 22].

2. Реализовать поддержку различных архитектур приложений. В настоящее время существует множество различных архитектур от классической двухзвенной «файл-сервер» до многозвенной архитектуры с поддержкой «распределенных объектов» [4, 6]. Выбор оптимальной структуры во многом зависит от масштаба реализуемой КИС и аппаратных возможностей, имеющихся у предприятия. Выполнение данного КО позволит своевременно и с минимальными затратами масштабировать приложения при возрастающих потребностях организации.

3. Использовать современный язык программирования и современной СУБД. Популярной парадигмой сегодня является объектно-ориентированное программирование, поэтому использование соответствующего языка программирования является наиболее целесообразным [4]. Для хранения информации в настоящее время чаще всего используют реляционные системы управления базами данных (РСУБД). Применение современных технологий позволяет увеличить срок эксплуатации системы и сократить затраты на ее владение [2].

4. Реализовать инфраструктуру безопасности. Разграничение прав доступа к данным и отображение минимального набора информации, достаточного лишь для выполнения должностных обязанностей являются ключевыми особенностями разработки современных КИС [1]. Имеет смысл предусмотреть возможность двух способов проверки подлинности пользователей: 1) внутренней – основанной на учетных записях, имеющихся в КИС (чаще всего используется для Веб-приложений) и 2) встроенной – основанной на механизмах, предоставляемых операционной системой.

5. Обеспечить поддержку механизмов расширения функциональных возможностей среды раз-

работки. Современные языки программирования предоставляют обширную метаинформацию [4], что позволяет реализовать «плагинную» архитектуру, поддерживающую установку модулей расширений, представляющих собой dll-библиотеки [17]. Этот подход позволяет безгранично увеличивать функционал системы и её гибкость.

После того как представлены ключевые КОср можно перейти к описанию внутренних особенностей реализации.

2. Структура метамодели объектной системы

В широком смысле слова метамодель — это модель, позволяющая описать другие модели. В нашем понимании метамоделью является объектная модель, которая описывает модель прикладной предметной области. В основе рассматриваемой среды разработки положена единая метамодель системы, реализованная в виде иерархии классов языка C#. Отметим, что сама метамодель хранится в виде таблиц РБД [7-14, 16-17].

При реализации были сформулированы следующие критерии оптимальности метамодели (КОмм) [16]:

1. Разработать единую иерархию, позволяющую представить ключевые элементы объектно-ориентированной парадигмы, такие как классы, атрибуты, наследование (одиночное и множественное), ассоциации и т.п.

2. Предусмотреть возможность предоставления базовых (системных) классов, в которых реализован наиболее общий функционал

3. Разработать иерархию атомарных литеральных типов, представляющих наиболее распространенные типы данных современных объектно-ориентированных языков программирования.

Рассмотрим требования более подробно. Реализация требования КОмм1 позволит унифицировано обрабатывать все элементы с помощью различных структур данных, например с помощью итераторов при формировании цикла. При этом пользователю предоставляется возможность применять хорошо зарекомендовавшие себя объектно-ориентированные подходы, как например наследование и организация ассоциаций.

КОмм2 требует наличия системных (встроенных) классов. В общем случае пользователи не могут модифицировать поведение этих классов, но имеют

возможность наследоваться от них. При этом наличие множественного наследования предоставляет широкие возможности моделирования задач прикладной предметной области.

Современные ООЯП имеют в своем арсенале богатый набор встроенных атомарных литеральных типов. Поэтому необходимо предоставить подобный функционал в разрабатываемой системе, что соответствует требованию КОмм3 [6, 18].

Идея представления метамодели объектной системы не нова. В работах, посвященных стандарту ODMG 3.0 [19-20], и в работах, описывающих SQL:2003 [21], имеются определённые решения. Но первый стандарт ориентирован лишь на удобство обработки метамодели с помощью встроенных коллекций и итераторов, а второй — на представление сложных типов данных в РСУБД. Поэтому было принято решение самостоятельно разработать иерархию классов, полностью удовлетворяющую поставленной задаче. Результат представленный в данной работе основывается на работах [7-8, 11-13, 16].

На *рис. 1*, в виде диаграммы классов UML [4, 22], представлена разработанная иерархия. Рассмотрим ее соответствие выделенным критериям оптимальности. Иерархия соответствует требованиям КОмм1, т.к. содержит все ключевые элементы ОО-парадигмы: классы, атрибуты, ассоциации и предоставляет возможность реализации. Выполнены и требования КОмм2, т.к. имеется возможность предоставления базовых (системных) классов, реализующих наиболее общий функционал. Разработанная иерархия классов представляет атомарные литеральные типы, которые описывают распространенные типы данных современных ООЯП, т.е. выполнены требования КОмм3. Следовательно, разработанная иерархия удовлетворяет всем выделенным критериям оптимальности метамодели.

