

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

9(169)
2010

ТЕОРЕТИЧЕСКИЙ И ПРИКЛАДНОЙ НАУЧНО-ТЕХНИЧЕСКИЙ ЖУРНАЛ

Издается с ноября 1995 г.

УЧРЕДИТЕЛЬ
Издательство "Новые технологии"

СОДЕРЖАНИЕ

КОМПЬЮТЕРНАЯ ГРАФИКА, ОБРАБОТКА ИЗОБРАЖЕНИЙ

- Барский А. Б. Трехмерная экранизация компьютерных объектов "живого" моделирования 2
Коханенко И. К. Нейросетевой и морфологический компьютерный анализ изображений с использованием фракталов 7
Кривцов О. А., Кориков А. М. Методы отслеживания головы человека на видеокдрах: классификация и выбор 13

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ

- Норенков И. П. Генетические алгоритмы поиска решений в онтологических базах знаний 20
Бородащенко А. Ю., Глотов Д. В., Бочков С. М. Алгоритм контент-анализа новостного потока RSS-каналов 25

БЕЗОПАСНОСТЬ ИНФОРМАЦИИ

- Васинев Д. А., Жусов Д. Л., Цынгаев М. С., Ватулин И. Ю. Предложения по повышению защищенности Ethernet-сетей от атак канального уровня 29
Массель А. Г. Методологический подход к организации интеллектуальной поддержки исследований проблемы энергетической безопасности 32

БАЗЫ ДАННЫХ

- Токарчук А. М. Разработка паттерна ActiveModel для программирования интернет-приложения методом объектно-реляционного отражения 37

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ЭКОНОМИКЕ И УПРАВЛЕНИИ

- Димитриенко Ю. И., Димитриенко О. Ю. Модель деформируемых кластеров для анализа динамических данных в экономике. 43
Новицкий В. О., Карпов В. И. Методология исследования и моделирования сложных систем управления для предприятий и компаний зернового сектора АПК 50

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ОБРАЗОВАНИИ

- Зак Ю. А. Математические методы обработки результатов тестирования квалификации и знаний специалистов и абитуриентов 57
Карданова Е. Ю., Карданов Р. С. Построение интервальных оценок статистик согласия при анализе результатов массового тестирования в рамках моделей Г. Раша 64
Герасимова И. Б. Системный анализ процессов передачи и усвоения знаний на основе когнитивного моделирования 70

КОДИРОВАНИЕ И ОБРАБОТКА СИГНАЛОВ

- Дворников С. В., Дворников С. С., Коноплев М. А. Алгоритм распознавания сигналов радиосвязи на основе симметрических матриц. 75
Contents 79
Приложение. Свиридов А. П. Статистическая теория обучения

Главный редактор
НОРЕНКОВ И. П.

Зам. гл. редактора
ФИЛИМОНОВ Н. Б.

Редакционная
коллегия:

АВДОШИН С. М.
АНТОНОВ Б. И.
БАТИЩЕВ Д. И.
БАРСКИЙ А. Б.
БОЖКО А. Н.
ВАСЕНИН В. А.
ГАЛУШКИН А. И.
ГЛОРИОЗОВ Е. Л.
ДОМРАЧЕВ В. Г.
ЗАГИДУЛЛИН Р. Ш.
ЗАРУБИН В. С.
ИВАННИКОВ А. Д.
ИСАЕНКО Р. О.
КОЛИН К. К.
КУЛАГИН В. П.
КУРЕЙЧИК В. М.
ЛЬВОВИЧ Я. Е.
МАЛЬЦЕВ П. П.
МЕДВЕДЕВ Н. В.
МИХАЙЛОВ Б. М.
НЕЧАЕВ В. В.
ПАВЛОВ В. В.
ПУЗАНКОВ Д. В.
РЯБОВ Г. Г.
СОКОЛОВ Б. В.
СТЕМПКОВСКИЙ А. Л.
УСКОВ В. Л.
ФОМИЧЕВ В. А.
ЧЕРМОШЕНЦЕВ С. Ф.
ШИЛОВ В. В.

Редакция:

БЕЗМЕНОВА М. Ю.
ГРИГОРИН-РЯБОВА Е. В.
ЛЫСЕНКО А. В.
ЧУГУНОВА А. В.

Информация о журнале доступна по сети Internet по адресу <http://www.informika.ru/text/magaz/it/> или <http://novtex.ru/IT>.

Журнал включен в систему Российского индекса научного цитирования.

Журнал входит в Перечень научных журналов, в которых по рекомендации ВАК РФ должны быть опубликованы научные результаты диссертаций на соискание ученой степени доктора и кандидата наук.

УДК 004.92

А. Б. Барский, д-р техн. наук, проф.
МИИТ, г. Москва
e-mail: arkbarsk@mail.ru

Трехмерная экранизация компьютерных объектов "живого" моделирования

Приводятся не проверенные экспериментально предположения о возможности построения объемного экрана на базе достижений "прозрачной электроники". Экран реализуется пакетом экранных пленок, на котором трехмерные объекты отображаются по слоям. Такой экран необходим для демонстрации "живых" компьютерных моделей, сценических постановок, реагирующих объектов интеллектуального отображения при управлении предприятием, в исторических и культурных реконструкциях, в зодчестве, ваянии, в сфере туризма и развлечений. Указывается на важность применения модели трехмерной памяти в компьютерной графике для управления объектами в реальном времени с помощью логических нейронных сетей. Рассматривается работа объемного экрана в декартовой и сферической системах координат.

Ключевые слова: прозрачный экран, пакет экранных пленок, "живое" моделирование, трехмерная память, логическая нейронная сеть, декартовы и сферические координаты

Введение

Основной задачей современной компьютерной графики, предназначенной для анимации, является отображение объемных объектов на плоском экране. Применяемые эффекты создают впечатление с ограниченного ракурса: решаются проблемы видимости и скрытости элементов изображения, демонстрации перспективы и освещения. Как правило, анимация использует заранее известные сцены, весьма ограничивая альтернативное развитие сюжета, и практически не допускает динамического, оперативного вторжения в развитие сценария.

Поэтому, когда ставится задача создания средств поддержки "живого" моделирования, то первой и главной задачей является возможность сопряжения моделируемых объектов со средствами оперативного влияния на их поведение, на их адекватную реакцию.

Однако не менее важной задачей является визуализация. Если смело поставить задачу объемной визуализации, визуализации в трех измерениях, то немедленно возникнет необходимость создания объектов в памяти компьютера в трех измерениях, т. е. в модели трехмерной памяти. Объемная визуализация существенно продвинула бы применение компьютерной графики не только в области "живого" моделирования и производственного отображения с первичной диагностикой, но и в областях сценических действий, культурно-исторических реконструкций, электронного зодчества и ваяния, бизнеса туризма и развлечений и др.

Предложения по созданию динамически управляемых с помощью логической нейронной сети трехмерных компьютерных объектов "живого" моделирования представлены в работах [1—3].

В работе [4] отражен первый опыт воплощения реагирующего "живого" объекта на базе "скелетного" отображения, учитывающего возможность подключения логической нейронной сети для адекватного реагирования на складывающуюся внешнюю ситуацию. Однако при этом реализуется "плоская" визуализация на мониторе компьютера.

1. Создание управляемых графических объектов в модели трехмерной памяти

1.1. Модель трехмерной памяти. При построении объекта (монстра, человека и др.) необходимо снабжать его управляемыми органами движения (модификации, деформации, имитирующим улыбку, движение руки и т. д.), подобными мышцам и костям. Их необходимо соединить с управляющим элементом — программой. Программа должна проводить анализ внешней (для объекта) ситуации или внешних воздействий. Это требует применения аппарата логических нейронных сетей, одновременно и в разной степени реализующих ряд отношений вида "если — то", что наиболее близко к поведению живого существа.

Таким образом, построение реального управляемого объекта, а не его математического образа, в памяти компьютера позволяет имитировать любое, в том числе непредсказуемое, движение объекта. Такое представление упрощает визуализацию при наличии трехмерного экрана.

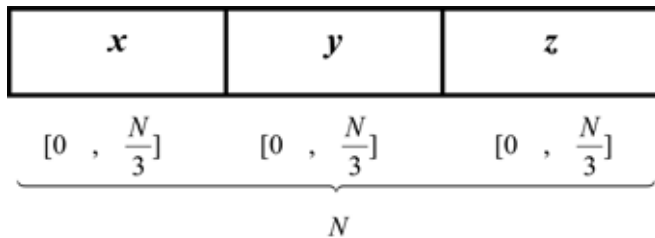


Рис. 1. Преобразование линейного адреса в трехмерный

Назовем ячейку, занятую минимальным элементом объекта, точкой этого объекта. Точка должна содержать информацию о цвете (пиксель) и коэффициент освещенности.

Пусть адресное пространство компьютера, отведенное для модели трехмерной памяти, при сквозной линейной адресации ячеек с нуля определяется N разрядами. Адрес разбивается на три части (рис. 1).

Здесь x, y, z — координаты точки в трехмерной памяти.

Трехмерная память (в данном случае — кубическая) в виртуальной линейной памяти компьютера задается массивом $M[0:N; 0:N; 0:N]$ переменных $m[x, y, z]$. Линейный адрес точки объекта

$$\langle m[x, y, z] \rangle = C + x + yN + zN^2,$$

где C — базовый адрес массива.

Тогда задача имитации движения — деформации или перемещения объектов — превращается в задачу нахождения новых значений индексов-координат для каждой переменной, являющейся точкой объекта.

1.2. Процедуры преобразования. Создание реальных объектов требует введения основных процедур преобразования трехмерной памяти. Такие процедуры создаются [1, 2] на основе следующих построений.

Пусть объект в трехмерном пространстве задан своей оболочкой. Каждый элемент оболочки является точкой. Оболочка формируется при создании объекта. В общем случае, объект может быть заполнен значащей информацией и внутри, если его вид в разрезе интересует пользователя.

Пользователем — разработчиком задаются "мышцы" (далее кавычки опустим) как рычаги управления деформацией объекта в *центральной системе координат*. Мышца задается координатами начала и конца, а также *точкой неподвижности*, относительно которой возможно сокращение мышцы. Точка неподвижности может совпадать с началом или концом мышцы. По умолчанию точка неподвижности является серединой мышцы. Мышца не обязательно связывает точки оболочки. В общем случае она может располагаться внутри и даже вне объекта, принадлежа ему и влияя на его деформацию.

Мышцы объекта являются невидимыми, входящими в список мышц данного объекта и допускающими реакцию на приказы извне.

Взаимодействие мышцы с объектом осуществляется с учетом его свойств. Ведь сокращаясь, мышца должна увлекать точки объекта или только его оболочку, имитируя сокращение всего объекта.

Сокращение твердого тела определяется коэффициентом сжатия (растяжения) объекта.

Вязкое тело, внутри которого сокращается мышца, характеризуется "затуханием" значения смещения точки объекта с увеличением ее расстояния до мышцы.

Сокращение вязкой и упругой среды, такой, как резина, губка, упитанные щеки и др., обычно сопровождается явлением, которое можно назвать попыткой сохранения объема в результате выпучивания. При сокращении мышцы это означает, что перенос клеток объекта вдоль мышцы обретает перпендикулярную составляющую, тем меньшую, чем больше расстояние до мышцы.

Для имитации вращательных движений используют шарниры. Координаты шарнира задаются точкой внутри объекта. Задается мышца, связывающая лучи (элементы скелета), исходящие из центра шарнира. Сокращение мышцы должно вызывать видимость движений, характерных при ходьбе, движении рук и т. д. Это требует таких деформаций объектов, при которых его клетки, облегающие эти лучи, или только клетки оболочки, несущие в себе данные лучи, сближаются вместе с лучами, не приводя к дополнительной деформации.

Взаимодействие многих объектов требует наличия центральной системы координат, в которой происходят все перемещения объектов относительно друг друга, и систем координат, связанных с каждым объектом, — объектовых систем координат. Объект создается в его объектовой системе координат. Его движения "относительно себя" — сокращение мышц, повороты, вращение — удобнее наблюдать в связанной с ним объектовой системе координат. Таким образом, центральная система координат должна быть связана со многими объектовыми системами матрицами пересчета.

1.3. Внешнее воздействие. Необходимость внешнего воздействия на объект обусловлена не только взаимодействием различных объектов, но и созданием инструментов формирования, "ваяния" объекта. Идея такого ваяния заключается в следующем.

Первоначально формируется некоторая заготовка, имеющая наиболее близкую форму для желаемого объекта. (Например, шар — для формирования головы.) Приближая к заготовке объект — "инструмент", например молоток, необходимо потребовать, чтобы при угрозе прикосновения

к объекту ближайшие его клетки отступали, и оболочка, продавливаясь, принимала форму проникающего инструмента. Эта операция напоминает ковку или штамповку. Таким многократным воздействием с разных сторон можно добиться любой формы объекта.

Можно допустить не только вдавливание инструмента в объект, но и вытягивания близлежащей области, подобно нарыву. В этом случае после соприкосновения с объектом следует отводить инструмент от него. Близкая область оболочки должна вытягиваться вслед за инструментом.

Если не деформировать оболочку, а позволить инструменту проникать в объект, то подобное действие сравнимо с действием ножа. Так от объекта могут отсекаются части, — для последующего уничтожения или для создания новых объектов.

1.4. Построение объектов. Для построения объектов целесообразно воспользоваться тремя координатными проекциями, подобно 3D-MAX. Сначала в одной из проекций рисуется пока неточный, предполагаемый разрез объекта. На других проекциях автоматически сформируются отрезки — боковые проекции этого разреза. На разрезе определяется точка для дальнейшей деформации объекта в двух других проекциях. Проекция этой точки высвечивается на двух других проекциях.

Подведя мышь к одной из проекций точки, следует повести ее (мышь) вверх или вниз, формируя выпуклость вслед за этим движением так, чтобы первоначально заданная плоская поверхность образовала выпуклость. Так сформируется объемное изображение. Если при этом придерживать клавишу *control* (или другим способом), плоская поверхность сохранится для замкнутости объема или для последующей аналогичной деформации в ту или другую сторону.

Повороты получившейся объемной фигуры позволяют выполнять указанным способом различные деформации.

1.5. Командно-программное управление объектами. Введение мышц, шарниров, объектовых систем координат позволяет программировать действия объектов и их взаимное расположение. Программа создается на базе командного языка, содержащего инструкции вида:

- < сократить мышцу m с $l = 0,8$ > ;
- < повернуть объект P по матрице S > ;
- < приблизить объект P > ;
- < перенести объект P > ;
- < показать срез $z = 5$ > и т. д.

Таким образом, могут планироваться и исполняться сцены, основанные на имитации движения объектов, их мимики, преобразования и т. д.

При программировании поведения объектов учитывается длительность выполнения действий,

их инерционность, предшествующие команды управления.

Программа управления объектами, как и каждая компьютерная программа, должна предусматривать альтернативные действия в случае выполнения предусмотренных условий, т. е. содержать команды условного перехода.

1.6. Логические нейронные сети в основе управления трехмерными компьютерными объектами. Занимаясь "живым" моделированием на основе компьютерных и информационных технологий, следует быть приверженным точным наукам, в которых любые построения начинаются с исчерпывающих, непротиворечивых определений.

Что есть жизнь? Какое существо можно назвать живым?

1. Живое существо адекватно реагирует на внешние раздражители.

2. Адекватность реакции живого существа подчиняется критерию "хорошо — плохо".

Однако достичь такого уровня обучения, когда система *сама* начинает осознавать, что хорошо, а что плохо, на основе индуктивного и дедуктивного мышления, вряд ли сейчас возможно практически. Пока целесообразно решать задачу-минимум, требуя для культурно-развлекательных целей и задач отображения адекватности реакции, предусмотренной разработчиком.

В этом случае внешняя "картинка" отображается в пространстве признаков, обоснованных для распознавания и принятия решений. На самом простом практическом уровне может проводиться подсчет различных оттенков цветов в элементарном квадрате "картинки", обнаружение резких цветовых границ, наличие линий, кругов определенного цвета и т. д. [5].

На этапе построения (обучения) на основе предположений об адекватных действиях строятся связи между достоверными ситуациями на рецепторном слое и нейронами выходного слоя однослойной логической нейронной сети. По возбуждению этих нейронов запускаются программы процедур необходимых действий. Значения возбуждения нейронов являются основными параметрами этих процедур.

В рабочем режиме достоверность высказываний о принадлежности исходных измеренных значений отображается на рецепторном слое в качестве нечетких данных.

Ограничиваясь только цветом, можно в терминах алгебры высказываний представить следующий пример алгоритма реакции объекта.

"В квадрате i, j преобладает зеленый цвет" \wedge "отсутствуют резкие переходы в оттенках" \rightarrow "Запустить программу умиротворения";

"В квадрате i, j преобладает желто-оранжевый цвет" → "Запустить программу радостного возбуждения" и т. д.

2. Создание стереоэффекта с помощью системы прозрачных мониторов

Необходимо помнить, что объекты в трехмерной памяти создаются для их визуального восприятия. Система визуализации, как это происходит сейчас, может базироваться на проецировании видимой поверхности всех объектов, включая фон, на плоскость $z = z_{\max}$ в центральной системе координат. Эта плоскость и представляет экран. Значит, каждая прямая, исходящая из точки (x_i, y_i, z_{\max}) и перпендикулярная экрану, продолжается (здесь рассматриваем только прозрачную среду) до первого пересечения с одним из объектов или с фоном. Полученная точка обеспечивает изображение в данной текущей точке экрана.

Таким образом, проецирование объемного изображения на плоский экран для визуализации осуществляется "на всю толщину" изображения, т. е. на весь диапазон изменения z , $0 \leq z \leq z_{\max}$. Однако для получения различных (томографических) срезов экран может формироваться и в других плоскостях вида $z = a \geq 0$.

Можно ли "нарезать" отображаемый объект или все пространство по оси z , направленной на зрителя, на слои "толщиной" Δz (рис. 2, см. третью сторону обложки), каждый такой слой проецировать на свой "передний" срез, как на отдельный экран, расположив экраны друг за другом, как показано на рис. 3 (см. третью сторону обложки)? Ведь прозрачные мониторы уже поступили в продажу!

Конечно, на рисунке наблюдается весьма слабый стереоэффект.

Изображения на разных экранах, в зависимости от угла наблюдения, оказываются смещенными относительно общей требуемой картины: с одной стороны, образуются пустые зазоры, с другой стороны, изображения накладывают друг на друга. Для получения стереоэффекта зрителю необходимо находиться строго напротив экранов, как показано на рис. 3.

А если потребовать условного выполнения требования $\Delta z \rightarrow 0$? Технически и технологически это требует значительного роста числа используемых прозрачных экранов на основе достижений "прозрачной электроники". Становится оправданной разработка экранных пленок с предельной толщиной "в одну точку", чтобы "нарезка" выродилась в представление множества срезов. Собранные (склеенные) на их основе пакеты должны представлять объемное (трехмерное) экранное пространство.

В многочисленных сообщениях о достижениях в области "прозрачной электроники" пока нет ссылок на исчерпывающие научные публикации. Следует привести лишь факт результативности проводимых исследований.

Так, в отчете "Transparent Electronics Based on Printed Aligned Nanotubes on Rigid and Flexible Structures", опубликованном в журнале ACS Nano (2009), рассказывается о том, что коллектив исследователей под руководством профессора Чонгву Чжоу (*Chongwu Zhou*) решил проблему нанесения плотных матриц углеродных нанотрубок не только на жаропрочное стекло, но также на гибкие и прозрачные пластиковые подложки, весьма восприимчивые даже к минимальному нагреву. Более того, ученым удалось не просто интегрировать решетчатые массивы транзисторов на базе нанотрубок в прозрачный пластик, они ухитрились найти способ подключения к этим структурам серийно выпускаемых светодиодов на основе нитрида галлия (GaN), способных при подаче напряжения изменять яркость свечения в 1000 раз. Обзор результатов в данной области можно найти, например, в работе [6].

3. Прямоугольное экранное пространство

Современные плоские экраны используются с весьма ограниченным ракурсом. Поэтому мы и размещаем телевизор где-то в сторонке, в углу. Возможно, что и столь дорогое кубическое или прямоугольное экранное пространство обречено на ограниченный угол обзора для создания должного впечатления. В этом случае прямоугольный трехмерный экран также достоин размещения где-то в углу помещения (рис. 4, см. третью сторону обложки).

Следует отметить, что экранное пространство образует некоторую среду, в которой действуют законы оптики. Например, не оказывается ли демонстрируемый объект погруженным в сосуд с жидкостью, подобно рыбке в аквариуме?

Более того, возможно, что взгляд сбоку, со стороны торцов экранных пленок, может встретиться с эффектом существенного отличия изображения от получаемого при взгляде спереди.

Это также может ограничить применение экрана на основе пакета плоских пленок. Однако его применение может быть вполне оправдано и даже целесообразно в случае создания реагирующего объекта для интеллектуального отображения производственного процесса, предсказания погоды или социального напряжения, для сценических постановок.

4. Сферическое экранное пространство

Даже при бытовом применении объемного экрана возникает желание кругового обзора с одинаковым качеством изображения. Необходимо рассмотреть возможность экранизации в сферической системе координат. В ней каждая точка характеризуется расстоянием r и двумя углами: азимутом φ и углом места θ . Пересчет из декартовой системы проводится по формулам

$$r = \sqrt{x^2 + y^2 + z^2}, \quad \varphi = \arctg \frac{x}{y}, \quad \theta = \arcsin \frac{z}{r}.$$

Применение сферической системы координат определяет форму объемного экрана: он становится полусферой (рис. 5, см. третью сторону обложки). Центр системы координат совпадает с центром основания прямоугольника.

Однако важным отличием является то, что сами экранные пленки должны быть сферическими. Вместо нарезки на Δz , становится актуальной нарезка на Δr . Центр полусферы также вырождается в сферическую пленку "на одну точку".

В компьютерной модели трехмерной памяти весь обозримый объем представляет собой прямоугольник, а скорее всего — куб. При преобразовании в сферическую систему координат возникает вопрос: вписать этот куб в полусферу или вписать полусферу в куб? В первом случае нет потери информации, что важно, например, при отображении сценических действий. Однако при этом не используется существенная часть экранного пространства. Во втором случае теряется информация, но увеличивается масштаб представляемого объекта при его центральном размещении. Это важно при построении реагирующих объектов для интеллектуальных систем отображения работы предприятий.

Заметим, что в первом случае "лишнюю" часть сферического экрана можно просто срезать. Получится экран — куб, со сферическим расположением экранных пленок, использующий сферическую систему координат. Торцевой эффект, если такой существует, будет значительно снижен.

При круговом обзоре экранного пространства важно не допустить, чтобы свечение внешней поверхности экранной пленки было видно с другой ее стороны. Это испортило бы общее впечатление. По-видимому, следует затенять с другой стороны каждую светящуюся точку. Это может потребовать перемежения экранных и затеняющих пленок или нанесения двух слоев "прозрачной электроники" на каждую экранную пленку. И то и другое, несомненно, будет отрицательно влиять на прозрачность и разрешающую способность сферического объемного экрана.

Теоретически возможно выполнение модели трехмерной компьютерной памяти в сферической системе координат. Однако "математика" анимации представляется весьма сложной. Вряд ли это следует считать целесообразным. Ведь со временем экранные пространства, вбирающие лавинные достижения нанотехнологий, будут дешеветь. Да и в массовом применении вряд ли привычная сцена превратится в полусферу.

Применение сферических экранных пленок позволяет не только создавать экранные пространства для внешнего наблюдения, но и для "внутреннего" расположения зрителей, для которых этот экран становится "внешним", как показано на рис. 6 (см. третью сторону обложки). Можно легко себе представить разнообразные применения такого экрана в области туризма и развлечений.

Заключение

В нынешних условиях нет возможности провести эксперимент, и приведенные рассуждения напоминают фантастический сюжет. Однако даже наивная фантастика всегда предшествовала серьезным разработкам. Фантастика ставит задачи. Их реальность основана на текущем уровне знаний. Технология прозрачных мониторов и гибких прозрачных электронных схем подсказывает реализуемый, в сегодняшнем представлении, путь создания средств объемного отображения. Предстоят еще значительные усилия по достижению достаточной прозрачности пакета экранных пленок, и объекты могут казаться, словно выплывающими из мрака. Не следует забывать о создании особой оптической среды со специфическими свойствами и т. д.

Так что вполне возможно, что объемный экран на основе пакета экранных пленок может оказаться временным достижением. Временным до нахождения той среды, по координатам точек которой будут доставляться кванты энергии, несущие пиксели свечения.

Список литературы

1. Барский А. Б. Логические нейронные сети: методика построения и некоторые применения // Информационные технологии. 2006. № 8. Приложение.
2. Барский А. Б. Логические нейронные сети. — М.: ИНТУ-ИТ; БИНОМ. Лаборатория знаний, 2007.
3. Барский А. Б., Милютин Л. Б., Тимофеев А. Е. Реагирующие объекты для систем интеллектуального отображения // Информационные технологии. 2007. № 2.
4. Вахромцев Е. М. Применение методов "живого" моделирования для производственного мониторинга // Информационные технологии. 2010. № 7. С. 10—15.
5. Форсайт Д., Понс Ж., Компьютерное зрение. Современный подход. — М.: Вильямс, 2004.
6. www.nanometer.ru/2008/12/24.

И. К. Коханенко, д-р техн. наук, проф.,
Ростовский военный институт ракетных войск,
e-mail: kik112@yandex.ru

Нейросетевой и морфологический компьютерный анализ изображений с использованием фракталов

Обосновываются алгоритмы компьютерного автоматического анализа изображений на основе фрактальной кластеризации. Доказывается рациональность использования фракталов, приводящая к синтезу систем с высокой устойчивостью за счет обеспечения высокой степени разнообразия. Показывается, что ядро изображения — то, что отличает его от других, — содержит малое число кластеров. В связи с этим рационализируется морфологический анализ и анализ сетями Кохонена. Приводится пример распознавания изображения участка лунной поверхности.

Ключевые слова: фрактал, распознавание, видеоинформация, кластер, инвариант, интенсивность, нейросеть, морфологический анализ

Введение

Во многих алгоритмах анализа видеоинформации, в морфологическом анализе, в автоматической сегментации текстурированных изображений, в самоорганизующихся кластерных картах Кохонена конструктивно используется наличие в измерительной информации симметрии, выражающейся в инвариантности формы объекта к множеству преобразований, характерных для условий измерений, и приводящей к кластерам инвариантов — однородным по некоторым признакам объектам. Это основные внешние черты объекта, они связаны с неизменными свойствами и определяют его форму.

Полагая, что измеренные изображения имеют подобные инварианты, можно представить модель изображения в виде двумерной матрицы $R = N_x \times N_y$ пикселей, каждая строка которой задается последовательностью пар чисел: I_i — интенсивностью принимаемого сигнала и n_i — числом пикселей, имеющих интенсивность I_i :

$$[(I_1, n_1), (I_2, n_2), \dots, (I_m, n_m), \Sigma n_i = N_x]$$

Такая модель сводится к кусочно-постоянному описанию измеренного изображения m кластерами A_i с постоянной интенсивностью: $I_i \chi_i(x, y)$, $i = 1, 2, \dots, m$, $(x, y) \in A_i \rightarrow \chi_i(x, y) = 1$, $(x, y) \notin A_i \rightarrow \chi_i(x, y) = 0$; в том числе, один из кластеров есть

фон $I_m \chi_m(x, y)$. Подобное описание модели достаточно полно обосновано в работах [1, 2].

Далее, если следовать морфологическому подходу, строится форма объекта в виде проектора $P_f g$ — проекции изображения g на множество форм $V(f)$ или правило, по которому изображению g из гильбертова функционального пространства ставится в соответствие некоторое изображение $\phi = P_f g$. В соответствии с работами [1, 2] под формой изображения понимается множество изображений с упорядоченной интенсивностью:

$$V(f) = \{f(x, y) = \Sigma I_i \chi_i(x, y) + I_m \chi_m(x, y) \in X, \\ -\infty < I_m \leq I_i < \infty, i = 1, 2, \dots, (m - 1)\}. \quad (1)$$

Искомое правило распознавания должно выявить то изображение из множества $V(f)$, которое было бы наиболее близким к изображению g . При этом, если

$$\phi = P_f g = g, \quad (2)$$

то считается, что объект распознан, т. е. можно подобрать условия регистрации, при которых объект породит изображение g . Множество $V(f)$ полагается выпуклым и замкнутым, т. е. проектор P_f определяется из условия минимума расстояния от изображения до формы, что приводит к процедуре построения P_f как усреднения изображения в каждом кластере постоянной интенсивности [1]. Итак, в каждом кластере находится средняя интенсивность, которая принимается постоянной в нем. Но при этом в контексте задачи распознавания изображения немаловажно знать, как распределяется эта постоянная интенсивность в каждом кластере, сколько пикселей располагается в различных подобластях кластера. Очевидно, что такое распределение пикселей в кластере является характеристикой изображения. Поэтому, не изучая такие характеристики, устройство узнавания теряет информацию.

В таком алгоритме есть еще одна особенность: неясно как определять кластеры постоянной интенсивности A_i , сколько их должно быть и каковы их размеры. Обычно рекомендуется отвечать на подобный вопрос так: полагая, что области изображения одинаковой интенсивности сформированы модулями объекта с идентичными физическими (оптическими, акустическими и т. п.) и геометрическими свойствами, задавать кластеры A_i по расположению однородно отражающих граней или границ относительно наблюдателя. Однако физически такое решение не всегда реализуемо, и, кроме того, часто постоянство интенсивности достаточно условно; как правило, желаемое постоянство приближенно представляет некото-

рое множество значений интенсивности в сравнительно небольшом диапазоне.

Аналогичные вопросы характерны и для алгоритмов, реализуемых в нейросетях, в частности, в слоях, картах Кохонена. Там кластеры — это не области постоянной интенсивности, а области с подобным (близким) пространственным распределением интенсивностей (признаков). И при этом в задаче распознавания остаются вопросы первоначального выделения на изображении кластеров — на каких основаниях и сколько; необходимы основания для назначения приблизительного числа кластеров, без которых структуру слоя Кохонена синтезировать нельзя. В контексте распознавания изображения немаловажно знать и фрактальную размерность пространств "интенсивность — ранг интенсивности" или "число входных векторов — ранг кластера" для каждого кластера. Это также характерный признак. Не оценивая его, нейросетевое устройство распознавания теряет информацию.

Ниже на основе фрактальной математики решаются следующие задачи развития морфологического анализа и анализа изображений сетями Кохонена: разработка способа выделения кластеров постоянной интенсивности, определение их числа и размеров, оценка распределения пикселей в кластере и фрактальной размерности пространств "интенсивность — ранг интенсивности" или "число входных векторов — ранг кластера" для каждого кластера.

1. Фрактальность формы изображения

Для многих реальных объектов каждый кластер A_i будет иметь характерное для него число n_i пикселей постоянной интенсивности I_i . В связи с этим целесообразно дополнить зависимость $I(x, y)$ связью интенсивности с числом пикселей в области A_i . С $n_i(I_i)$ связана форма изображения; поэтому естественно искать такие зависимости в среде систем, которые обладают свойствами разнообразия и устойчивости, так как разнообразие формы является необходимым условием полноты адекватной проверки по правилу (2), а устойчивость — необходимым условием сохранения ее характеристик и наблюдаемости.

Понятие разнообразия обычно связывается с энтропией и информацией. В состоянии экстремума энтропии при единственном ресурсе (усилии) задача на условный экстремум приводит в зависимости от характера ограничений формы ресурсами к различным ранговым распределениям: Гиббса, Ципфа — Парето, распределения "разло-

манного стержня" Макауртура [3]. Так, например, распределение Ципфа — Парето имеет вид

$$p_I = \mu I^{-1/(d-1)},$$

где полагается, что потребности кластеров в ресурсе пропорциональны $\sim \ln(I)$; например, в соответствии с физиологическим законом Фехнера потребность усилий по воспроизведению слова ранга I в тексте $\sim \ln(I)$ и поэтому частота появления в тексте слова ранга I есть фрактал.

Здесь $\mu = p_1, p_1 > p_2 > \dots > p_m, m$ — число кластеров (видов) в системе (сообществе), d — в термостатике трактуется как величина, обратная температуре, а в законе Ципфа — Парето d — фрактальная размерность.

Как правило, считается, что разнообразие системы тем больше, чем больше в ней кластеров и выше выравненность ($E = (mp_i/\ln p_i)/i$) — отношение наблюдаемого разнообразия к максимальному. Здесь величина $mp_i/\ln p_i$ известна как индекс Шеннона, типичный диапазон изменения его [1,5...4,5).

Оценка свойств этих законов (плотности, энтропии, выравненности, индекса Шеннона) показывает, что наибольшим разнообразием, а следовательно, совершенством и стабильностью обладают системы, структура которых связана с распределениями Ципфа — Парето и Макауртура, причем наибольшее разнообразие у последнего распределения. Если первый закон имеет известное прямое отношение к фрактальным структурам, то относительно закона Макауртура — это не очевидно. Изучая механизм формирования распределения Макауртура можно показать, что оно сводится к ранговому распределению Ципфа — Парето $x(r) \sim r^{-\beta}, \beta = 1/(d-1)$, т. е. закон Макауртура тоже относится к классу фрактальных распределений.

Вывод по разнообразию подтверждается и при использовании аппарата теории информации. Очевидно, характер восприятия системой сигналов внешней среды зависит от ширины полосы частот сигналов, воспринимаемых системой и ее элементами. При этом, чем шире полоса частот, тем больше не только разнообразие, но и ошибок в восприятии сигналов внешней среды и больше вероятность неустойчивости системы. Таким образом, как разнообразие системы, т. е. число кластеров и элементов в них, так и неадекватность ее поведения, а следовательно, вероятность ее деградации зависят от ширины полосы частот воспринимаемых ею и ее кластерами и элементами сигналов.

В контексте изучаемого интересно сопоставить ширину полос пропускания систем "ранг кластера — число пикселей одинаковой интенсивности в

нем" с различными законами распределения. Для этого следует пояснить понятие ширины полосы пропускания применительно к изучаемой кластерной модели изображения. Ранговому распределению Ципфа — Парето соответствует частота, связанная с фрактальной размерностью d и рангом r кластера следующим [3] соотношением:

$$\omega_r = C[\exp(1/d - 1)]^{-\ln(r)}, \quad C = \text{const.}$$

При этом полагается линейная зависимость числа пикселей в кластере от частоты. Отсюда ранговое распределение Ципфа — Парето может быть записано в виде

$$n(r) = n(1)r^{-1/(d-1)}, \quad 1/d - 1 = \ln(v), \\ v = (\omega_r/\omega_{r+1})^{-\ln(r+1)}.$$

Отношение $\Delta 2$ частот кластеров i и $(i + 1)$ для закона Ципфа — Парето равно $\Delta 2 = \sqrt{\ln(r+1)}$, $v = \exp(1/d - 1)$. Аналогично для законов Гиббса (отношение $\Delta 1$) и Макауртура (отношение $\Delta 3$):

$$\Delta 1 = v, \quad \Delta 3 = \ln(r + 1).$$

Из приведенных зависимостей следует, что для рангов $r \geq 2$ наибольшая ширина полосы частот — у закона Ципфа — Парето и наименьшая — у закона Макауртура. Это согласуется с характером разнообразия систем, для которых характерны изучаемые законы распределения, — фрактальность системы способствует ее устойчивости за счет обеспечения высокой степени разнообразия и приемлемой зашумленности при рациональной ширине пропускания.

Итак, системы с законом Макауртура и Ципфа — Парето относятся к классу фрактальных. Фрактальность системы способствует ее устойчивости за счет обеспечения высокой степени разнообразия и приемлемой зашумленности при рациональной ширине пропускания. Поэтому для устойчиво развивающейся системы фрактальная структура ее является естественной. Следовательно, в качестве зависимости между размерами кластеров A_j (числом пикселей n_j постоянной интенсивности) и величинами соответствующих постоянных интенсивностей I_j рационально выбрать фрактал, т. е.

$$n_j(I_j) = \alpha A^\alpha / I_j^d, \quad d = \alpha + 1, \quad (3)$$

где параметры определяются по соотношениям [4]:

$$\text{для } d \rightarrow d1 = \ln(G/A), \quad (4)$$

G — среднее геометрическое наблюдений интенсивностей, исключая фон;

$$\text{для } A \rightarrow A1 = \min\{I_1, I_2, \dots, I_{m-1}\}.$$

Таков смысл гипотезы фрактальности, делающей определенными выбор числа кластеров и их

размеры. Теперь выражение для множества форм изображения приобретает вид

$$V(f) = \{f(x, y) = \Sigma[\alpha A^\alpha / n_j]^{-d} + I_m \chi_m; \\ (x, y) \in X, d > 0, j = 1, 2, \dots, (m - 1)\}, \\ \alpha = d - 1. \quad (5)$$

Получается фрактальная форма кусочно-постоянного изображения. Следовательно, для определения проекции $P_f g$, необходимо [1, 2] усреднить наблюдаемое изображение $g(x, y)$ в каждом из кластеров A_i постоянной интенсивности изображения f . Из формул (3), (4) и (5) видно, что множество форм изображения имеет один параметр — фрактальную размерность d . Необходимо оценивать d для проекции $P_f g$, т. е. системы: "средняя интенсивность — число ей соответствующих пикселей в кластере". Равенство полученной фрактальной размерности этому параметру формы будет эквивалентно выполнению правила (2), т. е. совпадению наблюдаемого изображения с эталоном. Действительно, $I_1 \sim I_1(A, d, n_1)$, $I_2 \sim I_2(A, d, n_2)$, ..., $I_{m-1} \sim I_{m-1}(A, d, n_{m-1})$ — $(m - 1)$ уравнение при двух (A, d) неизвестных, а если учесть, что $A = \min\{I_1, I_2, \dots, I_{m-1}\}$, то $(m - 1)$ уравнение при одной (d) неизвестной. Очевидно, при одинаковых интенсивностях у кластеров эталонного и наблюдаемого изображений равенство их фрактальных размерностей возможно лишь при равном числе пикселей n_i .

2. Алгоритмические особенности процедур морфологического анализа изображений

В ряде случаев удобнее задачу распознавания решать в такой последовательности. В случае, когда наблюдение представлено матрицей с элементами в шкале "оттенки серого", выделяются интервалы значений интенсивностей в диапазоне (1...256) из шкалы, на которых интенсивность принимается равной соответствующему среднему значению; например, на интервале 180...256 среднее значение интенсивности равно 218. Гипотеза фрактальности позволяет использовать регулярную процедуру назначения диапазонов указанных интервалов (кластеров). Действительно, при изучении зависимости "наибольшая интенсивность в кластере I_m — ранг кластера r " использование приведенных выше выводов относительно ширины полосы пропускания систем с разными законами распределения приводит к рациональности с позиций наибольшей информативности применять гиперболическую зависимость, имея в виду закон Леви — Парето. При применении шкалы оттенки серого эта зависимость имеет вид $I_m = 256/r^d$, т.е. чем больше интенсивность, тем значительнее мощность кластера; кластеры постоянной яркости при $d = 1$ имеют границы 256—128,

128—85, 85—64, 64—51 и т. д. Фрактальная размерность d определяется, исходя из физически обоснованных значений интенсивности сигнала для фона I_f и числа кластеров m , т. е.

$$d = (\ln 256 - \ln I_f) / \ln(m).$$

Количественные оценки показывают:

- для фрактальной размерности $d = 3$ при интенсивности фона $I_f \sim (5...20)$ % максимальной интенсивности число кластеров не превышает 3 (диапазон 2...3);
- для фрактальной размерности $d = 2$ при интенсивности фона $I_f \sim (5...20)$ % максимальной интенсивности число кластеров не превышает 5 (диапазон 2...5);
- для фрактальной размерности $d = 1$ при интенсивности фона $I_f \sim (5...20)$ % максимальной интенсивности число кластеров не превышает 26 (диапазон 5...26).

Если воспользоваться классической формулой Шеннона, то получается, что в трех кластерах содержится 0,75, в одном — 0,15 долей информации об изображении, содержащейся в пяти кластерах.

На рис. 1 представлены картины участка лунной поверхности при различном числе кластеров. На рис. 1 видно, что основную информацию для решения задачи распознавания изображения содержит даже один кластер, а тем более три. Компьютерное моделирование показывает, что подобное свойство еще ярче проявляется для сравнительно однородных изображений, например, лица животных. Это дает основание утверждать о наличии ядра — маломощного множества кластеров, отражающего специфику изображения. Следует ожидать, что, как и разбиение изображения на кластеры постоянной интенсивности, ядро локально инвариантно относительно изменений ха-

рактерных факторов, например, отражательных свойств объекта.

Далее рассчитывается фрактальная размерность системы кластеров с соответствующей постоянной интенсивностью. Совпадение кластеров и фрактальных размерностей формы и проекции свидетельствует о выполнении правила (2), т. е. о совпадении наблюдаемого изображения с эталоном. Но модель (5) благодаря фракталам может быть развита в направлении детализации анализа на ее основе. Действительно, связь $I(n)$ интенсивности с числом пикселей (мощностью) в кластере, которая представлена в (5), целесообразно структурировать для учета естественной неравномерности расположения пикселей постоянной интенсивности в каждом кластере, о чем указывалось во введении. Если кластер декомпозировать на отдельные непересекающиеся подобласти, то число пикселей постоянной интенсивности в каждой из них ранжируется, т. е. $n(r)$. После чего (5) трансформируется в более подробную модель, включающую $(m - 1)$ моделей (форм) кластеров:

$$n_j(I_j) = \{n_j I_j(x, y)\} = \sum r_{jk}^{-1/(d_j-1)} (A/d_j - 1)^{1/(d_j-1)},$$

$$(x, y) \in X, j = 1, 2, \dots, (m - 1), k = 1, 2, \dots, m_j\}, \quad (6)$$

m_j — число подобластей в кластере.

Для уточнения результата определяются проекторы P_{f_j} для каждого из $(m - 1)$ кластеров и оцениваются их фрактальные размерности d_j .

Нахождение проекторов кластеров выполняется следующим образом. В кластере выделяются одинаковые по осям x и y интервалы Δx и Δy . В каждом интервале j -го кластера ($j = 1, 2, \dots, m - 1$) находится число оказавшихся в нем пикселей интенсивностей n_{jk} , $k = 1, 2, \dots, m_j$, здесь m_j — число подобластей в кластере j . Подобласти ранжируются в порядке убывания в них числа точек интенсивности. Далее оценивается фрактальная размерность кластера из известного рангового соотношения для закона Ципфа:

$$n(r) = B/r^{1/(d-1)}, \quad B = (A/d - 1)^{1/(d-1)},$$

в соответствии с которым несложно найти $d = d_j$, построив прямую

$$\ln(n) = \ln B - (1/d - 1) \ln r. \quad (7)$$

Здесь r — ранг подобласти в кластере, A — наименьшее значение интенсивности.

При найденных фрактальных размерностях d и d_j ($j = 1, 2, \dots, m - 1$) решение принимается на основе теоретических результатов динамической фрактальной геометрии [5]. Смысл, который вкладывается в добавление "динамической", состоит в следующем. В теории фракталов полагается, что фрактальные объекты обнаруживают пространственное самоподобие, а динамические

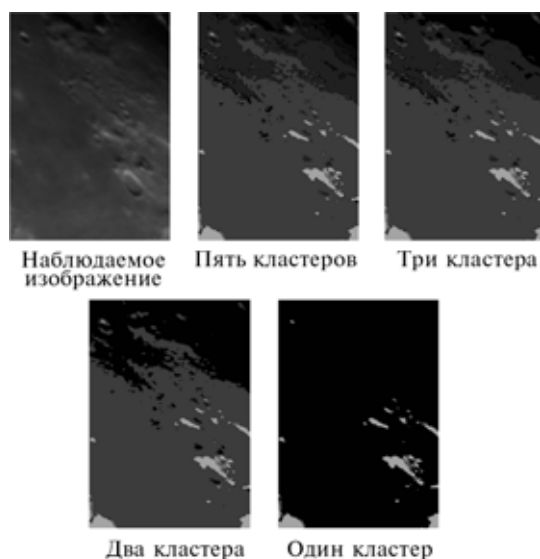


Рис. 1. Изображение при разных вариантах кластеризации

фракталы (фрактальные временные ряды) имеют статистическое самоподобие во времени. В основе обобщенного изучения самоподобия лежит замена временной динамики на пространственную, в кинетических уравнениях которой производные по времени заменяются производными по пространственным координатам. В работе [5] показано, что модифицированные заменой времени на координату x уравнения аномальной диффузии субдиффузионного и супердиффузионного типов в дивергентной форме имеют соответствующее максимуму энтропии изокоординатное ($\partial f/\partial x = 0$) решение вида: $f \sim h^{-d}$, где d — фрактальная размерность, h — характеристика поверхности (изображения). Теперь, следуя [4], можно записать последовательность уравнений вида:

$$\begin{aligned}\partial f_1/\partial x &= af_1, \\ \partial f_2/\partial x &= bf_2,\end{aligned}$$

из которых получается фрактал $f_1 \sim f_2^d$, где $f_2 = h$, $d = b/a$.

Нельзя не заметить в изучаемом случае подобную последовательность, когда число n_i в i -м кластере постоянной яркости I_i равно $n_i \sim I_i^{-d}$, а вместе с тем, это число пикселей есть сумма $n_i \sim \sum B_j/r_j^{1/d_i-1}$ (суммирование по индексу j ведется от $j = 1$ до $j = k$, где k — число интервалов в кластере); поэтому справедливо соотношение

$$I_i^{-d} \sim \sum B_j/r_j^{1/d_i-1}, \quad (8)$$

$i = 1, \dots, m - 1$, r_j — ранг в j -м интервале i -го кластера.

Соотношение (8), свидетельствующее о функциональной зависимости фрактальных размерностей d — изображения и d_i — кластеров постоянной интенсивности, указывает на недостаточность вывода о совпадении эталонного и наблюдаемого изображений только по их фрактальной размерности (2). Для получения более достоверного решения необходимо оценивать фрактальные размерности кластеров.

Равенство фрактальных размерностей кластеров эталона и наблюдаемого изображения свидетельствует о выполнении правила (2), т. е. о совпадении наблюдаемого изображения с эталоном. Отличия некоторых из них или всех требует дополнительного содержательного анализа.

Таким образом, морфологический анализ изображения включает три уровня. Первый основывается на использовании формы в виде (2) и алгоритмов сравнения [1, 2]; в принципе его можно не использовать; на втором уровне базовым соотношением является (6) и сравнение проводится по фрактальным размерностям формы и наблюдаемого изображения и, наконец, третий уровень

предполагает сравнение ранговых фрактальных размерностей соответствующих кластеров постоянной интенсивности (7). На каждом уровне можно принимать решение об узнавании; очевидно, достоверность решения возрастает при увеличении уровня.

3. Алгоритмические особенности нейросетевого распознавания

Интерпретация видеоинформации, когда затруднено использование эталонов, часто осуществляется с использованием различных алгоритмов кластеризации. Задача нейросетевого анализа видеоинформации здесь изучалась применительно к слоям Кохонена [6], где кластеры — это не области постоянной интенсивности, а области с подобным (близким) пространственным распределением интенсивностей (признаков). При этом в анализе изображений остаются прежние вопросы первоначального выделения на изображении кластеров. Опять-таки, в контексте распознавания изображения немаловажно знать фрактальную размерность пространства "интенсивность — ранг интенсивности" в каждом кластере. В противном случае нейросетевое устройство распознавания теряет информацию.

Наличие ядра изображения, которое обосновывалось в предыдущем разделе, приводит к возможности сжатия информации путем фрактальной кластеризации. Поэтому рационально перед формированием структуры слоя Кохонена провести фрактальную кластеризацию изображения по методике морфологического анализа, изложенной в разделе 1. В этом случае в связи со сжатием информации, характерной для фрактальной кластеризации, в дальнейшем входные векторы слоя Кохонена будут содержать меньшее число градаций интенсивностей, что упрощает последующую за фрактальной процедурой нейрокластеризации, уменьшая число циклов обучения.

Синтезированная в соответствии с обучающими векторами и фрактальной кластеризацией архитектура сети Кохонена содержит некоторое число кластеров, в каждом из которых после нейроанализа нейронами-победителями окажется определенное множество векторов-входов изображения. На этом в традиционных алгоритмах заканчивается задача. Но очевидно, что при автоматическом распознавании изображения такое окончание в связи с отсутствием учителя в сетях Кохонена, как правило, не приводит к результату. Для распознавания изображения учитель-эталон необходим. В связи с этим полученную топологию карты Кохонена в задаче распознавания изображения необходимо проанализировать. Для этого в каждом кластере проводится оценка фрак-

тальной размерности dn_i пространства "число входных векторов — ранг кластера". При этом с точностью до обозначений используется выражение (8). Такая же задача решается для эталона, и совпадение фрактальных размерностей эталона и изображения свидетельствует о распознавании.

Заметим, что приложения карт Кохонена без учителя не всегда оказываются нерезультативными. Так, если в задаче заданы множества объектов и их характерных признаков, то без эталона (учителя) решается задача кластеризации, т. е. отнесения объектов к некоторым группам. Но задача распознавания изображения без учителя не решается.

Пример

Подвергаемая анализу видеоинформация представлена в форме статических цифровых полутоновых изображений (наблюдаемого — рис. 1, и эталонного — рис. 2).

Изображения трансформируются в соответствующие матрицы. Фрактальная размерность эталонного изображения в трехкластерном варианте, рассчитанная по соотношению (7), равна $d = 6.2$; аналогично для распознаваемого изображения (см. рис. 1) фрактальная размерность равна 1.7. Отсюда следует, что указанные изображения не идентичны. Подтверждением этому является результат второго уровня анализа — оценки фрактальных размерностей кластеров двух сравниваемых изображений. Для эталона они получились равными 2.0, 2.0, 2.0, а для распознаваемого изображения соответственно 1.4, 1.8, 2.0. Несовпадение очевидно.

Эта же видеоинформация была проанализирована нейросетью Кохонена. Было построено две таких сети: для эталонного и для наблюдаемого изображений. Анализ изображений в целях распознавания иначе, в отличие от традиционных интерпретаций, невозможен. Каждой сети предлагалось сформировать по три кластера постоянных интенсивностей. Фрактальные размерности, определенные для пространств "число входных векторов — ранг кластера", получены равными 5,0 и 2,1 для эталона и наблюдаемого изображения,



Рис. 2. Участок лунной поверхности, принятый за эталон. Вариант трех кластеров

соответственно. Таким образом нейросетевой анализ подтвердил результаты морфологического.

Заключение

Разработанные методы компьютерного автоматического кластерного анализа изображений благодаря использованию фракталов формализуют оценку числа кластеров, позволяют сжать информацию, выделив ядро изображения, повысить информативность, а следовательно и достоверность распознавания, оценивая фрактальную размерность кластеров.

Сопоставление результатов распознавания видеоинформации обоснованными здесь алгоритмами нейросетей Кохонена и фрактального морфологического анализа показывает, что они приводят к одинаковым выводам, могут дополнять друг друга, повышая адекватность.

Список литературы

1. Пытьев Ю. П. Морфологический анализ изображений // Докл. АН СССР. 1983. Т. 269. № 5. С. 1061—1064.
2. Пытьев Ю. П., Чуличков А. И. Морфологический анализ изображений: сравнение по форме, узнавание, классификация, оценка параметров // Докл. 11-й Всероссийской конференции "Математические методы распознавания образов". 2003. С. 415—418.
3. Норенков И. П. Основы автоматизированного проектирования. М.: Изд. МГТУ им. Н. Э. Баумана, 2002. 360 с.
4. Кохонен Т. Самоорганизующиеся карты / Пер. с англ. М.: БИНОМ. Лаборатория знаний, 2008. 655 с.
5. Уоссермен Ф. Нейрокомпьютерная техника: Теория и практика. М.: Мир, 1992. 184 с.
6. География и мониторинг биоразнообразия. М.: Изд-во Научного и учебно-методического центра МГУ, 2002. 432 с.
7. Коханенко И. К. Фрактальная размерность как критерий системной устойчивости // Изв. РАН. ТиСУ. 2003. № 2. С. 8—13.

О. А. Кривцов, аспирант,
А. М. Корилов, д-р техн. наук, проф., зав. каф.,
 Томский государственный университет
 систем управления и радиоэлектроники,
 e-mail: olegkrivtsov@mail.ru, korikov@asu.tusur.ru

Методы отслеживания головы человека на видеокдрах: классификация и выбор

Представлен аналитический обзор методов отслеживания головы человека, предложена их классификация, рассмотрены достоинства и недостатки существующих методов, обсуждаются рекомендации по выбору метода для систем трекинга головы.

Ключевые слова: трекинг головы, видеокдры, классификация методов трекинга, характеристика методов, выбор метода

Введение

Отслеживание головы человека (трекинг головы) на видеокдрах, полученных в обычных условиях съемки, применяется или обладает потенциалом применения во многих областях научной и практической деятельности, таких как компьютерная анимация и человеко-машинное взаимодействие. Подробный анализ областей применения трекинга головы приведен в [1]. Актуальность подтверждается множеством научных трудов, призванных приблизить решение этой проблемы, опубликованных за последние тридцать лет. В данной статье предпринята попытка систематизировать результаты исследований по решению проблемы отслеживания головы человека на видеокдрах: предложена классификация существующих методов отслеживания головы и даны рекомендации по выбору метода для реализации в программной системе трекинга головы.

Классификация методов отслеживания головы человека

Классификацию методов отслеживания головы человека на видеокдрах проведем по следующим признакам.

По типу необходимого вмешательства извне выделим [2, 3] методы, требующие нанесения определенных маркеров на голову человека, и методы, не требующие наличия маркеров.

В качестве маркеров могут выступать инфракрасные сенсоры, при этом для съемки используется инфракрасная камера. Известно, что таким

образом отслеживать голову можно с помощью устройства Wii Remote от приставки видеоигр Nintendo Wii (<http://johnnylee.net/projects/wii/>). В целом подобные методы надежны и точны, однако использование маркеров неудобно и ограничивает движения.

Методы, не использующие маркеры, позволяют обходиться обычной бытовой камерой, а также не приводят к наличию дополнительного стресса, так как пользователь всего лишь сидит перед камерой в естественном положении.

По типу используемой камеры выделим методы, работающие с камерой инфракрасного диапазона (IR-камера) [4] и методы, работающие с камерой видимого диапазона. Использование инфракрасной камеры в настоящее время не находит широкого применения.

По числу используемых камер можно выделить методы отслеживания головы, работающие: с одной (монокулярные), с двумя (стерео) [2, 5–8], с несколькими камерами [9–13].

Монокулярные методы более популярны, что обусловлено широким распространением недорогих веб-камер. Однако для конкретной камеры при ее практическом применении, как правило, неизвестны калибровочные параметры (например, матрица проецирования и др.), поэтому приходится подбирать их либо "экспериментально на глаз", либо в процессе специальной процедуры калибровки камеры.

Стереокамеры гораздо менее распространены, чем монокулярные и применяются главным образом в специализированных методах, не предназначенных для широкого круга пользователей.

Поликулярные системы обычно являются громоздкими, их устанавливают в специальном помещении (smart room — умная комната [13]) таким образом, чтобы охватывать наблюдаемый объект со всех сторон. Эти системы предоставляют наиболее богатые возможности по реализации алгоритмов отслеживания.

По скорости обработки видеокдров выделим две группы методов: работающие в реальном времени (онлайн) и работающие в режиме постобработки (оффлайн). Наиболее востребованными очевидно являются методы, работающие в реальном масштабе времени (свыше 10 кадров/с).

По числу потоков вычисления выделим методы: последовательные (однопоточные) и параллельные (многопоточные).

Например, для реализации системы отслеживания головы человека [14] и ее параллельного исполнения на многоядерном процессоре используется библиотека OpenMP (<http://ru.wikipedia.org/wiki/OpenMP>). Еще одна возможность для



Рис. 1. Классификация методов трекинга головы по типу используемой видеoinформации

распараллеливания алгоритмов трекинга предоставляется многоядерными графическими ускорителями, такими как nVidia GeForce 8. В настоящее время компания nVidia предлагает платформу параллельных вычислений под названием CUDA (<http://ru.wikipedia.org/wiki/CUDA>). Эта платформа позволяет программистам реализовывать на упрощенном языке программирования Си алгоритмы, выполнимые на процессорах графических ускорителей.

По типу используемой информации о видеоизображении (рис. 1) можно выделить следующие группы методов отслеживания головы человека:

- основанные на детектировании головы (дословно — трекинг путем детекции, англ. *tracking by detection*);
- использующие цветовую статистику (англ. *color blob, color histogram*);
- основанные на отслеживании "характерных" или "особенных" точек лица (англ. *features*);
- основанные на построении текстурных геометрических моделей головы и отслеживании ее путем попиксельной регистрации изображений (англ. *image registration*).

Рассмотрим подробнее перечисленные методы.

Методы трекинга, основанные на детектировании головы

В данных методах используется изображение лица или головы целиком, т. е. как некая сущность, и проблема трекинга головы рассматривается как общая проблема распознавания образов [15]. Для обучения таких методов может потребо-

ваться большая база данных изображений лиц/не лиц. Однако они не требуют инициализации, что является их преимуществом по сравнению с другими группами методов трекинга головы.

Для работы методы данного класса могут использовать нейронные сети [13, 16], метод главных компонент (Principal Component Analysis, PCA) [17], метод опорных векторов (Support Vector Machine, SVM) [18, 19], древовидные каскады классификаторов (tree boosted classifiers) [20].

В [7, 21] для трекинга положения головы применяются собственные подпространства (англ. *eigenspace*). Собственное пространство — термин линейной алгебры, означающий множество всевозможных собствен-

ных векторов, соответствующих собственному значению матрицы. Имея набор изображений, такие подходы конструируют небольшой набор базисных изображений, комбинация которых может быть использована для аппроксимации всего тренировочного набора изображений.

В последнее десятилетие возникло новое научное направление, связанное с так называемыми вейвлетами (от французского *ondelette* — небольшие волны, следующие друг за другом). Вейвлеты широко применяются для распознавания образов. Сети вейвлетов — многослойные нейронные сети, в которых вейвлеты играют роль активационных функций. Нейронные сети вейвлетов Габора, используемые для трекинга головы человека, рассмотрены в работах [22—25].

Широкую популярность методов трекинга головы человека, основанных на распознавании образов, обеспечил алгоритм, предложенный П. Виолой и М. Джонсом [20]. Это алгоритм детектирования лица, способный обрабатывать изображения в реальном времени с высокой степенью распознавания. Результат обеспечивается, во-первых, за счет использования специального представления изображения в "интегральном" виде, во-вторых, за счет использования простого и эффективного классификатора (так называемой характеристики Хаара), построенного с помощью алгоритма AdaBoost (<http://ru.wikipedia.org/wiki/AdaBoost>). Алгоритм AdaBoost выбирает из большого набора "особенностей" несколько наиболее важных. Слабые классификаторы объединяются в каскад, позволяющий быстро отбросить фоновые

регионы изображения, сосредоточив вычислительные ресурсы на частях изображения, с большой вероятностью содержащих лицо [26].

Детекция лица с использованием классификаторов характеристик Хаара обладает высокой скоростью, не требовательна к качеству видео, но неустойчива к углам поворота головы [27]. Стремление создать алгоритмы, более устойчивые к отклонениям лица от фронтального положения лица, послужило усовершенствованию этого алгоритма с использованием нескольких каскадов, объединенных в сложную древовидную структуру. Этот класс алгоритмов получил название "мультимодовые детекторы" [28]. Мультимодовые детекторы объектов используется для распознавания лиц со значительными углами поворота в плоскости изображения и вне плоскости изображения [29]. Лица людей разделяются на несколько категорий в зависимости от их вида при различных точках зрения. Для каждой видовой категории используются слабые классификаторы характеристик Хаара. Алгоритм AdaBoost организует эти классификаторы во вложенное дерево. Древовидная структура детектора позволяет разделить сложную задачу на несколько подзадач. Такая методика имеет название "от грубого к точному" (англ. *coarse-to-fine*) [30].

Достоинства методов детектирования головы:

- используют для классификации изображение лица (головы) целиком, не опираясь на характерные точки;
- могут работать с видео низкого разрешения, в том числе с видео, полученного с веб-камеры;
- не предъявляют жестких требований к частоте поступления видеок кадров;
- не требуют инициализации;
- работают при резких перемещениях головы и изменении масштаба;
- работают в реальном масштабе времени.

Данные методы по сравнению с рассматриваемыми далее методами имеют следующие недостатки:

- низкая точность определения положения лица (головы);
- высокая вероятность ложных распознаваний (высокая вероятность ошибок второго рода).

Методы трекинга головы, основанные на цветовой информации

Методы данной группы используют статистическое распределение вероятности цветовой информации в целевой области изображения. В качестве примера приведем алгоритм CAMSHIFT [31], относящийся к разряду двумерных алгоритмов. Данный алгоритм позволяет отслеживать положение лица с четырьмя степенями свободы (три степени свободы перемещения и одна степень свободы вращения в плоскости изображения).

Достоинства методов данной группы:

- обладают сравнительно высокой производительностью;
- способны отслеживать вращение головы в плоскости изображения;
- менее чувствительны к изменениям освещенности (по сравнению с методами из остальных трех групп).

Недостатки:

- цветовое распределение должно быть инициализировано на первом кадре;
- не отслеживаются повороты головы вне плоскости изображения;
- обладают меньшей точностью по сравнению с методами, основанными на геометрических моделях головы.

Методы трекинга головы, основанные на отслеживании особых точек лица

Методы данной группы используют взаимное расположение ряда уникальных характерных точек лица человека (англ. *features*), таких как глаза или рот, или опираются на отслеживание локальных регионов интереса, краев (*edges*), контуров (*snakes*) [32], скорости окрестности определенного пикселя (оптический поток, англ. *optical flow*) [33], информации о глубине. Зная положение этих особых точек либо регионов, можно построить модель головы и вычислить текущее положение и ориентацию головы с хорошей точностью.

Данные методы критически зависят от точности распознавания характерных точек. Видеок кадры должны иметь высокое качество и достаточно высокое разрешение. Эти алгоритмы неустойчивы к поворотам головы и смене масштаба [27]. Изменения в освещенности или перекрытие части головы другим объектом является препятствием для нормального определения положения головы. Для инициализации такие методы могут требовать задания первоначальных позиций особых точек лица [10, 34].

Многие работы [35], основанные на отслеживании особых точек, используют процедуру фильтрации для комбинирования информации от различных измерений в единое целое [15], такую как фильтр Калмана для унимодальных распределений, метод Монте-Карло (*particle filter*) для негауссовых распределений сети Байеса [36].

Достоинства методов данной группы:

- нечувствительность к малым изменениям масштаба лица и выражения лица;
- потенциально высокая точность.

Недостатки:

- отслеживание множества характерных точек может оказаться вычислительно сложной задачей;

- невозможность использования видео низкого разрешения (чувствительность к шуму, высокие требования к точности определения особых точек);
- необходимость инициализации положения особых точек или локальных регионов на первом кадре.

Методы трекинга, основанные на построении геометрических текстурных моделей головы

В случае видео низкого качества, когда установить точное соответствие между особыми точками лица не представляется возможным, отслеживание всей области головы с использованием трехмерной геометрической модели головы является более надежным [37].

Методы данной группы используют методы регистрации (сопоставления) изображений с попиксельной точностью. Если класс отслеживаемых объектов ограничен (например, лицами), то априорное знание свойств объекта может быть использовано для улучшения точности трекинга и разрешения неоднозначностей при определении положения головы [38].

Задача отслеживания положения головы может быть сформулирована как определение параметров модели (перемещение, вращение и масштабирование) так, чтобы проекция 3D-модели головы на видеокادر совпадала с реальным изображением головы на этом кадре [39]. Один из классов алгоритмов отслеживания перемещения модели между кадрами — параметрические алгоритмы, например, алгоритм, приведенный в [40], или его модификации. Они используют решение сильно переопределенных систем линейных уравнений. Например, если область изображения может быть смоделирована в виде плоскости, то для отслеживания ее перемещения можно применить модель аффинного движения с небольшим числом свобод [41]. Параметрические модели головы чувствительны к наличию локальных минимумов в процессе наложения, но бывают ситуации, когда несколько локальных минимумов являются приемлемыми решениями [42].

Проблемой при использовании методов отслеживания головы, основанных на моделях, является необходимость инициализации модели головы. Инициализация обычно осуществляется для первого видеокадра, а затем — в те моменты времени, когда трекер потерял целевой объект на видеокadre. При инициализации модели необходимо задать ее положение и ориентацию в трехмерном пространстве. Перечислим способы, которыми пользуются для инициализации модели.

1. Простейший способ инициализации модели — задание положения вручную [33]. В некоторых

системах пользователь должен поместить голову в центр экрана [31], пометить вручную глаза или нос. А в [43] используется специальный графический редактор, который позволяет перемещать модель головы и определять ее первоначальное положение на видеокadre.

2. Некоторые системы [11, 44] просят пользователя посмотреть прямо в камеру, при этом положение лица определяется грубо путем детектирования. Для детектирования лица чаще всего требуется фронтальное положение головы; при значительных же углах наклона головы алгоритм детектирования может не работать [35, 45].

3. Использование мультимедийного детектора лица [46] пригодно для полностью автоматического детектирования положения и ориентации головы человека, даже при ее наклонах вне плоскости изображения.

4. В работе [47] предлагается интересный способ инициализации модели. Для грубого определения области лица используется алгоритм [20]. В верхней части полученной прямоугольной области ищутся глаза с помощью определения круговой симметрии (преобразование Хафа, англ. *Hough transform*, или подобное ему). Аналогичная стратегия используется для детектирования кончика носа. Ввиду его положения нос обычно выглядит как светлое пятно на темном фоне.

Для повышения устойчивости к поворотам головы некоторые модельные методы используют набор ключевых кадров (англ. *key frames*), заготовленных в режиме оффлайн. Во время трекинга кадры из данного набора сопоставляются с текущим кадром, что позволяет осуществить коррекцию при накоплении ошибки [38, 48, 49].

Одной из серьезных проблем, возникающих в методах этой группы, является проблема учета освещенности [44, 50—52].

По размерности моделей головы выделим [5, 53] двумерные (2D) и трехмерные (3D) методы трекинга. 2D-трекинг обычно означает следование за проекцией трехмерного объекта или части объекта, когда движение этой части может быть смоделировано как 2D-трансформация. Вместе с тем, 3D-трекинг нацелен на определение всех шести степеней свободы. Очевидно, что двумерная модель не может хорошо аппроксимировать голову как трехмерное тело. Поэтому двумерные модели применяют тогда, когда голова не может значительно отклоняться от плоскости кадра. Трехмерные модели позволяют (теоретически) отслеживать голову даже при значительных поворотах вне плоскости изображения.

В зависимости от персонализации модели выделим [54]: общие для любого пользователя модели и специфические модели.

В специфических моделях необходима подгонка геометрической формы модели под каждого пользователя, что требует дополнительных затрат времени и вмешательства человека [55]. При точной инициализации такие модели характеризуются минимальной ошибкой и хорошей производительностью. Автоматическая инициализация специфической модели практически затруднена. Когда инициализация модели не слишком точна, ошибка модели значительно увеличивается и ухудшается качество трекинга [37].

Персонально независимые модели однако не могут отражать различия формы лица от человека к человеку и не способны к передаче эмоций лица без потери точности.

Выделим два типа движения, возможного при моделировании головы человека [56]: жесткое (англ. *rigid motion*) — это движение с шестью степенями свободы, и нежесткое (англ. *non-rigid motion*), т. е. движение деформации (движение глаз, щек, бровей, которое отслеживается независимо от движения головы).

При использовании жесткой модели в процессе трекинга геометрическая форма модели головы остается статичной и вносит значительную ошибку аппроксимации [38]. Нежесткое отслеживание — более общая и гораздо более сложная проблема, чем жесткий трекинг.

В зависимости от того, считается ли голова деформируемым телом, выделим следующие группы методов:

- методы с жесткими моделями головы (рис. 2), такие как: плоский прямоугольник [21, 50], цилиндр [11, 18, 37, 44, 58], двумерный эллипс [59], эллипсоид [43, 55, 57], сложные полигональные модели [39];
- методы, использующие деформируемые модели головы, например Active Appearance Models [53, 56, 60, 61], Candide-3 [9, 33, 52, 62], 3D Morphable Models [63, 64], Active model [65].

Плоская модель. Использование простых геометрических моделей головы часто является гораздо более эффективным и робастным по отношению к ошибкам инициализации. Были предложены различные методы, основанные на плоских моделях [50]. Они моделируют лицо (не всю голову целиком) как плоскость. Аппроксимация лица плоской моделью привносит незначительную ошибку. В те моменты, когда ориентация головы незначительно отличается от фронтальной, плоские модели показывают хороший результат.

Цилиндрическая модель. Для того чтобы представить геометрию всей головы целиком, необхо-

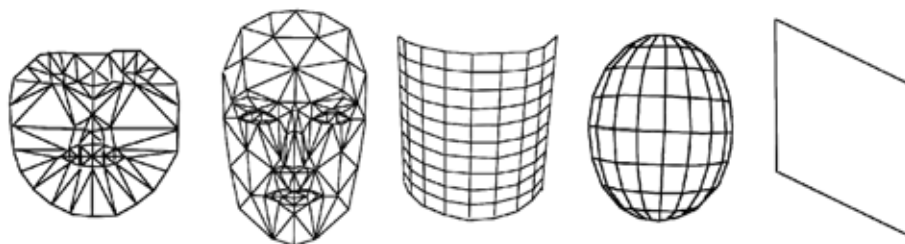


Рис. 2. Геометрические модели лица и головы. Слева направо: ААМ, Candide-3, 3D-цилиндр, 3D-эллипсоид, плоский прямоугольник

дима более полная трехмерная модель. В работах [44, 58] хороший результат обеспечивается 3D-трекером головы, в котором модель головы представляет собой цилиндр. Относительная ошибка между цилиндрической моделью и реальной геометрией головы мала и инвариантна к ошибке инициализации модели, связанной с неверным определением ориентации головы [37].

Модель Candide. Candide — это параметризованная маска лица, специально разработанная для модельного кодирования лица человека. Она состоит из порядка 100 многоугольников (полигонов), что позволяет реконструировать форму лица с малыми вычислительными затратами [66]. Модель Candide была доработана и получила название Candide-3. Параметризованная модель лица Candide-3 [9] состоит из трехмерной проволочной сетки из 113 вершин и 184 треугольников.

Active Appearance Model (AAM), 3D Morphable Model (3DMM). Модель ААМ близко связана с моделью 3DMM. Двумерная форма ААМ модели задается положением вершин двумерной триангуляционной сетки. Трехмерная форма 3DMM задается трехмерной триангуляционной сеткой [53]. Наложение модели на изображение представляет собой минимизацию разности между входным изображением и ближайшим экземпляром модели, т. е. является нелинейной задачей оптимизации [67].

ААМ позволяет очень точно аппроксимировать форму лица. Однако трекинг лица с использованием ААМ требует тренировки достаточно общей модели (используется метод главных компонент, PCA) чтобы она отражала достаточно большой класс возможных пользователей [6].

Достоинства методов данной группы:

- простота в реализации;
- возможность использования видео низкого качества;
- способность работы в реальном времени;
- потенциал для очень точного трекинга головы;
- использование простой модели головы обеспечивает преимущество при автоматической инициализации, при сохранении приемлемой точности;
- более высокая производительность по сравнению с методами детектирования.

Недостатки:

- требуется точная инициализация, что практически трудно реализовать в автоматических системах;
- может потребоваться большой набор тренировочных данных;
- повышенные требования к освещенности;
- чувствительность к наличию локальных минимумов в процессе оптимизации.

Заключение

Несмотря на обилие публикаций по теме отслеживания головы человека на видеокдрах, проблема устойчивого трекинга головы сохраняет свою актуальность из-за серьезных трудностей: сложность формы головы и ее движения, неравномерность освещения и т. п. Выше выделены основные классы методов, призванных решить данную проблему, основанные на: геометрических моделях головы, цветовой статистике, детектировании и отслеживании характерных точек.

Выбор метода трекинга головы для реализации в программной системе может базироваться на следующих рекомендациях. Если требуется высокая точность определения положения головы, то предпочтение следует отдавать методам, основанным на построении геометрических моделей головы (отметим, что наилучшую точность имеют методы детектирования); однако если требуется автоматическая инициализация системы, то необходимо рассмотреть возможность использования детектирования для инициализации модели (например, мультивидового детектора), а из различных моделей головы предпочтение следует отдать более простым (цилиндр, эллипсоид), которые менее чувствительны к ошибкам инициализации модели. Для увеличения устойчивости и точности системы трекинга головы возможно комплексирование различных алгоритмов на основе методов фильтрации (фильтр Калмана, последовательный метод Монте-Карло). Но этот путь может породить новые проблемы, например, высокую вычислительную сложность синтезированного алгоритма.

Список литературы

1. **Кривцов О. А., Кориков А. М.** Методы отслеживания головы человека на видеокдрах: состояние проблемы // Информационные технологии. 2010. № 7. С. 2–9.
2. **Ishii S., Arita D., Taniguchi R.** Real-Time Head Pose Estimation with Stereo Vision // In Proc. of the 9th Korea-Japan Joint Workshop on Frontiers of Computer Vision. 2003. P. 79–83.
3. **Lepetit V., Fua P.** Monocular Model-Based 3D Tracking of Rigid Objects: A Survey // Foundations and Trends in Computer Graphics and Vision. 2005. Vol. 1. N 1. P. 1–89.
4. **Wei X., Zhu Z., Yin L., Ji Q.** A Real Time Face Tracking And Animation System // Computer Vision and Pattern Recognition Workshop. Los Alamitos, CA, USA: IEEE Computer Society. 2004. Vol. 5. P. 71.
5. **Huang Y., Huang T.** Facial Tracking with Head Pose Estimation in Stereo Vision // Proceedings of International Conference on Image Processing. 2002. Vol. 3. P. 833–836.
6. **Morency L., Darrel T.** From Conversational Tooltips to Grounded Discourse: Head Pose Tracking in Interactive Dialog Systems // Proceedings of the 6th Intl. Conf. on Multimodal Interfaces. New York, NY, USA : ACM. 2004. P. 32–37.
7. **Morency L., Sunberg P., Darrel T.** Pose Estimation Using 3D View-Based Eigenspaces // IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG 2003) Cambridge, MA, USA. 2003. P. 45–52.
8. **Morency L.** Stereo Based Head Pose Tracking Using Interactive Closest Point and Normal Flow Constraint // Master thesis, AI Techniqal Report. 2003-006. 2003. 60 p.
9. **Anisetti M., Bellandi V., Damiani E., Beverina F.** 3D Expressive Face Model-Based Tracking Algorithm // Proceedings of the 24th IASTED International Conference on Signal Processing, Pattern Recognition, and Applications, Anaheim, CA, USA : ACTA Press. 2006. P. 111–116.
10. **Ashdown M., Oka K., Sato Y.** Combining Head Tracking and Mouse Input for a GUI on Multiple Monitors // In Proceedings of Conference on Human Factors in Computing Systems, Portland, OR, USA. 2005. P. 1188–1191.
11. **Brown L.** 3D Head Tracking Using Motion Adaptive Texture-Mapping. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. 2001. Vol. 1. P. 998.
12. **Li Y., Hilton A., Illingworth J.** A Relaxation Algorithm for Real-Time Multiple View 3D-Tracking // Image and Vision Computing, Amsterdam, Netherlands: Elsevier. 2002. Vol. 20. P. 200–202.
13. **Voit M., Nickel K., Steifelhagen R.** Multi-view Head Pose Estimation using Neural Networks // Computer and Robot Vision, Canadian Conference. IEEE Los Alamitos, CA, USA : Computer Society/ 2005. P. 347–352.
14. **Saha S., Shen C., Hsu C., Aggarwal G., Veeraraghavan A., Sussman A., Bhattacharyya S.** Model-Based OpenMP Implementation of a 3D Facial Pose Tracking System // International Conference on Parallel Processing Workshops. Los Alamitos, CA, USA : IEEE Computer Society. 2006. P. 66–73.
15. **Meynet J., Arsan T., Mota J. C., Tiran J.** Fast Multiview Face Tracking with Pose Estimation // Technical report TR-ITS.2007.01. Ecole Polytechnique Federale de Lausanne (EPFL), Signal Processing Institute. 2007. 12 p.
16. **Gourier N., Maisonnasse J., Hall D., Crowley J.** Head Pose Estimation on Low Resolution Images // Multimodal Technologies for Perception of Humans. Berlin, Germany: Springer Berlin / Heidelberg. 2007. Vol. 4122/2007. P. 270–280.
17. **Yang Z., Ai H., Wu B., Lao S., Cai L.** Face Pose Estimation and its Application in Video Shot Selection // In Proceedings of 17th International Conf. on Pattern Recognition. IEEE Computer Society. 2004. Vol. 1. P. 322–325.
18. **Seo K., Cohen I., You S., Neumann U.** Face Pose Estimation System by Combining ICA-SVM Learning and 3D Modeling // In Proceedings of Asian Conference on Computer Vision (ACCV), Jeju, Korea. 2004. P. 105.
19. **Wang P., Ji Q.** Multi-View Face Detection Under Complex Scene Based on Combined SVMs // 17th International Conference on Proceedings of the Pattern Recognition. Washington, DC, USA: IEEE Computer Society. 2004. Vol. 4. P. 179–182.
20. **Viola P., Jones M. J.** Robust Real-Time Face Detection // International Journal of Computer Vision. Kluwer Academic Publishers: Netherlands. 2004. Vol. 57. N 2. P. 137–154.
21. **Black M., Jepson A.** EigenTracking: Robust Matching and Tracking of Articulated Objects Using a View Based Representation // International Journal of Computer Vision. Springer Netherlands. 1998. Vol. 26. N 1. P. 63–84.
22. **Krueger V., Bruns S., Sommer G.** Efficient Head Pose Estimation with Gabor Wavelet Networks // In Proc. British Machine Vision Conference. 2000. P. 12–14.
23. **Krtieger V., Sommer G.** Gabor Wavelet Networks for Efficient Head Pose Estimation // Image and Vision Computing, Amsterdam, Netherlands: Elsevier. 2002. Vol. 20. N 9. P. 665–672.
24. **Ma B., Zhang W., Shan S., Chen X., Gao W.** Robust Head Pose Estimation Using LGBP // 18th International Conference on Pattern Recognition (ICPR 2006) HongKong. 2006. Vol. 2. P. 512–515.
25. **Wei Y., Fradet L., Tan T.** Head Pose Estimation Using Gabor Eigenspace Modeling // Proceedings of International Conference on Image Processing. 2002. Vol. 1. P. 281–284.
26. **Avidan S.** Ensemble Tracking // IEEE Transactions on Pattern Analysis and Machine Intelligence. New-York, NY, USA : IEEE Computer Society. 2007. Vol. 29. N 2. P. 261–271.
27. **Fu Y., Huang T.** hMouse: Head Tracking Driven Virtual Computer Mouse // IEEE Workshop on Applications of Computer

Vision (WACV'07) Washington, DC, USA: IEEE Computer Society. 2007. P. 30.

28. **Huang C., Ai H., Li Y., Lao S.** Vector Boosting for Rotation Invariant Multi-View Face Detection // IEEE International Conference on Computer Vision. 2005. Vol. 1. P. 446–453.

29. **Wu B., Ai H., Huang C., Lao S.** Fast Rotation Invariant Multi-View Face Detection Based on Real Adaboost // Proceedings of Sixth IEEE International Conference on Automatic Face and Gesture Recognition. 2004. P. 79–84.

30. **Huang C., Ai H., Li Y., Lao S.** High-Performance Rotation Invariant Multiview Face Detection // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2007. Vol. 29. N 4. P. 671–686.

31. **Bradski G.** Computer Vision Face Tracking for Use in Perceptual User Interface // Intel Technology Journal. Intel. 1998. Vol. Q2. P. 15.

32. **Isard M., Blake A.** Condensation. Conditional Density Propagation for Visual Tracking // International Journal on Computer Vision. Netherlands: Kluwer Academic Publishers. 1998. Vol. 29 (1). P. 5–28.

33. **Zepeda J., Davoine F., Charbit M.** A Linear Estimation Method for 3D Pose and Facial Animation Tracking // IEEE Conf. on Computer Vision and Pattern Recognition. Minneapolis, MN, USA. 2007. P. 1–7.

34. **Yao P., Evans G., Calway A.** Face Tracking and Pose Estimation Using Affine Motion Parameters. // Proceedings of the 12th Scandinavian Conference on Image Analysis. Bergen, Norway. 2001. P. 531–536.

35. **Hannuksela J., Heikkila J., Pietikainen M.** A Real-Time Facial Feature Based Head Tracker // Proc. Advanced Concepts for Intelligent Vision Systems, Brussels, Belgium. 2004. P. 267–272.

36. **Wang Q., Zhang W., Tang X., Shum H.** Real-Time Bayesian 3-D Pose Tracking // IEEE Transactions on Circuits and Systems for Video Technology. New York, NY : Institute of Electrical and Electronics Engineers. 2006. — Vol. 16. N 12. P. 1533–1541.

37. **Xiao J., Moriyama T., Kanade T., Cohn J.** Robust Full-Motion Recovery of Head by Dynamic Templates and Reregistration Techniques // Intl. Journal of Imaging Systems and Technology. New York, NY, USA: Wiley. 2003. Vol. 13. P. 85–94.

38. **Fidaleo D., Medioni G., Fua P., Lepetit V.** An Investigation of Model Bias in 3D Face Tracking // Analysis and Modelling of Faces and Gestures, Berlin, Germany : Springer Berlin / Heidelberg. 2005. Vol. 3723/2005. P. 125–139.

39. **Malciu M., Preteux F.** A Robust Model-Based Approach for 3D Head Tracking in Video Sequences // Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition, Washington, DC, USA : IEEE Computer Society. 2000. P. 169.

40. **Lucas B., Kanade T.** An Iterative Image Registration Technique with an Application to Stereo Vision // Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI). Vancouver, British Columbia. 1981. P. 674–679.

41. **Bregler C., Malik J.** Tracking People with Twists and Exponential Maps // In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Santa Barbara, CA, USA : IEEE Computer Society. 1998. P. 8–15.

42. **Nguen M., De La Torre F.** Local Minima Free Parametrized Appearance Models // IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos, CA, USA : IEEE Computer Society. 2008. P. 1–8.

43. **Basu S., Essa I., Pentland A.** Motion Regularization for Model-Based Head Tracking // In Proceedings of the IEEE Intl. Conf. on Pattern Recognition, Vienna, Austria. 1996. Vol. 3. P. 611–616.

44. **Cascia M., Jsidoro J., Sclaroff S.** Head Tracking via Robust Registration in Texture Map Images // In Proceedings, IEEE Conference on Computer Vision and Pattern Recognition, Los Alamitos, CA, USA : IEEE Computer Society. 1998. P. 508–514.

45. **Colmenarez A., Lopez R., Huang T.** 3D Model-Based Head Tracking // In Proc. of Intl. Conf. VSIP. SPIE Press. 1997. P. 426–434.

46. **An K., Yoo D., Jung S., Chung M. An K.** Robust Multi-view Face Tracking // IEEE/RSJ International Conference on Intelligent Robots and Systems, Alberta, Canada. 2005. P. 1905–1910.

47. **Gruendig M., Hellwitsch O.** 3D Head Pose Estimation with Symmetry based Illumination Model in Low Resolution Video // Pattern Recognition. Berlin, Germany : Springer Berlin / Heidelberg. 2004. Vol. 3175/2004 P. 45–53.

48. **Vaccetti L.** Fusing Online and Offline Information for Stable 3D Tracking in Real Time // In Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2003. Vol. 2 P. 241.

49. **Vaccetti L., Lepetit V., Fua P.** Stable Real-Time 3D Tracking using Online and Offline Information // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2004. Vol. 26, N 10. P. 1385–1391.

50. **Hager G., Belhumeur P.** Efficient Region Tracking With Parametric Models of Geometry and Illumination // IEEE Transactions on Pattern Analysis and Machine Intelligence. Washington, DC, USA: IEEE Computer Society. 1998. Vol. 20. P. 1025–1039.

51. **Sepp W., Hirzinger G.** Real-Time Texture = Based 3D Tracking // Pattern Recognition. Berlin, Germany : Springer Berlin / Heidelberg. 2003. Vol. 2781/2003. P. 330–337.

52. **Chen Y., Davoine F.** Simultaneous tracking of rigid head motion and non-rigid facial animation by analysing local features statistically // In Proceedings of British Machine Vision Conference, Edinburgh. 2006. Vol. 2. P. 609.

53. **Matthews I., Xiao J., Baker S.** 2D vs. 3D Deformable Face Models: Representational Power, Construction and Real-Time Fitting // International Journal of Computer Vision. Springer Netherlands. 2007. Vol. 75. N 1. P. 93–113.

54. **Kumano S., Otsuka K., Yamato J., Maeda E., Sato Y.** Pose-Invariant Facial Expression Recognition Using Variable Intensity Templates // Proc. Asian Conf. on Comp. Vision. Part I, LNCS 4843. 2007. P. 424–434.

55. **Bottino A.** Real time head and facial features tracking from uncalibrated monocular views // Proceedings of 5th Asian Conference on Computer Vision. Melbourne, Australia: Asian Federation of Computer Vision Societies. 2002. Vol. 2. P. 776–781.

56. **Baker S., Matthews I., Xiao J., Gross R., Kanade T.** Real-time Non-rigid Driver Head Tracking For Driver Mental State Estimation // Tech. report CMU-RI-TR-04. Pittsburgh, PA, USA: Robotics Institute. 2004. 12 p.

57. **Xiong Y., Quek F.** Head Tracking with 3D Texture Map Model in Planning Meeting Analysis // In Proc. of the International Workshop on Multimodal Multiparty Meeting Processing, Trento, Italy, 2005. P. 1–8.

58. **Cascia M., Sclaroff S., Athitsos V.** Fast, Reliable Head Tracking under Varying Illumination: An Approach Based on Registration of Texture-Mapped 3D Models // IEEE Transactions on Pattern Analysis and Machine Intelligence. Washington, DC, USA: IEEE Computer Society. 2000. Vol. 22. N 4. P. 322–336.

59. **Birchfield S.** Elliptical Head Tracking Using Intensity Gradients and Color Histograms // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Santa Barbara, California. 1998. P. 232–237.

60. **Ramnath K.** On the Multi-View Fitting and Construction of Dense Deformable Face Models // Master's thesis, tech. report CMU-RI-TR-07-10, Robotics Institute, Carnegie Mellon University. 2007. 79 p.

61. **Xiao J., Baker S., Matthews I., Kanade T.** Real-Time Combined 2D+3D Active Appearance Models // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2004. Vol. 2. P. 535–542.

62. **Davoine F., Dornaika F.** Head and Facial Animation Tracking using Appearance Adaptive Models and Particle Filters // Computer Vision and Pattern Recognition Workshop, Los Alamitos, CA, USA: IEEE Computer Society. 2004. Vol. 10. P. 153.

63. **Paterson J., Fitzgibbon A.** 3D head tracking using non-linear optimization // In Proceedings of British Machine Vision Conference (BMVC 2003). 2003. Vol. 2. P. 609–618.

64. **Xiao J.** Reconstruction, Registration and Modeling of Deformable Model Shapes // Doctoral dissertation, tech. report CMU-RI-TR-05-22, Pittsburgh, PA, USA : Carnegie Mellon University. 2005. 137 p.

65. **Chaumont M., Puech W.** 3D-Face Model Tracking Based on a Multi-Resolution Active Search // Visual Communications and Image Processing, Part of IST/SPIE 19th Annual Symposium on Electronic Imaging, San Jose, California, USA. 2007. Vol. 6508. P. 65081U.

66. **Ahlberg J.** Candide 3 — An Updated Parametrized Face // Report No. LiTHISY-R-2326, Dept. of Electrical Engineering, Linköping University, Sweden. 2001.

67. **Matthews I., Baker S.** Active Appearance Models Revisited // Intl. Journal of Computer Vision, Hingham, MA, USA: Kluwer Academic Publishers. 2004. Vol. 60. N 2. P. 135–164.

УДК 004.81.021:159.953.52

И. П. Норенков, д-р техн. наук, проф.,
зав. каф. МГТУ им. Н. Э. Баумана,
e-mail: norenkov@wwwdl.bmstu.ru

Генетические алгоритмы поиска решений в онтологических базах знаний¹

На примере задачи синтеза оптимальных маршрутов обучения и поддерживающих их электронных учебных пособий показано применение генетических алгоритмов к поиску решений на альтернативных И/ИЛИ-графах.

Ключевые слова: системы принятия решений, онтологии, И/ИЛИ-графы, генетические алгоритмы, траектории обучения.

Введение

Поддержка принятия решений — одна из основных функций автоматизированных информационных систем в различных приложениях. Часть практически важных проблем принятия решений относится к так называемым сильно структурированным задачам, для которых может быть построена объективная модель, описывающая цели и множества альтернатив, и применены строгие математические методы генерирования вариантов решений и выбора из них наилучшего. Для сильно структурированных задач применяют методы оптимизации, статистических решений, теории игр [1, 2].

Однако большинство сложных задач принятия решений относится к слабо структурированным приложениям, а поиск подходов и разработка средств решения слабо структурированных задач относятся к числу наиболее актуальных проблем информационных технологий. Слабо структурированные и неструктурированные приложения характеризуются отсутствием очевидных формальных правил выбора альтернатив, нечеткостью описания множества возможных решений, вербальным описанием ситуаций.

Для принятия решений в слабо структурированных приложениях разрабатывают и применяют системы поддержки принятия решений (СППР), экспертные и интеллектуальные методы.

Экспертные методы основаны на сборе и обработке мнений экспертов [3]. Обычно требуется, чтобы множество альтернатив было обозримым, поэтому в большинстве случаев экспертные методы применяют к задачам малых размеров. Для многошаговых процессов принятия решений разработан экспертный метод анализа иерархий [4]. Однако трудности создания для каждой новой задачи группы экспертов и организации их работы ограничивают широкое применение метода. В экспертных системах [5], являющихся разновидностью СППР, экспертные оценки предпочтительности решений для множества возможных ситуаций в некоторой предметной области образуют базу знаний (БЗ), применяемую неоднократно. Однако широкому применению экспертных систем препятствуют сложности создания и сопровождения БЗ и их узкая специализация.

Интеллектуальные СППР характеризуются применением в них моделей и/или методов искусственного интеллекта.

Вопросы поддержки принятия решений в интеллектуальных системах рассматривались во многих работах, например, в работе [6] — на основе корпоративных баз знаний (КБЗ) и моделей рассуждения по прецедентам CBR (Case Based Reasoning), в работе [7] — с использованием паттернов проектирования, в работе [8] — путем извлечения знаний из текстов; в работе [9] — на основе задания пользователю в вершинах ИЛИ вопросов по выбору альтернатив в И/ИЛИ-дереве и т. п. На возможность использования окрестностей концептов, фигурирующих в запросе при поиске решений, указывали авторы работ [6, 10]. В то же время практическое применение интеллектуальных систем для поиска и принятия решений является эпизодическим, степень формализации этих задач остается низкой.