3. Организация уровня доступа к данным

Общий процесс разработки корпоративной информационной системы предполагает создание экземпляров классов, образующих метамодель. Затем самой системой выполняется интерпретация этой информации и построения физического уровня доступа к данным, который реализуется в виде набора таблиц РСУБД. Это выполняется на основе реализации методов объектно-реляционного отображения (ОРО).

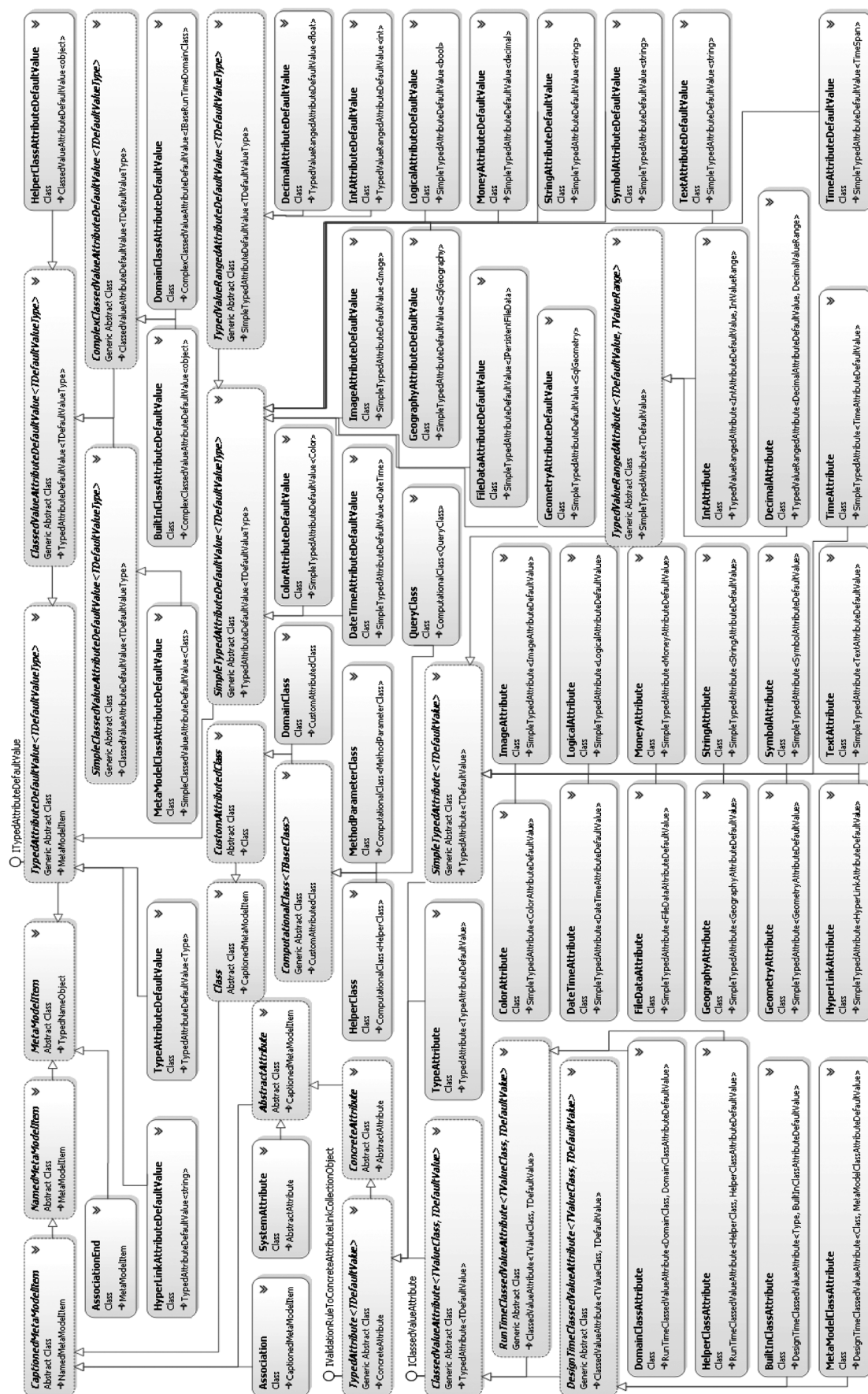


Рис. 1. Иерархия классов метамодели объектной системы

Рассмотрим основные методы ОРО, используемые для представления иерархии классов в виде таблиц РБД (рис. 2) [5-6, 15, 17]. Метод Наследование с одной таблицей (Single Table Inheritance) физически представляет иерархию наследования классов в виде одной таблицы реляционной базы данных, столбцы которой соответствуют атрибутам всех классов, входящих в иерархию и позволяет отобразить структуру наследования и минимизировать количество соединений, которые необходимо выполнить для извлечения информации. В данном методе каждому экземпляру класса соответствует одна строка таблицы. При создании объекта значения заносятся только в те столбцы таблицы, которые соответствуют атрибутам класса, а все остальные остаются пустыми (имеют null-значения).