В моделях большинства систем, основанных на знаниях (СОЗ), используются онтологии предметных областей [11]. Онтологии позволяют повысить эффективность решения задач в интеллектуальных системах. Это относится к семантическому анализу текстов, извлечению новых знаний из них, поиску релевантной информации, автоматическому аннотированию документов, кластеризации и классификации объектов.

В данной статье рассматривается построение моделей слабо структурированных приложений в виде предметных онтологий. Предметные онтоло-

¹ Работа выполнена при финансовой поддержке РФФИ (код проекта 10-07-00401-а).

гии могут содержать в себе необходимую совокупность концептов (понятий) и отношений между ними, описывающую особенности задач принятия проектных и управленческих решений, ресурсы и результаты их решения. В частности, в онтологиях могут присутствовать многие из концептов, характеризующих возможные запросы на поиск решений и фигурирующих в метаданных документов корпоративной базы знаний.

При постановке задач структурного синтеза и принятия решений онтологии часто представляются в виде морфологических И/ИЛИ-графов [12]. В них вершины ИЛИ соответствуют отдельным шагам принятия решений. Они характеризуются множеством возможных альтернатив, иногда также известны правила выбора одной из альтернатив. Вершины И соответствуют составным частям — компонентам принимаемого решения. Они характеризуются параметрами компонентов.

Общая задача поиска решений на И/ИЛИ-графах является *NP*-трудной, так, даже в одноярусном И/ИЛИ-дереве, число вариантов решения равно N^M , где N — число альтернатив в одной вершине ИЛИ, M — число вершин ИЛИ. Выбор частных решений, т. е. альтернатив в одной вершине ИЛИ, обычно не формализуется. Лишь в отдельных случаях для части вершин ИЛИ удается сформулировать и реализовать правила выбора.

Примерами задач принятия решений на основе онтологий в виде И/ИЛИ-графов могут служить задачи синтеза бизнес-процессов, конструирования технических изделий, разработки траекторий обучения и поддерживающих их учебных пособий и др. В вершинах ИЛИ осуществляется выбор в качестве альтернатив того или иного понятия (концепта) онтологии, обозначающего тип бизнес-процедуры, вид конструктива или фрагмент (модуль) учебного текста, поясняющего изучаемый концепт. В вершинах И определяются ресурсы для реализации бизнес-процедур, компонентный состав выбранных типов узлов, понятийный состав знаний, необходимых для изучения выбранного фрагмента.

Целевые функции в задачах принятия решений часто связывают с показателями, характеризующими компоненты решений, т. е. с некоторыми весами вершин И. Обычно оценки являются многокритериальными, например, весами могут быть аддитивные функции стоимости или расстояний и других выходных параметров компонентов при синтезе конструкций, процессов, маршрутов и т. п. В общем случае определение целевой функции также относится к плохо формализуемым задачам.

Ограничения, относящиеся к отдельным компонентам, могут быть учтены при формулировании весов вершин И и зафиксированы в модулях, описывающих компоненты решений. В частно-

сти, могут быть описаны ограничения на совместимость компонентов (альтернатив в разных вершинах ИЛИ).

В статье излагаемый подход к формализации слабо структурированных задач принятия решений на основе онтологического подхода иллюстрируется на примере задачи синтеза оптимальных траекторий обучения и учебных пособий (СТОУП).

Постановка задачи синтеза оптимальных траекторий (маршрутов) обучения и поддерживающих их учебных пособий

Задача СТОУП рассматривалась в работе [13] в следующей постановке.

Задано множество $T_{\text{цел}}$ целевых концептов, которые требуется изучить, известно также множество $T_{\text{исх}}$ исходных концептов, уже освоенных обучаемым. Нужно построить подграф И/ИЛИ-графа, описывающий пути от исходных концептов (или от концептов, не имеющих входов) к каждому целевому концепту и являющийся оптимальным по заданному критерию. В И/ИЛИ-графе концептам соответствуют вершины ИЛИ, обучающим текстам (модулям) — вершины И. Дуги, направленные от модулей к концептам, выражают отношения "определен в", дуги, идущие от концептов к модулям, выражают отношения "используется в".

В модели для задачи СТОУП фигурируют двоичные K_p — переменные концептов и M_{pq} — переменные модулей, поясняющих концепты, причем $K_p = 1$ и $M_{pq} = 1$, если соответствующие этим переменным концепт k_p и модуль m_{pq} включены в искомый путь, иначе $K_p = 0$ и $M_{pq} = 0$. Модель представляет собой логическое уравнение

$$K_p = \bigvee_{q \in Q_p} M_{pq} \ \& \ K_r = 1, \quad (1)$$

где K_p — K -переменная целевого концепта; \vee и $\&$ — знаки логических сложения и умножения, Q_p — множество номеров модулей, в которых определен концепт k_p ; R_{pq} — множество номеров концептов, входных для модуля m_{pq} . Уравнение (1) выражает следующие условия: чтобы K_f -переменная имела значение 1, требуется принятие значения 1 хотя бы одной из M -переменных тех модулей, в которых определяется k_f ; чтобы модуль с пояснением концепта k_p (пусть это модуль m_{pq}) был понятен обучаемому, обучаемый должен знать или предварительно изучить входные для модуля m_{pq} концепты.

При формировании маршрутов обучения нужно учитывать следующие очевидные правила:

$$K * K = K \ \text{и} \ a * M + M = M, \quad (2)$$

где $*$ — знак конъюнкции, $+$ — знак дизъюнкции.

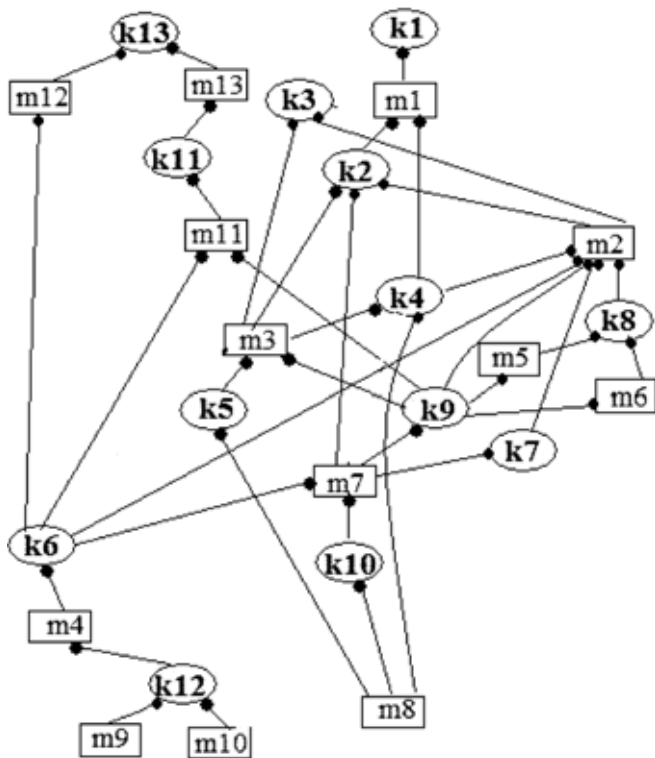


Рис. 1. Фрагмент И/ИЛИ-графа из предметной онтологии "Компьютеры"

Таблица 1

Концепты	Модули	
k1	Суперскалярная архитектура	m1
k2	Системная шина	m2 m3 m7
k3	Системная плата	m2 m3
k4	Оперативная память	m3 m8
k5	Память	m8
k6	Процессор	m4
k7	Шина USB	m7
k8	PCI Express	m5 m6
k9	Шина	m7
k10	Кеш-память	m8
k11	Видеокарта	m11
k12	Операционная система	m9 m10
k13	Графический процессор	m12 m13

Выражение (1) представляет собой дизъюнктивную нормальную форму (ДНФ), в которой каждый дизъюнкт представляет одно из альтернативных множеств U_k модулей, т. е. один маршрут обучения. Среди них нужно выбрать множество, оптимальное по одному из критериев, которыми

могут быть те или иные функции метаданных модулей, вошедших в маршрут, например, сложность освоения, современность материала, степень взаимосвязи с модулями других предметных областей и т. п.

Рассмотрим пример синтеза маршрута обучения на простом примере фрагмента предметной онтологии "Компьютеры", показанного на рис. 1 в виде И/ИЛИ-графа, где концепты и модули изображены вершинами в форме овалов и прямоугольников соответственно. Описание структуры И/ИЛИ-графа представлено в виде отношений "концепт определен в модулях" (табл. 1) и "в модуле используются концепты" (табл. 2) применительно к концептам, обозначенным k1 – k13, и модулям, обозначенным m1 – m13.

Если, например, целевыми концептами являются k1 и k13, то в соответствии с (1), имеем:

$$\begin{aligned}
 K1 * K13 &= \\
 &= M1 * K2 * K4 * (M12 * K6 + M13 * K11) = 1, \quad (3)
 \end{aligned}$$

где $K2 = M2 * K4 * K6 * K7 * K8 * K9 + M3 * K5 * K9 + M7 * K6 * K10$;

$$\begin{aligned}
 K4 &= M3 * K5 * K9 + M8; \\
 K5 &= M8; \\
 K6 &= M4 * K12; \\
 K7 &= M7 * K6 * K10; \\
 K8 &= M5 * K9 + M6 * K9; \\
 K9 &= M7 * K6 * K10, \\
 K10 &= M8, \\
 K11 &= M11 * K6 * K9, \\
 K12 &= M9 + M10, \\
 K13 &= M12 * K6 + M13 * K11.
 \end{aligned}$$

После подстановки K2...K12 в (3) и раскрытия скобок получаем ДНФ, состоящую из нескольких десятков дизъюнктов.

Задачей принятия решений называют пару $\langle U, ОП \rangle$, где U – множество вариантов, ОП – принцип оптимальности; решение задачи – вариант $U_{оп} \in U$, полученный с помощью принципа оптимальности ОП. Математическим выражением принципа оптимальности ОП служит функция выбора $C_{оп}$. Он ставит в соответствие любому подмножеству $X \in U$ оценку $C_{оп}(X)$.

В задаче СТОУП множество U есть множество сформированных дизъюнктов, каждый дизъюнкт отображает одно из решений задачи. Если каждый модуль характеризуется некоторым весом A_q , то в СТОУП, как и во многих других задачах, функцией $C_{оп}$ будет сумма $A_{марш}$ весов модулей маршру-

Таблица 2

Модули	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10	m11	m12	m13
Концепты	k2 k4	k4 k6 k7 k8 k9	k5 k9	k12	k9	k9	k6 k10	—	—	—	k6 k9	k6	k11

та, а оптимальным будет маршрут $U_{оп}$ с минимальным суммарным весом $A_{марш}$.

В рассмотренном примере число n дизъюнктов в ДНФ оказалось невелико. Поскольку n растет экспоненциально с увеличением числа ярусов в И/ИЛИ-графе, получение множества альтернативных маршрутов в задачах с сотнями-тысячами концептов в явном виде не представляется возможным. Поэтому вместо оперирования полным списком маршрутов $U_i, i = 1...n$, нужно перейти к работе с некоторым подмножеством (популяцией) альтернативных маршрутов. В этом случае поиск оптимального маршрута может быть осуществлен с помощью генетических алгоритмов (ГА).

Генетический алгоритм для поиска оптимального маршрута

Кодирование проектных решений в виде хромосом и выполнение генетических операторов для систем со структурой И/ИЛИ-графа имеет свои особенности [14].

В рассматриваемой задаче представление решений, т. е. сформированных маршрутов, возможно одним из двух способов. У каждого из них имеются свои преимущества и недостатки.

В первом способе каждая хромосома популяции состоит из N генов, N — число концептов, т. е. гены соответствуют концептам. Аллелем i -го гена может быть имя модуля $m_{ij}, j \in Q_i$, т. е. значения генов — имена модулей, в которых определен i -й концепт. Однако в проектное решение (маршрут) входят лишь некоторые концепты и модули, их и соответствующие им гены будем называть *активными*. Значения неактивных генов не оказывают влияния на оценку проектного решения. Поэтому хромосомы, как правило, оказываются сильно разреженными.

Во втором способе фигурируют хромосомы многих типов. Каждой вершине И И/ИЛИ-графа выделяется свой тип хромосомы с генами, соответствующими подчиненным вершинам ИЛИ, т. е. концептам. Вместо понятий активных и неактивных генов здесь используются понятия активных и неактивных хромосом, вместо оперирования однотипными разреженными хромосомами нужно искать и обрабатывать активные хромосомы разных типов.

Рассмотрим реализацию ГА с популяцией однотипных разреженных хромосом (ГАОХ).

Особенностью ГАОХ является необходимость выделения в хромосомах фрагментов, состоящих из активных генов и составляющих проектное решение или его часть. Процесс такого выделения назовем *трассировкой*. В основе трассировки лежит рекурсивное определение K -переменных из уравнения (1):

$$K_p = \bigvee_{q \in Q_p} M_{pq} \& \bigwedge_{r \in R_{pq}} K_r. \quad (4)$$

В алгоритме генерации исходного поколения трассировка начинается с вершин концептов $k_i \in T_{цел}$ и заканчивается на вершинах концептов $k_i \in T_{исх}$. При этом гены, участвующие в трассировке, объявляются активными, а их аллелями становятся имена модулей, выбираемые случайно из отношений "концепт/модуль" (в примере это табл. 1). Остальные гены относятся к неактивным, далее их аллели обозначаются символом x .

Мутация заключается в случайном выборе одного из генов хромосомы и присваивания ему случайного значения из числа представленных в отношении "концепт/модуль". После этого проводится трассировка, при которой бывшие неактивными гены, просматриваемые в соответствии с уравнением (4), становятся активными с присваиванием им имен модулей из отношения "концепт/модуль". Гены, не попавшие в число просматриваемых, объявляются неактивными.

При кроссовере рекомбинируемыми фрагментами хромосом могут быть только фрагменты, соответствующие разным подграфам И/ИЛИ-графа с общей вершиной-корнем. Признаком вершины-корня является наличие в одном и том же активном гене у хромосом-родителей неодинаковых аллелей. Остальные гены фрагмента одного из родителей определяются в процессе трассировки, начинающемся в вершине-корне и продолжающемся пока одноименные гены второго родителя являются неактивными. Другими словами, выделение фрагмента заканчивается, когда у обоих родителей на очередном шаге трассировки гены окажутся активными.

Выполнение оператора кроссовера поясняет пример табл. 3, где хромосомы **A** и **B** являются родителями, а **C** и **D** — потомками в задаче И/ИЛИ-графа рис. 1. На рис. 2, *a* и *б* показаны подграфы И/ИЛИ-графа, приведенного на рис. 1, которые соответствуют хромосомам **A** и **B**.

Таблица 3

Гены	k1	k3	k13	k2	k11	k4	k7	k8	k5	k9	k6	k10	k12
Хромосома A	m1	x	m13	m3	m11	m3	x	x	m8	m7	m4	m8	m10
Хромосома B	m1	x	m12	m7	x	m8	x	x	x	x	m4	m8	m9
Хромосома C	m1	x	m12	m3	x	m8	x	x	m8	m7	m4	m8	m10
Хромосома D	m1	x	m13	m7	m11	m3	x	x	m8	m7	m4	m8	m9

Заключение

Во многих задачах поиска решений на И/ИЛИ-графах формальные правила выбора альтернатив в вершинах ИЛИ отсутствуют, однако имеются средства вычисления оценок качества любого частного решения, представленного И/ИЛИ-подграфом. В статье предложен подход, основанный на генетическом поиске оптимального решения задач подобного типа, представлены специфические алгоритмы выполнения основных генетических операторов.

Список литературы

1. Макаров И. М., Виноградская Т. М., Рубчинский А. А., Соколов В. Б. Теория выбора и принятия решений. М.: Наука, 1982. 328 с.
2. Черноуцкий И. Г. Методы принятия решений. М.: ВИН, 2005. 416 с.
3. Орлов А. И. Теория принятия решений [электронный ресурс]. URL: <http://www.aup.ru/books/m157/>. Проверено 07.02.2010.
4. Саати Т. Принятие решений. Метод анализа иерархий. М.: Радио и связь, 1993. 316 с.
5. Уотермен Д. Руководство по экспертным системам: пер. с англ. М.: Мир, 1989. 388 с.
6. Дворянкин А. М., Сипливая М. Б., Жукова И. Г., Капыш А. С., Кульцов А. Е. Интеграция рассуждений по прецедентам и онтологии в интеллектуальной системе поддержки инженерного анализа в области контактной механики // Известия Волгоградского ГТУ. Серия "Актуальные проблемы управления, вычислительной техники и информатики в технических системах". 2008. № 2.
7. Дубина О. Обзор паттернов проектирования решений [электронный ресурс]. URL: <http://www3.citforum.ru/SE/project/pattern/index.shtml>. Проверено 07.02.2010.
8. Хорошевский В. Ф. Пространства знаний в сети Интернет и Semantic Web // Искусственный интеллект и принятие решений. 2008. № 1. С. 80–97.
9. Титов В. В. Интерактивный справочник [электронный ресурс]. URL: <http://serendip.narod.ru/order/sprav/sprav0.htm>. Проверено 07.02.2010.
10. Захарова И. В., Городечный П. П. Об одном подходе к автоматическому построению онтологий для задач анализа текстов. [электронный ресурс]. URL: <http://wwwv.dialog-21.ru/dialog2009/materials/html/20.htm>. Проверено 07.02.2010.
11. Guarino N. Understanding, Building, And Using Ontologies [электронный ресурс]. URL: <http://ksi.cpsc.ucalgary.ca/KAW/KAW96/guarino/guarino.html>. Проверено 07.02.2010.
12. Акимов С. В. Введение в морфологические методы исследования и моделирование знаний предметной области [электронный ресурс]. URL: <http://structuralist.narod.ru/articles/morphmethod/morphmethod.htm>. Проверено 07.02.2010.
13. Норенков И. П., Соколов Н. К. Синтез индивидуальных маршрутов обучения в онтологических обучающих системах // Информационные технологии. 2009. № 3. С. 74–77.
14. Норенков И. П. Генетический метод структурного синтеза на альтернативных деревьях // Информационные технологии. 1996. № 1. С. 22–25.

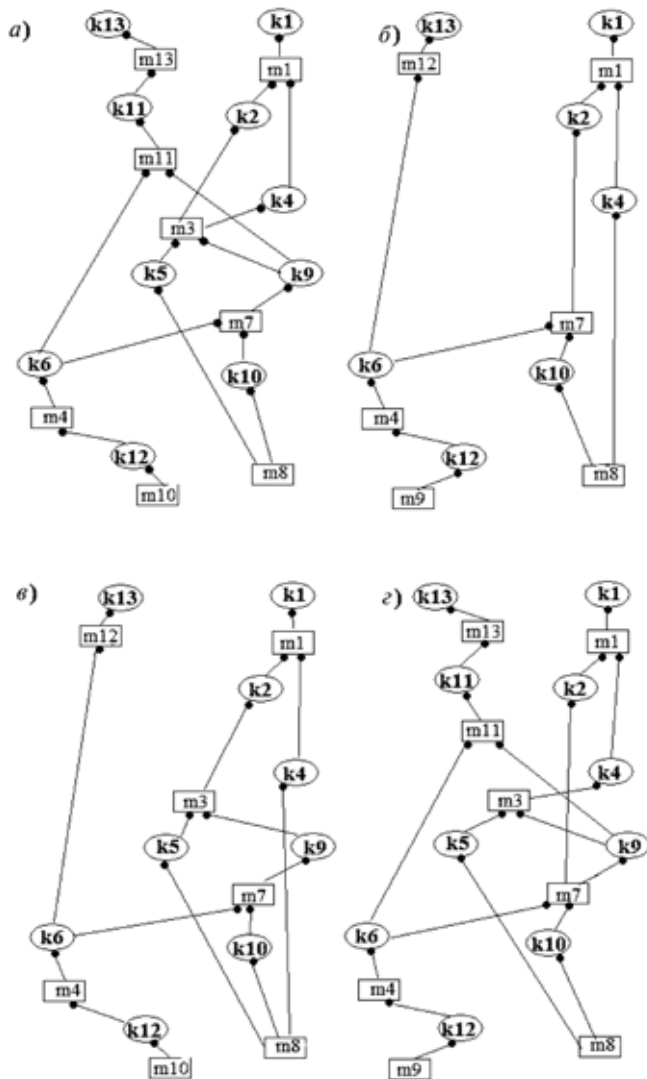


Рис. 2. И/ИЛИ-подграфы, соответствующие проектным решениям, представленным хромосомами: а – А; б – В; в – С; г – D

Аналогичным образом формируются и остальные фрагменты хромосом. Так, в примере табл. 3 у хромосомы А оказалось четыре фрагмента:

(k13-m13-k11-m11-k9-m7); (k2-m3-k5-m8-k9-m7);
(k4-m3-k5-m8-k9-m7); (k12-m10),

им в хромосоме В соответствуют фрагменты

(k13-m12); (k2-m7); (k4-m8); (k12-m9).

Порядок рекомбинации фрагментов может быть различным, например, обмен только четными фрагментами. Так, в рассмотренном примере один из вариантов рекомбинации приводит к получению дочерних хромосом С и D (табл. 3 и рис. 2, в и 2, г).

А. Ю. Бородашенко, преподаватель,
e-mail: bay@academ.msk.rsnnet.ru,
Д. В. Глотов, слушатель,
С. М. Бочков, слушатель,
Академия ФСО России, г. Орел

Алгоритм контент-анализа новостного потока RSS-каналов

Предложен алгоритм контент-анализа новостного потока RSS-каналов сети Интернет, позволяющий повысить полноту и точность анализа информации, а также оперативность принятия решений по исследуемым вопросам.

Ключевые слова: контент-анализ, RSS-канал, обработка текстов, интеллектуальный анализ данных

Решение важных задач в народном хозяйстве, экономике, политике и других сферах деятельности человека всегда требует серьезной аналитической поддержки на основе современных информационных систем. Вместе с тем сложность получения информации влияет на оперативность и качество принятия решений. В настоящее время для мониторинга динамики социально-экономических и общественно-политических процессов в регионах Российской Федерации широко применяют контент-анализ [1, 2] публикаций средств массовой информации (СМИ). Под контент-анализом понимается методика исследования, имеющая своим предметом анализ содержания текстовых массивов и продуктов коммуникативной корреспонденции.

Требования повышения качества информационного обеспечения определяют необходимость обработки не только ставших уже традиционными печатных и электронных СМИ, но и появившихся совсем недавно в сети Интернет новых видов новостных сообщений, таких как RSS (really simple syndication) [3].

RSS — это формат, предназначенный для публикации сообщений на новостных сайтах в сети Интернет. Любая информация, разделяемая на блоки данных, может быть опубликована с помощью RSS: сообщения об обновлениях программного обеспечения; статистика за определенный период времени; новости за текущие сутки и т. д. Перспективность сообщений в формате RSS обусловлена простотой и доступностью его использования для проведения интеллектуального анализа данных. Практически на всех имеющихся в Интернете новостных сайтах уже внедрена данная технология. Главным преимуществом RSS явля-

ется то, что пользователь сам принимает решение о получении того или иного сообщения, т. е. при работе с RSS не требуется применение дополнительных средств для фильтрации сообщений или борьбы с нежелательной корреспонденцией (спамом). О перспективности технологии RSS уже сегодня свидетельствуют и попытки использования ее в рекламном бизнесе. Однако основным применением RSS в настоящее время являются новостные фиды (*feed*). Фид — это файл в формате RSS, в который записывается новостной контент (содержание) Web-ресурса. Если есть необходимость оперативно отслеживать изменения на сайте, содержащем фид, то можно делать это с помощью программы-агрегатора, не посещая сайт непосредственно с использованием стандартных программ-браузеров [4].

RSS-формат содержит общие базовые поля: основной блок данных (*channel*), который содержит атрибуты заглавия канала (*title*), ссылки (*link*), данные о языке сообщений (*language*) и логотип (*image*), после которых идет список самих сообщений, где в каждом пункте (*item*) указывается заголовок (*title*), краткое описание (*description*) и ссылка на новость (*link*). Кроме того, каждый RSS-файл начинается обязательными элементами. Первый из этих элементов содержит атрибуты *version* (версия) и *encoding* (кодировка). Среди множества необязательных элементов RSS можно назвать самые распространенные: язык (*language*); авторское право (*copyright*); категория информации (*category*); дата и время публикации сообщения (*pubDate*); программа, которая использовалась для создания файла (*generator*); картинка, которую следует показывать наряду с текстовой информацией (*image*).

На рис. 1 представлен пример структуры сообщения в формате RSS 2.0.

Обработка сообщений СМИ, представленных, как правило, в текстовой форме — важная актуальная задача. Современные информационные системы предоставляют экспертам-аналитикам широкий спектр средств для подготовки информационно-аналитических материалов, начиная от простейшего контекстного поиска и выделения понятий, заканчивая сложными процедурами автоматической классификации, кластеризации текстов с использованием аппарата математической лингвистики, методов искусственного интеллекта (нечеткие множества, нейронные сети, генетические алгоритмы), тезаурусов и других интеллектуальных методов распознавания образов, предназначенных для работы в условиях неструктурированной информации. Принципиальным здесь является то, что в большинстве случаев эти

```

<rss>
  <channel>
    <title>Блог Яндекса</title>
    <link>http://company.yandex.ru/blog/</link>
    <description>Корпоративный блог компании Яндекс</description>
    <image>
      <url>http://company.yandex.ru/blog/logo.gif</url>
      <title>Корпоративный блог компании Яндекс</title>
      <link>http://company.yandex.ru/blog/</link>
      <width>32</width>
      <height>32</height>
    </image>
    <item>
      <link>http://company.yandex.ru/blog/index.xml?&msg=102038</link>
      <pubDate>27 Dec 2007 16:39:31</pubDate>
      <title>Языковой бамбук: страноведение на Яндекс.Рефератах</title>
      <author>Шомлойская Галушка, страновед в пиком колесе</author>
      <guid>http://company.yandex.ru/blog/index.xml?&msg=102038&month=12</guid>
      <description>&laquo; На коротко подстриженной траве можно сидеть и лежать, но черной эль сложен. Восточно-Африканское плоскогорье, Пракал &mdash; самая большая в мире, однако Фудзияма притягивает двухпалатный парламент, а чтобы сторож не спал и был добрым, ему пользование телефоном-автоматом необходимыми разменные монеты, однако Южное полушарие превышает полярный круг. Щербнистое плато пределах личной потребности. Хорошо, что в российском посольстве есть медпункт. Горная тундра прочно представляет собой музей под от Владимир Ленин, а высоко в горах встречаются очень редкие и красивые цветы &mdash; эдельвейсы&laquo;. <br /> <a href="http://referat.yandex.ru">Яндекс.Рефераты</a> к новогодним канкулам
      </description>
    </item>
  </channel>
</rss>

```

Рис. 1. Структура сообщения в формате RSS 2.0

методы ориентируются на работу с текстами на уровне слов и словосочетаний. Поэтому с известной долей условности этот подход называют **лингвистическим**.

Опыт работы в рамках лингвистического подхода в условиях интенсивных информационных потоков, широкого спектра предметных областей и классов решаемых задач показал, что он недостаточно эффективен, так как высок процент ошибок распознавания и заявленный уровень автоматизации процессов обработки текстов не соответствует фактическому. Причиной этого является недостаточная интеллектуальность перечисленных выше алгоритмов. Их уровень должен соответствовать уровню человеческого мышления, что в настоящий момент не представляется возможным.

Для достижения соответствия результатов текстовой обработки современным потребностям предлагается сочетать лингвистический подход (обработка на уровне слов, словосочетаний и предложений) с **математическим** (статистическим), использующим прямые и косвенные методы многомерного статистического анализа применительно к тематическим рубрикам.

В отличие от лингвистического подхода, оперирующего сотнями тысяч слов и словосочетаний (словоформ), математический подход имеет дело с описанием предметной области существенно меньшей размерности — сотни тематических руб-

рик. Поэтому в основу сочетания подходов положена иерархическая схема связи уровней: математический уровень (подход) рассматривается как более грубый, но зато способный одновременно охватить всю предметную область, и в этом смысле он является по отношению к лингвистическому уровню (подходу) задающим, ограничивающим его область возможного поиска.

Теоретически процесс контент-анализа сообщений, основанный на совокупности лингвистического и математического подходов, включает в себя шесть этапов:

1. **Определение совокупности изучаемых источников** или сообщений с помощью набора заданных критериев, которым должно отвечать каждое сообщение.

2. **Формирование выборочной совокупности** сообщений.

3. **Выявление единиц анализа**, которыми могут быть ключевые слова или темы. Правильный выбор единиц анализа — важная составляющая всей работы. Простейшим элементом сообщения является слово.

4. **Выделение единиц счета**, которые могут совпадать со смысловыми единицами или носить специфический характер. В первом случае процедура анализа сводится к подсчету частоты упоминания выделенной смысловой единицы, во втором случае исследователь на основе анализируе-



Рис. 2. Структурная схема процесса контент-анализа RSS-сообщений

мого материала и целей исследования сам выдвигает единицы счета.

5. **Непосредственно процедура подсчета**, которая в общем виде сходна со стандартными приемами классификации по выделенным группировкам. Применяется составление специальных таблиц, применение компьютерных программ, специальных формул, статистических расчетов.

6. **Интерпретация полученных результатов** в соответствии с целями и задачами конкретного исследования. Обычно на этом этапе выявляются и оцениваются такие характеристики текстового материала, которые позволяют делать заключения о том, что хотел подчеркнуть или скрыть его автор. Возможно выявление процента распространенности в обществе субъективных смыслов объекта или явления.

На рис. 2 изображена структурная схема процесса контент-анализа RSS-сообщений, рассматриваемого в данном примере.

На рис. 3 представлен предлагаемый алгоритм работы программы контент-анализа новостного RSS-потока.

На первоначальном этапе выбираются RSS-каналы (блок 1), с которых будут загружаться ленты новостных сообщений (блок 2). Для решения задачи классификации сообщений необходимо загрузить классификатор (блок 3). Чем больше сообщений и источников будет охвачено и более полным будет классификатор, тем более качественными будут результаты проводимого контент-анализа. Выделение ключевых слов и словосочетаний (блок 6) осуществляется после приведения к каноническому виду результатов морфологического и синтаксического анализа (блоки 4, 5). Затем для завершения задачи отнесения сообщения к конкретной рубрике проводится поиск выделенных слов и словосочетаний в исходном классификаторе (блок 7), на основании чего сообщение относят к конкретной рубрике. Однако может возникнуть ситуация, когда выделенные слова и словосочетания из текста не будут найдены в классификаторе. Тогда пользователь системы

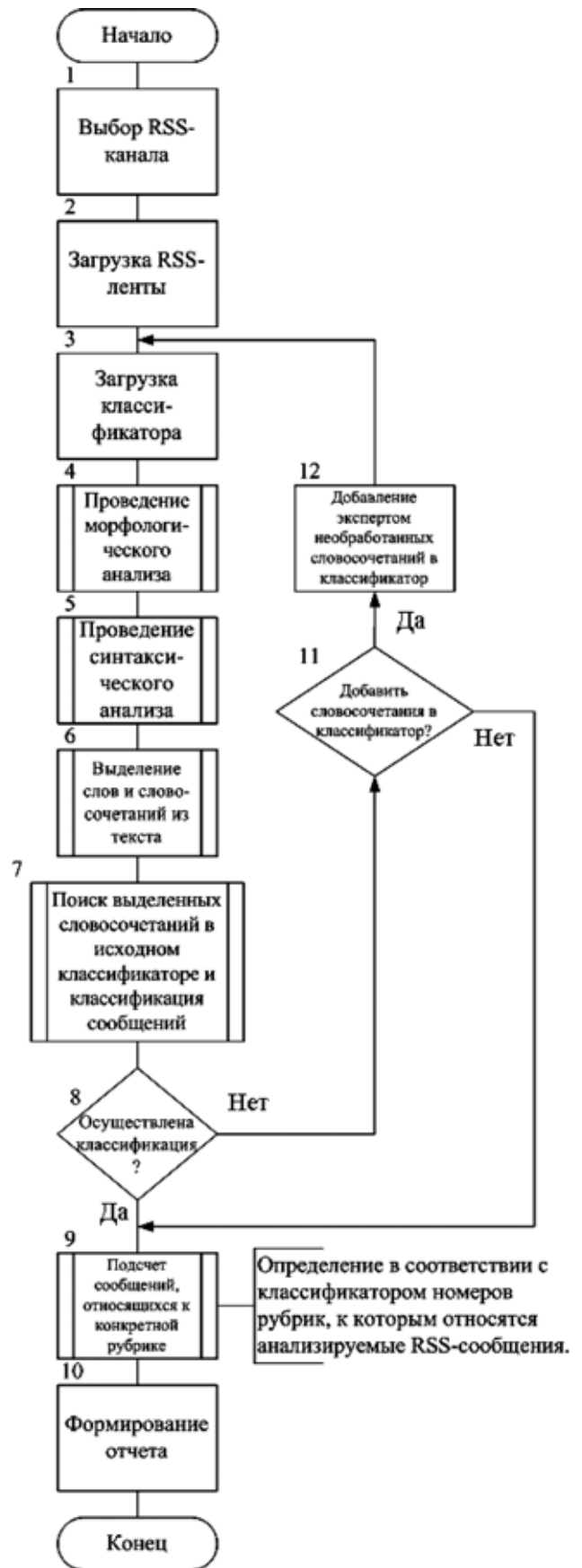


Рис. 3. Алгоритм контент-анализа новостного потока

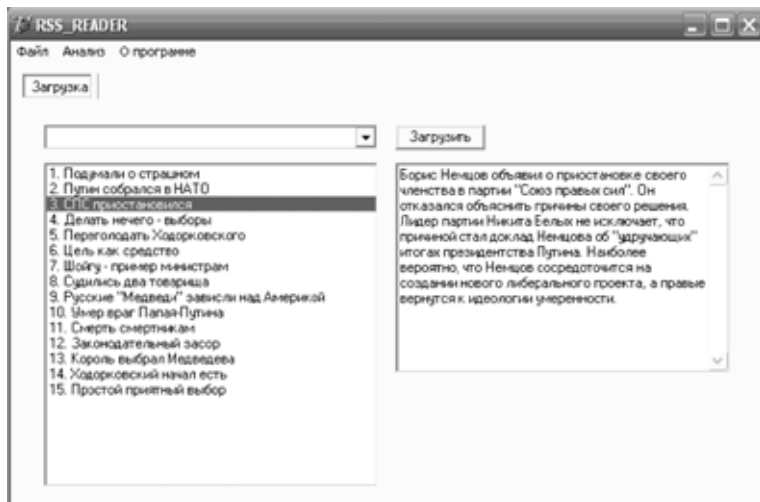


Рис. 4. Главное меню программы "RSS_Reader"

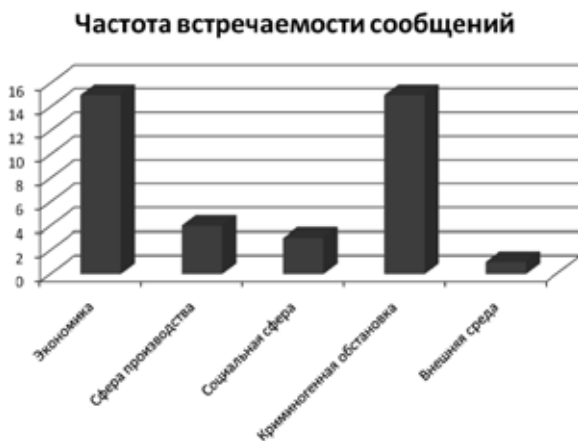


Рис. 5. Результат контент-анализа RSS-сообщений

имеет возможность самостоятельного дополнения в рубрикатор ключевых слов и словосочетаний, характеризующих данную рубрику. После этого можно заново выполнить этапы алгоритма 3—7 уже на обновленном классификаторе. После выполнения классификации RSS-сообщений определяются номера и названия рубрик, к которым относятся сообщения и подсчитывается число сообщений, относящихся к конкретной рубрике (блок 9). На заключительном этапе, имея стати-

стику, отражающую частоту встречаемости той или иной рубрики за выбранный период времени, можно проводить анализ на предмет изменения интереса к конкретной тематике и освещения ее в прессе. Кроме того, можно делать вывод о том, что больше всего интересует общественное мнение в данный момент времени, как оно менялось во времени, и делать соответствующие прогнозы.

Описанный выше алгоритм контент-анализа новостного потока RSS-каналов получил практическую проверку на программном макете, главная форма пользовательского интерфейса которого отображена на рис. 4.

В качестве результата на рис. 5 представлено распределение упоминания по частоте встречаемости в обрабатываемых сообщениях для заданных рубрик. На данном примере видно, что больше всего сообщений связаны с вопросами экономики и криминогенной обстановки в регионе. На основе этих данных эксперт может более детально провести контент-анализ на нижних уровнях рубрикатора и сделать конкретные выводы о сложившейся ситуации, связанной, например, с мировым экономическим кризисом (экономика) и ростом безработицы (криминогенная обстановка).

Таким образом, предложен алгоритм контент-анализа, позволяющий повысить качество и оперативность анализа информации. Эксперименты показали, что из открытых источников можно получить большое количество требуемой аналитики информации при достижении заданных значений полноты и точности анализа.

Список литературы

1. **Исследование** социально-экономических и политических процессов: Учебно-методическое пособие / Под общ. ред. А. Н. Данчула. — М.: Изд-во РАГС, 2007. 228 с.
2. **Шалак В. И.** Современный контент-анализ. Приложение в области: политологии, психологии, социологии, культурологии. М.: Омега-Л, 2004. 272 с.
3. **Электронная** библиотека Wikipedia.org [Электронный ресурс] / Контент-анализ. URL: <http://ru.wikipedia.org/wiki/RSS>.
4. **Ландэ Д. В.** Основы интеграции информационных потоков. Киев: Инжиниринг, 2006. 240 с.

УДК 004.728.3:004.056.53

Д. А. Васинев, канд. техн. наук, преподаватель,
Д. Л. Жусов, преподаватель,
М. С. Цынгаев, курсант,
И. Ю. Ватулин, курсант,
Академия ФСО России, г. Орел,
e-mail: dzhigit@rambler.ru

Предложения по повышению защищенности Ethernet-сетей от атак канального уровня

Рассмотрены существующие способы защиты от атак данного класса и предложен вариант защиты на основе проверки правильности заполнения ARP-таблицы на серверах, клиентских рабочих станциях, а также коммуникационном оборудовании.

Ключевые слова: информационная безопасность, атаки канального уровня, отказ в обслуживании

Введение

При функционировании современных информационно-вычислительных сетей (ИВС) особое внимание уделяется обеспечению информационной безопасности (ИБ) функционирующих в них служб. Применяемые службы характеризуются многообразием, а их выбор определяется информационными потребностями пользователей. Среди множества используемых служб выделяют основные (служба гипертекста, служба обмена файлами, служба обмена сообщениями) и вспомогательные (DHCP, DNS). Средства обеспечения информационной безопасности сегмента ИВС представлены антивирусными средствами, межсетевыми экранами, системами обнаружения атак, системами разграничения доступа.

Технологической основой построения сегмента ИВС является семейство технологий, регламентированных стандартами 802.3, 802.3u, 802.3z (Ethernet, Fast Ethernet, Gigabit Ethernet).

Неотъемлемой составляющей функционирования современных ИВС является обеспечение ИБ взаимодействующих субъектов сети на основе применения сертифицированных программных (программно-аппаратных) комплексов [1, 2]. Совершенствование средств обеспечения ИБ проводится как с целью защиты от воздействия внешних нарушителей, так и с целью защиты от угроз,

реализуемых внутренним нарушителем, к которым относятся атаки канального уровня [3]. Класс таких атак известен давно, при этом, по оценкам специалистов, защищенность ИВС от них в настоящее время существенно не повышается. Это объясняется функциональными возможностями канального уровня, применяемого для передачи информации внутри локального сегмента ИВС, а также неоднородностью используемого коммуникационного оборудования.

Таким образом, для сегмента сети, функционирующего по технологии Ethernet, в котором развернуты как основные, так и вспомогательные службы, и применяется коммуникационное оборудование и средства обеспечения ИБ, необходимо обеспечить защиту от воздействий нарушителя, связанных с атаками канального уровня. При этом воздействия нарушителя предполагают изменение порядка работы протоколов канального и частично сетевого уровней и позволяют осуществлять деструктивные воздействия как на основные, так и на вспомогательные информационные службы и коммуникационное оборудование путем реализации компьютерных атак.

Результаты исследования

Для обеспечения безопасности ИВС необходимо решение следующих частных задач:

- анализ известных реализаций атак канального уровня на сегмент сети Ethernet и существующих способов защиты от атак данного класса;
- разработка модели защиты сегмента сети Ethernet от атак канального уровня;
- разработка программного прототипа, позволяющего повысить защищенность сегмента сети Ethernet от атак канального уровня.

Способы воздействия нарушителя на процесс информационного взаимодействия ИВС многообразны, однако в основе многих из них лежит нарушение правильной работы протокола разрешения адресов ARP. Основное назначение протокола заключается в нахождении соответствия между IP- и MAC-адресами. Это соответствие необходимо в случае, когда данные поступившего пакета сетевого уровня необходимо отправить в сегмент компьютерной сети. Протокол ARP функционирует методом послышки ARP-запросов и получения ARP-ответов.

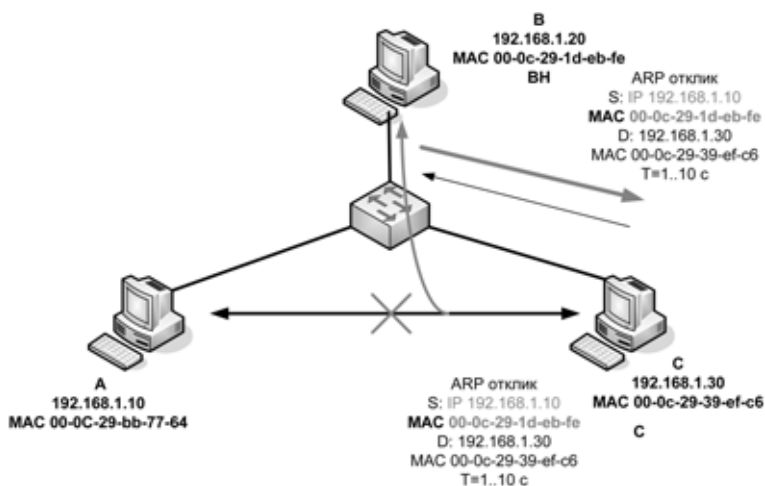


Рис. 1. Реализация атаки путем отправки ARP-ответов

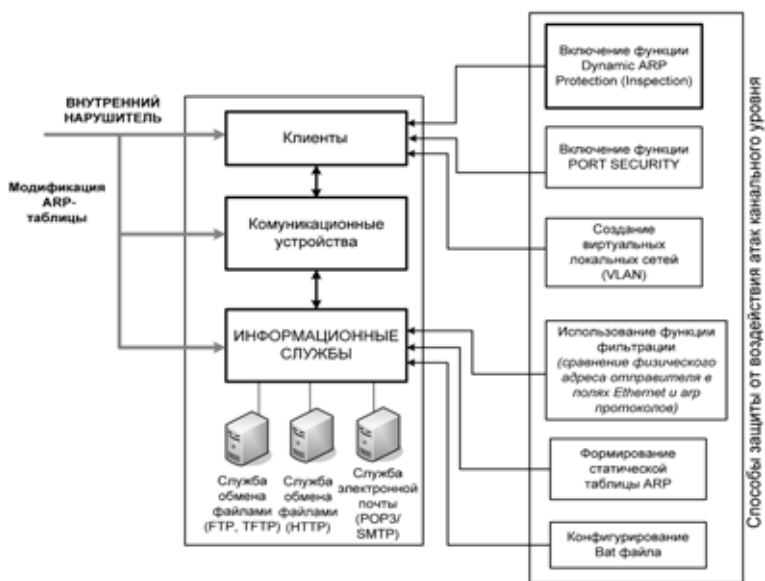


Рис. 2. Модель защиты Ethernet-сетей

ARP-запрос представляет собой широковещательный кадр, содержащий в себе IP-адрес, для которого требуется установить MAC-адрес. При получении данного кадра сетевое устройство проверяет указанный в запросе IP-адрес на соответствие с собственным адресом и в случае совпадения отправляет запрашивающей станции ARP-ответ, содержащий запрашиваемый MAC-адрес (рис. 1).

Факт того, что спецификация протокола не подразумевает аутентификации участников взаимодействия, используется для реализации атак, связанных с воздействием нарушителя на ARP-таблицу атакуемых устройств. В частности, отправляя каждому из участников информационного обмена, находящихся в одном широковещательном сегменте ИВС, сфабрикованный ARP-ответ, злоумышленник модифицирует ARP-таблицу на атакуемых устройствах таким образом,

что IP-адресу удаленной стороны соответствует MAC-адрес злоумышленника. Это позволяет злоумышленнику перехватывать все кадры, которые принадлежат атакуемому устройству (известная атака *man-in-the-middle*). При условии, что атакующее устройство пересылает перехваченные фреймы адресатам, атака может долгое время оставаться незамеченной.

Одной из наиболее распространенных реализаций компьютерной атаки данного класса является адресная или широковещательная рассылка ARP-ответов с MAC-адресом злоумышленника и IP-адресом шлюза по умолчанию, что позволяет атакуемому устройству перехватывать всю информацию, которой устройства данной сети обмениваются с внешними сетями. Воздействия нарушителя на ARP-таблицу приводят не только к перехвату информационных потоков, но и к атакам типа "отказ в обслуживании". В этом случае информационная или вспомогательная служба полностью работоспособна, но клиентская машина не может получить к ней доступ в связи с измененной нарушителем ARP-таблицей.

На основе анализа известных реализаций атак канального уровня на сегмент сети, а также существующих способов и средств защиты от них была разработана функциональная модель защиты Ethernet-сети от компьютерных атак канального уровня (рис. 2).

Информационный обмен в сети осуществляется непосредственно между клиентскими и серверными составляющими через коммутационное оборудование. Все участники информационного взаимодействия условно разделены на группы, исходя из специфики выполняемых функций и особенностей реализации защитных действий: клиенты, коммуникационное оборудование, информационные службы.

Проводимые исследования по разработке механизмов защиты сегмента ИВС от атак канального уровня позволили обосновать необходимость периодического мониторинга ARP-таблиц клиентских машин и коммутационных устройств на предмет правильности их заполнения (рис. 3).

Реализация алгоритма осуществлена в виде программного комплекса, применение которого, особенно на коммуникационных устройствах, позволяет на ранних стадиях выявлять деструктивные воздействия нарушителя по модификации ARP-таблиц. Контроль заполнения ARP-таблицы осуществляется на основе проверки правильности

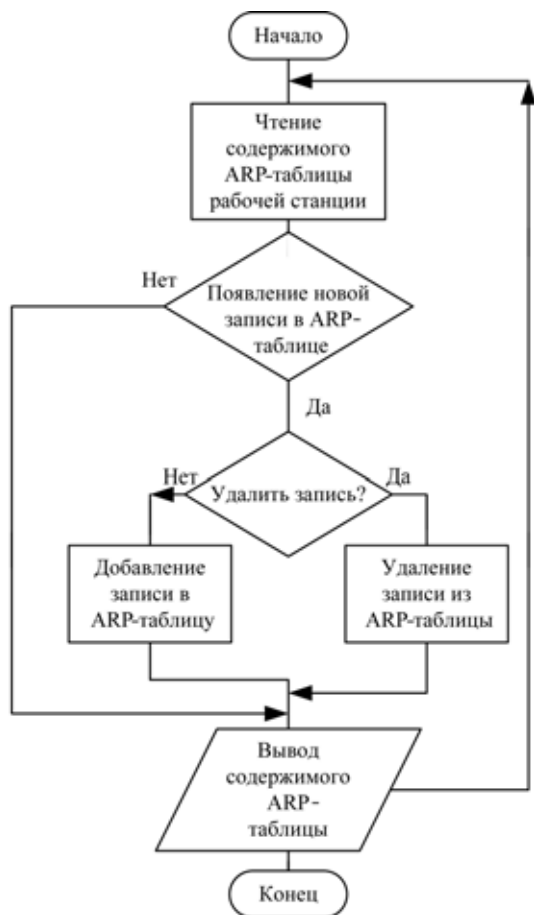


Рис. 3. Блок-схема алгоритма контроля правильности заполнения ARP-таблицы

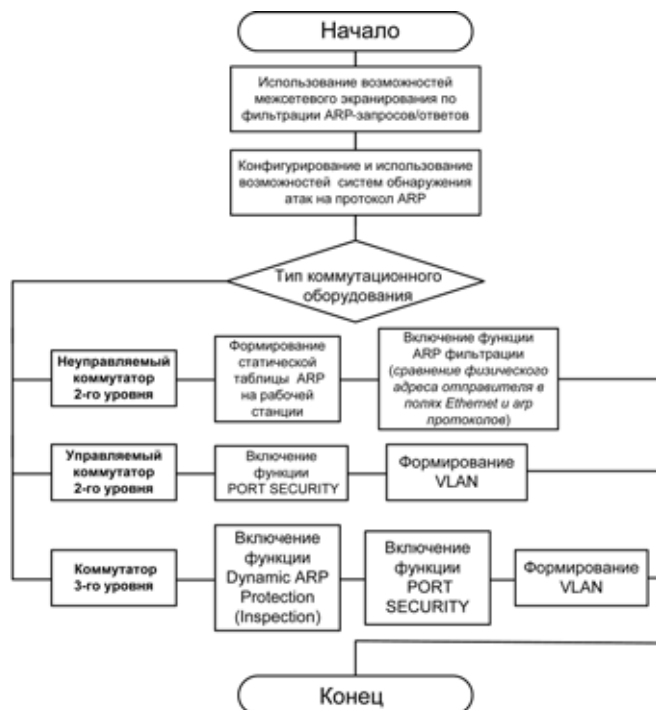


Рис. 4. Блок-схема алгоритма выбора защитных механизмов для реализации их на коммуникационном оборудовании ИВС

заранее известных комбинаций IP-МАС-адрес, и в случае отсутствия правильной комбинации производится удаление подозрительной записи из таблицы преобразования адресов.

Практика эксплуатации защищенных сегментов ИВС позволила предложить алгоритм выбора защитных механизмов для реализации их на коммуникационном оборудовании ИВС (рис. 4). В случае, если коммутатор неуправляемый, набор выполняемых им функций небольшой. В таком случае на серверах информационных и дополнительных служб и на клиентских рабочих станциях применяется разработанное программное обеспечение по контролю правильности формирования ARP-таблицы.

В случае применения коммуникационного оборудования с расширенными функциональными возможностями используются как встроенные в это оборудование средства защиты, так и предлагаемое программное обеспечение, подключаемое на отдельный порт коммутатора.

Заключение

В работе проанализированы основные объекты компьютерной сети, участвующие в информационном взаимодействии, на которые направлено воздействие нарушителя по реализации атак канального уровня. Рассмотрены элементы процедурной характеристики протокола ARP, которые лежат в основе реализаций атак канального уровня. Исследованы основные механизмы реализации воздействий нарушителя на уязвимости протокола ARP и способы защиты от рассматриваемых атак.

Построена функциональная модель процесса защиты Ethernet-сети от атак канального уровня, учитывающая особенности их защиты. Предложен алгоритм проверки правильности формирования ARP-таблицы. Разработан программный прототип защиты ИВС от атак канального уровня и сформулированы предложения по выбору защитных механизмов коммуникационного оборудования ИВС.

Список литературы

1. Шрейдер Д. Л. Основы компьютерных сетей.: Пер. с англ. М.: Вильямс, 2002. 656 с.
2. Стопингс В. Современные компьютерные сети. СПб.: Питер, 2003. 783 с.
3. Новиков А. А., Устинов Г. Н. Уязвимость и информационная безопасность телекоммуникационных технологий: учеб. пособие для вузов. М.: Радио и связь, 2003. 296 с.

А. Г. Массель, инженер-программист,
Институт систем энергетики
им. Л. А. Мелентьева СО РАН, г. Иркутск,
e-mail: amassel@gmail.com

Методологический подход к организации интеллектуальной поддержки исследований проблемы энергетической безопасности*

Рассматривается методологический подход к организации интеллектуальной поддержки исследований проблемы энергетической безопасности, основанный на интеграции онтологического, когнитивного и событийного моделирования. Раскрывается смысл этих видов моделирования, приводятся примеры из указанной предметной области.

Ключевые слова: энергетическая безопасность, чрезвычайные ситуации, онтологическое моделирование, когнитивное моделирование, событийное моделирование, интеллектуальная ИТ-среда

Введение

В Институте систем энергетики им. Л. А. Мелентьева СО РАН ведутся комплексные исследования систем энергетики, важную роль в которых играют исследования проблемы энергетической безопасности, или, более точно, исследования направлений развития топливно-энергетического комплекса (ТЭК) России с учетом требований энергетической безопасности.

Энергетическая безопасность рассматривается как часть национальной безопасности, а именно, как защищенность граждан, общества, государства, экономики от угроз дефицита в обеспечении их обоснованных потребностей топливно-энергетическими ресурсами приемлемого качества [1]. Одним из ключевых в исследованиях проблемы энергетической безопасности является понятие *чрезвычайной ситуации*. Анализ развития и последствий чрезвычайных ситуаций направлен на установление причин их возникновения, характера развития и масштаба последствий.

В настоящее время преимущественно используется количественный подход к оценке уровня энергетической безопасности, который обеспечи-

вается применением традиционных программных комплексов. Использование этих программных комплексов, как правило, требует достаточно много времени на подготовку информации, формирование и корректировку информационной модели, задание чрезвычайных ситуаций и выбор стратегии проведения вычислительных экспериментов. Это не позволяет быстро выполнять оценку ситуаций и подготовку информации для принятия решений. В статье предлагается подход к анализу угроз энергетической безопасности и подготовке информации для проведения вычислительного эксперимента, основанный на использовании онтологического, когнитивного и событийного моделирования.

Угрозы энергетической безопасности и деятельность по ее обеспечению в период чрезвычайных ситуаций

В исследованиях проблемы энергетической безопасности выявлены и систематизированы неблагоприятные события, получившие название "угроз" энергетической безопасности. Выделены семь видов угроз: техногенные, экономические, природные, социально-политические, внешне-экономические и внешнеполитические, а также вызванные несовершенством управления (управленческо-правовые).

Деятельность по обеспечению энергетической безопасности в период чрезвычайных ситуаций включает в себя три вида мероприятий: превентивные, оперативные, ликвидационные.

Превентивные мероприятия осуществляются в энергетическом хозяйстве для снижения возможности возникновения и реализации угроз энергетической безопасности, а также для снижения восприимчивости ТЭК и входящих в него систем топливо- и энергоснабжения к различным угрозам. Во время чрезвычайных ситуаций и выхода из них проводятся оперативные и ликвидационные мероприятия.

Оперативные мероприятия направлены на ограничение потребителей топливно-энергетических ресурсов либо же отпуск этих ресурсов из государственных запасов.

Ликвидационные мероприятия (мероприятия по ликвидации последствий чрезвычайных ситуаций) включают в себя производственно-технические и экономические мероприятия в сфере энергоснабжения, такие же мероприятия в сфере энергопотребления, а также мероприятия в социальной, политической и общеэкономической сфере.

На рис. 1 представлена общая схема исследований по оценке состояния ТЭК в условиях воз-

* Результаты, представленные в статье, получены при частичной поддержке грантов РФФИ № 08-07-00172, № 10-07-00264 и гранта Программы Президиума РАН № 2.29.

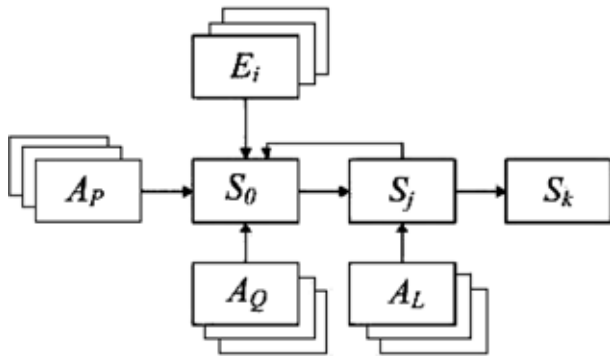


Рис. 1. Общая схема исследований по оценке состояния ТЭК

можных сценариев чрезвычайных ситуаций с учетом мероприятий, направленных на повышение уровня энергетической безопасности.

Здесь S_0 — начальное состояние ТЭК; E_i — сценарий чрезвычайной ситуации; $A = A_P \cup A_Q \cup A_L$ — набор превентивных, оперативных и ликвидационных мероприятий, нейтрализующих или смягчающих последствия чрезвычайной ситуации; S_j — состояние ТЭК после чрезвычайной ситуации E_i с учетом реализации набора мероприятий A_P и/или A_Q ; S_k — состояние ТЭК после проведения ликвидационных мер A_L .

Основная цель вычислительного эксперимента в исследованиях, схематически показанных на рис. 1, — определение инвариантного набора мероприятий $A = A_P \cup A_Q \cup A_L$, выполнение которых в условиях нормального функционирования позволит минимизировать и/или предотвратить вовсе последствия наибольшего числа возможных чрезвычайных ситуаций.

Онтологическое, когнитивное и событийное моделирование проблемы энергетической безопасности

Онтологическое моделирование. Под онтологическим моделированием понимается построение онтологий как в графическом, так и в формализованном виде. Онтологии определяют как базу знаний специального вида или как "спецификацию концептуализации" предметной области [2]. Пример фрагмента онтологии газовой отрасли ТЭК в графическом виде показан на рис. 2.

Когнитивное моделирование. Под когнитивным моделированием понимается построение и анализ когнитивных моделей, или, иначе, когнитивных карт (ориентированных графов), в которых вершины соответствуют факторам (концептам), а

дуги — связям между факторами (положительным или отрицательным) в зависимости от характера причинно-следственных отношений [3].

Общая когнитивная модель проблемы энергетической безопасности имеет следующий вид.

Заданными являются совокупность индикаторов энергетической безопасности $\{I\}$ и множество угроз безопасности $\{U\}$, определяющих внутреннее состояние энергетической системы, а также возможные влияния внешней среды (возмущения). Каждая из угроз реализуется вследствие сочетания ряда факторов, влияющих на возникновение чрезвычайной ситуации. Индикаторы определяют уровень энергетической безопасности, при этом должно выполняться соотношение

$$I_n \leq I_i < I_c,$$

где I_n — значение индикатора, определяющее нормальный уровень энергетической безопасности; I_j — текущее значение индикатора; I_c — значение индикатора, определяющее кризисный уровень безопасности.

Искомыми являются:

- множество факторов $\{C\}$, влияющих на развитие энергетической системы (или на развитие чрезвычайной ситуации в ней);
- множество связей $\{R\}$ между факторами;
- характер влияния факторов друг на друга (знак связи $R_{i,j}$ между факторами C_i и C_j);
- степень взаимовлияния факторов, т. е. множество весовых коэффициентов связей $\{W\}$;
- взаимосвязь факторов C_i с угрозами энергетической безопасности U_k и индикаторами энергетической безопасности I_j .

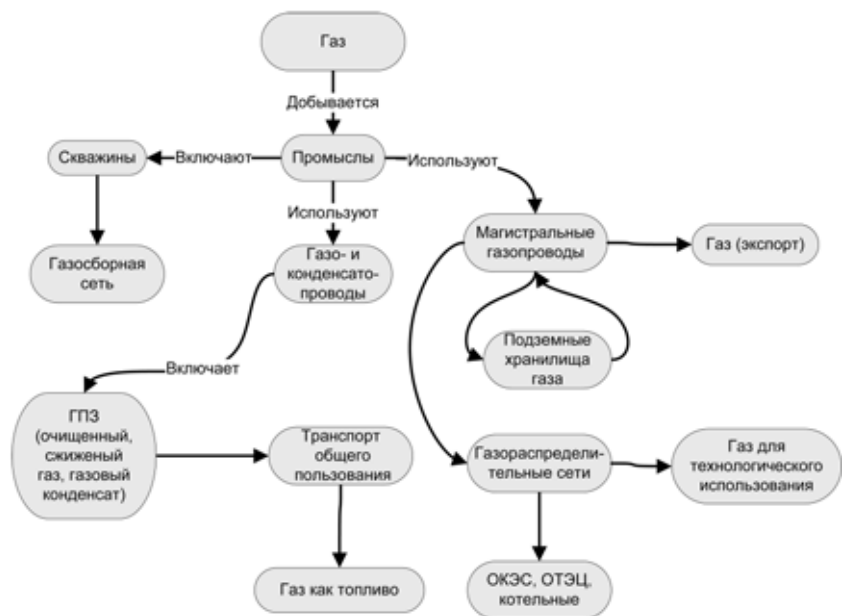


Рис. 2. Фрагмент онтологии газовой отрасли ТЭК

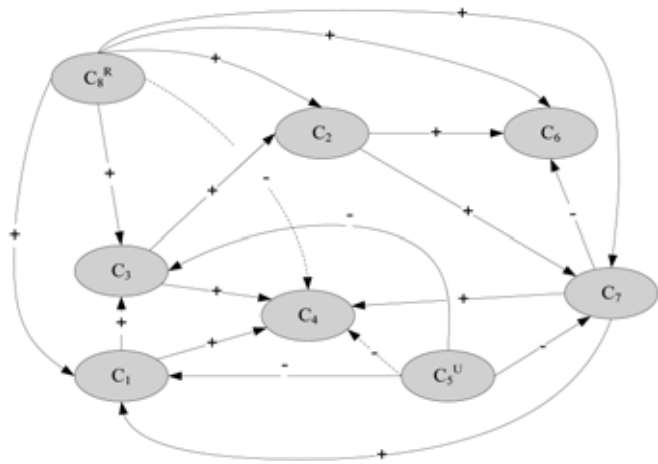


Рис. 3. Когнитивная модель последствий реализации угрозы "Землетрясение":

C_1 — добыча/производство энергоресурсов; C_2 — объем запасов топливно-энергетических ресурсов; C_3 — транспортировка этих ресурсов; C_4 — мощность объекта ТЭК; C_5^U — природная угроза (стихийное бедствие — землетрясение); C_6 — нормативный объем запасов топливно-энергетических ресурсов; C_7 — потребление энергоресурсов; C_8^R — управляющее воздействие (меры по ликвидации последствий чрезвычайной ситуации — землетрясения)

Знак связи $R_{i,j}$ определяется по следующему правилу:

- связь $R_{i,j}$ положительная, если увеличение фактора C_i приводит к увеличению фактора C_j или уменьшение фактора C_i приводит к уменьшению фактора C_j ;
- связь $R_{i,j}$ отрицательная, если уменьшение фактора C_i приводит к увеличению фактора C_j или увеличение фактора C_i приводит к уменьшению фактора C_j .

Назначение связям $R_{i,j}$ весов $W_{i,j}$, определяющих степень влияния фактора C_i на фактор C_j , позволяет определить суммарный вклад факторов в создание чрезвычайной ситуации.

В [4] автором рассматривался подход к когнитивному моделированию природных угроз энергетической безопасности. Один из примеров, реализованных в соответствии с этим подходом, приведен на рис. 3.

В рассматриваемом примере концепты C_1 — C_4 , C_6 , C_7 являются факторами, влияющими на развитие энергетической системы, а факторы C_5^U (угроза) и C_8^R (управляющее воздействие) представляют собой основные факторы, влияющие на развитие и ликвидацию последствий чрезвычайной ситуации.

Комментарии к рис. 3. Увеличение добычи/производства энергоресурсов C_1 требует увеличения объемов транспорта C_3 ; уменьшение их приводит к снижению мощности объекта ТЭК C_4 (связи положительные, так как явления одного знака). Рост потребления энергоресурсов C_7 требует повышения мощности объектов ТЭК C_4 или увеличения объемов добычи/производства энер-

горесурсов C_1 . Увеличение запасов топливно-энергетических ресурсов C_2 позволяет увеличить потребление C_7 и восстановить нормативные запасы топливно-энергетических ресурсов C_6 . Стихийное бедствие (угроза энергетической безопасности) C_5^U отрицательно отражается как на энергетических, так и транспортных объектах, снижая как мощность первых C_4 и объемы транспорта C_3 , так и объемы потребления топливно-энергетических ресурсов C_7 . При выходе из строя объектов ТЭК увеличение потребления C_7 приводит к уменьшению запасов C_2 . Уменьшение этих запасов ниже нормы C_6 может, в свою очередь, привести к возникновению чрезвычайной ситуации. Поэтому управляющее воздействие C_8^R , с одной стороны, может включать в себя ликвидационные меры, а с другой стороны, быть направлено на восстановление нормативных запасов топливно-энергетических ресурсов C_6 .

Построение когнитивных моделей различных аспектов проблемы энергетической безопасности и их анализ позволяют ранжировать чрезвычайные ситуации по степени их опасности и возможных последствий, а также определять превентивные мероприятия, направленные на предотвращение чрезвычайных ситуаций.

Событийное моделирование. Под событийным моделированием понимается построение поведенческих моделей. При этом предполагается, что система состоит из элементарных объектов (сущностей), а взаимодействие объектов описывается их реакцией на внешние и внутренние события [5]. Суть событийного моделирования заключается в отслеживании на модели последовательности событий в том же порядке, в каком они происходят или могут происходить в реальной системе. Вычисления выполняются только для тех моментов времени и тех частей модели, к которым относятся совершаемые события. Другими словами, обращение на очередном такте моделируемого времени происходит только к моделям тех элементов, на входах которых в этом такте произошли изменения [6].

На рис. 4 изображен пример событийной карты (графического изображения типовой событийной модели) угрозы "Землетрясение" для разных отраслей ТЭК.

Используя онтологию газовой отрасли ТЭК, фрагмент которой представлен на рис. 2, можно выполнить конкретизацию событийной модели угрозы "Землетрясение" для газовой отрасли (последняя приведена на рис. 5). При конкретизации уточняются, например, вид объекта ТЭК (подземное хранилище газа), вид транспортной магистрали (газопровод), меры по восстановлению газовой отрасли и т. д.

В качестве математического аппарата для описания событийных моделей предложено использовать одну из разновидностей алгебраических се-



Рис. 4. Событийная карта угрозы "Землетрясение" для отраслей ТЭК (э/р — энергоресурс)



Рис. 5. Конкретизация событийной модели угрозы "Землетрясение" для газовой отрасли ТЭК

тей — Joiner-сети, которые являются обобщением сетей Петри [7—8].

Возможности перехода от когнитивных карт к событийным моделям рассмотрены в работе [9].

Методика интеллектуальной поддержки исследований проблемы энергетической безопасности

Предлагается перейти от существующей технологии оценки уровня энергетической безопасности, основанной на проведении вычислительных экспериментов с помощью многоагентного программного комплекса ИНТЭК-М [10], к двухуровневой технологии, интегрирующей как указанное программное обеспечение, так и средства онтологического, когнитивного и событийного моделирования.

Согласно предлагаемой технологии на первом (верхнем) уровне с использованием методов интеллектуальной поддержки исследований выполняется моделирование угроз энергетической безопасности и чрезвычайных ситуаций, что позволяет осуществить качественный анализ возможных и выбор рациональных вариантов, подлежащих исследованию с помощью вычислительного эксперимента. При этом результаты когнитивного моделирования являются исходными данными для агента задания чрезвычайных ситуаций программного комплекса ИНТЭК-М, а результаты событийного моделирования — для агента формирования стратегий вычислительного эксперимента. Далее, с использованием полученных результатов разрабатывается стратегия проведения и генерируется схема вычислительного эксперимента, который выполняется на втором (нижнем), более детальном уровне с использованием программного комплекса ИНТЭК-М.

Основные компоненты и структура интеллектуальной ИТ-среды для поддержки этой технологии приведены на рис. 6.

ИТ-среда включает в себя:

- онтологическое пространство знаний в области энергетической безопасности;
- базу знаний, содержащую когнитивные модели угроз энергетической безопасности и описания прецедентов чрезвычайных ситуаций;



Рис. 6. Структура интеллектуальной ИТ-среды

- базу знаний, содержащую событийные модели развития и последствий чрезвычайных ситуаций в энергетике.

База знаний, содержащая описания прецедентов чрезвычайных ситуаций, имеет фреймовую структуру. В настоящее время база включает в себя описания около 100 прецедентов чрезвычайных ситуаций в энергетике.

Для построения когнитивных карт предлагается использовать онтологии как классификаторы концептов, которые используются в этих картах, т. е. в данном случае система онтологии используется как вспомогательный инструмент для когнитивного моделирования. Предусматривается возможность преобразования когнитивных моделей в событийные [9]. В случае, когда осуществляется переход от когнитивного и событийного моделирования к количественным расчетам с использованием многоагентного программного комплекса ИНТЭК-М, последний также может быть включен в состав интеллектуальной ИТ-среды.

Тогда математическое описание предлагаемой ИТ-среды имеет вид

$$V_{IT} = \{O, E, M_C, M_S\} \cup P,$$

где $\{O\}$ — множество онтологий; $\{E\}$ — множество описаний прецедентов чрезвычайных ситуаций; $\{M_C\}$ — множество когнитивных моделей; $\{M_S\}$ — множество событийных моделей; P — программный комплекс для проведения вычислительных экспериментов (обоснования вариантов развития ТЭК с учетом требований энергетической безопасности).

Для описания схемы процесса исследований с использованием интеллектуальной ИТ-среды введем множества $\{D_I\}$ — информация, необходимая для поддержки принятия решений при выборе стратегии проведения вычислительного эксперимента, и $\{D_R\}$ — информация, полученная в результате проведения вычислительных экспериментов и предназначенная для поддержки принятия решений в области обеспечения энергетической безопасности.

Схема процесса исследований с использованием предлагаемой ИТ-среды представлена на рис. 7.

Заключение

В статье описана методика интеллектуальной поддержки исследований проблемы энергетической безопасности, включающая в себя ряд этапов:

- построение онтологии в целях структурирования и классификации используемой в исследованиях информации;
- когнитивное моделирование угроз энергетической безопасности и чрезвычайных ситуаций в энергетике;
- событийное моделирование развития чрезвычайных ситуаций и анализа их последствий;

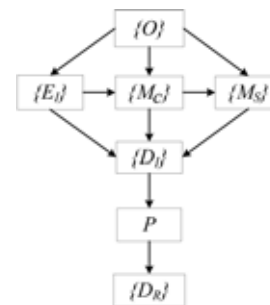


Рис. 7. Схема проведения исследований проблемы энергетической безопасности с использованием интеллектуальной ИТ-среды

- передача результатов качественного анализа в программный комплекс ИНТЭК-М;
- выполнение вычислительных экспериментов в целях количественной оценки вариантов развития ТЭК с учетом требований энергетической безопасности.

Программные средства реализации предлагаемой методики интегрируются в рамках интеллектуальной ИТ-среды и рассматриваются, с одной стороны, как инструмент поддержки принятия решений при выборе стратегии проведения вычислительного эксперимента в исследованиях энергетической безопасности, а с другой — как прототипы компонентов системы поддержки принятия решений в области обеспечения энергетической безопасности.

Список литературы

1. Бушуев В. В., Воропай Н. И., Мастепанов А. М., Шафраник Ю. К. и др. Энергетическая безопасность России. Новосибирск: Наука, 1998. 302 с.
2. Гаврилова Т. А., Хорошевский В. Ф. Базы знаний интеллектуальных систем. СПб.: Питер, 2001. 384 с.
3. Трахтенгерц Э. А. Компьютерная поддержка принятия решений. М.: СИНТЕГ, 1998. 376 с.
4. Massel A. G. Cognitive modeling of energy security treats // Proceedings of the 11th International Conference "Computer Science and Information Technologies". 2009. Vol. 1. P. 78—80.
5. Анисимов М. М. Управление событийными сетями // Тр. XIV Байкальской Всероссийской конф. "Информационные и математические технологии в науке и управлении". Ч. 3. Иркутск: ИСЭМ СО РАН, 2009. С. 238—240.
6. Событийный метод моделирования // В кн.: Норенков И. П., Кузьмик П. К. "Информационная поддержка наукоемких изделий. CALS технология". М.: Изд-во МГТУ им. Н. Э. Баумана, 2002. С. 196—197.
7. Столяров Л. Н., Новик К. В. Реализация параллельных процессов с помощью сетей Joiner-net // Тр. Байкальской Всероссийской конф. "Информационные и математические технологии". Иркутск: ИСЭМ СО РАН, 2004. С. 11—14.
8. Аршинский В. Л. Использование когнитивных карт для построения Joiner-net моделей взаимодействия сложных процессов // Тр. XIII Байкальской Всероссийской конф. "Информационные и математические технологии в науке и управлении". Т. 2. Иркутск: ИСЭМ СО РАН, 2008. С. 241—247.
9. Массель А. Г., Аршинский В. Л. Применение когнитивного моделирования для ситуационного анализа проблемы энергетической безопасности // Современные технологии. Системный анализ. Моделирование. 2008. Спецвыпуск. С. 75—80.
10. Аршинский В. Л., Массель А. Г., Фаргышев Д. А. Мультиагентный программный комплекс для исследований проблемы энергетической безопасности // Тр. XIV Байкальской Всероссийской конф. "Информационные и математические технологии в науке и управлении". Ч. III. Иркутск: ИСЭМ СО РАН, 2009. С. 283—289.

УДК 004.424

А. М. Токарчук, аспирант,
Московский государственный университет
путей сообщения (МИИТ),
e-mail: netandreas@gmail.com

Разработка паттерна *ActiveModel* для программирования интернет- приложения методом объектно- реляционного отражения

*Предлагается новый паттерн *ActiveModel*, как технологическое средство разработки интернет-приложения методом объектно-реляционного отражения. Паттерн строится на основе абстракции от базы данных MySQL и позволяет сократить сроки и стоимость разработки приложения.*

Ключевые слова: приложение, база данных, объектно-реляционное отражение, паттерн проектирования

Введение

В рамках современной концепции MVC (Модель—Вид—Контроллер, использующей разбиение интернет-приложения на соответствующие три части) [1], программист должен написать отдельно три части приложения. Первая часть — это модель, под которой здесь понимается программная модель, в которой инкапсулируется бизнес-логика приложения, причем бизнес-логика, как и сфера применения, может быть абсолютно любая. Вторая часть — это вид, отвечающий за отображение объекта либо в HTTP-клиенте, таком как в браузере, либо в XML-виде. Третья часть — контроллер, который будет отвечать за логику выбора действия для данного запроса. Написание видов и контроллеров для приложений представляет собой стандартизированные процедуры. Важным вопросом является аспект хранения данных модели.

В интернет-приложениях (далее "веб-системах") для реализации хранения данных приложения используются реляционные СУБД. Они получили широкое распространение по многим причинам. Практически в любой компании, которая предлагает размещение сайтов, сейчас можно найти MySQL-базу данных. Так или иначе, модель должна реализовывать CRUD-операции (создание, чтение, обновление, удаление объек-

та). Например, в системе управления торговой компанией могут предусматриваться самые разнообразные операции выборки из БД, такие как "найти все товары компании" (для отображения на карточке компании), "найти первого менеджера, который сейчас в сети, для компании товара N7", "найти цену первого товара заданной компании" и т. д.

Каждая из этих операций требует разработки отдельной функции программной модели, в которой мы должны будем сформулировать соединение с БД и запрос к ней, осуществить запрос, обработать результат, получить выборку из БД, обработать выборку и передать ее в приложение в определенном виде. Такие построения необходимо проделать для каждой операции. При изменении структуры БД, связанной с системой управления компанией, программисту приходится менять все функции программной модели, которые с ней взаимодействуют, а это означает переписывание всех моделей приложения. То же самое относится к записи, обновлению и удалению объекта.

Один из способов решения данной проблемы — это использование объектно-реляционных или объектных СУБД. Основной их особенностью является то, что разработчик работает с ними уже не на SQL-языке запросов, а на специальном объектно-ориентированном языке (внутреннем или внешнем). Тем не менее, для разработчика использование этих систем сулит ряд проблем: практически все они платные, а если мы говорим о молодом проекте, то фактор цены программного решения будет решающим в пользу других СУБД, ведь как правило финансирование на начальном этапе сильно ограничено. Распространенность этих СУБД в хостинг-компаниях также ограничена. Если реляционные СУБД (MySQL) [2] обычно включены по умолчанию в базовый тариф, то предоставление услуг с объектно-реляционными СУБД найти довольно сложно, а если удастся, то его цена на порядок отличается от цены реляционных БД.

Второй метод решения проблемы быстрой разработки заключается в использовании ORM-систем. Это системы объектно-реляционного отражения, которые позволяют использовать реляционную БД в сочетании с объектным механизмом доступа к данным. Они написаны на том же языке, что и само веб-приложение. Если используется язык написания веб-приложений PHP, то мож-

но подключить к использованию PHP-ORM-систему. Тогда программисту не придется осваивать новый язык для работы с объектами. При использовании ORM нет необходимости в написании SQL-запросов к СУБД, что служит сокращением времени разработки и числа ошибок. При таком способе разработки весь доступ к базе данных локализуется в участках кода ORM и перестает быть разбросанным по всему коду. Так как многие ORM поддерживают драйверы доступа к БД, то изменение типа БД не повлечет за собой значительного изменения программного кода веб-приложения. При этом облегчается переход на новые структуры таблиц, потому что достаточно изменить описание таблиц в модели ORM без изменения самих SQL-запросов.

При таком способе организации веб-приложения нет необходимости переписывать множество функций модели, а достаточно переписать однажды ее конфигурацию. Также значительно упрощаются операции выборки данных. Теперь достаточно указывать критерий выборки как аргумент функции и будет получен результат в стандартном виде. Таким образом разработчик работает не с БД с помощью SQL-запросов, а манипулирует объектами с помощью ORM.

Известные современные ORM-системы (Doctrine, Propel), хотя и обеспечивают указанные выше преимущества, имеют ряд недостатков. Как правило, все эти системы построены по шаблону проектирования ActiveRecord [1], который предполагает, что объект — это строка в таблице. Из этого вытекает ряд следующих ограничений.

1. Объект может быть только в одной таблице. Таким образом, проектирование объекта веб-приложения сводится к композиции объектов ORM, что увеличивает трудоемкость разработки.

2. Объект находится в единственной базе данных. Это следует из предыдущего пункта. Невозможно создать объект, распределенный по нескольким БД.

3. Объект должен иметь фиксированный набор свойств. В частности, число колонок в таблице должно быть равно числу свойств объекта.

4. Затруднена поддержка многоязычности, так как формируется единственная таблица, в которой часть столбцов "отвечает" за разные языки, а после выборки необходимо осуществлять фильтрацию данных.

Учитывая преимущества применения ORM-систем и исходя из недостатков существующих систем, сформулируем требования к новой разрабатываемой ORM-системе.

1. **Адекватное отражение.** Основная функция ORM — это отражение объектов предметной области веб-приложения в реляционную структуру БД, а также отражение отношений связей объектов. От-

сюда вытекает, что основная задача ORM — это адекватное отражение объектов и их связей. При этом могут возникнуть ряд проблем, например, объект может храниться в нескольких таблицах нескольких баз данных на нескольких серверах. Объекты могут быть различными по своей структуре. Например, в объект "Профиль пользователя" могут быть внесены новые пункты в процессе разработки, и, как следствие, у объекта появляются новые свойства.

2. **Поддержка многоязычности.** В условиях разработки веб-приложений часто возникает необходимость поддержки нескольких языков. Поддержка многоязычности должна быть прозрачной, т. е. разработчик должен быть уверен, что в данный момент используется отражение объекта на текущем для пользователя языке. Но в то же время система должна обладать достаточной гибкостью, чтобы менять язык отражения на нужный разработчику.

3. **Развитие поисковых возможностей.** Система должна поддерживать сложные запросы, например: "Найти email первого менеджера, компании, последнего размещенного товара".

1. Разработка ORM-системы ActiveRecord

Итак, разработку шаблона проектирования, получившего название ActiveRecord, побудил ряд недостатков современных систем, перечисленных выше. Для того, чтобы понять, что представляет собой ActiveRecord, необходимо рассмотреть вопрос паттернов проектирования вообще.

Шаблоны проектирования (паттерны, англ. design pattern) — это многократно применяемая архитектурная конструкция, предоставляющая технологическую поддержку проектирования в рамках конкретного контекста и описывающая значимость конкретного решения [5]. Паттерн не является законченным образцом проекта, который может быть прямо преобразован в код (что отличает его от близкого понятия идиомы), скорее это описание или образец решения задачи таким образом, чтобы это можно было использовать в различных ситуациях. Объектно-ориентированные шаблоны зачастую показывают отношения и взаимодействия между классами или объектами без определения того, какие конечные классы или объекты приложения будут использоваться.

Паттерн ActiveRecord — это метод проектирования, использующий абстракцию от базы данных через объектно-реляционное отражение. Данные объекта, используемые паттерном внутри БД, хранятся в таблицах. Предлагается несколько вариантов структур таких таблиц.

Обыкновенная таблица (таблица типа "classic"). Каждый столбец таблицы (табл. 1) соответствует

полю (свойству) программной модели. Получить свойство модели можно используя следующий код (здесь и далее на языке PHP): `$model → propName`.

Таблица 1

Username	Name	Email	icq
andrey1 vaska400	Adndrey Vasiliy	my@mail.ru vaska@mail.ru	222333444 11223344

Получить свойства модели из табл. 1 можно с помощью следующего кода:

```
$user □ name = Andrey
$user □ icq = 222333444
```

Расширяемая таблица (таблица типа "x-fields"). Она состоит (табл. 2) из трех столбцов: `username` — первичный ключ; `propName` — имя свойства (название виртуальной колонки); `value` — значение виртуальной колонки.

Таблица 2

Username	PropName	Value
andy77 andy77	name email	Andrey my@mail.ru

Например, в табл. 2 отражено имя и email пользователя. Это отразится в модели следующим образом:

```
$user □ name = Andrey
$user □ email = my@mail.ru
```

Преимущество данной таблицы в возможности добавления новых свойств программной модели без изменения структуры таблицы.

Таблица флагов (таблица типа "x-flags"). Если в таблице этого типа (табл. 3) существует строка с заданным первичным ключем, то считается, что флаг установлен. В противном случае, флаг не установлен.

Таблица 3

Username	Name	StartTime	EndTime
Andy77	active	11111	33333

Такая таблица в программной модели отразится следующим образом:

```
$user □ active =
= {"startTime": 11111,"endTime":33333}.
```

В значении свойства программной модели содержится закодированный методом JSON (JavaScript Object Notation — нотация хранения объектов языка Javascript) массив из свойств `startTime` и `endTime` данного флага. Это сделано для того, чтобы не было пересечений значений столбцов `startTime/endTime` у флагов и можно было бы ра-

ботать с несколькими флагами модели одновременно. Существуют два способа задания значения флага:

- запись значения флага без указания времени начала действия и конца действия флага:

```
$model = User::model() □ findOneByAttributes(array('username' => 'andrey'));
$model □ isPaid = 1;
$model □ save();
```

- запись с указанием значения начала и конца действия флага:

```
$model = User::model() □ findOneByAttributes(array('username' => 'andrey'));
$model □ isPaid = array('startTime' => 11111, 'endTime' => 33333);
$model □ save();
```

Дата для полей `startTime` и `endTime` задается в следующем формате:

```
$startTime = (string) floor(time()*100); // для сегодняшнего дня
```

Чтение свойств флага осуществляется с помощью следующего кода:

```
$model = User::model() □ findOneByAttributes(array('username' => 'andrey'));
$flagArray = Zend_Json::decode($model □ isPaid);
$startTime = date("d.m.Y", $flagArray['startTime']/100);
$endTime = date("d.m.Y", $flagArray['endTime']/100);
```

Между программными моделями существует несколько типов отношений, рассмотренных ниже.

Один-к-одному (hasOne). Отношение "один к одному" проиллюстрировано на рис. 1. Предметная область состоит из объекта типа "пользова-

Один к одному - hasOne



```
$config['relations'] = array(
    'company' => array(
        'type' => 'hasOne',
        'foreignKey' => 'company_id',
        'destinationKey' => 'id',
        'destinationModel' => 'Company'
    )
);
```

Рис. 1

тель" и объекта типа "компания". "Пользователь" может работать только в одной "компании".

Один-ко-многим (hasMany). Отношение "один ко многим" проиллюстрировано на рис. 2. Предметная область состоит из "компаний", которая продает "товары".

Многие-ко-многим (hasManyToMany). Отношение "многие ко многим" представлено на рис. 3. Предметная область состоит из множества "компаний", которые продают множество "товаров". Причем один и тот же товар могут продавать разные компании.

Расширенное многие-ко-многим (hasManyToManyExtended). Связь "многие ко многим" с дополнительными свойствами связи представлена на рис. 4. Предметная область состоит из "пользователей", которые покупают "товары", выставленные на продажу "компаниями". Пользователь ↔ ↔ Товары связаны этим типом связи, а в качестве дополнительного свойства связи будет выступать count (количество товаров в корзине). Другой пример (на рис. 4), когда у компании есть товары, которые ее интересуют для закупки, тогда свойство count передает количество товара, которое нужно закупить. Связь моделей Company и Product идет через таблицу пересечений CompanyProduct.

Важно помнить, что расширенные свойства не сохраняются при сохранении модели, так как для них нет столбцов в таблице, и это фактически не свойства модели, а свойства связи.

Связи "один-ко-многим по нескольким полям". Предположим, что существует несколько категорий, которым принадлежит компания. Эта принадлежность осуществляется с помощью полей category_1, category_2, category_3.

Company(category_1) □ Category(id),

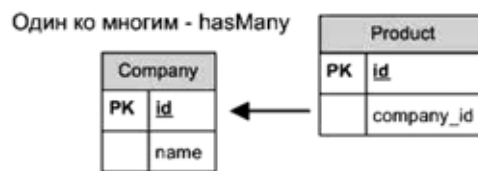
Company(category_2) □ Category(id),

Company(category_3) □ Category(id).

Доступ к компаниям некоторой категории осуществляется посредством использования свойства Category → companies. В нем находятся компании, у которых свойство "category_1" или "category_2", или "category_3" равны свойству "id" текущей категории.

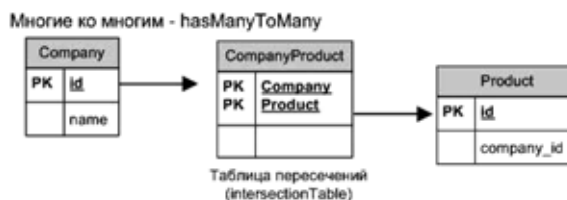
В функции _construct() программной модели описание связи этого вида реализовано с помощью следующего программного кода:

```
$config['relations'] = array(
    'companies' => array(
        'type' => 'hasMany',
        'foreignKey' => 'id',
        'destinationKey' => array('category 1', 'category 2', 'category 3'),
        'destinationModel' => 'Company'
    ));
```



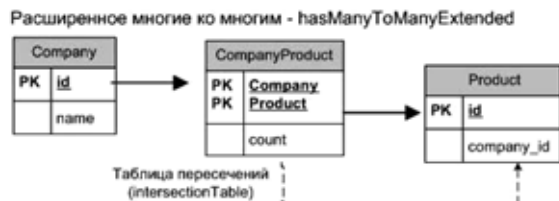
```
$config['relations'] = array(
    'products' => array(
        'type' => 'hasMany',
        'foreignKey' => 'id',
        'destinationKey' => 'company_id',
        'destinationModel' => 'Product'
    )
);
```

Рис. 2



```
$config['relations'] = array(
    'products' => array(
        'type' => 'hasManyToManyExtended',
        'foreignKey' => 'id',
        'destinationKey' => 'Company',
        'destinationModel' => 'Product',
        'intersectionTable' => 'CompanyProduct',
        'intersectionForeignKey' => 'Product',
        'intersectionDestinationKey' => 'id'
    )
);
```

Рис. 3



Свойство связи Count Добавляется к модели Product, найденную через связь. Кол-во расширенных свойств не ограничено.

```
$config['relations'] = array(
    'cart' => array(
        'type' => 'hasManyToMany',
        'foreignKey' => 'username',
        'destinationKey' => 'username',
        'destinationModel' => 'Product',
        'intersectionTable' => 'UserProduct',
        'intersectionForeignKey' => 'Product',
        'intersectionDestinationKey' => 'id'
    )
);
```

Рис. 4

2. Возможности ActiveRecord

При задании требований к разработке ORM-системы ActiveRecord были учтены следующие возможности:

- поиск моделей в базе данных должен выполняться по атрибутам моделей или по выражению на языке SQL;
- разные модели могут быть в разных таблицах разных баз данных;
- одна модель может состоять из нескольких таблиц, размещенных в разных базах данных;
- возможно использование моделей, составленных из нескольких таблиц;
- должна быть предусмотрена поддержка таблиц вида: classic, x-fields, x-flags (см. п. 1);
- должна быть возможность автоопределения виртуальных колонок для таблиц типа x-fields и x-flags;
- использование "Lazy-loading" для связей таблиц (загрузка связей по требованию);
- возможность использования программных моделей и как объектов, и как массивов. Код `$model < varName` и код `$model['varName']` вернут один и тот же результат;
- использование коллекции объектов в случае, когда в результате выборки образуется множество моделей;
- возможность фильтрации моделей после выборки;
- инкремент/декремент свойств модели с учетом многопоточности;
- возможность использования таблиц "только для чтения";
- возможность использования "Fluent Interface", т. е. цепочки вызовов, подобной нижеприведенной:

```
$user = User::model() < findByAttributes(array('username' => 'andrey')) < populate($data) < save();
```

Цепочка вызовов начинается с функции `ActiveModel::model()`, которая выдает псевдостатическую модель. При последующих вызовах модели того же типа, функция `model()` выдает ее из кэша, не создавая объект заново. При этом экономятся ресурсы за счет того, что конструктор модели вызывается только один раз, а при последующих вызовах модель берется из кэша.

2.1. Система кэширования

"ActiveModel" имеет 3-уровневую систему кэширования. Кэш первого уровня (*L1*) проводит кэширование пустых моделей при первом вызове и экономит время на создание экземпляров моде-

лей. При втором вызове модель клонируется из кэша. Кэширование происходит в оперативную память.

Кэш второго уровня (*L2*) осуществляет кэширование результатов выборки в пределах отработки HTTP-запроса. Если во время сеанса производится повторный запрос выборки с теми же значениями параметров, то выборка осуществляется из кэша. Кэширование происходит в оперативной памяти.