Метод имеет достоинства:

- ♦ В структуру БД присутствует лишь одна таблица, представляющая все классы иерархии.
- ♦ Для извлечения экземпляров классов иерархии нет необходимости выполнять соединение таблиц.
- ♦ Перемещение полей из базового класса в производный (так же из производного в базовый) не

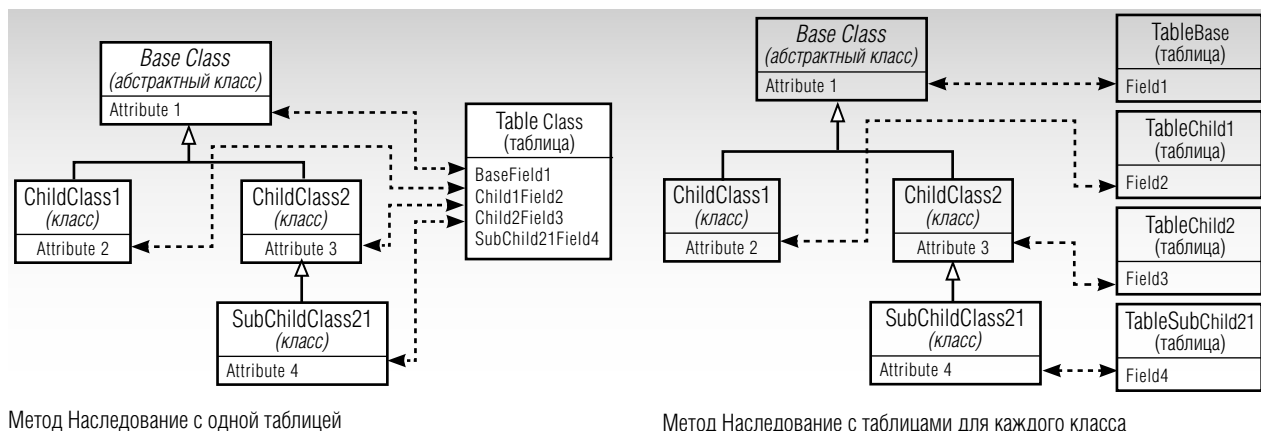
требует внесения изменений в структуру таблицы.

Методу присущи недостатки:

♦ При изучении структуры таблиц БД могут возникнуть проблемы, ведь не все имеющиеся столбцы таблицы предназначены для описания каждого класса предметной области. Это усложняет процесс доработки системы в будущем.

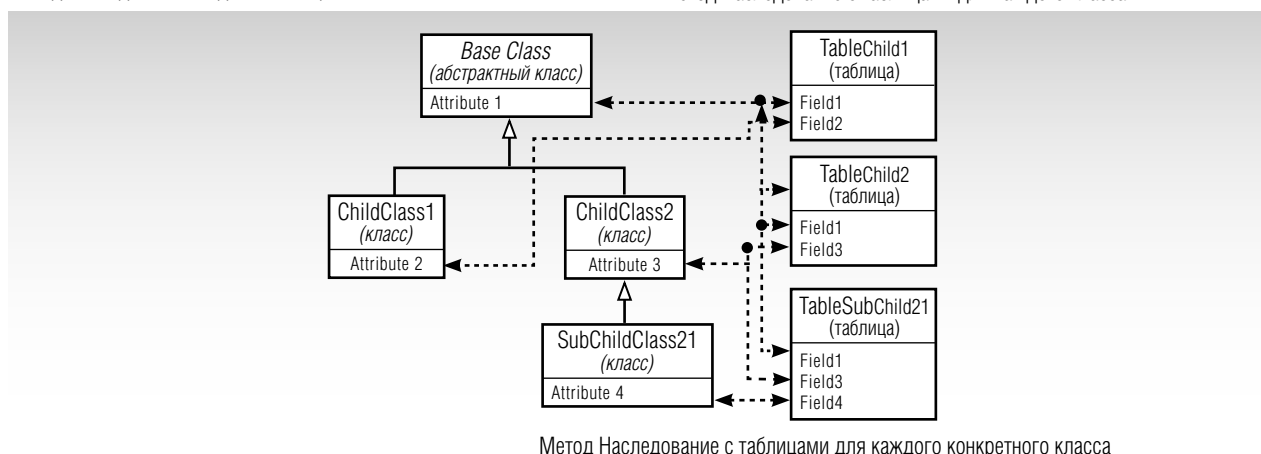
♦ При наличии глубокой иерархии наследования с большим количеством атрибутов многие столбцы могут иметь пустые значения (null-маркеры). Это приводит к неэффективному использованию свободного пространства в БД. Однако современные СУБД способны сжимать строки, содержащие большое количество отсутствующих значений.