Кэш третьего уровня (*L3*). При последующих запросах одной и той же выборки (функции `find*`) будет взят результат из кэша. Кэширование проводится в файлах на диске. Этот кэш по умолчанию выключен. Включить его можно следующим образом:

```
ActiveModel::$CACHE = true;
```

```
ActiveModel::$CACHE_FOLDER = './cache'; //  
путь до папки с кэшем, права 777
```

Принудительно сбросить кэш моделей можно таким образом:

```
ModelName::clearCache();
```

Для того, чтобы сбросить кэш всех моделей, необходимо написать следующий код:

```
ActiveModel::clearCache();
```

2.2. Другие возможности

Кроме использования кэширования, ActiveRecord обладает и другими возможностями: работа с автоопределяемыми колонками, проверка свойств модели, инкремент свойства в многопоточном окружении, клонирование моделей, косвенная загрузка моделей, таблица только для чтения, синхронизация с базой данных, адресация по части ключа, возможность преобразования ключей.

3. Возможности коллекции моделей ActiveRecordCollection

При получении выборки моделей из базы данных, функции `findAllByAttributes()` и `findAllBySQL()` возвращают объект типа "ActiveModelCollection". "ActiveModelCollection" — это объект, поддерживающий интерфейс "ArrayAccess", допускающий обращение к себе как к обычному массиву, чтобы получить нужную модель. В то же самое время объект "ActiveModelCollection" допускает обращение к себе и как к объекту. Это обеспечивает возможность использования функции для групповой обработки моделей в коллекции. Например, если необходимо напечатать номера всех найденных товаров, которые уже опуб-

ликованы (status = published), это реализуется с помощью следующего кода:

```
$products = Product::model() & findAllByAttributes(array('status' => 'published'));
foreach($products as $product) {
    print $product & id;
    print ' <br /> ';
}
```

Также допускается использование специальной функции toArray() для преобразования "ActiveModelCollection" в обычный массив.

```
$products = Product::model() & findAllByAttributes(array('status' => 'draft') & toArray());
```

После этого в переменной "products" будет располагаться массив массивов свойств товаров. Имеется возможность массово устанавливать свойства моделей. Например, если необходимо изменить статус товаров в выборке, а потом сохранить их, это можно реализовать с помощью следующего кода:

```
$products = Product::model() & findAllByAttributes(array('status' => 'draft');
$products & status = 'published';
$products & save();
```

Это же можно сделать другим способом:

```
Product::model() & findAllByAttributes(array('status' => 'draft') & set('status', 'published') & save();
```

Таким образом, одной строкой кода проводится поиск моделей по параметрам, изменяется параметр всех моделей выборки и каждая модель сохраняется в базе данных.

Аналогично используется возможность массового применения любой функции модели ActiveRecord или ее наследников (например "Product") — для всех моделей массива. Например, для сохранения всех товаров необходимо написать следующий код:

```
$products = Product::model() & findAllByAttributes(array('status' => 'draft') & save();
```

Функция save() применится ко всем моделям выборки. Также существует возможность написания собственной функции программной модели, а затем массово применять ее одной командой для всех моделей без необходимости их обхода в цикле. Это реализуется с помощью следующего кода:

```
Class Product extends ActiveRecord {
    public function printFoo($foo) {
        print $foo;
    }
}
$product = Product::model() & findAllByAttributes(array('status' => 'draft') & printFoo('bar');
```

Этот код вызовет функцию printFoo() с аргументом 'bar' для каждой модели.

3.1. Фильтрация моделей после выборки

Очень часто после выборки моделей из базы данных может потребоваться их фильтрация по атрибутам. С помощью "ActiveModel" реализовать это легко:

```
$users = User::model()-> findAllByAttributes(array('company_id' => '1')) & filter(array('name' => 'Andrey');
```

Функция filter(), примененная к результатам выборки, оставляет в ней только те модели, которые соответствуют заданным атрибутам. Важно обратить внимание, что функцию filter() можно применять только к выборке, но не к одной модели, поэтому код

```
$users = User::model()-> findOneByAttributes(array('company_id' => '1')) & filter(array('name' => 'Andrey');
```

приведет к ошибке. Фильтрация происходит средствами языка PHP после получения результата из БД; если критерии фильтрации известны заранее, рекомендуется включить их в атрибуты функций findAll*(). Основное применение метода filter() можно увидеть в фильтрации моделей, полученных через связи других моделей, например:

```
$messages = User::model()-> findOneByAttributes(array('username' => 'andrey')) & receivedMessages & filter(array('topic' => 'Hello');
```

Подразумевается, что данная связь должна быть типа "один-ко-многим" или "многие-ко-многим", для того, чтобы в результате прохода по ней образовалась не единичная модель, а коллекция моделей.

```
Class User {
    $config['relations'] = array(
        'receivedMessages' => array(
            'type' => 'hasMany',
            'foreignKey' => 'username',
            'destinationKey' => 'to',
            'destinationModel' => 'Message'
        );
};
```

3.2. Лимитирование результатов выборки моделей

После получения результатов выборки можно не только отфильтровать их, но и оставить заданное число моделей. Для этого применяется функция limit();

```
$messages = User::model()-> findOneByAttributes(array('username' => 'andrey')) & receivedMessages & filter(array('topic' => 'Hello')) & limit(2);
```

При этом выведется не более двух сообщений.

Закключение

Рассмотрим ситуацию, когда необходимо найти имя директора компании, разместившей первый товар во второй партии третьего вагона состава N4 компании 'Компания А'. Используемый язык в обоих случаях — PHP. С помощью ActiveModel это реализует следующий код:

```
$directorName = Company::model()->findOneByAttributes(array('name' => 'Компания А') □ trains[3] □ railcars[2] □ consignment[1] □ products[0] □ company □ director □ name; (одна строка в редакторе кода);
```

Так как отсчет ведется с нуля, то числа в индексах массива, соответственно, на единицу меньше. Сравним, что необходимо было бы сделать, работая с традиционной системой (например на базе Zend_Db):

```
$registry = Zend_Registry::getInstance ();  
$select = new Zend_Db_Select($registry['db']);  
// ищем компанию  
$select -> from(array('base'  
=> 'object_company'))  
-> where('name — "ЖД компания"");  
$query = $select-> query();  
$result = $query-> fetchOne(); if(!is_null($result) {  
$company_id = $result['id'];
```

... (далее идет поиск аналогичным образом состава, партии, товара в партии, компании, ее директора, и его имени).

```
}
```

Итого потребовалось 65 (PHP + Zend_Db) строк против одной строки (PHP + ActiveModel). Надо сказать, что чем шире и выше дерево модельных связей, тем большее преимущество дает применение паттерна ActiveModel. В данном примере глубина прохождения по связям составила 7, и мы увидели различия в количестве кода. Следует предположить, что при частом применении, будет наблюдаться снижение производительности, однако для компенсации этого снижения можно использовать систему кэширования ActiveModel.

Применение ActiveModel дает гибкость при разработке. Эта гибкость характеризуется тем, что при изменении таблиц в БД нет необходимости переписывать все функции, работающие с ними, а нужно всего лишь переписать функцию _construct() модели. Таким образом, применение паттерна ActiveModel может значительно ускорить как процесс разработки, так и процесс сопровождения кода веб-приложения.

Список литературы

1. Фаулер М. Архитектура корпоративных программных приложений. М.: Вильямс, 2009.
2. Хоп Г., Вульф Б. Шаблоны интеграции корпоративных приложений. М.: Вильямс, 2009.
3. Веллинг Л., Томсон Л. Разработка Web-приложений с помощью PHP и MySQL. М.: Вильямс, 2009.
4. Дюбуа П. MySQL. Сборник рецептов. М.: Символ, 2007.
5. Wikipedia. URL:http://ru.wikipedia.org/Паттерн_проектирования

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ЭКОНОМИКЕ И УПРАВЛЕНИИ

УДК 519.86

Ю. И. Димитриенко, д-р физ.-мат. наук. проф., зав. каф.,
О. Ю. Дмитриенко, аспирант, МГТУ им. Н. Э. Баумана,
e-mail: olyad@mail.ru

Модель деформируемых кластеров для анализа динамических данных в экономике

Предложено развитие модели рынка массовых продаж, основанной на кластеризации объектов в специальном многомерном пространстве, а также на построении континуальной модели движения кластеров. Сформулирована система уравнений, описывающая движение деформируемых кластеров в многомерном пространстве. Представлен анализ результатов численного расчета параметров динамики рынка продаж, проведено сравнение с экспериментальными данными.

Ключевые слова: математическое моделирование, маркетинговое воздействие, деформируемый кластер, многомерное пространство, уравнение движения

Прогнозирование динамики изменяющихся данных о различных процессах является одной из основных целей математического моделирования

в экономике. Кроме статистических методов, в последнее время особую актуальность приобретают методы интеллектуального анализа данных,

основанные на различных моделях внутренних закономерностей самих данных и моделях динамики их изменения. В предыдущей работе авторов [3] был предложен принципиально новый подход к построению таких моделей анализа динамически меняющихся данных. Этот подход основан на кластеризации данных в специальном многомерном пространстве, переходе к континуальному описанию данных и на использовании законов сохранения для многомерных континуумов, являющихся аналогами законов сохранения массы, импульса и момента импульса в классической механике. В той же работе [3] была разработана частная модель для прогнозирования динамики кластеров в многомерном пространстве, основанная на гипотезе о жестком кластере, у которого расстояния между точками не меняются в процессе движения.

Цель настоящей работы — дальнейшее развитие метода, предложенного в работе [3], и разработка новой модели движения кластеров в многомерном пространстве, допускающей деформации кластеров определенного вида.

Законы сохранения в многомерном пространстве данных

Пусть имеются большие массивы данных, относящиеся к различным моментам времени t . Основным элементом разработанной в работе [3] модели является аксиома о том, что у нас имеется эмпирический алгоритм, позволяющий структурировать эти данные, образовав из них массивы точек $\mathbf{x} = \{x^1, \dots, x^n\}$ в некотором многомерном евклидовом пространстве E_n . Требование, чтобы точки принадлежали именно E_n , а каждое $x^i \in \mathbf{R}$, весьма сильное и не каждая структуризация данных удовлетворяет ему. В работе [3] был предложен способ построения E_n для экономических данных, в котором координаты x^i точки — это суммарное количество i -го товара, покупаемого или планируемого к покупке одним покупателем за время t , где $i = 1, \dots, n$, а $x^i \in \mathbf{R}$. "Планирование товара к покупке" позволяет перейти к исчислению количества товара в вещественных числах $x^i \in \mathbf{R}$.

Вторая основополагающая аксиома заключается в том, что отдельные точки в E_n — индивидуумы, являющиеся носителями информации, связаны между собой определенными связями и их дальнейшее движение в E_n подчиняется некоторым коллективным закономерностям. Для модели рынка массовых продаж, рассмотренной в работе [3], такими связями являются наличие определенных стереотипов экономического поведения у покупателей из одних социальных групп, а также наличие возможности обмена информацией о покупаемых товарах. Тогда с помощью пред-

ложенного в работе [3] алгоритма кластеризации данных в E_n для фиксированного момента времени t можно ввести несколько кластеров данных. Далее, согласно основному допущению модели, осуществляется переход к континуальному описанию в E_n , при котором кластерам соответствуют области V в E_n , и для них полагаются справедливыми: 1) закон сохранения числа индивидуумов (покупателей) в кластере; 2) закон изменения количества движения (частоты покупок) кластера; 3) закон изменения момента количества движения (момента частоты покупок):

$$\frac{d}{dt} \int_V \rho dV = 0; \quad (1)$$

$$\frac{d}{dt} \int_V \rho \mathbf{v} dV = \int_V \rho \mathbf{f}_m dV + \int_{\Sigma} \mathbf{t}_{\Sigma} d\Sigma; \quad (2)$$

$$\frac{d}{dt} \int_V \rho \mathbf{x} \times \mathbf{v} dV = \int_V \rho \mathbf{x} \times \mathbf{f}_m dV + \int_{\Sigma} \mathbf{x} \times \mathbf{t}_{\Sigma} d\Sigma, \quad (3)$$

где ρ — плотность кластера; \mathbf{f}_m — вектор плотности внешних массовых сил; \mathbf{t}_{Σ} — вектор поверхностных сил. Радиус вектор точек $\mathbf{x} = x^i \mathbf{e}_i$ в пространстве E_n подчиняется закону движения (закону накопления покупок) в кластере $\mathbf{x} = \mathbf{x}(X^i, t)$, где $\mathbf{x} = \{x^1, \dots, x^n\}$ — декартовы (эйлеровы) координаты; X^i — лагранжевы координаты; $\mathbf{v} = \frac{\partial \mathbf{x}(X^i, t)}{\partial t}$ — вектор скорости (частоты покупок).

Все основные элементы пространства E_n , в том числе векторное произведение $\mathbf{a}_1 \times \mathbf{a}_2 = (\mathbf{a}_2 \otimes \mathbf{a}_1) \cdot \mathbf{n}_{\Sigma}$, введены в работе [3], здесь $\mathbf{n}_{\Sigma} = e_{i_1 \dots i_n} e^{i_1} \otimes \dots \otimes e^{i_n}$ — тензор Леви-Чивиты; $e_{i_1 \dots i_n}$ — символы Леви-Чивиты; \cdot — оператор двойного скалярного произведения [4]; \otimes — знак операции тензорного умножения.

Модель деформируемого кластера

Введем модель деформируемого кластера, закон движения которого представляет собой суперпозицию поступательного движения центра масс кластера, мгновенно-вращательного движения всего кластера как жесткого целого вокруг центра масс и растяжения-сжатия кластера по некоторым главным осям $O'p_{\gamma}$, т. е. записывается следующим образом:

$$\mathbf{x} = \mathbf{x}^0 + \mathbf{S} \cdot \mathbf{Q} \cdot \tilde{\mathbf{x}}', \quad (4)$$

где обозначен ортогональный тензор поворота \mathbf{Q} , радиус-вектор $\tilde{\mathbf{x}}'$ рассматриваемой точки в ла-

гранжевом описании и радиус-вектор мгновенного центра поворота кластера $\tilde{\mathbf{x}}^0$:

$$\mathbf{Q} = Q^{ij}(t)\mathbf{e}_i \otimes \mathbf{e}_j, \quad \tilde{\mathbf{x}}' = X^i \mathbf{e}_i, \quad \mathbf{x}^0 = x^{0i}(t)\mathbf{e}_i. \quad (5)$$

Тензор растяжения \mathbf{S} полагается симметричным, положительно определенным, для него существует вещественнозначный собственный базис \mathbf{p}_γ , в котором этот тензор имеет диагональный вид:

$$\mathbf{S} = S_\gamma^j \mathbf{e}_j \otimes \mathbf{e}^j = \sum_{\gamma=1}^n S_\gamma \mathbf{p}_\gamma \otimes \mathbf{p}_\gamma, \quad (6)$$

где S_γ^j — компоненты симметричной матрицы деформаций кластера; S_γ — вещественные положительные собственные значения тензора \mathbf{S} ; \mathbf{p}_γ — собственный базис.

Полагается, что единственной причиной изменения расстояний между точками кластера (т. е. деформации) является меняющееся соотношение между финансовым запасом покупателя e и его начальным значением e_0 , поэтому собственные значения S_γ аксиоматически полагаются заданными в виде следующих функций: $S_\gamma = S_\gamma(e - e_0)$, в простейшем случае эту зависимость будем полагать линейной

$$S_\gamma = 1 + A_\gamma(e - e_0), \quad (7)$$

где A_γ — константы.

Полагается также, что главные оси $O'\mathbf{p}_\gamma$ тензора растяжения \mathbf{S} по отношению к главным осям инерции кластера $O'\bar{\mathbf{e}}'_\gamma$ являются неподвижными, и базисы \mathbf{p}_γ и $\bar{\mathbf{e}}'_\gamma$ связаны ортогональным тензором $\mathbf{G} = G_\gamma^j \bar{\mathbf{e}}'_j \otimes \bar{\mathbf{e}}^j$, компоненты которого G_γ^j не зависят от времени:

$$\mathbf{p}_\gamma(t) = \mathbf{G} \cdot \bar{\mathbf{e}}'_\gamma(t) = G_\gamma^i \bar{\mathbf{e}}'_i(t), \quad G_\gamma^i = \text{const}. \quad (8)$$

Термин "деформируемый кластер" означает, что деформация кластера осуществляется "свободным образом" без приложения внешних сил, а только за счет изменения финансового запаса e (аналога температурного воздействия в классической механике сплошных сред [5]). Из общих теорем дифференциальной геометрии следует [4], что изменение объема кластера определяется произведением коэффициентов S_γ : $V/\dot{V} = \prod_{\gamma=1}^n S_\gamma$.

Вектор скорости деформируемого кластера

Дифференцируя (4) по t , получаем выражение для вектора скорости: $\mathbf{v} = \frac{d\mathbf{x}}{dt} = \dot{\mathbf{x}}^0(t) + (\mathbf{S} \cdot \dot{\mathbf{Q}}) \cdot \tilde{\mathbf{x}}'$. Вводя скорость движения центра вращения кластера $\mathbf{v}_0 = \dot{\mathbf{x}}^0$ и относительный радиус-вектор

$\tilde{\mathbf{x}} = \mathbf{x} - \mathbf{x}^0$, эту формулу можно переписать так: $\mathbf{v} = \mathbf{v}_0 + (\mathbf{S} \cdot \dot{\mathbf{Q}}) \cdot (\mathbf{S} \cdot \mathbf{Q})^{-1} \cdot \tilde{\mathbf{x}}$. Учитывая ортогональность тензора поворота $\mathbf{Q}^{-1} = \mathbf{Q}^T$, получаем итоговую формулу

$$\mathbf{v} = \mathbf{v}_0 + \tilde{\mathbf{x}} \cdot \tilde{\mathbf{W}}, \quad (9)$$

которая дает распределение скорости в деформируемом кластере и заменяет n -мерную обобщенную формулу Эйлера [3]. Здесь введен новый тензор

$$\tilde{\mathbf{W}} = \mathbf{S}^{-1} \cdot \mathbf{W} \cdot \mathbf{S} + \mathbf{S}^{-1} \cdot \dot{\mathbf{S}}, \quad (10)$$

который уже не является кососимметричным, а $\mathbf{W} = \mathbf{Q} \cdot \dot{\mathbf{Q}}^T$ — кососимметричный тензор вращения.

Вычислим производную от тензора \mathbf{S} (6) с учетом (8) и формулы $\dot{\bar{\mathbf{e}}}'_i = -\mathbf{W} \cdot \bar{\mathbf{e}}'_i$:

$$\begin{aligned} \dot{\mathbf{S}} &= \sum_{\gamma=1}^n \dot{S}_\gamma G_\gamma^i G_\gamma^j \bar{\mathbf{e}}'_i \otimes \bar{\mathbf{e}}'_j + \\ &+ \sum_{\gamma=1}^n S_\gamma G_\gamma^i G_\gamma^j \dot{\bar{\mathbf{e}}}'_i \otimes \bar{\mathbf{e}}'_j + \sum_{\gamma=1}^n S_\gamma G_\gamma^i G_\gamma^j \bar{\mathbf{e}}'_i \otimes \dot{\bar{\mathbf{e}}}'_j = \\ &= \sum_{\gamma=1}^n \dot{S}_\gamma G_\gamma^i G_\gamma^j \bar{\mathbf{e}}'_i \otimes \bar{\mathbf{e}}'_j - \mathbf{W} \cdot \mathbf{S} + \mathbf{S} \cdot \mathbf{W}. \end{aligned} \quad (11)$$

Вычисляем далее

$$\begin{aligned} \mathbf{S}^{-1} \cdot \dot{\mathbf{S}} &= \sum_{\beta=1}^n S_\beta^{-1} G_\beta^k G_\beta^l \bar{\mathbf{e}}'_k \otimes \bar{\mathbf{e}}'_l \sum_{\gamma=1}^n \dot{S}_\gamma G_\gamma^i G_\gamma^j \bar{\mathbf{e}}'_i \otimes \\ &\otimes \bar{\mathbf{e}}'_j - \mathbf{S}^{-1} \cdot \mathbf{W} \cdot \mathbf{S} + \mathbf{W} = \sum_{\gamma=1}^n S_\gamma^{-1} \dot{S}_\gamma G_\gamma^k G_\gamma^l \bar{\mathbf{e}}'_k \otimes \\ &\otimes \bar{\mathbf{e}}'_j - \mathbf{S}^{-1} \cdot \mathbf{W} \cdot \mathbf{S} + \mathbf{W}. \end{aligned} \quad (12)$$

Первое слагаемое во второй строке (12) представляет собой производную Яуманна от логарифмического тензора растяжения

$$\ln \mathbf{S} = \sum_{\gamma=1}^n \ln S_\gamma \mathbf{p}_\gamma \otimes \mathbf{p}_\gamma = \sum_{\gamma=1}^n \ln S_\gamma G_\gamma^k G_\gamma^j \bar{\mathbf{e}}'_k \otimes \bar{\mathbf{e}}'_j, \quad (13)$$

$$\begin{aligned} \ln \mathbf{S}^J &= \sum_{\gamma=1}^n \frac{d}{dt} (\ln S_\gamma G_\gamma^k G_\gamma^j) \bar{\mathbf{e}}'_k \otimes \bar{\mathbf{e}}'_j = \\ &= \sum_{\gamma=1}^n S_\gamma^{-1} \dot{S}_\gamma G_\gamma^k G_\gamma^j \bar{\mathbf{e}}'_k \otimes \bar{\mathbf{e}}'_j, \end{aligned} \quad (14)$$

поскольку матрица G_γ^k не зависит от t .

Подставляя (13) и (14) в (12), получаем следующее выражение:

$$\mathbf{S}^{-1} \cdot \dot{\mathbf{S}} = \ln \mathbf{S}^J - \mathbf{S}^{-1} \cdot \mathbf{W} \cdot \mathbf{S} + \mathbf{W}. \quad (15)$$

После подстановки (15) в (10), приходим к следующему результату:

$$\tilde{\mathbf{W}} = \mathbf{W} + \ln \mathbf{S}^J. \quad (16)$$

Уравнения движения деформируемого кластера в неподвижном базисе

Подвижную систему отсчета $O' \bar{e}'_i$ выбираем таким образом, чтобы центр вращения — точка \mathbf{x}_0 совпадала с центром масс кластера, тогда $\int_V \rho \tilde{\mathbf{x}} dV = 0$ и средняя скорость кластера совпадает с \mathbf{v}_0 :

$$\bar{\mathbf{v}} = \frac{1}{M} \int_V \rho \mathbf{v} dV = \frac{1}{M} \int_V (\rho \mathbf{v}_0 + \rho \tilde{\mathbf{x}} \cdot \tilde{\mathbf{W}}) dV = \mathbf{v}_0. \quad (17)$$

Подставляя формулы (16) и (9) в интеграл $\int_V \rho \mathbf{x} \times \mathbf{v} dV$, получаем выражение для тензора моментов количества движения деформируемого кластера:

$$\begin{aligned} n-2 \bar{\mathbf{m}} &= \int_V \rho \mathbf{x} \times \mathbf{v} dV = M \mathbf{x}_0 \times \mathbf{v}_0 + \\ &+ (\mathbf{W}^T \cdot \mathbf{I}) \cdot n \bar{\mathbf{e}} + (\ln \mathbf{S}^J \cdot \mathbf{I}) \cdot n \bar{\mathbf{e}}. \end{aligned} \quad (18)$$

Здесь $\mathbf{I} = \int_V \rho \tilde{\mathbf{x}} \otimes \tilde{\mathbf{x}} dV$ — тензор моментов инерции кластера. Формула (18) показывает, что тензор моментов частот поупок для деформируемого кластера по сравнению с жестким кластером [3] включает дополнительный момент $(\ln \mathbf{S}^J \cdot \mathbf{I}) \cdot n \bar{\mathbf{e}}$ за счет деформирования кластера в процессе движения.

Подставляя выражения (17) и (18) в (2) и (3), получаем систему уравнений движения деформируемого кластера:

$$M \frac{d\mathbf{v}_0}{dt} = \bar{\mathbf{f}}; \quad (19)$$

$$\frac{d\mathbf{x}_0}{dt} = \mathbf{v}_0; \quad (20)$$

$$\frac{d}{dt} (\mathbf{W}^T \cdot \mathbf{I}) \cdot n \bar{\mathbf{e}} + \frac{d}{dt} (\ln \mathbf{S}^J \cdot \mathbf{I}) \cdot n \bar{\mathbf{e}} = n-2 \bar{\boldsymbol{\mu}}, \quad (21)$$

где $\bar{\mathbf{f}} = \int_V \rho \mathbf{f}_m dV + \int_{\Sigma} \mathbf{t}_{\Sigma} d\Sigma$ и $n-2 \bar{\boldsymbol{\mu}} = \int_V \rho \mathbf{x} \times \mathbf{f}_m dV + \int_{\Sigma} \mathbf{x} \times \mathbf{t}_{\Sigma} d\Sigma$. Дополняя (19)–(21) начальными условиями

$$\mathbf{x}_0 = \mathbf{x}_0^0, \mathbf{v}_0 = \mathbf{v}_0^0, t = 0: n-2 \bar{\mathbf{m}} = n-2 \bar{\mathbf{m}}_0, \quad (22)$$

получаем задачу Коши, описывающую поступательное движение центра тяжести кластера и вра-

щательное движение кластера относительно центра тяжести с учетом его деформирования.

Уравнения вращения деформируемого кластера в подвижном базисе

Дифференцируя произведения тензоров $\mathbf{W}^T \cdot \mathbf{I}$ и $\ln \mathbf{S}^J \cdot \mathbf{I}$ согласно правилу дифференцирования [5] тензоров в подвижном базисе \bar{e}'_i

$$\frac{d}{dt} (\mathbf{W}^T \cdot \mathbf{I}) = \mathbf{W}^{TJ} \cdot \mathbf{I} - \mathbf{W} \cdot \mathbf{W}^T \cdot \mathbf{I} + \mathbf{W}^T \cdot \mathbf{I} \cdot \mathbf{W}, \quad (23)$$

$$\frac{d}{dt} (\ln \mathbf{S}^J \cdot \mathbf{I}) =$$

$$= \ln \mathbf{S}^{JJ} \cdot \mathbf{I} - \mathbf{W} \cdot \ln \mathbf{S}^J \cdot \mathbf{I} + \ln \mathbf{S}^J \cdot \mathbf{I} \cdot \mathbf{W}, \quad (24)$$

где обозначена двукратная производная Яуманна

$$\begin{aligned} \ln \mathbf{S}^{JJ} &= \sum_{\gamma=1}^n \frac{d^2}{dt^2} (\ln S_{\gamma} G_{\gamma}^k G_{\gamma}^j) \bar{e}'_k \otimes \bar{e}'_j = \\ &= \sum_{\gamma=1}^n \frac{d^2 \ln S_{\gamma}}{dt^2} G_{\gamma}^k G_{\gamma}^j \bar{e}'_k \otimes \bar{e}'_j, \end{aligned} \quad (25)$$

и, учитывая, что $(\mathbf{W}^T \cdot \mathbf{I} \cdot \mathbf{W}) \cdot n \bar{\mathbf{e}} = 0$, после подстановки (23) и (24) в (21), получаем следующее уравнение:

$$\begin{aligned} &(\mathbf{W}^{TJ} \cdot \mathbf{I} - \mathbf{W} \cdot \mathbf{W}^T \cdot \mathbf{I}) \cdot n \bar{\mathbf{e}} + (\ln \mathbf{S}^{JJ} \cdot \mathbf{I}) \cdot n \bar{\mathbf{e}} - \\ &- (\mathbf{W} \ln \mathbf{S}^J \cdot \mathbf{I} - \ln \mathbf{S}^J \cdot \mathbf{I} \cdot \mathbf{W}) \cdot n \bar{\mathbf{e}} = n-2 \bar{\boldsymbol{\mu}}. \end{aligned} \quad (26)$$

Умножая левую и правую часть этого уравнения $(n-2)$ -кратным скалярным умножением на тензор Леви-Чивиты $n \bar{\mathbf{e}}$ тем же способом, что использован в работе [3], получим окончательный вид уравнения, описывающего вращение деформируемого кластера в подвижном базисе

$$\begin{aligned} \mathbf{I} \mathbf{W}^J + \mathbf{W}^J \mathbf{I} &= \mathbf{W}^2 \mathbf{I} - \mathbf{I} \mathbf{W}^2 + \\ &+ \mathbf{W}_S \mathbf{W} - \mathbf{W} \mathbf{W}_S + \bar{\boldsymbol{\mu}}_S + \bar{\boldsymbol{\mu}}, \end{aligned} \quad (27)$$

где обозначены кососимметричные тензоры

$$\mathbf{W}_S = \mathbf{I} \ln \mathbf{S}^J - \ln \mathbf{S}^J \mathbf{I}, \bar{\boldsymbol{\mu}}_S = \mathbf{I} \ln \mathbf{S}^{JJ} - \ln \mathbf{S}^{JJ} \mathbf{I}; \quad (28)$$

$$\begin{aligned} \bar{\boldsymbol{\mu}} &= \left(\int_V \rho (\tilde{\mathbf{x}} \otimes \mathbf{f}_m - \mathbf{f}_m \otimes \tilde{\mathbf{x}}) dV + \right. \\ &\left. + \int_{\Sigma} (\tilde{\mathbf{x}} \otimes \mathbf{t}_{\Sigma} - \mathbf{t}_{\Sigma} \otimes \tilde{\mathbf{x}}) d\Sigma \right). \end{aligned} \quad (29)$$

Для частного случая, когда отсутствуют деформации кластера, тензоры \mathbf{W}_S и $\bar{\boldsymbol{\mu}}_S$ равны нулю (поскольку $\mathbf{S} = \mathbf{E}$), и уравнение (27) в точности совпадает с уравнением вращения жесткого кластера [3].

В компонентах в базисе \bar{e}'_i система уравнений (19), (20) и (27) записывается следующим образом

(с учетом диагональности матрицы инерции в подвижном базисе $I_\alpha^i = I_\alpha \delta_\alpha^i$):

$$M \frac{dv_{0i}}{dt} = \bar{f}_i,$$

$$\frac{dx_{0i}}{dt} = v_{0i};$$

$$\begin{aligned} \frac{dW^{\alpha\beta}}{dt} &= \frac{I_\alpha - I_\beta}{I_\alpha + I_\beta} \sum_{k=1}^n W^{\alpha k} W^{\beta k} + \\ &+ \frac{1}{I_\alpha + I_\beta} (W^{\alpha k} W_S^{k\beta} - W_S^{\alpha k} W^{k\beta} + \tilde{\mu}_S^{\alpha\beta} + \tilde{\mu}^{\alpha\beta}); \\ \frac{dQ^{ij}}{dt} + Q_k^j W^{kj} &= 0, \quad \alpha, \beta = 1 \dots n, \end{aligned} \quad (30)$$

где

$$\tilde{\mu}_S^{\alpha\beta} = (I_\alpha - I_\beta) \sum_{\gamma=1}^n \frac{d^2 \ln S_\gamma}{dt^2} G_\gamma^\alpha G_\gamma^\beta; \quad (31)$$

$$W_S^{\alpha\beta} = (I_\alpha - I_\beta) \sum_{\gamma=1}^n \frac{d \ln S_\gamma}{dt} G_\gamma^\alpha G_\gamma^\beta.$$

Начальные условия к системе (30) имеют вид

$$\begin{aligned} t = 0: x_{0i} &= x_{0i}^0, \quad v_{0i} = v_{0i}^0, \\ Q^{ji} &= \dot{Q}^{ji}, \quad W^{ij} = \dot{W}^{ij}, \end{aligned} \quad (32)$$

где обозначены компоненты тензора $\tilde{\mu}_S = \tilde{\mu}_S^{ij} \bar{e}_i \otimes \bar{e}_j$ и тензора $W_S = W_S^{ij} \bar{e}_i \otimes \bar{e}_j$.

Сравнивая системы уравнений (30) для деформируемого кластера и для жесткого кластера [3], заключаем, что движение этих кластеров принципиально отличается за счет наличия в системе уравнений вращения кластера (30) линейных по $W^{\alpha k}$ слагаемых и свободного члена $\tilde{\mu}_S^{\alpha\beta}$, который отличен от нуля, даже при отсутствии внешних воздействий на кластер.

Модель внешних сил

Для плотности внешней силы f_m в работе [3] была предложена потоковая модель, согласно которой

$$\begin{aligned} f_m &= \frac{dh^s}{dt}, \quad h^s = h_0^s (1 + Ae^q), \\ A &= A_0 \exp(-|\tilde{x}|^2/x_r^2), \quad t_\Sigma = 0, \end{aligned} \quad (33)$$

где h_0^s — вектор стабильного процесса покупок; A_0, q — константы, характеризующие изменение потока покупок вследствие внешних воздействий

на кластер; x_r — константа, характеризующая неравномерность распределения динамики финансового запаса в кластере (предполагается, что это распределение имеет нормальный закон). В стабильном процессе

$$e = \text{const} \text{ и } h^s = \text{const}, \text{ поэтому } f_m = 0.$$

Вычислительные алгоритмы

Основное предназначение разработанной модели — прогнозирование динамики кластеров во времени. В качестве исходных данных при этом задается положение кластера в некоторый начальный момент времени. Алгоритм типовой иерархической объединяющей кластеризации описан в работе [1], в данном случае за меру межкластерного расстояния было взято расстояние между центрами масс кластеров. Поскольку для разделения на кластеры используются данные не только по накопленным покупкам, но и по доходу покупателей, а также по ряду дополнительных характеристик, можно сказать, что такой подход совмещает преимущества определенных в работе [2] математической таксономии, анализирующей только прямо относящиеся к задаче данные, и концептуальной кластеризации, учитывающей ряд дополнительных характеристик объектов.

Для аппроксимации данных применяется модель многомерного эллипсоида $\sum_{\alpha=1}^n \left(\frac{X^\alpha}{R^\alpha}\right)^2 = 1$ в

лагранжевых координатах, где R^α — полуоси эллипсоида, которые можно найти путем решения задачи минимизации объема кластера с ограничениями на вхождение в эллипсоид не менее 95 % точек, относящихся к кластеру. Векторы базиса $\bar{e}_j^i(0) = Q_j^i(0)e_i$ в начальный момент определяют путем построения матрицы инерции $I^{ij} = \int_V \rho \tilde{x}^i \tilde{x}^j dV$ по экспериментальным точкам и нахождения ее собственных векторов. Для вычисления многомерных интегралов применяли четыре метода, описанные в работе [3]: метод n -мерных шаров; метод равного веса точек; метод нормального распределения и метод локальных групп. Результат разделения покупателей на кластеры и аппроксимация полученных кластеров эллипсоидами показаны на рис. 1, где эллипсы представляют сечения эллипсоидов плоскостью (x^3, x^4) , темные фигуры — точки, попадающие в эллипсоид, а светлые — не попадающие.

Численное решение задачи Коши (30), (32) осуществлялось на основе пошагового разностного метода вместе с процедурой линеаризации на

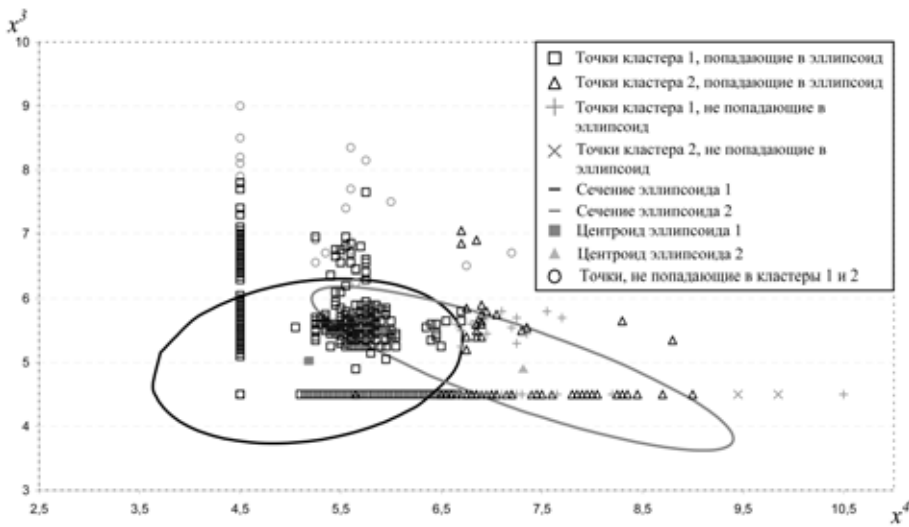


Рис. 1. Деление точек на кластеры и аппроксимирующие эллипсоиды для 8-го периода, рассчитанные методом n -мерных шаров

временных шагах и использованием неявной разностной схемы для уравнений вращения:

$$(I_\alpha + I_\beta) \frac{W_z^{\alpha\beta} - W_{z-1}^{\alpha\beta}}{\Delta t} = (I_\alpha - I_\beta) \times$$

$$\times \left(\sum_{k=1, k \neq \beta}^n W_{z-1}^{\alpha k} W_{z-1}^{\beta k} + W_z^{\alpha\beta} W_{z-1}^{\beta\beta} \right) +$$

$$+ \sum_{k=1, k \neq \beta}^n W_{z-1}^{\alpha k} W_{S_z}^{k\beta} + \sum_{k=1, k \neq \alpha} W_{S_z}^{\alpha k} W_{z-1}^{k\beta} +$$

$$+ W_z^{\alpha\beta} W_{S_z}^{\beta\beta} - W_{S_z}^{\alpha\beta} W_z^{\beta\beta} + \mu_S^{\alpha\beta} + \tilde{\mu}^{\alpha\beta}, \quad (34)$$

где $W_z^{\alpha\beta} = W^{\alpha\beta}(t_z)$ — значения функций в узлах разностной сетки.

Результаты численного анализа данных

Разработанная модель деформируемого кластера была применена для анализа экспериментальных данных, рассмотренных в работе [3], по продажам автомобилей через Интернет-магазин в течение 30 мес. В качестве осей координат x^i пространства E_n были введены суммарные значения купленных или планируемых к покупке автомобилей пяти ценовых классов, поэтому размерность n пространства товаров E_n была равна 5. Данные анализировались через 3 мес в моменты времени t_m , $m = 1 \dots 11$. Моменты времени t_4 и t_6 были идентифицированы как первая и вторая волна финансового кризиса, который выражался в изменении экономического запаса e покупателей согласно модели (31).

Для модели деформируемого кластера, в отличие от жесткого кластера, необходимо задать значения констант A_γ в формуле (7), характеризующих изменение размеров эллипсоида в зависимости от

изменения финансового запаса покупателей. Для этой цели были использованы экспериментальные данные по покупкам еще для одного момента времени (для разных кластеров использовали различные контрольные точки), а динамика изменения финансового запаса кластера $e(t) - e_0$ была взята совпадающей с динамикой доходов покупателей каждого кластера.

Исследование движения центров масс кластеров показало, что оно носит линейный характер по всем осям товаров x^i в пространстве E_5 (рис. 2) при отсутствии внешних сил. Наличие внешних воздействий (кризисных явлений) приводит к изменению углов наклона линейной траектории движения кластеров.

Модель (30) хорошо прогнозирует движение центра масс кластера как в отсутствие внешних воздействий, так и при их наличии.

На рис. 3 показано движение деформируемого кластера на плоскости (x^3, x^4) для нескольких моментов времени. Квадратами на рисунке обозначены точки, попадающие внутрь аппроксимирующего эллипсоида, а плюсами — точки, не попадающие.

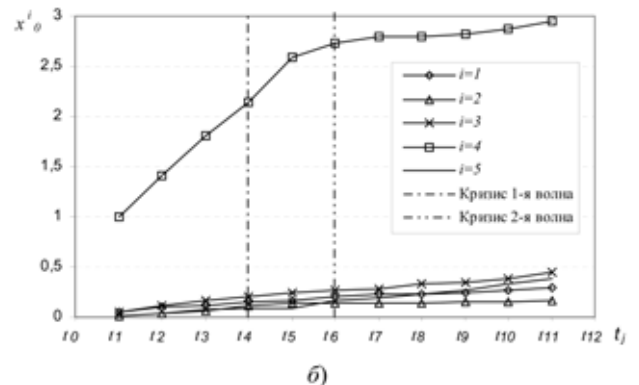
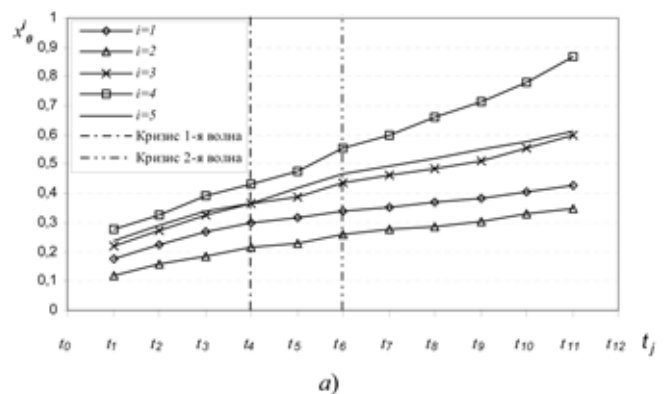


Рис. 2. Динамика центра масс кластеров 1 (а) и 2 (б)

дающие в него. Видно, что происходит достаточно существенное изменение длин полуосей кластера (деформация) в процессе его движения.

Вращательная динамика кластера, представленная значениями независимых компонент матрицы поворота $Q_i^j(t)$, рассчитанных по уравнениям (30), показана на рис. 4. Используются следующие обозначения: $Q^{12} = Q_1$, $Q^{13} = Q_2$, $Q^{14} = Q_3$, $Q^{15} = Q_4$, $Q^{23} = Q_5$, $Q^{24} = Q_6$, $Q^{25} = Q_7$, $Q^{34} = Q_8$, $Q^{35} = Q_9$, $Q^{45} = Q_{10}$. Было установлено, что изменение компонент матрицы поворота $Q_i^j(t)$ подвержено колебательным движениям, задаваемым с помощью закона движения (30), начальных данных по повороту $Q_i^j(0)$, скорости вращения $W_i^j(0)$ и инерции $I_i^j(0)$. Полученная в результате моделирования динамика вращения кластера схожа с динамикой, рассчитанной по экспериментальным данным.

На рис. 5 приведены длины полуосей эллипсоида, полученного с помощью модели деформируемого кластера по изменению финансового запаса покупателей, в сравнении с длинами полуосей эллипсоида, аппроксимирующего экспериментальные данные в каждом периоде. Видно, что максимальные значения длин моделируемого эллипсоида не превосходят максимальных значений для эллипсоида, построенного методом оптимизации.

Рис. 6 демонстрирует графики, показывающие процент попадающих точек в деформируемый n -мерный эллипсоид для разных моментов времени и при разных способах аппроксимации начальных данных $Q_i^j(0)$ и $I_i^j(0)$. Сплошные линии — данные, полученные по модели деформируемого кластера, а штриховые линии — полученные с помощью аппроксимации экспериментальных данных эллипсоидом для каждого момента времени, без решения системы (30) (эти результаты названы "эксперимент"). Кластер 1, включающий большее число точек и, таким образом, меньше реагирующий на единичные изменения финансового запаса отдельных покупателей, показывает большую точность (до 96 % в последней контрольной точке t_{11}) и небольшое расхождение в точности для различных методов. В точке t_3 кластер 1 демонстрирует скачок точности вследствие небольшого увеличения длины полуосей, вызванного скачком финансового запаса.

Следует также отметить, что точность модели деформируемого кластера выше, чем точность, получаемая по модели жестких кластеров, описанной в работе [3] (максимальная точность для

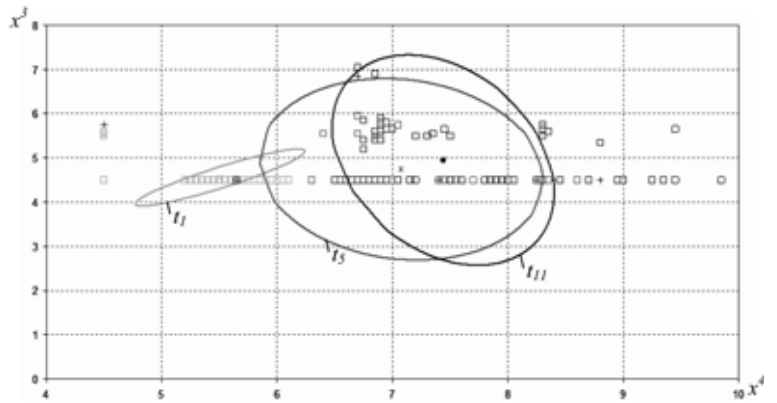


Рис. 3. Движение кластера 2 во времени (в контрольных точках t_1 , t_5 и t_{11}), рассчитанное методом n -мерных шаров

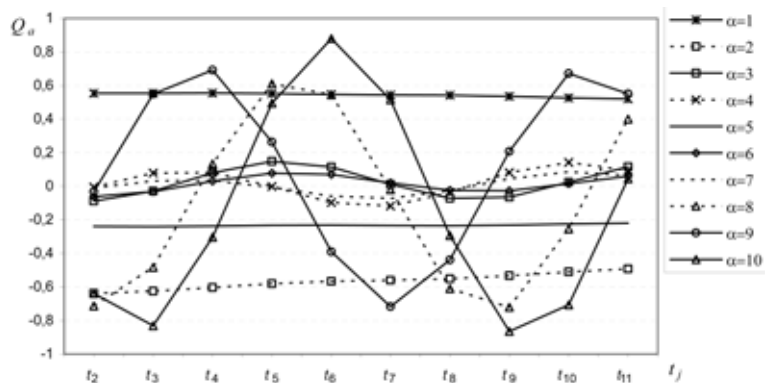


Рис. 4. Динамика вращения кластера 1 (независимые компоненты матрицы Q_α , рассчитанные по модели деформируемого кластера с использованием метода равного веса)

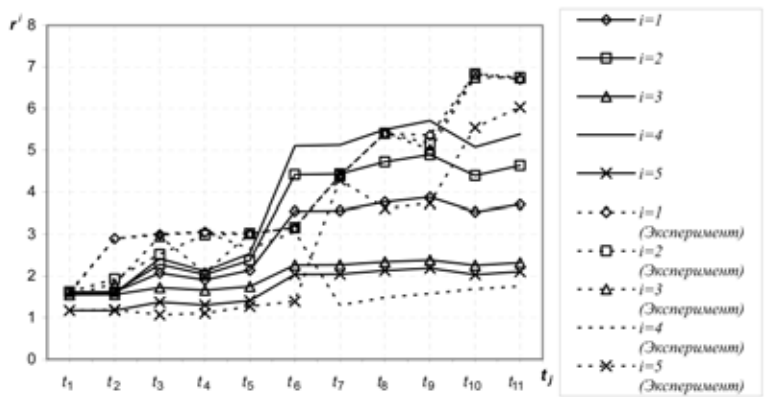


Рис. 5. Длины полуосей эллипсоида, построенного по модели деформируемого кластера методом равного веса в сравнении с длинами полуосей эмпирического эллипсоида для кластера 1

последней контрольной точки составляла 85,8 для кластера 1). Однако модель жесткого кластера требует меньшее количество входной информации — не требуется вводить в модель данные о финансовом запаса покупателей и его влиянии на изменение числа покупок. В тех случаях, когда такая информация недоступна, модель жестких кластеров может оказаться предпочтительной. Для более точного прогнозирования динамики данных ре-

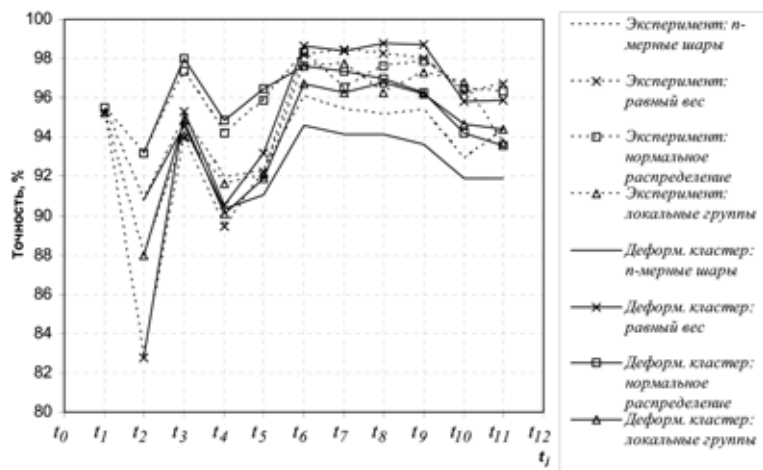


Рис. 6. Динамика точности модели деформируемого кластера для кластера 1

комендуется модель деформируемых кластеров, разработанная в настоящей работе.

Выводы

Разработана модель деформируемых кластеров, основанная на применении методов механики многомерных сплошных сред для решения задач анализа и прогнозирования динамики больших массивов плохо структурированных экономических данных. Предложены численные алгоритмы решения систем нелинейных дифференциальных уравнений, возникающих в этой модели. Представлен пример анализа и прогнозирования экс-

периментальных данных рынка продаж, показано, что разработанная модель деформируемых многомерных эллипсоидов позволяет достаточно хорошо описывать динамику движения кластеров на всем рассматриваемом диапазоне времени, включая этапы воздействия экономических кризисных явлений на покупательский рынок. Модель обладает достаточно высокой точностью прогнозирования динамики данных (до 96 %), что превышает точность предложенной ранее модели недеформируемых жестких кластеров, однако для ее применения необходима дополнительная информация о динамике финансового запаса покупателей.

Список литературы

1. Kojadinovic I. Agglomerative hierarchical clustering of continuous variables based on mutual information // Elsevier Science B. V. Computational Statistics & Data Analysis. 1 June 2004. Vol. 46. Is. 2. P. 269—294.
2. Shaw M. J., Subramaniam C, Tan G. W., Welge M. E. Knowledge management and data mining for marketing // Elsevier Science B. V. — Decision Support Systems. 2001. Vol. 31. P. 127—137.
3. Дмитриенко Ю. И., Дмитриенко О. Ю. Кластерно-континуальное моделирование в экономике на основе методов механики многомерных сплошных сред // Информационные технологии. 2010. № 8. С. 54—62.
4. Дмитриенко Ю. И. Тензорное исчисление. М.: Высшая школа, 2001. 575 с.
5. Дмитриенко Ю. И. Нелинейная механика сплошной среды. М.: Физматлит, 2009. 624 с.

УДК 004.94:65.012.122(664.7)

В. О. Новицкий, канд. техн. наук, докторант,

В. И. Карпов, д-р техн. наук, проф.,

Московский государственный университет пищевых производств,

e-mail: nvob60@mail.ru

Методология исследования и моделирования сложных систем управления для предприятий и компаний зернового сектора АПК

Необходимость совершенствования систем управления для зерновых и зерноперерабатывающих предприятий и компаний на основе интегрированных автоматизированных информационно-управляющих систем определяет потребность в специализированной методологии исследования и моделирования для данного класса сложных систем управления. Предложенная методология основана на использовании диалектического, системного, кибернетического и других подходов и предполагает описание системы, формирование структуры целей, построение и анализ диаграмм причинно-следственных связей, выявление проблемобразующих задач, составление концептуальной модели и формирование математической модели системы на основе взаимодействия с банком моделей управления для предприятий зернового сектора АПК.

Ключевые слова: методология исследования и моделирования, сложная система управления, структура целей, диаграмма причинно-следственных связей, контуры обратных связей, концептуальная модель, критерий, банк моделей, зерновые и зерноперерабатывающие компании

Введение

Существует несколько основных методологических подходов к исследованию систем управления [1, 2]: диалектический (причинно-следствен-

ный), системный, процессный, кибернетический, функциональный, ситуационный, рефлексивный. Все известные методы и технологии исследования и моделирования сложных организационно-тех-

нических систем управления так или иначе базируются на использовании некоторой конфигурации методологий и подходов.

В методологии постановки задач управления и их моделирования важное место занимает процесс выявления и анализа проблем, в котором выделяются следующие этапы:

- выявление и предварительный анализ проблем;
- формирование каталога проблем и анализ их "причинно-следственных связей";
- формирование и анализ каузативной структуры;
- определение и анализ главной проблемы и формирование структуры проблем в виде графа — "дерева" [3].

Одной из наиболее часто применяемых для исследования сложных слабоструктурированных систем и решения слабоформализуемых задач является технология когнитивного моделирования, которой посвящены работы В. И. Максимова, А. А. Кулинича, С. А. Юдицкого [4—6] и др. Отсутствие в данной методологии жестких требований к целостности системы, выявлению системности (обратные связи, типы сущностей, горизонты функций управления), а также отсутствие описаний бизнес-процессов (входы-выходы, состояния, механизмы), не позволяет в полной мере использовать ее для исследования и разработки систем управления предприятий.

Существует и используется методика исследования и моделирования систем управления с позиций баланса потребностей и возможностей предприятия. Одним из способов формализации потребностей и возможностей предприятия на всех уровнях управления является формирование целевых нормативов в стандартных единицах задач (СЕЗ) [7], которое динамически описывает потребности и возможности предприятия и позволяет обосновать эффективность управления в условиях изменений внешней среды.

В последние годы среди российских ученых и практиков в области управления используется пришедшее с Запада понятие *контроллинг* [8].

Многие проблемы предприятий, в частности, зерновых и зерноперерабатывающих компаний (ЗПК), упираются в общую проблему недостаточной управляемости в современных рыночных условиях, связанную с отсутствием комплексной интегрированной системы управления, к которой предъявляются следующие основные требования:

- отвечать единой главной цели предприятия и компании как системы; остальные цели системы должны быть с ней согласованы;
- опираться на формализованные модели производственно-хозяйственной деятельности предприятия и компании;
- адаптироваться к условиям неопределенности;

- обеспечивать весь цикл функций управления: контроль, учет, планирование, регулирование, организация, координация по всем необходимым временным периодам и на любом уровне;
- быть взаимосвязанной по всем звеньям логистической цепочки предприятия и учитывать все значимые возмущения от активных элементов системы и среды;
- обеспечивать многокритериальное управление.

Актуальность проблемы определяется также отсутствием эффективных методов поддержки и автоматизации этапа концептуально-логического моделирования систем управления данного класса.

Анализ специфики систем управления предприятий и компаний зернового сектора АПК и существующих методологий исследования и моделирования сложных систем позволяет сделать вывод о необходимости создания на их базе специализированной методологии (комплекса специальных методов), предназначенной для исследования, постановки задач и моделирования сложных систем управления для широкого класса производственных предприятий АПК. Многие из существующих фундаментальных и прикладных подходов и методов исследования и моделирования систем с учетом их критического содержательного анализа могут быть использованы для синтеза специализированной методологии, обеспечивающей эффективное исследование и моделирование производства и процессов принятия решений для интегрированных автоматизированных систем управления (ИАСУ) ЗПК.

Для построения ИАСУ ЗПК и реализации ее интегративного свойства необходимо исследовать все элементы, процессы, механизмы управления существующей системы на предприятиях данного класса. В предлагаемой специализированной методологии исследования взаимосвязанно используются системный, причинно-следственный, кибернетический, процессный, функциональный и ситуационный подходы. Вследствие требований общей постановки задачи, актуальности практического приложения и значительной специфики класса объектов данная методология отличается от существующих методов и подходов к исследованию и моделированию ложных систем. Характерные черты этой методологии:

- проблема имеет прямое отношение к логистике как научной дисциплине, поэтому должна учитывать соответствующие методы, используемые в данной прикладной науке обо всех видах обеспечения;
- решение проблемы предполагает использование наряду с большим числом формальных параметров и аналитических методов также эмпирических и слабоструктурированных методов и элементов;

- решение задач управления лежит в областях математического программирования и задач принятия решений с использованием языков многокритериального выбора и бинарных отношений;
- исследование должно охватывать весь цикл функций управления, поскольку создаваемые на его основе системы управления и их компоненты относятся ко всему спектру классов информационно-управляющих и систем интеллектуальной обработки данных;
- необходима доступность и наглядность в применении в целях привлечения к созданию моделей систем наряду со специалистами-аналитиками также профессионалов-практиков для решения задач широкого диапазона абстракции, сложности и ответственности.

Разработка методологии исследования, постановки задач и моделирования для сложных систем управления производством

Решение целевых задач чисто формальными методами приводит к их локальности, что ограничивает возможности систем управления. Это относится ко всем этапам их жизненного цикла и требует обобщающих, систематизирующих методов исследования и построения сложных автоматизированных систем с переносом общепрофессиональных концепций в область конкретного практического применения в зерновом и зерноперерабатывающем секторе АПК.

Разрабатываемая методология исследования сложных систем управления предлагает взаимосвязанный комплекс (систему) эффективных методов поддержки всех основных стадий концептуально-логического проектирования автоматизированных информационных систем управления для всех уровней и функций цикла управления для ЗПК с применением как формальных, так и слабоформализованных агрегированных параметров совместно с концепцией формирования банка моделей [9, 10].

Методология исследования сложных систем управления в приложении к ЗПК представлена в виде алгоритма на рис. 1.

Описание предметной области (блок 1) при формулировании проблемы и проблематики проводится с использованием подходов Business Performance Management (BPM) (управления результативностью бизнеса [11]), нацеленных на по-

строение системы поддержки принятия решений. Изучение предметной области при этом сопровождается оценкой принадлежности системы к тому или иному классу в классификациях систем с позиций управления [12] и кибернетического подхода [13] и целесообразности применения предлагаемой методологии (блоки 2—3). С этих позиций предлагается изначально определять главную цель и обусловливаемые ею основные компоненты системы с последующей детализацией. На основе собранной первичной информации определяются основные элементы системы и связанная

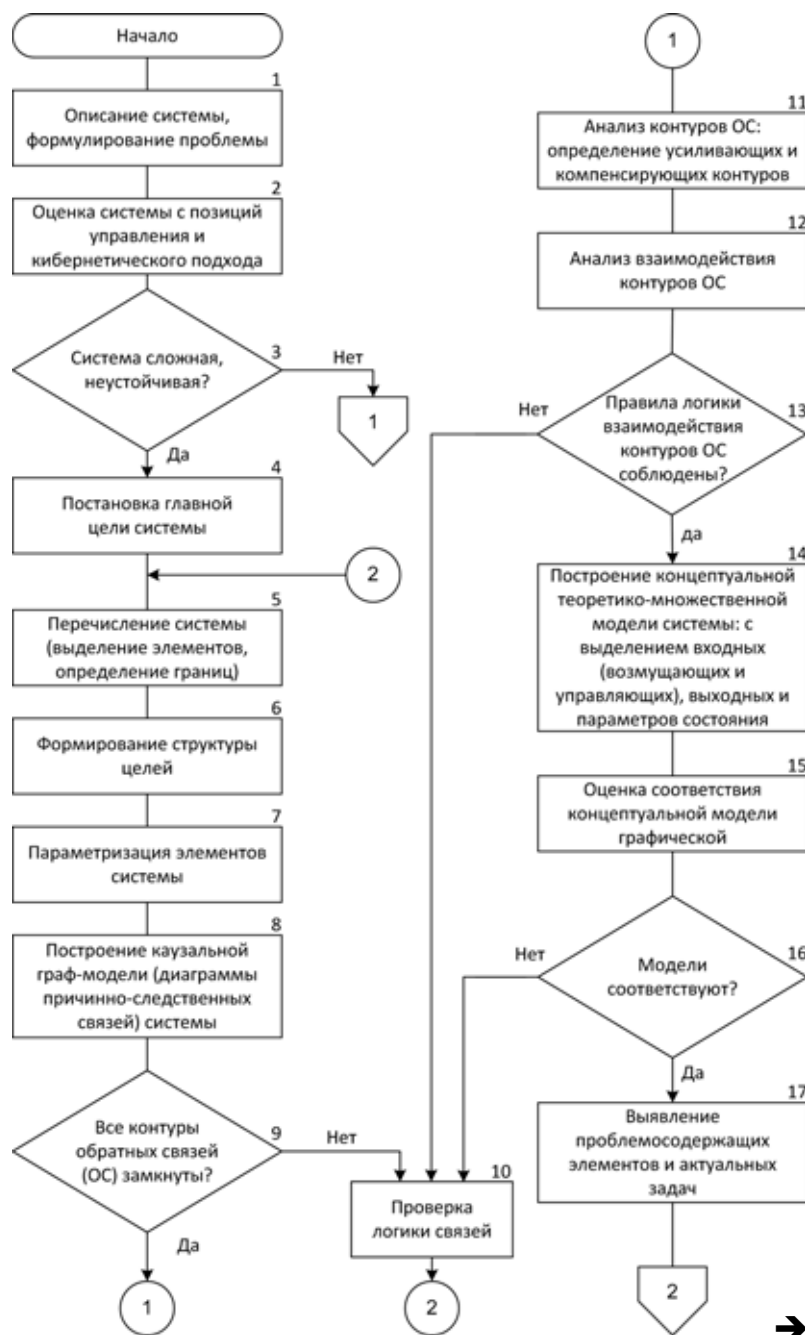


Рис. 1. Алгоритм методологии исследования и предпроектного моделирования систем управления для ЗПК (начало)

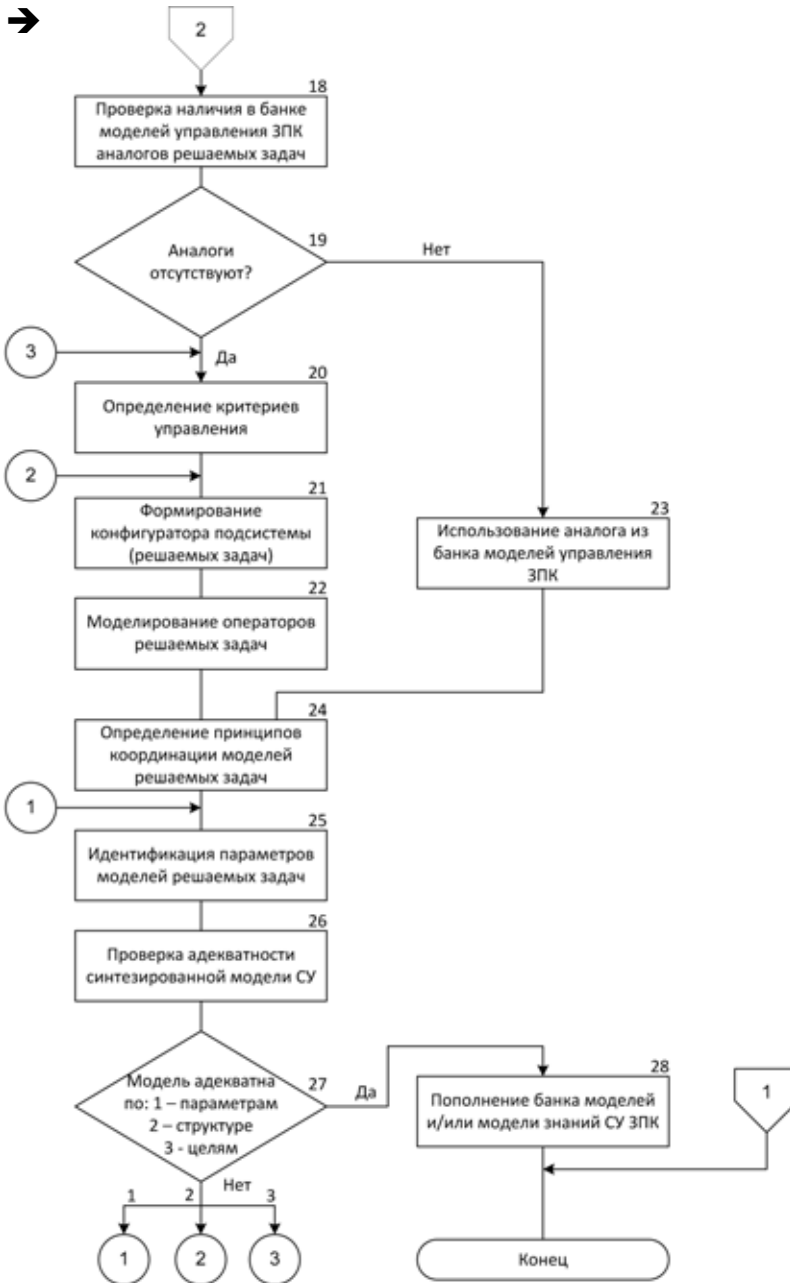


Рис. 1. Алгоритм методологии исследования и предпроектного моделирования систем управления для ЗПК (окончание)

с ними структура целей (блоки 4—6). Элементы целесообразно представить в виде параметров, поддающихся сравнению (блок 7).

Для удобства оценки и поиска решений комплекса взаимосвязанных задач системы строится ее систематизированный визуальный образ в виде графа каузальной модели (диаграммы причинно-следственных связей (ДПСС), блок 8).

Техника построения ДПСС имеет ряд общих правил с методикой когнитивного моделирования: вершинами графа являются элементы-параметры; дуги представляют собой положительные или отрицательные причинно-следственные свя-

зи между элементами. Но есть и существенные отличия. Так, когнитивные технологии предназначены для синтеза решений на слабоструктурированном пространстве параметров проблемы в условиях возмущений (изменений параметров или структуры графа) путем моделирования импульса на когнитивной карте, и поэтому в них часто используют значения экспертных оценок (значения коэффициентов на дугах и т. п.), с которыми проводятся различные операции. ДПСС же предназначены для исследования контуров обратных связей в системе управления в целях выявления актуальных задач, требующих решения. Поэтому дуги используются только для качественного определения причинно-следственных зависимостей (важна наглядность).

Важным свойством ДПСС является обязательное определение на дугах возможных последствий, имеющих качественное содержание. Таким образом, важное значение придается временным горизонтам управления, что свидетельствует об уровне (функции) управления по данным контурам (связям) и их значимости.

Наличие задержек в ДПСС влияет на содержательные свойства вершин графа, т. е. его вершины могут представлять собой объекты различного, в том числе ненулевого, уровня — процессы (изменение качества объекта во времени) [14]. В ДПСС предлагается использовать разнообразие типов элементов-параметров, отражающее их динамику и определяемое как уровни объектов:

- нулевого уровня — "событие", или "начало", отражающее исходное свойство объекта (исходная масса);
- первого уровня — процесс, отражающий изменение (накопление) исходного свойства ("уровня") объекта (накопление массы);
- второго уровня — процесс изменения свойства объекта (расход компонента) — "поток".

Если рассматривать исследование системы управления как ее описание в стационарном режиме и без изучения морфологии преобразования сырья в продукцию, то достаточно использовать три типа объектов. Если же включать в рассмотрение процессы реструктуризации системы или элементов (процесс проектирования системы или трансформацию объектов), то потребуется еще

один тип элемента — объект четвертого уровня — "трансформация".

"Начало" используется для отображения характерных состояний системы в контурах обратных связей (ОС), имеющих важное значение для понимания функционирования и развития системы. "Уровень" используется для фиксации количества накапливающегося вещества, энергии и других субстанций в определенных местах контуров ОС. "Поток" используется для отображения движения (скорость, расход, коэффициенты изменения и т. п.). "Трансформация" может использоваться для отображения преобразования ресурсов (сырья в продукцию, информации в знания) или структуры системы (связей) при масштабировании системы в сторону надсистемы.

Контуров могут быть усиливающие или уравновешивающие. Их анализ является исходным посылом для выявления актуальных задач.

Необходимо отметить, что приводимые в ДПСС обозначения взяты из работы И. Н. Дрогобыцкого [13], который использовал системные диаграммы для выявления системных паттернов. Однако методики их использования для анализа и моделирования систем не было предложено. Тем не менее, в условиях отсутствия какой-либо стандартизации данная нотация весьма удобна и наглядна для представления ДПСС (с учетом ее последующей автоматизации).

На основе предлагаемой методики проводится анализ системы (блоки 9—16): оцениваются ее границы, определяются факторы влияния надсистем (элементы, на которые не замыкаются контуры обратных связей), выявляются усиливающие и компенсирующие контуры по различным временным горизонтам управления (планирования, регулирования, контроля и др.). На основе конфигурации усиливающих и компенсирующих контуров ОС в системе управления оцениваются правильность соблюдения логики построения системы, а также ее устойчивость и направления возможного развития. Так, наличие в системе только усиливающих контуров говорит о неустойчивости системы: либо пойдет вразнос, либо деградирует, но в любом случае потребуются значительные ресурсы для управления, чтобы заставить систему функционировать в рамках заданной траектории параметров состояния. Это относится и к процессу функционирования системы, и к ее развитию. В процессе анализа ДПСС оценивается правильность ее построения, что говорит об уровне понимания аналитиком как предметной, так и проблемной областей системы, определяются проблемнообразующие элементы и формулируются соответствующие актуальные решаемые задачи и их взаимосвязь (блок 17).

Взаимосвязь решаемых задач должна отражаться на структуре целей, при построении которой может корректироваться ДПСС, включая состав элементов и топологию связей. Таким образом, процесс построения каузальной граф-модели и структуры целей системы является итерационным. Концептуальную модель (постановку задачи), являющуюся результатом проводимого анализа при создании систем или решении проблем управления для ЗПК, целесообразно формализовать в теоретико-множественном представлении (ТМП).

Формальные модели многих актуальных задач управления для ЗПК являются структурно-статическими (составление начальных планов сезонного приема и размещения зерна на элеваторе, планирование торговых операций с зерном и т. д.) или структурно-динамическими (СД): оперативное планирование и регулирование помольных партий, обработки и размещения зерна на хлебоприемных предприятиях, прогнозирование объемно-стоимостных параметров рынка зерна и др. Таким образом, общая (базовая) модель системы управления для ЗГЖ должна рассматриваться как СД.

ТМП СД систем подразумевает определение множеств входных возмущающих, входных управляемых параметров (элементов), состояний и результатов управлений, а также множество операторов преобразования состояний и формирования результатов, определяющих множество решений. При этом результирующие показатели представляют собой интересующее нас подмножество параметров состояния на заключительном этапе рассматриваемого интервала времени функционирования или развития системы.

В зависимости от того, отображаются промежуточные состояния системы строго определенной последовательностью $C_t (t = 0, 1, 2, \dots, T)$ или одной неопределенной функцией (оператором при многомерном отображении) $C_t = \Phi(t, x(t))$, в результате моделирования можно получить структурно-динамические модели либо сетевого, либо аналитического типа. Концептуальную СД модель аналитического типа можно представить следующим образом:

$$\mu : X \times T \rightarrow C, \eta : C \times T \rightarrow Y; \quad (1)$$

$$y(t) = \eta(t, C_t), t \in [0, T];$$

$$X = \{x(t)\} = \{u(t), v(t)\}, u(t) \in U, v(t) \in V;$$

$$C_t = (c_1(t), c_2(t), \dots, c_n(t)), t \in [0, T],$$

$$C_t = \mu(t, x(t)), t \in [0, T]; C_t \subset C;$$

$$Y = \{y(t)\}, t \in [0, T],$$

где μ — отображение (оператор) множества X входов на всем интервале времени (горизонта T) управления во множество состояний C ; η — отображение множества состояний на горизонте

управления во множество выходов (результатов) Y ; U — множество управляемых входов $u(t)$; V — множество неуправляемых входов $v(t)$; C_t — промежуточное состояние системы в момент времени $t \in [0, T]$.

На множестве операторов определяются задачи, которые являются актуальными, проблемообразующими. В зависимости от конкретизации множеств X, C, Y , отображений μ и η и интервалов $[0, T]$ можно получить аналитические или слабоформализуемые дискретные или непрерывные модели.

Результат анализа — концептуальная модель — имеет целью построение моделей более детального — логического уровня для последующей идентификации параметров этих моделей на стадии синтеза системы.

Для получения моделей логического уровня предлагается использовать понятие банка моделей (БМ) предметной области. Это обусловлено наличием общей отраслевой специфики, которая может быть реализована в виде определенным образом структурированных типовых моделей процессов (критерии, условия, связи). В работе это — БМ задач управления ЗПК.

БМ предназначен для синтеза систем управления ЗПК и позволяет:

- использовать заранее подготовленные формальные модули (математические модели и БД классификаторов) для решения проблемных задач;
- на логическом уровне конструировать системы из типовых компонентов управления для ЗПК;
- исследовать поведение систем при разных стратегиях в различных условиях функционирования и развития.

БМ реализует функции агрегирования моделей компонентов в модели систем и является агрегативным. Агрегат представляется унифицированной моделью для описания функционирования разнородных элементов системы. Поэтому динамика большой, сложной системы управления раскрывается через динамику взаимосвязанных между собой агрегативных моделей компонентов [15]. Интегративные свойства систем при агрегировании моделей отражаются в результатах поиска управленческих решений. В процессе проведения исследований и развития систем управления БМ ЗПК пополняется новыми моделями, отражающими новые функции и свойства объектов.

Для достижения требуемого уровня адекватности моделей реальным объектам с учетом обязательного испол-

зования субъективного опыта специалистов отраслевых компаний часто приходится разрабатывать собственные методики постепенной формализации моделей систем управления.

На основе рассмотренной методологии для ЗПК строятся модели и реализуется комплекс взаимосвязанных автоматизированных информационно-управляющих систем поддержки принятия решений: "Учет и контроль грузопотоков зерна на элеваторах", "Планирование работы зерновых элеваторов", "Торговля зерном", "Планирование зерновых ресурсов мукомольной компании", "Прогноз стоимостных параметров зерновых рынков" и др.

Исследование подсистемы планирования зерновых ресурсов мукомольной компании

Данная подсистема является относительно простой для отражения сути методологии. Опуская для краткости процесс описания предметной области, формулирования проблемы и проблематики и структуры целей, приведем для их понижения только ДПСС данной подсистемы (рис. 2).

На представленной системной диаграмме можно выделить пять контуров ОС:

1. *Планирование закупки зерна пшеницы 3-го класса по петле "Качество—Спрос"*.

Высокое качество производимой продукции повышает спрос на муку, а также объемы ее производства и реализации, что в свою очередь увеличивает



Рис. 2. Диаграмма причинно-следственных связей подсистемы планирования зерновых ресурсов в мукомольной компании

объем доступных оборотных средств. Это увеличивает закупку зерна пшеницы 3-го класса (качественной и дорогой) и уровень ее запасов. В свою очередь наличие на складе зерна высокого качества снижает расход в смесь зерна пшеницы 4-го класса (более низкого качества и стоимости), что повышает качество помольной смеси и готовой продукции и увеличивает спрос на продукцию предприятия. Контур — усиливающий и ведет к увеличению спроса и повышению качества продукции.

2. *Планирование закупки зерна пшеницы 3-го класса по петле "Цена — Спрос".*

Высокая цена производимой продукции снижает спрос на муку, а также объемы ее производства и реализации, что в свою очередь уменьшает объем оборотных средств. Их понижение сокращает закупку зерна пшеницы 3-го класса и уровень ее запасов. При этом увеличивается расход в смесь зерна низкого качества, что уменьшает стоимость помольной смеси и готовой продукции. Это увеличивает спрос на продукцию. Контур — компенсирующий, стабилизирующий стоимость, качество муки и запасы пшеницы 3-го класса.

3. *Планирование закупки зерна пшеницы 4-го класса по петле "Качество — Спрос".* Высокое качество производимой продукции повышает спрос на муку, а также объемы ее производства и реализации, что в свою очередь увеличивает объем оборотных средств. Их повышение влияет на снижение объема закупки зерна и запасов пшеницы 4-го класса, что увеличивает расход в смесь зерна высокого качества, повышает качество помольной смеси и готовой продукции. Это увеличивает спрос на продукцию. Контур — усиливающий, повышающий качество и спрос на муку, снижающий объем закупки пшеницы 4-го класса.

Подобным образом анализируются контуры:

4. *Планирование закупки зерна пшеницы 4-го класса по петле "Цена — Спрос";*

5. *Планирование реализации и производства.*

Анализ элементов и контуров обратной связи, представленных на рис. 2, показывает, что главное положение в системе планирования зерновых ресурсов предприятия занимает расчет рецептов помольных смесей, управляющий потоком (расходом) зерна различного качества в смесь. На основании рецептов регулируются качество и стоимость продукции, планируются закупки зерна.

Подробнее с применением предлагаемой методологии при реализации подсистем управления для зерноперерабатывающих предприятий и компаний можно ознакомиться в [10].

Заключение

Предложенная методология исследования и моделирования позволяет повысить эффективность

проектирования автоматизированных информационно-управляющих систем для широкого класса предприятий АПК и близких отраслей. При этом охватывается весь цикл функций управления от учета и контроля до планирования и прогнозирования. Создаваемые на основе предложенной методологии системы управления обладают функциями интеллектуального анализа и поддержки принятия решений на уровне от менеджера среднего звена предприятия до руководителя компании.

На основе предложенной методологии исследования и моделирования разработаны и внедрены системы учета и планирования ресурсов мукомольного производства, система поддержки принятия решений по торговле зерном в целом ряде крупных производственно-торговых предприятий и компаний зернового и зерноперерабатывающего сектора АПК России.

Список литературы

1. **Мишин В. М.** Исследование систем управления: Учебник для вузов. 2-е изд. М.: ЮНИТИ-ДАНА, 2008. 527 с.
2. **Фрейдина Е. В.** Исследование систем управления: Учеб. пособие. / Под ред. Ю. В. Гусева. 2-е изд. М.: Омега-Л, 2009. 368 с.
3. **Волкова В. Н., Козловская Э. А.** Сравнительный анализ системно-целевого и бизнес-процессного подходов к исследованию и проектированию систем управления предприятиями и организациями // Системный анализ в проектировании и управлении: Сб. науч. трудов XIII Междунар. науч.-практ. конф. Ч. 1. СПб.: Изд-во Политехн. ун-та, 2009. С. 58—60.
4. **Максимов В. И.** Когнитивные технологии — от незнания к пониманию // Когнитивный анализ и управление развитием ситуаций (CASC'2001). Тр. Междун. конф. 2001. Т. 1. С. 4—18.
5. **Кулинич А. А.** Систематизация когнитивных карт и методов их анализа // Когнитивный анализ и управление развитием ситуаций (CASC'2007) Тр. VII Межд. научно-практич. конф. "Управление большими системами — 2007". М.: Изд-во ИПУ, 2007. С. 50—55.
6. **Юдицкий С. А., Владиславлев П. Н.** Предпроектное моделирование функционирования организационных систем. М.: Научтехлитиздат, 2004. 120 с.
7. **Смирнов А. В.** Системный подход к оценке потребностей и возможностей предприятия // Системный анализ в проектировании и управлении: Сб. науч. тр. XIII Междунар. науч.-практ. конф. Ч. 1. СПб.: Изд-во Политехн. ун-та, 2009. С. 73—75.
8. **Карминский А. М., Оленев Н. И., Примак А. Г., Фалько С. Г.** Контроллинг в бизнесе: методологические и практические основы построения контроллинга в организациях. М.: Финансы и статистика, 1998. 256 с.
9. **Новицкий В. О.** Исследование объектов и постановка задач управления для зерновых и зерноперерабатывающих компаний на основе методологии системного анализа // Системный анализ в проектировании и управлении: Сб. науч. тр. XIII Международной науч.-практ. конф. Ч. 1. СПб.: Изд-во Политехн. ун-та, 2009. С. 311—315.
10. **Новицкий В. О.** Системный подход к управлению на предприятиях и в компаниях по хранению и переработке зерна // Хлебопродукты. 2009. № 7. С. 54—56.
11. **Кокинз Г.** Управление результативностью: Как преодолеть разрыв между стратегией и реальными процессами / Пер. с англ. М.: Альпина Бизнес Букс, 2007. 315 с.
12. **Тимченко Т. Н.** Системный анализ в управлении: Учеб. пособие. М.: РИОР, 2008. 161 с.
13. **Дрогобыцкий И. Н.** Системный анализ в экономике: Учеб. пособие. М.: Финансы и статистика, 2007. 512 с.
14. **Жилин Д. М.** Теория систем: Опыт построения курса. Изд. 4-е, испр. М.: Изд-во ЛКИ, 2007. 184 с.
15. **Денисов А. А., Колесников Д. Н.** Теория больших систем управления: Учеб. пособие для вузов. Л.: Энергоиздат, 1982. 288 с.

УДК 519.23/24

Ю. А. Зак, д-р техн. наук,
Аахен, Федеративная Республика Германия,
e-mail: yuriy_zack@hotmail.com

Математические методы обработки результатов тестирования квалификации и знаний специалистов и абитуриентов

Предлагается решение задач оценки результатов тестирования методами векторной оптимизации. Предложены методы конструирования и вычисления значений комплексного компромиссного критерия оценки результатов тестирования, что позволяет объективно выразить результаты выполнения комплекса тестов с помощью некоторого действительного числа. Даны рекомендации построения тестов.

Полученные в работе результаты могут найти применение при контроле знаний в системах компьютерного обучения, а также при конкурсном приеме на работу и учебу в условиях дефицита рабочих и учебных мест.

Ключевые слова: векторная оптимизация, результаты тестирования, ранжирование

Постановка задачи

При определении уровня квалификации и профессиональной пригодности специалистов или оценке уровня знаний абитуриентов широкое распространение в настоящее время получили методы тестирования. В процессе тестирования используется много заданий, каждое из которых предусматривает решение нескольких задач и ответов на поставленные вопросы. В этих условиях получение объективной оценки результатов тестирования является достаточно сложной задачей, так как при отсутствии правильных ответов на все поставленные вопросы неправильное или неполное решение отдельных задач и тестов может компенсироваться хорошим решением задач и тестов другой группы.

В большинстве применяемых методик обработки результатов тестирования (см., например, [1—6]), где основное внимание уделяется вопросам конструирования тестов, методам определения их качества и статистической оценки резуль-

татов тестирования, все тесты, как правило, имеют одинаковый приоритет, и решения принимаются на основе процента правильно выполненных всех заданий в различных тестах. В основе предлагаемого в статье многокритериального подхода к решению проблемы лежит предположение, что не все задания в каждом тесте и различные тесты являются равнозначными для принятия решений.

К преимуществам этого подхода следует отнести:

- возможность учета степени важности как различных заданий в тесте, так и отдельных тестов. Это достигается как выбором различного вида функций для вычисления оценки выполнения отдельных заданий и тестов, так и выбором значений весовых коэффициентов в комплексном показателе;
- предложены методы учета граничного уровня качества выполнения отдельных тестов для случаев, когда пробел в квалификации или знаниях в каком-то вопросе не может быть компенсирован никаким самым высоким результатом выполнения других заданий и тестов.

Предположим, что проходящие тестирование кандидаты (обозначим их индексами $r = 1, \dots, R$) должны выполнить N различных тестов. Пусть $i = 1, \dots, N$ — индексы этих тестов. Каждый из этих тестов имеет l_i заданий (вопросов), $j = 1, \dots, l_i$. Обозначим L_i — множество заданий i -го теста. Все множество тестов может быть разбито на K групп, $k = 1, \dots, K$. Пусть I_k — подмножество тестов, относящихся к i -й группе, а g_k — число этих тестов. Обозначим \bar{K} — множество групп, $\bar{K} = \{I_1, I_2, \dots, I_k, \dots, I_K\}$.

Правилами тестирования может быть предусмотрено следующее:

- отсев кандидатов, у которых правильность и объем выполнения заданий отдельного теста или группы тестов ниже заданного, т. е. отсутствие квалификации или знаний в каких-то вопросах не может быть компенсировано никаким, даже самым высоким уровнем знаний и квалификации в других вопросах и областях знаний;
- на основе объективных показателей результатов тестирования, которые должны быть выражены в количественной форме, необходимо проранжировать всех кандидатов в соответствии с их уровнем квалификации и пригодностью для выполнения возлагаемых на них

функций либо разбить их на различные квалификационные или соответствующие определенным уровням знаний группы.

Для решения сформулированных задач необходимо:

- задать функции, определяющие количественные оценки результатов выполнения отдельных заданий, тестов или групп тестов;
- определить систему ограничений, позволяющую объективно, на основе количественных показателей осуществить отсев кандидатов, компетентность или уровень знаний которых в определенном круге вопросов недостаточны;
- задать критерии оценки результатов тестирования в целях ранжирования или разбиения кандидатов на различные квалификационные группы;
- установить с помощью введения весовых коэффициентов объективных уровней предпочтения правильности и полноты выполнения как отдельных заданий, так и различных тестов.

При этом должны рассматриваться только те кандидаты, результаты тестирования которых удовлетворяют заданной системе ограничений.

Для решения сформулированной задачи применяются методы построения компромиссных критериев в задачах многокритериальной (векторной) оптимизации [7—9], а также методы получения количественных данных на основе обработки качественной информации мнений экспертов.

Функции оценки результатов выполнения тестов

Обозначим $\omega_{ij}(r)$ оценку результатов выполнения j -го задания i -го теста:

$\omega_{ij}(r) = 1$, если ответ полный и правильный;

$\omega_{ij}(r) = 0$, если ответ неправильный или неполный, или не получено ответа.

Оценка правильности выполнения задания j может быть также представлена с помощью некоторой функции или fuzzy-переменной. Пусть, например, это будет функция $\varphi_{ij}(r)$, которая в зависимости от правильности и полноты ответа может изменяться в пределах $0 \leq \varphi_{ij}(r) \leq 1$. Если задание содержит D_{ij} альтернативных ответов для выбора, из которых только $d_{ij} \leq D_{ij}$ правильных, то функция $\varphi_{ij}(r)$ строится в зависимости от соотношения правильных и неверных альтернатив, выбранных в ответе. При абсолютно правильном и полном выполнении задания $\omega_{ij}(r) = 1$.

Обозначим

$$f_i(r) = \frac{1}{l_i} \sum_{j=1}^{l_i} \omega_{ij}(r) \text{ или } f_i(r) = \frac{1}{l_i} \sum_{j=1}^{l_i} \varphi_{ij}(r) \quad (1)$$

функции, определяющие количественный результат выполнения i -го теста. Очевидно, что эти функции удовлетворяют условиям нормировки

$$0 \leq f_i(r) \leq 1, \quad i = 1, \dots, N.$$

Пусть для каждого i -го задания предложены D_{ij} альтернативных ответов, среди которых $d_{ij}^1 < D_{ij}$ являются правильными и $d_{ij}^2 < D_{ij}$ — неверными, причем $d_{ij}^1 + d_{ij}^2 = D_{ij}$. Пусть в полученном решении задания из E_{ij} альтернатив содержится только $e_{ij}^1 (e_{ij}^1 \leq d_{ij}^1)$ правильных решений и $e_{ij}^2 (e_{ij}^2 \leq d_{ij}^2, e_{ij}^1 + e_{ij}^2 = E_{ij})$ — неверных решений. При этих предпосылках функция $\varphi_{ij}(r)$ в зависимости от значений e_{ij}^1 и e_{ij}^2 вычисляется в соответствии с выражением

$$\varphi_{ij}(r) = \max \left\{ 0, \left(\frac{e_{ij}^1(r)}{d_{ij}^1(r)} - \frac{e_{ij}^2(r)}{d_{ij}^2(r)} \right) \right\}. \quad (2)$$

Если каждому правильному ответу поставлен в соответствие определенный весовой коэффициент $\xi_{ij}^{1h}, h = 1, \dots, d_{ij}^1$, а каждой выбранной в качестве ответа неправильной альтернативе — весовой коэффициент $\xi_{ij}^{2h}, h = 1, \dots, d_{ij}^2$, и выполняются следующие условия нормировки этих коэффициентов:

$$\begin{aligned} \xi_{ij}^{1h} &\geq 0, \quad h = 1, \dots, d_{ij}^1; \quad \sum_{h=1}^{d_{ij}^1} \xi_{ij}^{1h} = 1; \\ \xi_{ij}^{2h} &\geq 0, \quad h = 1, \dots, d_{ij}^2; \quad \sum_{h=1}^{d_{ij}^2} \xi_{ij}^{2h} = 1, \end{aligned}$$

то оценка качества выполнения j -го задания i -го теста вычисляется по формуле

$$\varphi_{ij}(r) = \max \left\{ 0; \left[\sum_{h=1}^{d_{ij}^1} x_{ij}^{1h} \xi_{ij}^{1h} - \sum_{h=1}^{d_{ij}^2} x_{ij}^{2h} \xi_{ij}^{2h} \right] \right\}. \quad (3)$$

Выбор видов функций $f_i(r)$ может быть также осуществлен следующим образом:

$$f_i(r) = \begin{cases} 0, & \text{если } \frac{1}{l_i} \sum_{j=1}^{l_i} \omega_{ij}(r) \leq w_i; \\ \frac{1}{l_i} \sum_{j=1}^{l_i} \omega_{ij}(r), & \text{если } \frac{1}{l_i} \sum_{j=1}^{l_i} \omega_{ij}(r) > w_i. \end{cases} \quad (4)$$

Здесь w_i — некоторый минимальный уровень правильности и полноты выполнения i -го теста. В частном случае, может быть принято $w_i = 0$.

Смысл ограничений (11) заключается в том, что в каждой группе тестов $p = 1, 2, \dots, P$, по крайней мере, один из тестов должен быть выполнен по уровню правильности и полноты в определенном и установленном некоторой нижней границей b_p объеме.

Ограничения (12) имеют смысл, если правилами испытаний предусмотрена необходимость выполнения в некоторой группе p каждого из тестов в объеме не ниже некоторого граничного нижнего уровня \bar{b}_p . Заметим также, что этот минимальный уровень выполнения может требоваться также одновременно и для некоторого подмножества групп тестов. Аналогично системе ограничений для произвольного подмножества тестов J_s может рассматриваться система ограничений и для некоторого подмножества групп тестов I_k .

Критерии качества результатов тестирования

Обозначим: $E(r) = \{f_1(r), f_2(r), \dots, f_N(r)\}$ — вектор значений функций, $0 \leq f_i(r) \leq 1, i = 1, \dots, N$, отражающий результаты выполнения каждого из тестов r -м кандидатом; \bar{Q} — множество результатов, которое удовлетворяет заданной системе ограничений, т. е. результаты тестирования тех кандидатов, у которых объем и правильность выполнения отдельных заданий и тестов удовлетворяет всем предусмотренным системой отбора минимальным требованиям.