♦ Итоговая таблица может оказаться слишком большой и содержать огромное число столбцов. Основной способ оптимизировать запрос (сократить время выполнения) — создать покрывающий индекс. Однако множество индексов и большое количество запросов к одной таблице способны привести к частым блокировкам, что будет оказывать негативное влияние на производительность программного приложения.



Метод Наследование с одной таблицей

Метод Наследование с таблицами для каждого класса



Метод Наследование с таблицами для каждого конкретного класса

Рис. 2. Методы объектно-реляционного отображения

Альтернативным является метод Наследование с таблицами для каждого класса (Class Table Inheritance), представляющий иерархию классов по одной таблице для любого класса (как абстрактного так и реализованного). Атрибуты класса отображаются непосредственно на столбцы соответствующей таблицы. При использовании данного метода ключевой является задача соединения соответствующих строк нескольких таблиц БД, представляющих единый объект предметной области.

Методу присущи преимущества:

- ◆ В каждой таблице присутствуют лишь поля, соответствующие атрибутам определённого класса. Поэтому таблицы легки в понимании и занимают мало места на жёстком диске.

- ◆ Взаимосвязь между объектной моделью и схемой базы данных проста и понятна.

Однако имеются недостатки:

- ◆ При создании экземпляра определённого класса необходимо выполнить загрузку данных из нескольких таблиц, что требует либо их естественного соединения, либо множества обращений к базе данных с последующим соединением результатов в оперативной памяти.

- ◆ Перемещение полей в производный класс или суперкласс требует изменения структуры нескольких таблиц.

- ◆ Таблицы суперклассов могут стать слабым местом в вопросах производительности, поскольку доступ к таким таблицам будет осуществляться слишком часто, что приведёт к множеству блокировок.

- ◆ Высокая степень нормализации может стать препятствием к выполнению незапланированных заранее запросов.

Метод Наследование с таблицами для каждого конкретного класса (Concrete Table Inheritance) представляет иерархию наследования классов, использующий по одной таблице для каждого реализованного (неабстрактного) класса этой иерархии. С практической точки зрения данный метод предполагает, что каждый экземпляр класса (объект), который находится в оперативной памяти, будет отображён на отдельную строку таблицы. При этом каждая таблица в нашем случае содержит столбцы, соответствующие атрибутам как конкретного класса, так всех его предков.

Преимуществами являются то, что:

- ◆ Каждая таблица не содержит лишних полей,

вследствие чего ее удобно использовать в других приложениях, в которых не использован инструмент объектно-реляционного отображения.

- ◆ При создании объектов определённого класса в памяти приложения и извлечения данных из реляционной базы данных выборка выполняется из одной таблицы, т.е. не требуется выполнять реляционные соединения.

- ◆ Доступ к таблице осуществляется только в случае доступа к конкретному классу, что позволяет снизить количество блокировок, накладываемых на таблицу и распределить нагрузку на систему.

При этом имеются недостатки:

- ◆ Первичные ключи могут быть неудобны в работе.

- ◆ Отсутствует возможность моделировать отношения (связи) между абстрактными классами.

- ◆ Если атрибуты классов перемещаются между суперклассами и производными классами, требуется изменять структуру нескольких таблиц. Эти изменения будут не так часты, как в случае наследования с таблицами для каждого класса, однако их нельзя игнорировать (в отличие от метода Наследование с одной таблицей, в котором эти изменения вообще отсутствуют).

- ◆ Если в суперклассе изменить определение хотя бы одного атрибута (например, поменять тип данных), это потребует изменить структуру каждой таблицы, представляющей производный класс, поскольку поля суперкласса дублируются во всех таблицах его производных классов.

- ◆ При реализации метода поиска данных в абстрактном классе потребуется просматривать все таблицы, представляющие экземпляры производных классов. Это требует большого количества обращений к базе данных.

Выбор паттерна ОРО зависит от опций, указанных разработчиком КИС в метамодели.

4. Реализованные проекты

Проект SharpArchitect RAD Studio разрабатывался в последние несколько лет и был использован как для научных целей, так и с целью практической разработки приложений. Для тестирования функциональных особенностей и корректности работы механизма ОРО создана соответствующая модель, представленная в работе [15]. В среде разработано 12 дипломных проектов, выполненных в качестве квалификационных работ в государственных

ВУЗах РФ: в ШИ (ф) ЮГТУ (НПИ) и ИГХТУ. Каждый программный продукт представляет из себя учетную систему, содержащую от 30 до 50 классов предметной области, каждый из которых хранит не менее 100 экземпляров.

В качестве первой крупной реализации создана информационная система проведения научных конференций SharpArchitect Scientific Conference Manager [23], которая используется при проведении конференции «Объектные системы» (www.objectsystems.ru). Работа ведется с 2010 года и реализованы все ключевые модули, позволяющие отслеживать жизненный цикл научной статьи от пересылки ее автором в Оргкомитет, рецензирования и до отслеживания статуса отосланного сборника научных трудов по почте [24-25]. В настоящий момент в системе хранится информация более чем о 300 авторах, 200 статьях и 7 проведенных конференций.

Вторым крупным проектом является реализация платформы создания и управления торговыми биржевыми роботами, архитектура которой представлена в работах [26]. На базе платформы создано более 20 высокочастотных торговых стратегий, одновременно торгующих в режиме реального времени на бирже РТС. Работа системы протестирована на основе реализации алгоритмов техниче-

ского анализа, методов статистического анализа и различных их комбинаций. В качестве тестового массива данных использованы тиковые котировки, составляющие более 1 Гб данных в сутки.

Заключение

Все реализованные проекты представляют собой выполняемые Windows-приложения, предоставляющие развитый интуитивно-понятный пользовательский интерфейс. При этом использована возможность создания классических двухзвенных приложений на основе архитектуры «клиент-сервер» и трехзвенных, с выделенным сервером приложений. В настоящее время все большую популярность получают Веб-технологии. Поэтому ведутся работы по реализации Веб-модуля, позволяющего генерировать соответствующие приложения.

Также разрабатывается официальный Веб-сайт www.sharparchitect.com, который будет посвящен реализованным проектам в SharpArchitect RAD Studio. Планируется разработать мультимедиа руководство по функциональным возможностям и по принципам разработки приложений. Первый обучающий видео-ролик доступен заинтересованному читателю по адресу <https://www.youtube.com/watch?v=BtfEGdSKHEo>. ■

Литература

1. Олейник П. П. Корпоративные информационные системы: Учебник для вузов. Стандарт третьего поколения. — СПб.: Питер, 2012.
2. Коннолли Т., Бегг К. Базы данных. Проектирование, реализация и сопровождение. Теория и практика. 3-е издание: Пер. с англ. — М.: Издательский дом «Вильямс», 2003.
3. Дейт К.Дж. Введение в системы баз данных, 7-е издание: Пер. с англ. — М.: Издательский дом «Вильямс», 2001.
4. Грэхем И. Объектно-ориентированные методы. Принципы и практика. 3-е издание: Пер. с англ. — М.: Издательский дом «Вильямс», 2004.
5. Ambler S.W. Agile Database Techniques — Effective Strategies for the Agile Software. — John Wiley & Sons, 2003.
6. Флауер М. Архитектура корпоративных программных приложений: Пер. с англ. — М.: Издательский дом «Вильямс», 2004.
7. Олейник П.П. Представление метамодели объектной системы в реляционной базе данных // Известия высших учебных заведений. Северо-Кавказский регион. Спецвыпуск «Математическое моделирование и компьютерные технологии», 2005. — С. 3-8.
8. Олейник П.П. Организация метамодели объектной системы на основе реляционной СУБД // Научное творчество молодежи: Материалы X Всероссийской научно-практической конференции (21-22 апреля 2006 г.) Ч. 1. — Томск: Изд-во Томского ун-та, 2006. — С. 87-88.
9. Олейник П.П. Критерии построения унифицированных хранимых процедур выборки в реляционной БД // Математические методы и информационные технологии в экономике, социологии и образовании: Сборник статей XIX Международной научно-технической конференции, г. Пенза, 24-25 мая 2007 г. / Пензенская гос. технолог. акад. — Пенза, 2007. — С. 167-170.