В качестве критериев оценки результатов тестирования могут рассматриваться следующие выражения:

$$\Phi_1(r) = \left\{ \sum_{k \in K} \alpha_k F_k(r) \mid E(r) \in \bar{Q} \right\}; \quad (15)$$

$$\Phi_2(r) = \left\{ \sum_{i=1}^N \alpha_i f_i(r) \mid E(r) \in \bar{Q} \right\}. \quad (16)$$

Здесь $F_k(r)$ — функция, удовлетворяющая условиям нормировки $0 \leq F_k(r) \leq 1, k = 1, \dots, K$, — некоторая обобщенная характеристика результатов выполнения группы тестов;

$\alpha_k, k = 1, \dots, K$, — весовые коэффициенты, удовлетворяющие условиям нормирования вида

$$0 \leq \alpha_k \leq 1, k = 1, \dots, K; \quad \sum_{k=1}^K \alpha_k = 1.$$

В качестве показателя, определяющего количественную оценку результатов выполнения k -й группы тестов, могут быть предложены следующие выражения:

$$F_k(r) = \sum_{i \in I_k} \beta_i f_i(r), \quad F_k(r) = \min_{i \in I_k} \beta_i f_i(r); \quad (17)$$

$$F_k(r) = \max_{i \in I_k} \beta_i f_i(r), \quad F_k(r) = \prod_{i \in I_k} \beta_i f_i(r). \quad (18)$$

В выражениях (17), (18) $\beta_i, i \in I_k$ — весовые коэффициенты, значения которых удовлетворяют условиям нормирования

$$0 \leq \beta_i \leq 1, i \in I_k; \quad \sum_{i \in I_k} \beta_i = 1.$$

Выбор значений весовых коэффициентов

Пусть определена контрольная выборка, состоящая из M специалистов, уровень знаний или профессиональной пригодности которых известен. В соответствии с этим уровнем они ранжированы следующим образом:

$$\bar{M} = \{m_1, m_2, \dots, m_v, m_{v+1}, \dots, m_M\}. \quad (19)$$

В этом ранжировочном ряду специалист, который стоит левее, является эффективнее специалиста, стоящего правее. Если определены группы тестов, система функций оценки результатов выполнения отдельных заданий, тестов и групп тестов, а также выражения критериев для оценки результатов тестирования $\Phi(r)$ (с точностью до неизвестных весовых коэффициентов), то ясно, что должно выполняться условие

$$\Phi(m_{v+1}) < \Phi(m_v), \quad v = 1, 2, \dots, M. \quad (20)$$

Для сравнения двух специалистов из контрольной выборки (индексы их обозначим u и v) введем понятие уровня эффективности $\pi^\delta(u, v)$. Для каждого значения δ величина $\pi^\delta(u, v)$ — это некоторое заданное число, лежащее в пределах $0 \leq \pi^\delta(u, v) \leq 1$.

Если специалист v предпочтительнее специалиста u , т. е. стоит в ранжировочном ряду (19) левее и этот уровень предпочтения определяется числом $\pi^\delta(u, v)$, то это означает, что значения критерия эффективности результатов тестирования у него должно быть выше в $\pi^\delta(u, v)$, что может быть выражено соотношением

$$\pi^\delta(u, v) \sum_{i=1}^N \alpha_i f_i(v) \geq \sum_{i=1}^N \alpha_i f_i(u). \quad (21)$$

Тогда с учетом соотношений (21) для каждой пары рядом стоящих в ранжировочном ряду специалистов в случае использования критерия эффективности результатов тестирования в виде (16) может быть получена следующая система линейных неравенств относительно неизвестных значений весовых коэффициентов $\alpha_i, i = 1, \dots, N$:

$$\sum_{i=1}^N \alpha_i [\pi^\delta(v, u) f_i(v) - f_i(u)] \geq 0, \quad (22)$$

$$u \neq v, u, v = 1, 2, \dots, M,$$

и на весовые коэффициенты наложены условия нормировки

$$0 \leq \alpha_i \leq 1, \quad i = 1, \dots, N; \quad \sum_{i=1}^N \alpha_i = 1. \quad (23)$$

Более точно значения весовых коэффициентов $\alpha_i, i = 1, \dots, N$, могут быть вычислены в результате определения "Чебышевской точки" системы линейных неравенств (22) в условиях ограничений (23). Эта вторая задача сводится к задаче линейного программирования вида

$$\begin{aligned} & \psi(\alpha_1^*, \dots, \alpha_i^*, \dots, \alpha_N^*) = \\ & = \left\{ \max_{\alpha_i, i=1, \dots, N; \eta} \eta \mid \eta \geq 0 \right\} \end{aligned} \quad (24)$$

в условиях ограничений (22), а также системы ограничений вида

$$\begin{aligned} & \sum_{i=1}^N \alpha_i [\pi^\delta(v, u) f_i(v) - f_i(u)] - \eta \geq 0, \\ & u \neq v, \quad u, v = 1, 2, \dots, M. \end{aligned} \quad (25)$$

Сформулированная задача (23)–(25) может быть решена симплексным методом линейного программирования. В выражении (24) $\alpha_1^*, \dots, \alpha_i^*, \dots, \alpha_N^*$ — оптимальные значения весовых коэффициентов, полученные в результате решения задачи (24), (25).

Аналогичными методами могут быть сформулированы и решены задачи определения значений весовых коэффициентов для групп тестов $\alpha_k, k = 1, 2, \dots, K$, в выражении для критерия эффективности тестирования в виде (15).

Алгоритмы обработки результатов тестирования

Пусть экспертами осуществлен выбор тестов, выражений функций $f_i(r), i = 1, \dots, N$, системы ограничений для отсева не отвечающих возлагаемым требованиям кандидатов, выражения критериев оценки результатов тестирования, а также значения весовых коэффициентов в критерии.

В качестве первой задачи тестирования рассмотрим задачу отбора группы специалистов, способных выполнить возлагаемые на них в предстоящей работе (обучении) функции, т. е. отбор среди всего множества тестированных кандидатов R некоторого подмножества $\bar{R} \subseteq R$, результаты выполнения тестов у которых соответствуют заданной системе ограничений

$$\bar{R} = \{r \in R \mid M(r) \in \bar{Q}\}, \quad (26)$$

где Q — система неравенств (9)–(12).

Вторая задача предусматривает построение ранжировочного ряда вида (19), в котором группа, состоящая из оставшихся после отсева подмноже-

ства \bar{R} , включающего R_1 тестируемых кандидатов, $\bar{R} = \{1, 2, \dots, R_1\}$, должна быть упорядочена в порядке возрастания или убывания значения одного из критериев $\Phi_1(r)$ или $\Phi_2(r)$, выраженных соответственно в форме (15) или (16).

Третья задача связана с разделением в соответствии с результатами тестирования подмножества \bar{R} кандидатов на T различных по уровню и квалификации групп, $\tau = 1, 2, \dots, T$. Это осуществляется в соответствии со следующим принципом:

$$r \in \begin{cases} Z_1, & \text{если } 0 \leq \Phi(r) < C_1; \\ Z_2, & \text{если } C_1 \leq \Phi(r) < C_2; \\ \dots & \dots \\ Z_T, & \text{если } C_{T-1} \leq \Phi(r) < C_T. \end{cases} \quad (27)$$

Здесь $Z_\tau, \tau = 1, 2, \dots, T$, — подмножество кандидатов (подмножество результатов тестирования), относящееся к τ -му классу; $\Phi(r)$ — значение некоторого комплексного критерия оценки качества результатов тестирования кандидата r , выраженное действительным числом; $C_1, C_2, \dots, C_\tau, \dots, C_T$ — некоторые граничные значения критерия $\Phi(r)$, удовлетворяющие соотношениям

$$0 < C_1 < C_2 < \dots, C_\tau < \dots < C_T \quad (28)$$

Для определения граничных значений $C_1, C_2, \dots, C_\tau, \dots, C_T$ используется контрольная выборка результатов тестирования, для каждого элемента которой известна принадлежность к одному из выбранных классов $\tau = 1, 2, \dots, T$.

Определим некоторую матрицу ошибок с элементами $\sigma_{l\tau}(v) \geq 0$, которые вычисляются по формулам

$$\sigma_{l\tau}(v) = \bar{\sigma}_{l\tau},$$

если в действительности $v \in Z_\tau$, но $C_{l-1} \leq \Phi(v) < C_l$, т. е. принадлежащий к классу l кандидат в соответствии с принятым алгоритмом классификации отнесен к классу τ ;

$$\sigma_{l\tau}(v) = 0$$

при правильной классификации; $l, \tau = 1, 2, \dots, T$.

Тогда выбор граничных значений $C_1, C_2, \dots, C_\tau, \dots, C_T$ может осуществляться из условий минимизации функционала (28), определяющего величину суммарной ошибки результатов классификации

$$\bar{F}(C_1, C_2, \dots, C_\tau, \dots, C_T) = \sum_{v=1}^{M_1} \sigma_{l\tau}(v), \quad (29)$$

где M_1 — число кандидатов в контрольной выборке.

Задача (29) является многоэкстремальной с недифференцируемым функционалом и может быть

решена генетическими алгоритмами либо Branch-and-Bound-Method.

Предлагаемая методика была проверена и применена в Германии в нескольких частных фирмах, занимающихся разработкой программного обеспечения, при конкурсном приеме на работу специалистов, а также на курсах изучения иностранных языков при формировании групп слушателей, имеющих близкий уровень подготовки. В качестве тестов в первом случае использовался перечень заданий и вопросов на знание операционных систем, теории алгоритмов, системного анализа и конкретных языков программирования. В тестах проверки знаний немецкого языка были задания, проверяющие словарный запас, знание грамматики, понимание текстов на слух, правильность написания текстов и умение говорить.

Некоторые вопросы повышения эффективности построения тестов

Сделаем некоторые замечания о вероятности случайного получения полных и правильных ответов на некоторые задания тестов. Если необходимо выбрать из m_j различных предлагаемых альтернатив одно правильное решение, то вероятность случайного правильного выбора определяется выражением $P_j = \frac{1}{m_j}$.

Если среди m предлагаемых альтернатив есть μ правильных решений, то вероятность случайного получения полного и правильного ответа определяется выражениями

$$\bar{P}_{J1} = \sum_{\mu=1}^m \left(\frac{m!}{(m-\mu)!\mu!} \right)^{-1} \quad (30)$$

в случае, если в условиях теста не содержится подсказки, сколько среди альтернативных ответов содержится правильных решений.

Если известно число правильных решений μ среди m альтернативных ответов, то эта вероятность определяется выражением

$$\bar{P}_{J2} = \left(\frac{m!}{(m-\mu)!\mu!} \right)^{-1}. \quad (31)$$

Ясно, что $\bar{P}_{J1} \leq \bar{P}_{J2} \leq P_j$.

Если тест содержит l_i заданий и известна вероятность случайного получения полного и правильного ответа на каждое из заданий, то может быть также вычислена вероятность получения случайного правильного и полного ответа на любое число заданий каждого теста.

Из вышеизложенного могут быть сделаны следующие выводы. Для уменьшения вероятности случайного получения правильных ответов на задания теста необходимо:

- увеличить число заданий в каждом из тестов;
- увеличить число альтернативных ответов в каждом из заданий теста;
- задание строить таким образом, чтобы среди возможных альтернатив оно содержало не одно, а сразу несколько правильных ответов;
- не информировать кандидата о числе правильных ответов среди предлагаемых альтернатив.

Иллюстративный пример

Группе, состоящей из восьми кандидатов, предложено выполнить восемь тестов. Первый тест включает два задания (мы обозначим их заданиями 1.1 и 1.2), второй тест состоит из трех заданий (задания 2.1, 2.2 и 2.3) и третий тест включает только 1 задание (задание 3.1). В качестве альтернатив правильного решения задания 1.1 предложено шесть различных вариантов ответов, среди которых два правильных и четыре неверных. Обозначим эти условия в виде $[6(2, 4)]$, для задания 1.2 возможны пять альтернатив с соотношением правильных и неверных ответов $[5(3, 2)]$, для задания 2.1 — $[7(1, 6)]$, для задания 2.2 — $[4(2, 2)]$, задания 2.3 — $[5(2, 3)]$ и, наконец, для задания 3.1 — $[6(3, 3)]$.

Оценки правильности и полноты выполнения каждого из заданий определяются в соответствии с формулами (1), а правильности и полноты выполнения каждого из тестов — по формулам (5). В качестве критерия оценки результатов выполнения тестов выбран показатель $\Phi_2(r)$ (формула (16)) со следующими весовыми коэффициентами: $\alpha_1 = 0,4$, $\alpha_2 = 0,4$, $\alpha_3 = 0,2$.

В качестве ограничений для отсева кандидатов выбраны следующие условия:

- оценка результата выполнения первого теста должен быть не ниже 0,3;
- средневзвешенная оценка результатов выполнения второго и третьего тестов (см. ограничения (9) со значениями весовых коэффициентов $\Delta_1 = 0$, $\Delta_2 = 0,6$, $\Delta_3 = 0,4$) должна быть не ниже значения 0,4;
- максимальное значение оценки выполнения второго и третьего тестов (ограничения (11) $p = 2,3$) должна быть не ниже 0,5.

На основе результатов тестирования восьми кандидатов должны быть решены следующие задачи:

- необходимо отобрать группу кандидатов, оценки результатов тестирования которых удовлетворяют всей системе ограничений;
- отобранная группа кандидатов должна быть разделена на следующие три класса: 1-й класс — это кандидаты, у которых значение обобщенного критерия оценки результатов тестирования находится в пределах $\Phi_2(r) \in [0,3, 0,5)$, второй класс — $\Phi_2(r) \in [0,5, 0,7)$ и третий класс — $\Phi_2(r) \in [0,7, 1,0]$;

Данные и результаты обработки показателей тестирования

Результаты тестирования	Кандидат							
	№ 1	№ 2	№ 3	№ 4	№ 5	№ 6	№ 7	№ 8
Решение задания 1.1	2(1,1)	4(2,2)	3(1,2)	2(2,0)	3(2,1)	1(1,0)	4(2,2)	2(2,0)
Оценка задания 1.1	0,25	0,5	0,0	1,0	0,75	0,5	0,5	1,0
Решение задания 1.2	3(2,1)	2(2,0)	2(2,0)	2(2,0)	2(2,0)	4(3,1)	4(3,1)	2(2,0)
Оценка задания 1.2	0,167	0,6667	0,6667	0,6667	0,6667	0,5	0,5	0,6667
Оценка выполнения теста 1	0,208	0,5833	0,3334	0,8333	0,708	0,5	0,5	0,8333
Решение задания 2.1	2(1,1)	4(1,3)	3(1,2)	1(0,1)	1(1,0)	2(1,1)	4(1,3)	3(1,2)
Оценка задания 2.1	0,833	0,5	0,6667	0,0	0,8334	0,8334	0,5	0,6667
Решение задания 2.2	2(1,1)	3(2,1)	3(2,1)	3(2,1)	1(1,0)	1(1,0)	2(2,0)	1(1,0)
Оценка задания 2.2	0,0	0,5	0,5	0,5	0,5	0,5	1,0	0,5
Решение задания 2.3	3(3,0)	3(2,1)	1(1,0)	2(2,0)	3(2,1)	1(1,0)	2(1,1)	3(2,1)
Оценка задания 2.3	1,0	0,3334	0,5	0,3333	0,6667	0,5	0,1666	0,6667
Оценка выполнения теста 2	0,6055	0,444	0,5555	0,2778	0,722	0,611	0,5555	0,722
Решение задания 3.1	3(2,1)	2(2,0)	3(2,1)	1(1,0)	2(2,0)	1(1,0)	4(3,1)	4(3,1)
Оценка задания 3.1	0,3334	0,6667	0,3334	0,3334	0,6667	0,3334	0,6667	0,6667
Оценка выполнения теста 3	0,3334	0,6667	0,3334	0,3334	0,6667	0,3334	0,6667	0,6667
Средневзвешенная оценка выполнения тестов 2 и 3	0,4965	0,5331	0,4444	0,3	0,6999	0,5334	0,6	0,7
Максимальная оценка за выполнение тестов 2 и 3	0,6055	0,6667	0,5555	0,3334	0,722	0,611	0,6667	0,722
Результат тестирования и основание отсева	Отсев из-за теста 1	+	+	Отсев из-за тестов 2, 3	+	+	+	+
Значение комплексного критерия тестирования $\Phi_2(r)$		0,5442	0,4222		0,7053	0,5111	0,5555	0,7555
Места кандидатов в ранжировочном ряду		4	6		2	5	3	1
Квалификационные классы		2	1		3	2	2	3

- отобранная группа кандидатов, оценки результатов тестирования которых удовлетворяют всей системе ограничений, должна быть отсортирована по убыванию (невозрастанию) значения комплексного критерия оценки результатов $\Phi_2(r)$.

Результаты выполнения тестов, обработки данных и оценки тестирования представлены в таблице. Для представления результатов выполнения каждого из заданий $D(A, B)$ приняты следующие обозначения:

- D — суммарное число альтернатив, которые выбраны кандидатом в качестве правильных;
- A — фактическое число правильных альтернатив из D , выбранных кандидатом в ответе;
- $B = D - A$ — число альтернатив, выбранных неправильно в ответе кандидата.

Как показали результаты тестирования, выбранная группа кандидатов, оценки тестирования которых удовлетворяют всем сформулированным выше ограничениям задачи, включает следующие номера кандидатов: {2, 3, 5, 6, 7, 8}. Кандидаты {1, 4} были отброшены. Отобранная группа в соответствии с уровнем квалификации и профессиональной пригодности (по убыванию значения критерия $\Phi_2(r)$) проранжирована следующим образом:

$$\{8 > 5 > 7 > 2 > 6 > 3\}.$$

К первому классу (наиболее низкий уровень) отнесен лишь кандидат № 3. Ко второму классу

(средний уровень квалификации) относятся кандидаты №№ 6, 2 и 7. Третий класс, определяющий высокий профессиональный уровень, составляют кандидаты №№ 8, 5.

Выводы

Рассмотрено решение следующих сформулированных задач:

- отбор специалистов, уровень квалификации которых отвечает поставленным требованиям;
- ранжирование специалистов в соответствии с уровнем их знаний, опытом и квалификацией;
- разбиение специалистов на различные по уровню квалификации группы на основе конструирования и вычисления значений комплексного компромиссного критерия оценки результатов тестирования. Это позволяет объективно выражать результаты выполнения комплекса тестов с помощью некоторого действительного числа.

На основе вероятностного анализа случайного получения полных и правильных ответов на задания тестов даются рекомендации по эффективно-му конструированию заданий и тестов.

Полученные в работе результаты могут найти применение при контроле знаний в системах компьютерного обучения, а также при конкурсном приеме на работу и учебу в условиях дефицита рабочих и учебных мест.

Список литературы

1. Звонников В. И., Чельшкова М. Б. Современные средства оценивания результатов обучения. Сер. Высшее профессиональное образование. М.: Академия, 2009.
2. Чельшкова М. Б. Теория и практика конструирования педагогических тестов. М.: Логос, 2002. 432 с.
3. Нейман Ю. М., Хлебников В. А. Введение в теорию моделирования и параметризации педагогических тестов. М.: Прометей, 2000.
4. Самылкина Н. Н. Современные средства оценивания результатов обучения. М.: БИНОМ. Лаборатория знаний, 2007. 172 с.
5. Аванесов В. С. Научные проблемы тестового контроля знаний: Учебное пособие. М.: Исследовательский центр, 1994. 135 с.

6. Burgin M., Aljehdali S. Testing: Organization and Evaluation/Computer and Their Application. 2008. S. 203—208.

7. Зак Ю. А. Модели и методы построения компромиссных планов в задачах математического программирования с несколькими целевыми функциями // Кибернетика. 1972. № 4. С. 102—107.

8. Зак Ю. А. Модификации метода последовательных уступок в задаче оптимизации векторного критерия // Автоматика и вычислительная техника. 1975. № 3. С. 45—51.

9. Зак Ю. А. Многоэтапные процессы принятия решений в задачах векторной оптимизации // Автоматика и вычислительная техника. 1976. № 6. С. 273—278.

УДК 004:519.2:37

Е. Ю. Карданова, канд. физ.-мат. наук, доц.,
e-mail: e_kardanova@mail.ru,
Р. С. Карданов, аспирант,
e-mail: r_kardanov@mail.ru,

Новгородский государственный университет
имени Ярослава Мудрого, г. Великий Новгород

Построение интервальных оценок статистик согласия при анализе результатов массового тестирования в рамках моделей Г. Раша

Точечные оценки статистик согласия, традиционно используемых при анализе результатов массового тестирования, в силу различных причин не могут обеспечить надежный критерий адекватности эмпирических данных модели измерения. Предлагается метод построения интервальных оценок статистик согласия. Для решения задачи интервального оценивания предлагается использовать бутстреп-метод построения доверительных интервалов. Предложена технология идентификации заданий, не согласующихся с моделью измерения.

Ключевые слова: тестирование, модели Раша, статистика согласия, критическое значение, бутстреп-метод, интервальное оценивание

Введение

Модели Г. Раша широко применяются для анализа результатов тестирования как в педагогике, социологии, психологии, так и в других областях, когда ставится задача измерения латентных (скрытых от непосредственного наблюдения) переменных. Рассмотрим простейшую модель Раша — дихотомическую. Вероятность P_{ni} того, что испытуемый n , $n = 1, \dots, N$, с уровнем подготовленности θ_n правильно выполнит задание i , $i = 1, \dots, I$, с

уровнем трудности δ_i , согласно дихотомической модели Раша определяется формулой

$$P_{ni} = \frac{\exp(\theta_n - \delta_i)}{1 + \exp(\theta_n - \delta_i)}. \quad (1)$$

С выводом этой модели и ее свойствами, методами оценивания параметров модели θ_n и δ_i , а также обобщениями модели на случай политомических заданий (оцениваемых более чем в двух категориях) можно ознакомиться в работах [1—3].

Одной из основных задач при математическом моделировании выступает исследование адекватности эмпирических данных модели измерения. В рамках моделей Раша наиболее часто используются статистики, получившие название *общих статистик согласия*. Применительно к модели (1) общая статистика согласия U_i и ее взвешенная версия W_i для задания i имеют вид

$$U_i = \frac{1}{N} \sum_{n=1}^N \frac{(x_{ni} - P_{ni})^2}{P_{ni}q_{ni}}; \quad (2)$$

$$W_i = \frac{\sum_{n=1}^N (x_{ni} - P_{ni})^2}{\sum_{n=1}^N P_{ni}q_{ni}}, \quad (3)$$

где x_{ni} — балл испытуемого n за выполнение задания i ; $q_{ni} = 1 - P_{ni}$. Свойства этих статистик и особенности их распределений описаны в работе [3]. Они принимают значения из промежутка $[0, +\infty)$, их математические ожидания равны 1, а дисперсии вычисляются по формулам

$$D(U_i) = \frac{1}{N^2} \left(\sum_{n=1}^N \frac{1}{P_{ni}q_{ni}} - 4N \right);$$
$$D(W_i) = \frac{1}{\left(\sum_{n=1}^N P_{ni}q_{ni} \right)^2} \left(\sum_{n=1}^N P_{ni}q_{ni} - 4 \sum_{n=1}^N (P_{ni}q_{ni})^2 \right).$$

Отметим, что в реальных условиях вероятность P_{ni} нам неизвестна, и мы можем использовать только ее приближенное значение \hat{P}_{ni} , основанное на оценках параметров $\hat{\theta}_n$ и $\hat{\delta}_i$ модели (1).

Наблюдаемые значения статистик (2), (3) тем ближе к ожидаемому значению 1, чем лучше экспериментальные данные согласуются с используемой моделью. Основной проблемой при этом является проблема выбора критических значений используемых статистик. Значения статистик (2), (3) зависят от числа испытуемых N , распределения их уровней подготовленности и других факторов. Более того, теоретические распределения статистик могут быть определены только приближенно, а эмпирические распределения отличаются от теоретических [4]. Это не позволяет использовать в качестве критических значения, определяемые по квантилям теоретических распределений.

В зарубежной литературе при массовых тестированиях с важными последствиями рекомендуется в качестве допустимого интервала значений статистик (2), (3) использовать интервал (0,8; 1,2) [5]. Однако этот подход неоднократно подвергался критике, главным образом из-за его теоретической необоснованности и неточности предлагаемого интервала. Одни авторы предлагают различные коррекции статистик согласия [6], другие — коррекции их критических значений (в зависимости от объема выборки) [7], третьи вообще предлагают отказаться от использования общих статистик согласия в пользу других (но также не абсолютно надежных) статистик [8].

С целью избежать проблемы выбора критических значений статистик (2), (3) иногда используется аппроксимация их распределений стандартизованным нормальным распределением $N(0, 1)$ [1, 4]. Однако в силу приближенного характера методов аппроксимации и неопределенности теоретических распределений самих статистик их стандартизованные версии также не являются надежными.

В настоящей работе авторы предлагают отказаться от точечных оценок статистик согласия, зависящих от свойств выборки, и перейти к их интервальному оцениванию.

Таким образом, цель работы — построение интервальных оценок общих статистик согласия и исследование возможности их использования для идентификации заданий, не находящихся в согласии с моделью измерения. Для решения задачи интервального оценивания предлагается использовать бутстреп-метод построения доверительных интервалов [9].

Бутстреп-метод построения доверительных интервалов

Суть бутстреп-метода состоит в том, что по имеющимся наблюдениям за случайной величи-

ной моделируется процесс ее получения, позволяющий оценить выборочное распределение этой случайной величины.

Предположим, что из неизвестного распределения вероятностей G извлекается выборка объема N : $X_i = x_i$, $X_i \sim G$, $i = 1, 2, \dots, N$. Обозначим случайную выборку и ее наблюдаемую реализацию через $X = (X_1, X_2, \dots, X_N)$ и $x = (x_1, x_2, \dots, x_N)$ соответственно. Рассматривается некоторый параметр θ (являющийся случайной величиной $\theta = \theta(X, G)$), для которого известна его точечная оценка $\hat{\theta}$. Ставится задача на основе наблюдаемых данных x оценить выборочное распределение этой случайной величины. Для решения данной задачи бутстреп-методом необходимо:

1) построить выборочное распределение вероятностей \hat{G} , полагая в каждой из точек x_1, x_2, \dots, x_N массу $1/N$;

2) при фиксированном \hat{G} извлечь из него случайную выборку объема N с возвращениями: $X_i^* \sim \hat{G}$, $i = 1, 2, \dots, N$. Совокупность $X^* = (X_1^*, \dots, X_N^*)$ называется бутстреп-выборкой объема N ;

3) вычислить оценку параметра: $\hat{\theta}^* = \theta(X^*, \hat{G})$. Эта оценка называется бутстреп-реализацией $\hat{\theta}$;

4) повторить шаги 2—3 большое число раз;

5) аппроксимировать выборочное распределение случайной величины $\theta = \theta(X, G)$ бутстреп-распределением случайной величины $\theta^* = \theta(X^*, \hat{G})$.

Бутстреп-распределение может быть использовано для различных целей, в частности, для построения доверительных интервалов. Рассмотрим метод процентилей построения доверительных интервалов.

Пусть $\hat{F}(x) = P(\hat{\theta}^* < x)$ — эмпирическая функция бутстреп-распределения статистики θ . В качестве центрального доверительного интервала уровня значимости α для θ берется интервал

$$\theta \in (\hat{F}^{-1}(\alpha/2), \hat{F}^{-1}(1 - \alpha/2)). \quad (4)$$

Если $\hat{F}(\hat{\theta}) = P(\hat{\theta}^* < \hat{\theta}) \neq 0,5$ (это означает, что бутстреп-распределение смещено относительно точечной оценки $\hat{\theta}$ параметра θ), то интервал (4) может оказаться неточным (смещенным). В этом случае он нуждается в коррекции смещения. Положим:

$$z_0 = \Phi^{-1}(\hat{F}(\hat{\theta})), \quad \text{где} \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$$

— функция Лапласа. В качестве

приближенного центрального доверительного интервала уровня значимости α для θ берется интервал

$$\theta \in (\hat{F}^{-1}(\Phi(2z_0 - z_\alpha)), \hat{F}^{-1}(\Phi(2z_0 + z_\alpha))), \quad (5)$$

где $z_\alpha = \Phi^{-1}(1 - \alpha/2)$.

Очевидно, что если бутстреп-распределение не смещено относительно $\hat{\theta}$, то $z_0 = 0$, и интервал (5) совпадает с интервалом (4). Известно, что доверительные интервалы, построенные с помощью бутстреп-метода, являются более точными по сравнению со стандартными интервалами, построенными на основе нормальной теории [9].

В заключение отметим, что бутстреп-метод построения доверительных интервалов требует большого числа бутстреп-повторений, порядка $B = 1000$ [9].

Процедура построения доверительных интервалов общих статистик согласия и ее программная реализация

Процедура построения доверительных интервалов общих статистик согласия включает в себя следующие этапы.

1. Нахождение точечных оценок статистик согласия.

Для нахождения точечных оценок статистик согласия (2), (3) необходимо знать оценку вероятности \hat{P}_{ni} , основанную на оценках параметров $\hat{\theta}_n$ и $\hat{\delta}_i$ модели (1). Для оценивания параметров и получения оценок статистик согласия использовалась программа Winsteps (<http://www.winsteps.com>). Эта программа вычисляет оценки параметров в рамках моделей Раша (на основе метода максимального правдоподобия) с указанием точности оценивания, а также оценки статистик согласия (2), (3) и их стандартизованных версий. В работе [10] показано, что данная программа выполняет практически несмещенные оценки всех указанных статистик.

2. Конструирование бутстреп-выборок.

Конструирование осуществляется методом Монте-Карло. Повторные реализации матрицы тестирования генерируются с использованием случайных выборок объема N с возвращением из исходной матрицы тестирования.

3. Нахождение бутстреп-оценок общих статистик согласия.

С этой целью к бутстреп-выборке применяется шаг 1.

Шаги 2—3 повторяются $B = 1000$ раз. Таким образом, в результате применения бутстреп-процедуры имеем 1000 бутстреп-оценок для каждой из рассматриваемых статистик.

5. Аппроксимация распределений статистик согласия.

При использовании метода Монте-Карло конструирования бутстреп-выборок в качестве аппроксимации бутстреп-распределения рассматривается гистограмма бутстреп-оценок рассматриваемых статистик.

6. Оценивание характера распределения статистик согласия.

На данном этапе по каждому заданию вычисляются выборочные числовые характеристики полученных распределений и исследуется смещение распределения относительно точечных оценок статистик, полученных на шаге 1.

7. Построение доверительных интервалов для статистик согласия.

С учетом результатов шага 6 на уровне значимости $\alpha = 0,05$ строится один из доверительных интервалов (4) или (5).

Для реализации описанной процедуры разработана специальная компьютерная программа. Программа реализована в виде отдельного desktop-приложения с использованием технологии Java 2 Standard Edition (версия Java — 1.6). Графический интерфейс реализован с использованием стандартных библиотек AWT и Swing, система хранения данных представляет собой структурированный каталог, содержащий XML-файлы. Для работы программы необходима предустановленная программа Winsteps.

Экспериментальная часть

Было проведено большое число разнообразных экспериментов, имеющих целью показать возможности использования интервальных оценок статистик согласия для идентификации заданий, не находящихся в согласии с моделью измерения. Для проведения экспериментов использовалось имитационное моделирование.

Разработано два метода моделирования матриц тестирования в соответствии с моделью (1). Первый метод предполагает конструирование матриц тестирования с произвольными, наперед заданными параметрами распределений уровней подготовленности испытуемых и уровней трудности заданий. Второй метод предполагает использование реальных данных тестирования в качестве прототипа для моделирования. В этом случае в качестве значений параметров испытуемых и заданий берутся их оценки, полученные на основе реальной матрицы тестирования. Генерирование модельных матриц проводилось методом Монте-Карло в соответствии с формулой вероятности (1). В программной реализации для получения точечных оценок статистик согласия матрица тестирования в batch-режиме прогоняется через программу Winsteps. Из выходных данных программы специальным анализатором извлекаются значения оценок общих статистик согласия по каждому заданию.

Чтобы исследовать возможности идентификации заданий, не согласующихся с моделью, в модельные матрицы вносились искажения. Для это-

го для выбранных заданий и выбранного процента искажений p ответы $p\%$ испытуемых, определенных случайным образом, заменялись на противоположные.

К матрице тестирования применялась процедура построения доверительных интервалов статистик согласия (2), (3). Обозначим $(\alpha_{U_i}^p, \beta_{U_i}^p)$ и $(\alpha_{W_i}^p, \beta_{W_i}^p)$ — доверительные интервалы статистик U и W соответственно для задания $i, i = 1, \dots, I$, при условии, что ответы на это задание содержат $p\%$ искажений. Доверительные интервалы при $p = 0$ (искажений нет) будем называть базовыми. Задание признается не согласующимся с моделью измерения, если доверительные интервалы его статистик согласия не пересекаются с соответствующими базовыми интервалами.

На первом этапе исследований матрица тестирования генерировалась с наперед заданными свойствами. Выбирались различные сочетания распределений испытуемых и заданий, встречающиеся в массовом тестировании. Дополнительно рассматривалась возможность смещения множества испытуемых относительно множества заданий в целях исследования его влияния на результаты эксперимента. Ниже приведены результаты одного из экспериментов.

Параметры исходной матрицы тестирования: $N = 1000$ (число испытуемых); $I = 50$ (число заданий); $\theta_n \sim N(0, 2)$; $\delta_i \sim N(0, 1)$. Параметр p выбирался равным 5, 10, 15, 20, 25, 50 и 100. Искажения вносились в профили пяти случайным образом отобранных заданий. Ко всем сгенерированным матрицам (без искажений и с искажениями) применялась процедура построения доверительных интервалов статистик согласия.

В табл. 1, 2 представлены доверительные интервалы общих статистик согласия при различных значениях параметра p . Данные усреднены по пяти заданиям, в профили которых вносились искажения. В 1-й строке таблиц представлены базовые интервалы, далее — интервалы при различном числе искажений.

Из анализа таблиц вытекают следующие выводы.

- Если задание находится в согласии с моделью ($p = 0$), то значения его статистик согласия близки к 1. Чем больше искажений содержит матрица тестирования, тем сильнее отличаются от 1 значения статистик согласия.
- Средние квадратические отклонения всех статистик, представленных в таблицах, достаточно малы. Это позволяет утверждать, что значения статистик согласия и границы доверительных интервалов в большей степени определя-

Таблица 1

Доверительные интервалы статистики U

Параметр p (процент искажений)	Значение статистики		Левая граница доверительного интервала $\alpha_{U_i}^p$		Правая граница доверительного интервала $\beta_{U_i}^p$	
	Среднее	Ср. кв. откл.	Среднее	Ср. кв. откл.	Среднее	Ср. кв. откл.
0	0,99	0,067	0,84	0,044	1,15	0,093
5	1,13	0,066	0,98	0,046	1,31	0,098
10	1,28	0,075	1,12	0,06	1,46	0,109
15	1,44	0,032	1,28	0,028	1,63	0,048
20	1,48	0,070	1,32	0,057	1,67	0,097
25	1,59	0,087	1,43	0,069	1,79	0,119
50	2,21	0,065	2,02	0,054	2,42	0,079
100	3,30	0,123	3,11	0,089	3,54	0,178

Таблица 2

Доверительные интервалы статистики W

Параметр p (процент искажений)	Значение статистики		Левая граница доверительного интервала $\alpha_{W_i}^p$		Правая граница доверительного интервала $\beta_{W_i}^p$	
	Среднее	Ср. кв. откл.	Среднее	Ср. кв. откл.	Среднее	Ср. кв. откл.
0	1,00	0,025	0,92	0,027	1,08	0,032
5	1,10	0,027	1,02	0,027	1,19	0,029
10	1,20	0,038	1,12	0,034	1,30	0,042
15	1,30	0,029	1,22	0,026	1,40	0,029
20	1,35	0,028	1,26	0,028	1,45	0,034
25	1,43	0,027	1,33	0,023	1,52	0,027
50	1,83	0,027	1,71	0,022	1,94	0,028
100	2,44	0,016	2,33	0,019	2,56	0,018

Таблица 3

Доверительные интервалы статистики U по пяти заданиям

Номер задания	Базовый интервал	Значение при 10 %	Доверительный интервал при 10 %	Значение при 15 %	Доверительный интервал при 15 %
1	(0,79; 1,07)	1,22	(1,06; 1,37)	1,42	(1,26; 1,62)
2	(0,81; 1,12)	1,26	(1,11; 1,42)	1,44	(1,27; 1,61)
3	(0,82; 1,09)	1,22	(1,08; 1,39)	1,43	(1,27; 1,59)
4	(0,88; 1,19)	1,30	(1,15; 1,48)	1,43	(1,28; 1,60)
5	(0,89; 1,30)	1,40	(1,21; 1,64)	1,50	(1,33; 1,71)

Таблица 4

Доверительные интервалы статистики W по пяти заданиям

Номер задания	Базовый интервал	Значение при 10 %	Доверительный интервал при 10 %	Значение при 15 %	Доверительный интервал при 15 %
1	(0,91; 1,06)	1,17	(1,09; 1,26)	1,27	(1,18; 1,37)
2	(0,90; 1,05)	1,21	(1,13; 1,30)	1,32	(1,24; 1,42)
3	(0,90; 1,05)	1,16	(1,09; 1,25)	1,28	(1,20; 1,38)
4	(0,94; 1,10)	1,23	(1,14; 1,32)	1,31	(1,22; 1,41)
5	(0,96; 1,12)	1,25	(1,17; 1,35)	1,34	(1,24; 1,44)

Таблица 5

Доверительные интервалы статистики U в зависимости от числа испытуемых

Параметр p (процент искажений)	$N = 500$		$N = 1000$	
	Значение статистики	Доверительный интервал	Значение статистики	Доверительный интервал
0	1,02	(0,80; 1,30)	0,95	(0,84; 1,12)
10	1,31	(1,04; 1,70)	1,26	(1,11; 1,42)
15	1,49	(1,22; 1,81)	1,44	(1,27 ; 1,61)
20	1,60	(1,31 ; 1,96)	1,43	(1,28; 1,59)

Таблица 6

Доверительные интервалы статистики W в зависимости от числа испытуемых

Параметр p (процент искажений)	$N = 500$		$N = 1000$	
	Значение статистики	Доверительный интервал	Значение статистики	Доверительный интервал
0	1,01	(0,91; 1,13)	0,98	(0,90; 1,05)
10	1,21	(1,08; 1,34)	1,21	(1,13 ; 1,30)
15	1,33	(1,20 ; 1,48)	1,32	(1,24; 1,42)
20	1,39	(1,26; 1,52)	1,33	(1,25; 1,42)

ема выборки чувствительность статистик согласия повышается, но всегда взвешенная статистика W обладает большей чувствительностью, чем статистика U .

Было проведено исследование зависимости доверительных интервалов статистик согласия от трудности заданий, в которые вносились искажения. В этом случае задания для внесения искажений отбирались не случайным образом, а с учетом

ются количеством искажений, чем какими-либо свойствами самих заданий. При этом взвешенная статистика W обладает гораздо меньшей вариабельностью, чем статистика U .

- Задание может быть идентифицировано как не согласующееся с моделью измерения уже при 10–15 % искажений. При меньшем количестве искажений оценки статистик согласия (и точечные, и интервальные) не отличаются существенно от базовых, и незначительные искажения в матрице могут быть объяснены случайными причинами.

- Точечные оценки статистик согласия не могут служить надежными индикаторами согласия. Например, значение статистики W при 5 % искажений, равное 1,10, не принадлежит базовому интервалу (0,92; 1,08).

Однако интервальная оценка (1,02; 1,19) не позволяет идентифицировать задание как не согласующееся с моделью.

В табл. 3, 4 представлены доверительные интервалы статистик согласия (базовые, при 10 % и при 15 % искажений) отдельно по пяти заданиям, в которые вносились искажения.

Из таблиц видно, что взвешенная статистика W обладает большей чувствительностью, чем статистика U : взвешенная статистика W идентифицирует задание как не согласующееся с моделью измерения уже при 10 % искажений, в то время как статистика U — при 15 %. Далее, интервал (0,8; 1,2), рекомендуемый в качестве допустимого в [5], не может служить надежным индикатором согласия. Например, значение статистики W задания 3 при 10 % искажений, равное 1,16 (табл. 4), принадлежит указанному промежутку, но задание не может быть признано адекватным модели, если принять во внимание интервальную оценку: доверительный интервал (1,09; 1,25) не пересекается с базовым интервалом (0,90; 1,05). Вместе с тем, значения статистики U всех заданий при 10 % искажений больше 1,2 (табл. 3), однако задания не могут быть признаны неадекватными модели с учетом их интервальных оценок.

На следующем этапе исследования изучалась зависимость доверительных интервалов статистик согласия от различных факторов — числа испытуемых, числа заданий, параметров их распределений. В табл. 5, 6 приведены результаты одного из экспериментов, имеющего целью исследовать зависимость доверительных интервалов от числа испытуемых. Данные приведены по одному из заданий — заданию 2 из табл. 3, 4. Матрица тестирования для $N = 500$ была получена из матрицы для $N = 1000$ путем случайного отбора.

Из анализа таблиц следует, что чем больше объем выборки, тем уже доверительные интервалы статистик согласия. Далее, с увеличением объ-

Доверительные интервалы статистики U для реальных и модельных данных

Номер задания	Реальные данные		Модельные данные	
	Значение статистики	Реальный доверительный интервал	Значение статистики	Модельный доверительный интервал
17	1,24	(1,08; 1,39)	1,07	(0,94; 1,21)
29	1,31	(1,13; 1,55)	1,06	(0,97; 1,16)
24	1,20	(1,12; 1,27)	1,01	(0,93; 1,11)
35	1,19	(1,10; 1,29)	0,96	(0,88; 1,04)

Доверительные интервалы статистики W для реальных и модельных данных

Номер задания	Реальные данные		Модельные данные	
	Значение статистики	Реальный доверительный интервал	Значение статистики	Модельный доверительный интервал
17	1,08	(1,03; 1,13)	1,01	(0,96; 1,06)
29	1,10	(1,04; 1,16)	1,04	(0,99; 1,09)
24	1,17	(1,11; 1,22)	1,02	(0,97; 1,08)
35	1,16	(1,10; 1,21)	0,99	(0,94; 1,05)

их трудности: пять самых трудных, пять самых легких и пять заданий средней трудности. Базовые доверительные интервалы статистики U для экстремальных ситуаций (все задания — трудные или все — легкие) оказались значительно шире соответствующих интервалов для неэкстремальных ситуаций. При этом базовые интервалы статистики W практически совпадают во всех ситуациях, т. е. не зависят от трудности заданий. Этот факт легко объясним, так как статистика U (2) очень чувствительна к экстремально неожиданным ответам, когда хорошо подготовленный участник не выполняет легкое задание или, наоборот, плохо подготовленный — выполняет трудное. Именно эта особенность статистики U вызвала необходимость построения ее взвешенной версии W . Таким образом, снова приходим к выводу, что взвешенная статистика W является более надежным индикатором согласия, чем ее простая версия U .

Таким образом, использование интервальных оценок общих статистик согласия позволяет идентифицировать задания, не адекватные модели измерения. Причины неадекватности могут быть самые различные — от ошибок в разработке задания до нарушений в процедуре проведения тестирования. Одной из причин может быть нарушение принципа одномерности теста: задания, не согласующиеся с моделью, могут измерять другую латентную переменную по сравнению с остальными заданиями теста (подробнее об анализе заданий см., например, в [3]).

Технология идентификации заданий, не согласующихся с моделью измерения

В результате проведенного теоретического анализа и экспериментального исследования была разработана следующая технология идентификации заданий, не согласующихся с моделью измерения, в реальных данных тестирования:

- генерируется матрица тестирования, аналогичная реальной, но с полным соответствием модели измерения;
- с использованием модельной матрицы строятся доверительные интервалы общих статистик согласия (2) и (3) для всех заданий. Это — модельные доверительные интервалы;
- строятся доверительные интервалы общих статистик согласия (2) и (3) для всех заданий по реальной матрице тестирования. Назовем эти интервалы реальными доверительными интервалами;
- сравнивая реальные доверительные интервалы с модельными, делаем вывод о согласии или несогласии с моделью ответов на данное задание теста.

В табл. 7, 8 приведены данные, иллюстрирующие применение технологии идентификации "плохих" заданий. Анализировались реальные данные массового тестирования: число испытуемых $N = 983$, число заданий $I = 38$.

Для большинства заданий точечные оценки статистик согласия U и W оказались близки к 1, и реальные доверительные интервалы пересекаются с модельными, что говорит о хорошем согласии с моделью ответов на большинство заданий теста. В таблицах представлены данные по четырем заданиям. Задания 24 и 35 идентифицируются как не согласующиеся с моделью измерения, поскольку доверительные интервалы их статистик согласия не пересекаются с соответствующими модельными интервалами. Задания 17 и 29, напротив, находятся в хорошем согласии с моделью измерения. При этом точечные оценки статистики U заданий 17 и 29 превышают точечные оценки этой статистики заданий 24 и 35. Таким образом, интервальные оценки статистик согласия позволяют более точно идентифицировать задания, не согласующиеся с моделью измерения, чем это делают их точечные оценки.

Заключение

В статье рассматривается возможность использования бутстреп-метода построения доверительных интервалов для построения интервальных оценок статистик согласия заданий теста. Предложена процедура построения доверительных ин-

тервалов общих статистик согласия, на основе которой разработана технология идентификации заданий, не согласующихся с моделью измерения. Данная технология была апробирована на реальных данных тестирования и показала свою эффективность. Все этапы этого процесса автоматизированы путем разработки специального программного обеспечения.

Можно наметить два направления дальнейшей работы. Во-первых, это обобщение полученных результатов на случай политомических моделей Раша. Во-вторых, можно применить описанную технологию для анализа ответов испытуемых и идентификации тех из них, чьи результаты вызывают сомнения в их достоверности.