10. Олейник П.П. Об одном способе реализации процедур выборки данных // Международная научно-практическая конференция ИТО Поволжье – 2007. Российский научный семинар «Методы информационных технологий, математического моделирования и компьютерной математики в фундаментальных и прикладных научных исследованиях». (18-21 июня 2007 г.) // Материалы конференции и труды семинара. – Казань: Изд-во «Филиантъ», 2007. – С. 65-67.
11. Олейник П.П. К вопросу о необходимости проектирования иерархии атомарных литеральных типов для объектной системы, организованной в РСУБД // Информационно-вычислительные технологии и их приложения. IX Международная научно-техническая конференция: сборник статей. – Пенза: РИО ПГСХА, 2008. – С. 201-205.
12. Олейник П.П. Организация иерархии атомарных литеральных типов в объектной системе, построенной на основе РСУБД // Программирование. – 2009. – № 4. – С. 73-80.
13. Oleynik P.P. Implementation of the Hierarchy of Atomic Literal Types in an Object System Based of RDBMS // Programming and Computer Software. – 2009. – Vol. 35, No.4. – P. 235-240.
14. Олейник П.П. Организация серверного уровня доступа к данным объектной системы, построенной на основе РСУБД // Автоматизация и современные технологии. Ежемесячный межотраслевой научно-технический журнал. – 2009. – № 6. – С. 15-19.
15. Олейник П.П. Унифицированная модель для тестирования инструментов объектно-реляционного отображения // Объектные системы - 2011: материалы III Международной научно-практической конференции (Ростов-на-Дону, 10-12 мая 2011 г.) / Под общ. ред. П.П. Олейника. – Ростов-на-Дону, 2011. – С. 65-69.
16. Олейник П.П. Иерархия классов метамодели объектной системы // Объектные системы – 2012: материалы VI Международной научно-практической конференции (Ростов-на-Дону, 10-12 мая 2012 г.) / Под общ. ред. П.П. Олейника. – Ростов-на-Дону: ШИ ЮРГТУ (НПИ), 2012. – С. 37-40.
17. Олейник П.П. Методика построения унифицированных трёхзвенных объектно-ориентированных приложений. – Saarbrücken, Germany: LAP LAMBERT Academic Publishing, 2011.
18. Фаулер М., Бек К., Брант Д., Робертс Д., Апдайк У. Рефакторинг: улучшение существующего кода. – М: Символ-Плюс, 2008.
19. Cattell R.G., Barry D.K. The Object Data Standard: ODMG 3.0. – Morgan Kaufmann Publishers, 2000.
20. Джордан Д. Обработка объектных баз данных на C++. Программирование по стандарту ODMG: Пер. с англ.: Уч. пос. – М.: Издательский дом «Вильямс», 2001.
21. Calero C., Ruiz F. and oth. An ontological approach to describe the SQL:2003 object-relational features, Ontologies for Software Engineering and Software Technology. – Berlin Heidelberg: Springer, 2006. – P. 197-215.
22. Новиков Ф.А., Иванов Д.Ю. Моделирование на UML. Теория, практика, видеокурс. – СПб.: Профессиональная литература, Наука и Техника, 2010.
23. Олейник П.П. Информационная система проведения научных конференций SharpArchitect Scientific Conference Manager // Свидетельство об официальной регистрации программы для ЭВМ № 2013610565, 9 января 2013 г. – М.: ВНИИЦ, 2013.
24. Олейник П.П. Опыт применения инструментов объектно-реляционного отображения при разработке информационной системы каталогизирования научных работ // Информационные технологии в науке, экономике и образовании: материалы 5-й Всероссийской научно-практической конференции 2-3 сентября 2010 года / под. ред. О.Б. Кудряшовой; Алт. гос. техн. ун-т, БТИ. – Бийск: Изд-во Алт. гос. техн. ун-та, 2010. – С. 102-105.
25. Олейник П.П., Игумнов Е.А., Свечкарёв Е.А. Опыт проектирования информационной системы для каталогизирования научных работ при проведении международных конференций // Объектные системы - 2010: материалы II Международной научно-практической конференции. Россия, Ростов-на-Дону, 10-12 ноября 2010 г., Ростов-на-Дону, 2010. – С. 48-51.
26. Oleynik P.P. The implementation of the framework to create and manage exchange trading robots // Object Systems – 2011 (English session): Proceedings of the Fourth International Theoretical and Practical Conference. Rostov-on-Don, Russia, 10-12 November, 2011. Edited by Pavel P. Oleynik, P. 42-45.

METHODS OF PATTERN ANALYSIS IN STATICS AND DYNAMICS, PART 2: EXAMPLES OF APPLICATION FOR SOCIAL AND ECONOMIC PROCESSES ANALYSIS

Fuad Aleskerov, Veronika Belousov, Lyudmila Egorova, Boris Mirkin

Annotation

A survey of pattern analysis applications for dynamic macroeconomic analysis of countries, analysis of banking systems in Turkey and Russia, electoral behavior in the United Kingdom and Finland as well as long-term innovative development of regions in Russia is presented.

Key words: data patterns, dynamic pattern analysis, cluster analysis.

ON A METHOD OF VISUAL MODELS CLASSIFICATION

Tatiana Gavrilova, Dmitry Kudryavtsev, Irina Leshcheva, Yaroslav Pavlov

Annotation

Visualization is vitally important to enhance learning both in traditional and distant modes. Visual representation of knowledge boosts speed and quality of understanding of the material. Nowadays, you can find dozens of diverse visual languages. However, the choice of particular visual language is extremely difficult. The main reason for this is absence of clear guidelines for selecting the language. Our work is aimed at design and development of taxonomy of visual languages to cover this research gap. The type of knowledge that can be described by the visual language is the categorization principle of the taxonomy. Thus it can help to choose the visualization technique then the user understand what type of knowledge she wants to represent. This paper can be interesting for wide range of specialists, especially for professional teachers, communicators and business-analysts.

Key words: business modeling, visualization, taxonomy, knowledge, business education.

RISK ASSESSMENT INFORMATION SECURITY SYSTEMS ORGANIZATION WITH MATLAB SYSTEM

Sergey Glushenko

Annotation

The article explains the usefulness of fuzzy logic to evaluate the risk of information security and offers fuzzy production model (FPM). The article presents the implementation process of a fuzzy rule base modeling through the use of specialized package Fuzzy Logic Toolbox software tool MATLAB. Implementation of fuzzy inference algorithm is implemented on the basis of Mamdani.

Key words: risk, fuzzy set, term set, fuzzy production model, the linguistic variable, rule base, the membership function.

PRESENTATION AND COMPARISON METHODS FOR SEMANTICALLY DIFFERENT IMAGES

Georgy Kukharev, Ekaterina Kamenskaya, Nadegda Shchegoleva

Annotation

This paper discusses the methods of presentation and comparison for semantically unrelated images with assessment of their similarity in original feature space, and in Canonical Variables Space (CVS). The projection of the source images in CVS is implemented using two-dimensional canonical correlation analysis algorithm presented in this paper, and the measure of their similarity in CVS is based on the phase correlation.

Key words: comparison of semantically different images, canonical correlation analysis, phase correlation, SSIM.

APPLICATION OF THE EXPERT CLASSIFICATION PRINCIPLES FOR FORMAL CONCEPT ANALYSIS

Daria Stelmashenko, Sergey Soloviev

Annotation

The article is devoted to one of the methods of constructing a formal context lattice. The peculiarities of this method's algorithm are considered. It is suggested that the efficiency of this algorithm can be improved by the principle of expert classification.

Key words: formal concept analysis, expert classification method, principle of domination by specificity, specificity scale.

SCENARIO APPROACH TO THE RESEARCH INSTITUTE'S STAFF ALLOCATION WITH ATTENTION TO THE RISKS

Arkadiy Maron

Annotation

Formulation for the problem of optimal allocation of Research Institute's staff by type of work and customers is offered. It is appropriate to today's realities. The method of solution is developed. It is based on the theory of decision making under uncertainty.

Key words: distribution of staff, optimization, uncertainty, risk.

DYNAMIC MODEL OF ORGANIZATION OF CARGO TRANSPORTATION WITH A LIMITED CAPACITY OF THE DISTILLATION WAYS

Nerses Khachatryan

Annotation

In the paper [1] the model describing process of the organization of a cargo transportation with the given control system was investigated. Admissible modes of cargo transportation are studied, the question of stability of such modes is investigated, the domain of their stability is described. In this paper a modified model with additional resource restrictions is considered.

Key words: differential equations, stationary mode of cargo transportation, stability.

THE ELEMENTS OF DEVELOPMENT ENVIRONMENT FOR INFORMATION SYSTEMS BASED ON METAMODEL OF OBJECT SYSTEM

Pavel Oleynik

Annotation

In the paper the main structure elements of unified rapid development environment for information systems in various domains is presented. The structure of the data access layer, the method of application design and implemented projects are described.

Key words: information systems (IS), methods (pattern) of object-relational mapping, integrated development environment.

Тематические рубрики журнала
«БИЗНЕС-ИНФОРМАТИКА»

№	Рубрика
1	Математические модели социальных и экономических систем
2	Программная инженерия
3	Анализ данных и интеллектуальные системы
4	Математические методы и алгоритмы решения задач бизнес-информатики
5	Моделирование и анализ бизнес-процессов
6	Информационные системы и технологии в бизнесе
7	Электронный бизнес
8	Интернет-технологии
Дополнительные рубрики вне номенклатуры	
9	Тематические обзоры
10	Правовые вопросы бизнес-информатики
11	Стандартизация, сертификация, качество, инновации
12	Дискуссионный клуб / Опыт бизнеса

К СВЕДЕНИЮ АВТОРОВ

Редакция просит авторов при оформлении статей и тематических обзоров придерживаться следующих правил и рекомендаций:

1. Предоставляемый авторами материал должен соответствовать рекомендуемой структуре статей журнала.

2. Статья направляется в редакцию в электронном виде (в формате MS WORD версия 2003) и в виде бумажной копии, распечатанной на одной стороне листов А4. Первая страница оригинала подписывается всеми авторами статьи.

3. Ориентировочный объем статьи, предлагаемой к публикации, – 20–25 тыс. знаков (с пробелами) или 30–35 тыс. знаков – для обзорных статей по направлениям.