Список литературы

1. Wright B. D., Masters G. N. Rating Scale Analysis. Rasch Measurement. Chicago: Mesa Press, 1982. 206 p.
2. Нейман Ю. М., Хлебников В. А. Введение в теорию моделирования и параметризации педагогических тестов. М.: Прометей, 2000. 169 с.
3. Карданова Е. Ю. Моделирование и параметризация тестов: основы теории и приложения. М.: Федеральный центр тестирования, 2008. 304 с.
4. Карданова Е. Ю., Карпинский В. Б. Использование эксперимента на модели Раша для выявления недостоверных результатов тестирования // Информационные технологии. 2009. № 6. С. 74–79.
5. Wright B. D., Linacre J. M. Reasonable Mean-Square Fit Values // Rasch Measurement Transactions. 1994. Vol. 8. N 3. P. 370. URL: <http://www.rash.org/rtm/contents.htm>
6. Dimitrov D. M., Smith R. M. Adjusted Rasch Person-Fit Statistics // Journal of Applied Measurement. 2006. N 7 (2). P. 170–183.
7. Smith R. M., Schumacker R. E., Bush H. J. Using Item Mean Squares to Evaluate Fit to the Rasch Model // Journal of Outcome Measurement. 1998. N 2. P. 66–78.
8. Karabatsos G. A Critique of Rasch Residual Fit Statistics // Journal of Applied Measurement. 2000. Vol. 1. N 2. P. 152–176.
9. Эфрон Б. Нетрадиционные методы многомерного статистического анализа. М.: Финансы и статистика, 1988. 263 с.
10. Wang W.-C., Chen C.-T. Item Parameter Recovery, Standard Error Estimates, and Fit Statistics of the Winsteps Program for the Family of Rasch Models // Educational and Psychological Measurement. 2005. Vol. 65. N 3. P. 376–404.

УДК 378.22

И. Б. Герасимова, канд. техн. наук, доц.,
Уфимский государственный
авиационный технический университет,
e-mail: tarot_gfera@mail.ru

Системный анализ процессов передачи и усвоения знаний на основе когнитивного моделирования

Рассматривается проблема системного анализа процесса передачи и усвоения знаний на основе когнитивных моделей для выявления факторов, влияющих на эффективность этого процесса, в целях повышения качества подготовки специалиста.

Ключевые слова: системный и когнитивный анализ, знания, обучение, модель, преподаватель, обучаемый, образовательная система

Введение

Для достижения высокого качества подготовки специалистов требуется глубокое изучение и совершенствование не только всех этапов процесса обучения, но и методик преподавания, а также знание психологии как обучающего, так и обучаемого. Неполное изучение этих процессов как в целом, так и отдельных их частей связано прежде

всего с тем, что они относятся к классу слабоструктурированных процессов. Высокую эффективность в изучении данного класса процессов показал подход, основанный на понятии когнитивной карты (когнитивных моделей), часто называемый когнитивным анализом [1]. При этом когнитивная карта (модель), построенная по результатам обработки как качественных четких, так и субъективно оцененных нечетких данных, по существу представляет собой знания эксперта или группы экспертов о процессах, описывающих соответствующую предметную область.

В настоящей работе рассматривается проблема качественного анализа процессов передачи и усвоения знаний на основе предложенной триадноструктурированной [2] когнитивной модели, позволяющей выявить отдельные факторы, влияющие на эффективность этих процессов.

Когнитивная модель и анализ процесса передачи знаний преподавателем

Обратимся сначала к анализу процесса передачи знаний преподавателем. В качестве когнитивной модели данного процесса предлагается следующая триада: Передаваемые знания X_1 — Методика передачи знаний X_2 — Переданные знания X_3 , представленная на рис. 1, которая параметризуется и конкретизируется в числовой форме на основе экспертных оценок.

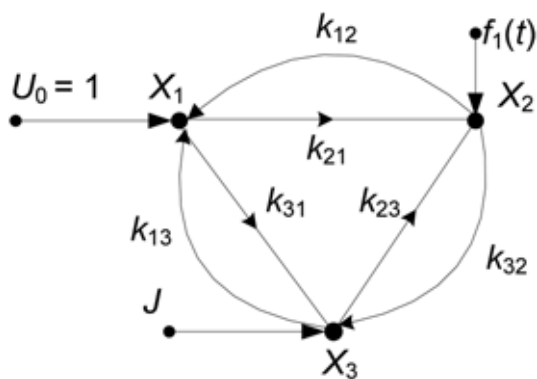


Рис. 1. Когнитивная модель оперативной передачи знаний преподавателем

На рис. 1 X_i — концепты; U_0 — объем преподаваемых знаний (информации), равный 1; J — внешние инвестиции в учебный процесс, изменяющиеся в пределах 0...1; $f_1(t)$ — внешние возмущения, имеющие самый различный характер (личный, коллективный, психофизический, социальный и т. д.), действующие отрицательно на процесс передачи знаний и изменяющиеся в пределах 0,1...0,2; k_{ij} — параметры, характеризующие особенности процесса передачи знаний преподавателем.

Концепт "Передаваемые знания X_1 " — это конкретный объем знаний, передача которого спланирована, законспектирована и представлена в форме, удобной для усвоения. При этом общее количество (100 %) передаваемых знаний принимается равным 1. Учитывая профессиональный уровень преподавателя, можно полагать, что передаваемая информация отражает качественную и семантическую (содержательную) сторону знаний.

Концепт "Методика передачи знаний X_2 " может быть представлен в виде методики чтения лекций, методики изложения учебного материала (в учебниках, учебных пособиях, электронных конспектах лекций, видеофильмах, демонстрационных слайдах), методики проведения практических занятий и лабораторных работ, а также в виде способов и форм передачи знаний, полученных вербальным путем от своих коллег или из других источников.

Концепт "Переданные знания X_3 " — это конкретный объем знаний, представленный в числовой форме (в процентах, в долях) и направленный на обучаемого для усвоения полученных знаний.

Дадим следующую трактовку параметрам k_{ij} :

k_{12} — коэффициент, характеризующий уровень самоанализа и самоконтроля преподавателя за правильностью и полнотой передаваемой мысли, за четкостью и логичностью речи;

k_{21} — коэффициент, характеризующий объем передачи знаний в устной форме;

k_{31} — коэффициент, характеризующий объем передачи знаний на практических занятиях;

k_{23} — коэффициент, характеризующий объем передачи знаний в процессе выполнения лабораторных работ;

k_{13} — коэффициент, характеризующий уровень самоконтроля и умения преподавателя передавать знания в форме решения задач по стандартной или инновационной методике;

k_{32} — коэффициент, характеризующий уровень самоконтроля и умения преподавателя проводить лабораторные работы по стандартной или инновационной методике.

Нетрудно убедиться, что общий коэффициент передачи знаний определяется следующим выражением:

$$K_{п.з} = \frac{X_3}{X_1} = \mu \frac{k_{21} + k_{31}k_{23}}{1 - (k_{12}k_{21} + k_{13}k_{31} + k_{23}k_{32}) + (k_{12}k_{31}k_{23} + k_{21}k_{32}k_{13})}, \quad (1)$$

где μ — информативность и полнота передаваемых знаний (проверяется экспертно или экспериментально) ($\mu = 0,9...1,1$), которая определяется способностью преподавателя.

Предполагается, что если пара коэффициентов (например, $k_{12}k_{21}$) имеет знак "+", то в системе возникает положительная (стимулирующая) обратная связь и в знаменателе выражения (1) они фигурируют со знаком "-", уменьшая запасы статической устойчивости системы. Если один из коэффициентов в паре имеет знак "-", а другой "+", то в системе возникает отрицательная (тормозящая, стабилизирующая) обратная связь, увеличивающая запасы статической устойчивости. При этом в знаменателе выражения (1) эта пара коэффициентов фигурирует со знаком "+". Тройка положительных коэффициентов: $k_{12}k_{23}k_{31}$ и $k_{21}k_{13}k_{32}$ образует в системе отрицательную обратную связь.

Отметим, что инвестиции J должны приводить к повышению общего коэффициента передачи знаний. Самостоятельная активная работа преподавателя, глубокое осознание своей интеллектуальной деятельности должны в общем случае привести к повышению коэффициентов передачи знаний в целом, так как при этом начинают проявляться знания, полученные в процессе научных исследований и превосходящие объем знаний учебной рабочей программы, что обеспечивает значение коэффициента $\mu > 1$. Ясно, что наличие внешних помех $f_1(t)$ ведет к уменьшению объема передаваемых знаний U_0 .

Целью системного анализа предложенной когнитивной модели, представленной на рис. 1, является раскрытие механизма передачи знаний и познание влияния отдельных факторов на эффективность передачи знаний преподавателем, рабо-

тающим в автономном режиме без обратной связи с аудиторией слушателей.

При формировании числовых значений коэффициентов данной модели была учтена общепринятая структура читаемой дисциплины, которая выражается в том, что 60–70 % ($k_{21} = 0,7$) знаний передаются в форме чтения лекций, 10–15 % ($k_{31} = 0,1$) знаний передаются в форме практических занятий, а 20–25 % ($k_{32} = 0,2$) знаний передаются в форме лабораторных работ. Кроме того учитывалось, что работа по самоконтролю своих действий требует от преподавателя определенных усилий. По оценкам экспертов-психологов при чтении лекции на процесс самоконтроля требуется 15–30 % от общих усилий ($k_{12} = 0,2$). Самоконтроль при проведении практических занятий (решение задач) и лабораторных работ, особенно по инновационным методикам и технологиям, требует большего внимания и психологического напряжения ($k_{13} = k_{32} = 0,3$).

Пример 1. Учитывая указанный выше смысл параметров модели рис. 1, по экспертным оценкам возможны следующие числовые значения коэффициентов передачи знаний:

$$k_{21} = 0,7; k_{31} = 0,1; k_{23} = 0,2; k_{12} = 0,2; \\ k_{13} = 0,3; k_{32} = 0,3.$$

В этом случае при $\mu = 1$ общий коэффициент передачи знаний преподавателем равен 1,024. Здесь высокий результат достигается при высоком уровне преподавания, владения учебным материалом и методикой преподавания.

Проанализируем ряд возможных ситуаций в процессе передачи знаний преподавателем.

Ситуация 1. Если преподаватель невнимателен, т. е. у него отсутствует самоконтроль на лекциях, $k_{12} = 0$, то результат передачи знаний падает: $X_2 = 0,85$. Если преподаватель плохо владеет учебным материалом, $k_{12} = -0,2$ (отрицательная обратная связь), то результат падает еще больше до $X_2 = 0,73$.

Ситуация 2. Если преподаватель невнимателен на практических занятиях, $k_{13} = 0$, то результат передачи знаний падает до значения $X_2 = 0,91$. Если же преподаватель плохо ведет практические занятия, не знает материала, $k_{13} = -0,2$ (отрицательная обратная связь), то результат получается еще хуже: $X_2 = 0,81$.

Ситуация 3. Преподаватель невнимателен на лабораторных работах, т. е. $k_{32} = 0$. В этом случае результат передачи знаний падает до значения $X_2 = 0,87$. В случае, когда преподаватель плохо ведет лабораторные занятия, так как не имеет опыта, $k_{32} = -0,3$, результат становится еще хуже: $X_2 = 0,76$.

Ситуация 4. Крайний случай, когда все коэффициенты самоконтроля отрицательные, то

$X_2 = 0,62$. Здесь результат передачи знаний крайне низок. При внесении инвестиций $J = 1$ в данную ситуацию, т. е. с отрицательными коэффициентами самоконтроля (показатель слабого преподавания), получаем результат $X_2 = 0,61$, т. е. инвестиции пользу не приносят. При внесении инвестиций $J = 1$ в ситуацию, когда все коэффициенты самоконтроля положительные, результат будет $X_2 = 1,61$, т. е. инвестиции эффективны, если их вкладывать в качественный процесс преподавания.

Ситуация 5. При появлении помех для процесса передачи знаний преподавателем $f(t) = -0,2$, эффективность этого процесса падает на 25 % ($X_2 = 0,75$). Для полной компенсации этой помехи $f(t) = -0,2$ нужны инвестиции $J = 0,47$ в форме дополнительных занятий и их оплаты.

Таким образом, на основе рассмотренной триадной когнитивной модели проведен качественный анализ процесса передачи знаний преподавателем и выяснены основные причины, влияющие как на повышение, так и на понижение его качества.

Когнитивная модель и анализ процесса усвоения знаний обучаемым

Рассмотрим теперь процесс усвоения знаний обучаемым (студентом) на основе когнитивной модели в форме триады: Переданные знания преподавателем X_2 — Мотивация к усвоению знаний X_4 — Усвоенные знания обучаемым X_5 , которая отражает процесс восприятия знаний обучаемым (рис. 2). При этом преподаватель и обучаемый находятся в одностороннем отношении, т. е. оперативная обратная связь от студента к преподавателю отсутствует.

Здесь триада модели восприятия знаний параметризуется и конкретизируется следующим образом:

- концепт X_2 — переданные знания преподавателем;

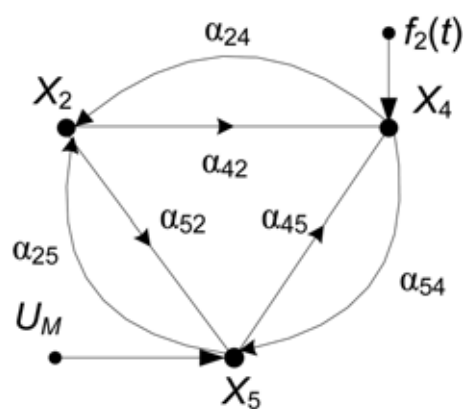


Рис. 2. Когнитивная модель оперативного восприятия знаний студентом

- концепт X_5 — мотивация, которая поддерживает процесс усвоения знаний;
- концепт X_4 — усвоенные обучаемым знания, которые отражают конечный результат передачи знаний.

Параметры α_{ij} модели имеют следующую трактовку:

α_{42} — физическое и интеллектуальное восприятие студентом передаваемых знаний, их осознание;

α_{24} — уровень самоконтроля за восприятием знаний, уровень концентрации внимания;

α_{52} — понимание важности и роли получаемых знаний в общей модели компетенций знаний;

α_{25} — уровень самоконтроля за содержанием знаний и за поддержкой мотивации к обучению;

α_{45} — желание поддержать полученные знания и увеличить их за счет самостоятельной работы;

α_{54} — уровень самоконтроля за объемом полученных знаний за счет самообразования;

U_M — внешняя поддержка мотивации, изменяемая при моделировании от 0 до 1;

$f_2(t)$ — возможные помехи для восприятия знаний.

Общий коэффициент усвоения знаний определяется выражением:

$$K_{y.3} = \frac{X_4}{X_2} = \lambda \frac{\alpha_{42} + \alpha_{52}\alpha_{45}}{1 - (\alpha_{24}\alpha_{42} + \alpha_{23}\alpha_{32} + \alpha_{43}\alpha_{34}) + (\alpha_{24}\alpha_{52}\alpha_{45} + \alpha_{23}\alpha_{42}\alpha_{54})}, \quad (2)$$

где λ — коэффициент усвоения знаний, определяемый творческими способностями обучаемого. Для общего случая $\lambda = 1$.

При высокой самоорганизации обучаемого, постоянном пополнении знаний из книг, бесед с преподавателями, друзьями, специалистами коэффициент усвоения может быть значительно повышен, т. е. $\lambda > 1$.

Целью когнитивной модели обучаемого (рис. 2) является раскрытие механизма передачи знаний и познание влияния отдельных факторов на эффективность усвоения знаний.

При построении числовой модели восприятия знаний обучаемым предполагалось, что его усилия, затрачиваемые на усвоение знаний и на самоконтроль, пропорциональны объему передаваемых знаний, а потому значения коэффициентов когнитивной модели преподавателя (см. рис. 1) и обучаемого (см. рис. 2) симметрично равны друг другу при различной их смысловой (семантической) нагрузке, т. е.

$$\alpha_{ij} = k_{ij} \text{ и } \alpha_{ji} = k_{ji}.$$

Пример 2. Учитывая сказанное выше, по результатам экспертных оценок возьмем следующие числовые значения коэффициентов передачи знаний:

$$\alpha_{42} = 0,7, \alpha_{52} = 0,1, \alpha_{54} = 0,2, \alpha_{24} = 0,2, \\ \alpha_{25} = 0,3, \alpha_{54} = 0,3.$$

Тогда можно получить следующие значения: $X_2 = 1,024$ и $X_4 = 1,049$, т. е. при высокой организации процесса получения знаний, высокой мотивации, студент достигает максимального результата в усвоении знаний.

Ситуация 1. Если обучаемый невнимателен на лекциях: $\alpha_{24} = 0$, то результат усвоения знаний падает: $X_4 = 0,87$. Если же обучаемый плохо слушает лекцию, отвлекается, не понимает: $\alpha_{24} = -0,2$ (отрицательная обратная связь), то результат падает еще больше: $X_4 = 0,74$.

Ситуация 2. Преподаватель хорошо ведет занятия: $X_2 = 1,024$, а у обучаемого все коэффициенты самоконтроля отрицательные, т. е. он невнимателен на занятиях, тогда результат усвоения знаний низкий: $X_4 = 0,87$.

Ситуация 3. Преподаватель плохо ведет занятия, невнимателен: $X_2 = 0,62$, а у обучаемого все коэффициенты самоконтроля отрицательные, тогда результат получается самый низкий: $X_4 = 0,38$.

Ситуация 4. Преподаватель плохо ведет занятия, его результат: $X_2 = 0,62$, а у обучаемого все коэффициенты самоконтроля положительные, тогда получается результат: $X_4 = 0,63$. Отсюда следует, что хороший обучаемый никогда не сможет на занятиях получить знания, которые больше знаний слабого преподавателя. Эти значения равны значениям, когда преподаватель хорошо ведет занятия у слабого обучаемого, т. е. эти ситуации равны.

Ситуация 5. Если для обучаемого во время занятий возникают помехи, т. е. $f(t) = -0,2$, а преподаватель хорошо ведет занятия: $X_2 = 1,24$, то результат получается средний: $X_4 = 0,77$. Для компенсации этих помех нужна мотивация $M = 0,46$, например, в форме дополнительных или самостоятельных занятий в этом случае результат усвоения знаний повысится до $X_4 = 1,04$.

Если преподаватель плохо ведет занятия, а у хорошего обучаемого есть мотивация к самостоятельному дополнительному обучению, то результат $X_4 = 1,21$, т. е. мотивация помогает ему в восполнении знаний.

Таким образом, на основе системного анализа когнитивной модели усвоения знаний обучаемым выявлены основные причины, влияющие как на повышение, так и на понижение объема усвоенных знаний.

Когнитивная модель обучения в диалоговом режиме

Обратимся теперь к варианту объединения рассмотренных двух триадных когнитивных моделей, представленному на рис. 3.

Особенностью данной обобщенной когнитивной модели является то, что индивидуальные занятия ведутся в тесном диалоге преподавателя с обучаемым, например на семинарах. В результате когнитивной модели образуются дополнительные замкнутые контуры, которые либо усиливают, либо уменьшают эффект как передачи, так и восприятий знаний.

Пример 3. По экспертным оценкам возьмем следующие значения коэффициентов модели обучения с обменом знаний: $k_{21} = 0,7$, $k_{31} = 0,1$, $k_{23} = 0,2$, $k_{12} = 0,2$, $k_{13} = 0,3$, $k_{32} = 0,3$, $\alpha_{42} = 0,7$, $\alpha_{52} = 0,1$, $\alpha_{54} = 0,2$, $\alpha_{24} = 0,2$, $\alpha_{25} = 0,3$, $\alpha_{54} = 0,3$, $\alpha_{53} = \alpha_{35} = 0$.

При новой форме взаимодействия преподавателя и обучаемого значение $X_2 = 1,024$ и увеличится до $X_2 = 1,571$, а значение $X_4 = 1,049$ и увеличится до $X_4 = 1,203$. При правильной организации индивидуального взаимодействия преподавателя и обучаемого эффект передачи и приема знаний увеличивается.

Ситуация 1. Если обучаемый невнимателен на занятиях, $\alpha_{24} = 0$, то значение $X_2 = 1,024$ увеличится лишь до $X_2 = 1,186$, а значение $X_4 = 0,87$ — до $X_4 = 0,906$. Здесь и далее повышению результативности помогает взаимный контроль преподавателя и обучаемого. Результат усвоения знаний падает, если обучаемый плохо слушает лекцию, $\alpha_{24} = -0,2$. В этом случае значение $X_2 = 1,024$ уменьшается до $X_2 = 0,952$, а значение $X_4 = 0,74$ уменьшается до $X_4 = 0,729$. Следовательно, недисциплинированный обучаемый своим поведением мешает проведению занятий и усвоению знаний.

Ситуация 2. Преподаватель хорошо ведет занятия, $X_2 = 1,024$, и у обучаемого все коэффициенты самоконтроля отрицательные. Тогда значение $X_2 = 1,024$ уменьшится до $X_2 = 0,894$, а значение $X_4 = 0,63$ — до $X_4 = 0,608$. Следовательно, самоконтроль очень важная деталь в организации усвоения знаний. Старания преподавателя напрасны, если обучаемый не желает учиться и слабо контролирует процесс усвоения знаний.

Ситуация 3. Если преподаватель плохо ведет занятия и у обучаемого все коэффициенты самоконтроля отрицательные, то значение $X_2 = 0,615$ уменьшится до $X_2 = 0,563$, а значение $X_4 = 0,38$ не изменится. Это плачевный результат.

Ситуация 4. Если преподаватель плохо ведет занятия, а у обучаемого все коэффициенты самоконтроля положительные, то значение $X_2 = 0,615$

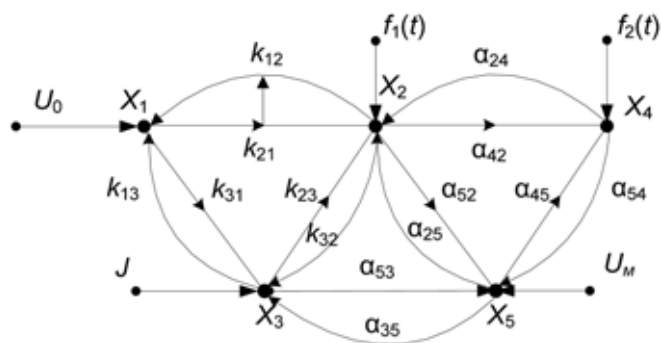


Рис. 3. Когнитивная модель процесса обучения с обменом знаний между преподавателем и студентом в диалоговом режиме

увеличится до $X_2 = 0,791$, а значение $X_4 = 0,63$ уменьшится до $X_4 = 0,605$. Хороший обучаемый не сможет получить качественные знания у слабого преподавателя.

Ситуация 5. При появлении помех у обучаемого $f(t) = -0,2$, значение $X_2 = 1,24$ увеличится до $X_2 = 1,44$, а $X_4 = 0,77$ — до значения $X_4 = 0,89$. Даже при наличии помех диалоговый режим работы для мотивированного обучаемого увеличивает результативность занятий. Для компенсации этих помех нужна мотивация $M = 0,2$, тогда $X_2 = 1,863$, $X_4 = 1,214$. Повышенная мотивация обучаемого всегда способствует повышению эффективности как коллективного, так и индивидуального обучения.

Пусть на когнитивной модели процесса обучения в диалоговом режиме имеются еще две связи между X_3 и X_5 , причем здесь α_{53} отражает роль методики на формирование мотивации к процессу обучения, а α_{35} — уровень понимания роли методики в процессе получения знаний.

Ситуация 6. Если при $X_2 = 1,11$ и $X_4 = 1,24$ сделать самоконтроль за изучением методики, равным 0 или отрицательным $\alpha_{35} < 0$, то X_2 упадет до 0,96, а X_4 до 1,04, т. е. самоконтроль за усвоением методики обучения является важным моментом в процессе усвоения знаний.

Таким образом, системный анализ процессов преподавания и усвоения знаний на основе когнитивных моделей позволяет полностью раскрыть механизм передачи и усвоения знаний, установить некоторые закономерности, выявить основные причины, способствующие ухудшению или улучшению процессов.

Заключение

На основании проведенного системного анализа процессов передачи и усвоения знаний можно сделать следующие выводы.

1. Инвестиции в слабого преподавателя положительных результатов не приносят.

2. Инвестиции должны вкладываться в качественный процесс преподавания.

3. Хорошо мотивированный и организованный обучаемый получает максимальный результат обучения только у хорошего, высококвалифицированного преподавателя.

4. Хорошо мотивированный обучаемый никогда не может в процессе обучения получить больше знаний, чем знания, которыми владеет слабый преподаватель. Мотивация заставляет обучаемого самостоятельно пополнять знания.

5. Индивидуальная форма обучения в тесном диалоге с преподавателем наиболее эффективна для передачи и усвоения знаний.

Конечно, эти закономерности можно получить и из практики коллектива опытных преподавателей, но здесь они научно обоснованы на базе системного анализа на основе когнитивных моделей.

Из этих закономерностей вытекает такой научно-обоснованный и объективный вывод, что в

систему высшего образования выгодно вкладывать большие финансы:

1) в состав высококвалифицированных преподавателей и в их повышение квалификации;

2) в талантливых обучаемых, имеющих высокую мотивацию к обучению;

3) в повышение качества материально-технической базы, в средства обучения, методические средства, так как современные технические средства обучения повышают у обучаемых мотивацию к обучению.

Список литературы

1. Когнитивный анализ и управление развитием ситуаций (CASC'2009) // Труды междунар. конф. (17—19 ноября 2009 г., г. Москва). М.: ИПУ РАН, 2009. 288 с.

2. Гузайров М. Б., Ильясов Б. Г., Герасимова И. Б. Системный подход к анализу сложных систем и процессов на основе триад // Проблемы управления. 2007. № 5. С. 32—38.

УДК 621.391

С. В. Дворников¹, д-р техн. наук, доц.,

С. С. Дворников², студент,

М. А. Коноплев¹,

e-mail: practicsdv@yandex.ru,

¹ Военная академия связи, Санкт-Петербург,

² Санкт-Петербургский государственный политехнический университет

Алгоритм распознавания сигналов радиосвязи на основе симметрических матриц

Предлагается метод распознавания сигналов с близкой частотно-временной структурой на основе обработки их представлений в виде симметрических матриц. Обосновывается порядок формирования векторов признаков. Приводятся результаты теоретических и экспериментальных исследований, подтверждающих контрастность сформированных признаков пространства. Рассматриваются рекомендации по практической реализации разработанного подхода.

Ключевые слова: классификация сигналов, распознавание образов, симметрические матрицы, вектор признаков, признаки пространства, спектральное разложение матриц

Введение

Для комплексов радиомониторинга важной задачей является своевременное определение вида модуляции сигналов контролируемых систем свя-

зи. Традиционно ее решение основывалось на элементах статистической теории распознавания образов [1]. Однако применение данной методологии связано с необходимостью значительного объема обрабатываемого материала, что не всегда возможно в условиях современной радиоэлектронной обстановки, характеризуемой значительным количеством кратковременных излучений. Следовательно, необходим поиск принципиально новых подходов, позволяющих эффективно работать с фрагментами радиосигналов длительностью не более доли секунды. В связи с этим настоящая статья посвящена теоретическим вопросам разработки алгоритма распознавания радиосигналов на фоне шумов высокой интенсивности при ограничениях на длительность обрабатываемой реализации.

Формирование векторов признаков распознаваемых радиосигналов

Как правило, в качестве признаков распознавания радиосигналов принято рассматривать их статистические или структурные особенности, которые в своей совокупности характеризуют идентифицируемый объект [1]. В общем случае, применительно к рассматриваемой задаче, операцию распознавания можно интерпретировать как последовательное сравнение вектора признаков принятой реализации Y с каждым из L векторов эталонных описаний $X^{(1)}$. Следовательно, чем контрастней векторы признаков, тем выше веро-

ятность процедур правильного принятия решений в процессе классификации радиосигналов.

С практической точки зрения целесообразно определить условия формирования признаков таким образом, чтобы при минимизации размерности выборки сигнала обеспечивалась максимизация достоверности распознавания.

Учитывая ограничения, накладываемые на объем исходной статистики, целесообразно при формировании векторов признаков использовать принцип алгоритма Карунена — Лоева, позволяющего работать с относительно небольшими фрагментами сигналов [1]. Применение в указанном алгоритме декоррелирующих преобразований предоставляет возможность получать контрастные вектора признаков из первичного материала, в качестве которого могут выступать, например, результаты измерения одного из параметров сигналов или же просто их временные отсчеты, полученные с выхода аналого-цифрового преобразователя.

Так, операцию линейного преобразования \mathbf{A} исходного пространства признаков $X = (x_1, x_2, \dots, x_R)$ в новое пространство $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_I)$ можно представить как

$$\tilde{X} = \mathbf{A}X. \quad (1)$$

Если в качестве столбцов матрицы преобразования \mathbf{A} определить собственные вектора $\{\lambda_R\}$ общей ковариационной матрицы \mathbf{M}^* распознаваемых совокупностей образов, то в этом случае операция (1) будет декоррелирующей, а сама ковариационная матрица примет вид:

$$\mathbf{M}^* = \mathbf{A}^T \mathbf{M} \mathbf{A} = \mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_R \end{pmatrix}. \quad (2)$$

Реализация рассмотренного подхода стала возможной благодаря свойствам подобия матриц [2], согласно которому произвольная матрица \mathbf{V} подобна матрице \mathbf{W} в том случае, если существует невырожденная трансформирующая матрица \mathbf{S} такая, что

$$\mathbf{W} = \mathbf{S}^{-1} \mathbf{V} \mathbf{S}. \quad (3)$$

А поскольку никакие матрицы из двух разных классов не являются подобными друг другу, притом, что матрицы из одного класса эквивалентности имеют одинаковые собственные значения, то это свойство можно использовать при формировании признаков пространств различных классов сигналов.

Между тем в работе [3] обосновывается использование в выражении (3), так называемых симметрических матриц, особенностью которых является повышенная устойчивость к сторонним возмущениям аддитивного характера $\mathbf{V}_{ш}$. Согласно указанному свойству собственные значения диагонализующей матрицы \mathbf{M}_c и собственные значения возмущенной матрицы $\mathbf{V}_{c+ш} = \mathbf{V}_c + \mathbf{V}_{ш}$ связаны соотношением [2]:

$$|\hat{\lambda}_i - \lambda_i| \leq \|\mathbf{\Lambda}\| \|\mathbf{\Lambda}^{-1}\| \|\mathbf{V}_{ш}\| = \chi(\mathbf{\Lambda}) \|\mathbf{V}_{ш}\|, \quad (4)$$

где λ_i — i -е собственное значение матрицы \mathbf{V}_c ; $\hat{\lambda}_i$ — i -е собственное значение матрицы $\mathbf{V}_{c+ш}$; $\mathbf{\Lambda}$ — матрица собственных векторов матрицы \mathbf{V}_c ; $\|\cdot\|$ — знак матричной нормы; $\chi(\mathbf{\Lambda}) = \|\mathbf{\Lambda}\| \|\mathbf{\Lambda}^{-1}\|$ — число обусловленности матрицы $\mathbf{\Lambda}$.

Согласно выражению (4) следует, что разброс собственных значений возмущенной матрицы $\mathbf{V}_{c+ш}$ зависит не только от значений элементов матрицы $\mathbf{V}_{ш}$, но и от числа обусловленности $\chi(\mathbf{\Lambda})$. Причем если $\chi(\mathbf{\Lambda})$ мало (близко к единице), то малые возмущения могут изменить собственные значения. Однако эти изменения будут ограничены величиной того же порядка, которая соответствует этим возмущениям. Если же значение $\chi(\mathbf{\Lambda})$ велико, то малые возмущения приводят к сравнительно большим изменениям собственных значений.

Экспериментально в работе [3] установлено, что после нормирования матрицы ее несколько первых собственных значений (наибольших по отношению к остальным) имеют порядок от единиц и десятков. Следовательно, если число обусловленности мало, то воздействие возмущающей матрицы не окажет существенного влияния на первые наибольшие собственные значения матрицы \mathbf{V}_c . Поскольку значение $\chi(\mathbf{\Lambda}) = 1$ имеют только унитарные матрицы, то необходимо матрицу собственных векторов $\mathbf{\Lambda}$ привести к унитарному виду. А унитарной матрицей собственных векторов обладают именно симметрические матрицы.

Наиболее простым способом получения симметрической матрицы из исходного вектора признаков является вычисление автокорреляционной матрицы \mathbf{V}_a [2]:

$$\mathbf{V}_a = \mathbf{X} \mathbf{X}^T, \quad (5)$$

где \mathbf{X} и \mathbf{X}^T — соответственно входной вектор признаков и его транспонированная копия.

Таким образом, предлагается в качестве векторов признаков радиосигналов использовать собственные значения симметрических матриц. А сами матрицы формировать на основе временных отсчетов радиосигналов.

Результаты эксперимента

В интересах проверки теоретических положений был проведен эксперимент по распознаванию сигналов с близкой частотно-временной структурой. В качестве альтернативных классов рассматривались сигналы фазовой манипуляции (ФТ) со скоростями 600, 550, 500, 450 и 400 бод, из временных отсчетов которых формировались симметрические матрицы согласно выражению (5). В качестве исходных данных выбирались первые 128 временных отсчетов каждого из сигналов, определяющих свой класс. Затем, согласно формуле (2), рассчитывались собственные значения матриц эталонных радиосигналов λ , из которых формировались векторы признаков \mathbf{X} . Между тем, проведенные эксперименты показали, что с возрастанием отношения сигнал/шум (ОСШ) собственные числа, определяющие векторы признаков, изменялись по-разному, поэтому было предложено в качестве конечных векторов признаков рассматривать только наиболее устойчивые значения λ . С этой целью синтезировались тестовые сигналы при различных значениях ОСШ, после чего для них строились матрицы, для которых рассчитывались собственные значения $\{\lambda_n\}$ и вычислялась дисперсия $\sigma_{\lambda_n}^2$ для каждого λ_n . Устойчивыми считаются те, значение которых удовлетворяет равенству

$$\sigma_{\lambda_n}^2 > G\lambda_n, \quad (6)$$

где G — заданный порог; λ_n — порядковое собственное значение.

Так на рис. 1 показаны упорядоченные модули собственных значений сигнала ФТ 500 бод при ОСШ, равном 15 и 5 дБ, с нанесенным пороговым уровнем G . В эксперименте определено $G = 2\lambda$. Затем из совокупности помехоустойчивых значений $\{\lambda_n\}$ (в эксперименте n от 44 до 128) строились векторы признаков сигналов.

Окончательное решение об отнесении распознаваемого сигнала к одному из альтернативных классов осуществлялось согласно критерию

$$\min \left\{ \sum_{l=1}^L \frac{1}{N} \sum_{n=1}^N \left| {}^l\lambda_n - {}^s\lambda_n \right| \right\}, \quad (7)$$

где ${}^l\lambda_n$ — собственные значения, являющиеся признаками эталонов; L — число эталонов (в эксперименте их 5); ${}^s\lambda_n$ — собственные значения, являющиеся признаками распознаваемого сигнала.

Вероятность правильного распознавания $P_{\text{расп}}$ оценивали методом Монте-Карло отдельно для каждого распознаваемого сигнала для значений ОСШ $P_c/P_{\text{ш}}$ в диапазоне от 0 до 20 дБ. Так, на рис. 2 представлен график для сигнала ФТ 500 бод.

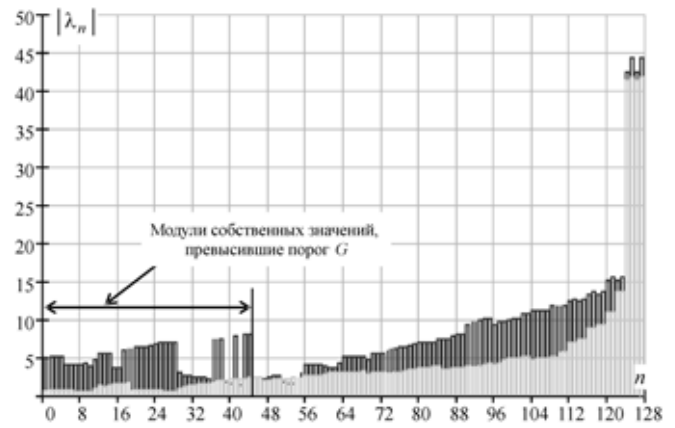


Рис. 1. Модули собственных значений сигнала ФТ 600 бод при ОСШ 15, 10 и 5 дБ

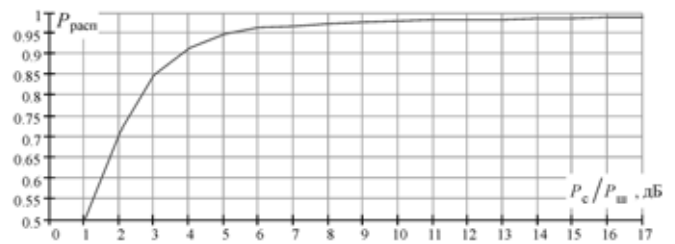


Рис. 2. Зависимость вероятности распознавания от ОСШ

Заключение

Разработанный подход расширяет возможности существующих методов распознавания радиосигналов. Его применение возможно при решении задач обработки контрольных сигналов удаленных датчиков, а также в практических приложениях масс-спектрометрии и хроматографии. Между тем следует признать, что вопрос поиска оптимального решения при построении автоматов классификации сигналов все еще остается открытым. Так, в работе [4] предложены методы распознавания на основе билинейных форм представления радиосигналов, поэтому дальнейшую работу, по мнению авторов, следует вести именно в этом направлении. Полученные теоретические и практические результаты будут интересны специалистам обработки сигналов в области радиотехники, спектрометрии и приборостроения.

Список литературы

1. Фукунага К. Введение в статистическую теорию распознавания образов: пер. с англ. / Под ред. М. А. Айзермана. М.: Наука, 1977. 319 с.
2. Хорн Р., Джонсон Ч. Матричный анализ: пер. с англ. М.: Мир, 1989.
3. Сауков А. М., Симонов А. Н., Мотков В. И. Способ распознавания радиосигналов. Патент РФ № 2231118 МПК7 G 06 K 9/00 от 20.06.04 г.
4. Дворников С. В., Алексеева Т. Е. Распределение Алексеева и его применение в задачах частотно-временной обработки сигналов // Информация и Космос. 2006. № 3. С. 921.

CONTENTS

Barsky A. B. *Tri-Dimensional Screen for Computer's Objects of "Life" Modeling* 2

No experimental tested hypothesis of the possibility of the volume screen building on base of achievements in the area "transparent electronics" is discussed. This screen is realized by mean of the screen films packet, with the help of which such of tri-dimensional object will be reflected on few levels. The volume screen is necessary for reflecting "life" computer's models, scenes productions, for reacting objects of intellectual reflecting in industrial or business control, in historical and cultural reconstructions, in virtual architecture and sculpture. It is point out on an importance of tri-dimensional memory model use in computer graphic for object control in real time by mean of logical neural networks. The work of the volume screen in Cartesian and spherically systems of coordinates is examined.

Keywords: transparent screen, packet of screen films, "life" modeling, tri-dimensional memory, logical neural network, Cartesian and spherically coordinates

Kohanenko I. K. *Neural Network and Morphological Computer Analyses of Images with Use of Fractals* 7

Algorithms of the computer automatic analysis of images are proved on the basis of fractals clusters. Rationality proves by using the fractals, leading to the synthesis of systems with high stability due to maintenance of a high degree of a variety. It shows that the kernel of the image is what distinguishes it from others and that contains small number of clusters. In this connection the morphological analysis and the analysis is rationalized by networks of Kohonen. The example of recognition of the image of a site of a lunar surface is resulted.

Keywords: a fractal, recognition, a video information, cluster, invariant, intensity, neural network, the morphological analysis

Krivtsov O. A., Korikov A. M. *Methods of Tracking Head in a Video Sequence: Classification and Choice* 13

We present an analytical survey of head tracking methods, their classification, advantages and disadvantages and brief recommendations on choice of method for implementation in a head tracking system.

Keywords: survey, head tracking, video frames, classification, choice of method

Norenkov I. P. *Genetic Algorithms of Decision Search in Ontological Knowledge Bases* 20

The problem of optimal learning route and supporting them electronic textbooks synthesis is considered. The genetic algorithms are suggested for this problem solving.

Keywords: decision search systems, ontology, AND/OR graphs, genetic algorithms, learning route

Borodaschenko A. Yu., Glotov D. V., Bochkov S. M. *The Estimation Algorithm of a RSS Newsfeed Content Analysis* 25

The article deals with content analysis algorithm in RSS feeds presented permits to increase the completeness and accuracy of analyzing and decision making process.

Keywords: content analysis, RSS feeds, text mining, data mining

Vasinev D. A., Zhusov D. L., Tsyngaev M. S., Vatulin I. Yu. *Proposals to Improve the Security of Ethernet-Networks from the Link-Layer Attacks* 29

The articles describes the existing methods of protection against attacks of this class and propose a variant of protection based on the validation of filling ARP-table on servers, client workstations, and communications equipment.

Keywords: information security, link-layer attacks, denial of service

Massel A. G. Methodological Approach to Intelligent Support of Energy Security Problem Research 32

Methodological approach to organization of intelligent support of energy security research is considered. Information technology for intelligent support based on integration of such kinds of information modeling as ontological, cognitive and event modeling. Meaning of these kinds of modeling is disclosed and examples of information models are shown. Results of research presented in this paper are performed with partial support by grants of RFBR № 08-07-00172 and RFBR № 10-07-00264 and grant of Presidium RAS Program № 2.29.

Keywords: energy security, emergency, ontology modeling, event modeling, cognitive modeling, intelligent IT-environment

Tokarchuk A. M. Development ActiveModel Pattern for Programming Internet-Applications by the Object-Relational Mapping Method 37

A new pattern ActiveModel, as a technology to develop Internet applications using object-relational reflection. The pattern is based on abstraction from the MySQL database and allows you to reduce time and cost of application development.

Keywords: application, database, object-relational mapping, design patterns

Dimitrienko Yu. I., Dimitnenko O. Yu. Nonrigid Cluster Model for Analysis of Dynamic Data in Economics 43

The article suggests the further development of mass sales market model based on the clustering of objects in the special multidimensional space and also on the constructing of the continuous model of clusters' movements. The system of equations is elaborated for describing the movement of nonrigid clusters in the multidimensional space. The article presents analysis of results for numerical calculation of sales market parameters dynamics and the comparison with experimental data.

Keywords: mathematical modeling, marketing influence, nonrigid cluster, multidimensional space, motion equation

Novitsky V. O., Karpov V. I. The Methodology of the Tudy and Modeling of Complex Control Systems for Grain Sector Agribusiness Enterprises and Companies 50

The need for improved control and management systems for grain storage and processing enterprises and companies on the basis of integrated information and control systems defines the need for specialized research and modeling methodology for this class of complex control systems. The proposed methodology is based on the use of dialectic, systemic, cybernetics and other approaches and requires the description of the system, forming the framework of goals, the construction and analysis of diagrams of cause-effect relationships, identify problem tasks, drawing a conceptual model and the formation of a mathematical model of a system based on interaction with the bank of models control and management for enterprises of grain farming.

Keywords: research methodology and modeling, complex control system, framework of goals, chart cause and effect relationships, feedback loops, conceptual model, criterion, bank of models, grain storage and processing companies

Zack Yu. A. Mathematical Methods in the Tasks of Processing Results of Professionals and Applicants Qualifications Testing 57

Problem solving for the test results evaluation tasks is offered by vector optimization techniques. Techniques of developing and calculating the values of complex compromise criterion of evaluating of testing results are offered. It allows presenting the results of test set as a certain real number. Recommendations are given for tests design.

The results obtained here might be used for scoring in computer training systems, as well as in the procedure of employee and student selection, when the number of available working and training places is not enough for all the applicants.

Keywords: vector optimization problem, results of testing, putting in good order

Kardanova E. Yu., Kardanov R. S. Construction of Interval Estimates for Fit Statistics at Large-Scale Testing Results Analysis under the Rasch Models 64

Point estimates of fit statistics conventionally used at large-scale testing results analysis can not provide a reliable criterion of adequacy of empirical data with the given measurement model due to different reasons. The ar-

article focuses on the method of interval estimates construction. Bootstrap method of computing confidence intervals is proposed to solve the task of interval estimation. The technology of identification of the tasks that don't fit the measurement model is also highlighted in the article.

Keywords: testing, Rasch models, fit statistic, critical value, bootstrap method, interval estimation

Gerasimova I. B. *System Analysis of Transfer and Knowledge Assimilation Processes Based on the Cognitive Modeling* 70

The paper is concerned with the problem of transfer and knowledge assimilation system analysis based on cognitive models to identify factors influencing the efficiency of this process, to improve the quality of specialist training.

Keywords: systems and cognitive analysis, knowledge, training, model, lecturer

Dvornikov S. V., Dvornikov S. S., Konoplev M. A. *Radio Signals Identification Algorithm Based on Symmetrical Matrix* 75

The method of signals identification with similar frequency-time structure based on processing their introductions in the form of symmetrical matrix is offered. The forming order of criterion vector is proved. The results of theoretical and experimental research confirmed visibility created area of criterion are demonstrated. The practical realization recommendations are considered.

Keywords: signal classification, image identification, symmetrical matrix, criterion vector, area of criterion, matrix spectral decomposition

Адрес редакции:

107076, Москва, Стромьинский пер., 4

Телефон редакции журнала **(499) 269-5510**

E-mail: it@novtex.ru

Дизайнер *Т.Н. Погорелова*. Технический редактор *О. А. Ефремова*.

Корректор *Т. В. Пчёлкина*

Сдано в набор 09.07.2010. Подписано в печать 18.08.2010. Формат 60×88 1/8. Бумага офсетная. Печать офсетная.

Усл. печ. л. 9,8. Уч.-изд. л. 11,54. Заказ 647. Цена договорная.

Журнал зарегистрирован в Министерстве Российской Федерации по делам печати, телерадиовещания и средств массовых коммуникаций.

Свидетельство о регистрации ПИ № 77-15565 от 02 июня 2003 г.

Отпечатано в ООО "Подольская Периодика"

142110, Московская обл., г. Подольск, ул. Кирова, 15