4. Кегль набора – 12 пунктов с полупторным интервалом. Нумерация страниц – сверху по центру. Поля: левое – 2,5 см, верхнее, нижнее и правое – по 1,5 см.

5. При наборе выключных и строчных формул должен быть использован редактор формул MS Equation. В формульных и символических записях греческие (русские) символы, а также математические функции записываются прямыми шрифтами, переменные аргументы функций в виде английских (латинских) букв записываются наклонным начертанием (курсивом), например, « $\cos a$ », « $\sin b$ », « $\min$ », « $\max$ ».

6. Формулы, таблицы и сноски (не концевые) оформляются стандартными средствами редактора MS WORD. Нумерация формул, рисунков и таблиц – сквозная, по желанию авторов допускается двойная нумерация формул с указанием структурного номера раздела статьи и – через точку – номера формулы в разделе.

7. Рисунки (графики, диаграммы и т.п.) оформляются средствами Word, Excel, Illustrator. Ссылки на рисунки в тексте обязательны и должны предшествовать позиции размещения рисунка. Допускается использование графического векторного файла в формате wmf/emf или cdr v.10. Фотографические материалы предоставляются в формате TIF или JPEG с разрешением не менее 300 dpi.

8. Библиографический список составляется в соответствии с требованиями ГОСТ. Нумерация библиографических источников – в порядке цитирования. Ссылки на иностранную литературу – на языке оригинала без сокращений.

Структура статей строится по правилам, рекомендованным журналом «Бизнес-информатика».

Плата с аспирантов
за публикацию рукописей не взимается.

Журнал «БИЗНЕС-ИНФОРМАТИКА»
Рекомендуемая структура статей

Журнал публикует исследовательские научные статьи, размещаемые в рубриках журнала, тематические обзоры, отражающие современное состояние проблем в области бизнес-информатики и сообщения, размещаемые в рубриках «Дискуссионный клуб» и «Опыт бизнеса».

Титульный лист рукописи начинается с указания Ф.И.О. авторов публикации с обязательным указанием учёной степени, учёного звания, должности, основного места работы и e-mail. Титульный лист должен быть подписан всеми авторами статьи.

I. Исследовательские научные статьи
(для размещения в тематических рубриках)

Редколлегия рекомендует авторам после названия статьи приводить **аннотацию**, в которой излагается краткое содержание статьи, её основные результаты и область применения. Авторам рекомендуется структурировать статью, выделяя **введение**, содержащее описание проблемы или задачи, обзор существующих подходов или методов решения, их недостатки, и основную цель статьи; **постановку задачи**, включающую допущения и ограничения; **содержательную часть** статьи, в которой предлагаемые решения должны быть аргументированы и сравниваться с существующими подходами или решениями; **заключение**, содержащее краткое изложение новых результатов, полученных в статье и область их применения; **библиографический список**, оформленный в соответствии с ГОСТ. Текст статьи должен содержать нумерованные ссылки на все указанные библиографические источники. Структурирование статьи и нумерация её разделов проводится по усмотрению авторов.

Возможный вариант структуры статьи:

- ◆ Ф.И.О;
- ◆ учёная степень, учёное звание, должность, основное место работы, e-mail;
- ◆ название статьи.
- ◆ аннотация;
- ◆ 1. Введение.
- ◆ 2. Постановка задачи.
- ◆ 3. Основная содержательная часть статьи.
- ◆ 4. Экспериментальные результаты (опционально).
- ◆ 5. Заключение.
- ◆ 6. Библиографический список.

II. Тематические обзоры по направлениям

Редколлегия рекомендует авторам структурировать обзор, выделяя аннотацию, содержащую тематику, краткое содержание обзора и область применения; **введение**, в котором даётся краткий исторический обзор тематики; **содержательную часть** обзора с критическим анализом существующих направлений; **заключение**, в котором отражаются перспективы развития в рамках обзора тематики и наиболее интересные направления с точки зрения научных и практических разработок и методов; **библиографический список**, оформленный в соответствии с ГОСТ.

Текст обзора должен содержать нумерованные ссылки на все указанные библиографические источники. Структурирование обзора и нумерация его разделов проводится по усмотрению авторов.

Возможный вариант структуры обзора:

- ◆ Ф.И.О;
- ◆ учёная степень, учёное звание, должность, основное место работы, e-mail;
- ◆ название обзора;
- ◆ аннотация;
- ◆ 1. Введение.
- ◆ 2. Основная содержательная часть обзора.
- ◆ 3. Заключение.
- ◆ 4. Библиографический список.

Редколлегия журнала проводит обязательное рецензирование рукописей. Статья принимается к публикации только после получения положительного заключения рецензента и одобрения на заседании редакционной коллегии журнала.