РОССИЙСКАЯ АКАДЕМИЯ НАУК

ВСЕРОССИЙСКИЙ ИНСТИТУТ НАУЧНОЙ И ТЕХНИЧЕСКОЙ ИНФОРМАЦИИ
(ВИНИТИ)

НАУЧНО • ТЕХНИЧЕСКАЯ
ИНФОРМАЦИЯ

Серия 2. ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ И СИСТЕМЫ

ЕЖЕМЕСЯЧНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ СБОРНИК

Издается с 1961 г. № 5 Москва 2012

ИНФОРМАЦИОННЫЕ СИСТЕМЫ

УДК 001.102 : [025.4.03 : 002.2]

О. Л. Голицына, Н. В. Максимов, О. В. Окропишина, В. И. Строгонов

Онтологический подход к идентификации информации в задачах документального поиска*

Предложен подход к определению онтологии и множества операций как инструмента формирования и количественно оцениваемого соотнесения идентифицирующих образов объектов предметной области в условиях диалектической взаимосвязи предметного, понятийного и знакового пространств. Онтологическое представление образа объекта в вычислительной среде соответствует объектно-ориентированному подходу и включает не только свойства, но и поведение. На практике такой подход позволит автоматизировать процессы динамического реформулирования и соотнесения поисковых образов запросов и документов на основе их приведения к общему понятийно-терминологическому контексту.

Ключевые слова: *онтологии, операции над онтологиями, информационный поиск, идентификация содержания, общая теория систем, теория графов*

ОНТОЛОГИИ КАК ИНСТРУМЕНТ ФОРМАЛИЗАЦИИ ЗНАНИЙ

Эффективность поиска и адекватность передачи смысла в системе кругооборота информации, обеспечивающей расширенное воспроизводство знаний, связаны с выбором или введением специальной терминологии. Одним из перспективных направлений в области формализации знаний, которое дает возможность их компьютерной обработки, являются онтологии. Наиболее общее их определение дано в [1]: «Онтология – это спецификация концептуализации». В [2] оно существенно уточняется: онтология определяется как «спецификация концептуализации на уровне эксплицитных знаний, зависящая от предметной области или задачи, для которой она предназначена». Под концептуализацией понимается строгое описание системы понятий, объектов и других сущностей, а также отношений, связывающих их друг с другом. С другой стороны, концептуализация – это абстрактное и, таким образом,

* Работа выполнена при поддержке РФФИ, грант № 11-09-13128 офи-м-2011-РЖД.

упрощенное видение мира, который мы хотим представить для достижения какой-то цели. Концептуализация расчленяет целостную область знаний, выделяет из этой области отдельные объекты и отношения, то, что с точки зрения конкретной цели будет адекватным операбельным объектом, заменяющим саму предметную область (ПрО) в процессе исследований.

В общем виде структура онтологии представляет собой набор элементов четырех категорий: понятия, отношения, аксиомы, отдельные экземпляры. Понятия рассматриваются как концептуализации класса всех представителей некой сущности. Понятия могут быть связаны различного рода отношениями. Аксиомы задают условия соотнесения категорий и отношений.

Однако в «инженерной» области онтологии сильно различаются. Например, с точки зрения предмета концептуализации выделяют прикладные онтологии, онтологии области знания, общие (родовые) и репрезентационные онтологии (онтологии метауровня) [3]. Отдельно выделяют лингвистические онтологии (проекты WordNet, MikroKosmos, РуТез и др.), фиксирующие понятия (слова) вместе с их языковыми свойствами и отношениями (синонимия, гипонимия и т. п.).

Существует и два подхода к концептуализации: интенсиональный и экстенсиональный. Экстенсиональный подразумевает, что каждое понятие и отношение может исчерпывающе описываться перечислением индивидуальных сущностей [4]. При интенсиональном подходе предлагается идентифицировать понятия не через их перечисление, а через их внутренние свойства и характеристики [3].

Показателен подход, предложенный в [5], где онтология – это файл с метаданными, предназначенными для аннотирования содержания ресурса и сервисов в нескольких аспектах, который должен включать следующие типы метаданных:

1) описание объектов, их свойств и отношений между ними (Domain ontologies);

2) описание задач и процессов, их свойств и отношений (Task ontologies);

3) описание атрибутов знания (Quality ontologies);

4) характеристика значимости контента (Value ontologies);

5) история обращений отдельного субъекта к информационным ресурсам (Personalization ontologies);

6) описание причин накопления, оценок использования (Argumentation ontologies).

Определение, конструктивное с точки зрения формализации, предложено в [6]: «Онтология – это набор определений (на формальном языке) фрагмента декларативных знаний, ориентированный на совместное многократное использование различными пользователями. В онтологии вводятся термины, типы и соотношения (аксиомы), описывающие фрагмент знания».

В задачах разметки текстов учет семантических категорий, описанных в онтологии, позволяет сделать идентификацию более точной, уменьшить неоднозначность. Такая семантическая разметка в дальнейшем позволяет проводить семантический анализ текста, статистические исследования, извлекать межъязыковые соответствия [7].

В целом, основная цель создания онтологий – сделать возможным *разделение (sharing)* и *повторное использование (reuse) знаний* и, в частности, обеспечить *интероперабельность* между несовместимыми инструментальными средствами в среде моделирования деятельности человека. Практическое применение онтологии – это соглашение об использовании общего словаря в согласованной и логичной манере [8].

В данной статье с общесистемных позиций будет рассмотрен семиотический подход к определению онтологии, позволяющий учесть следующую важную для идентификации обрабатываемых объектов особенность машинной обработки: в вычислительной среде как абстрактные, так и конкретные объекты, как классы, так и их экземпляры будут представлены знаками.

ВВЕДЕНИЕ В ЗАДАЧУ ИДЕНТИФИКАЦИИ ПрО

Любой целенаправленный процесс синтеза нового знания (как и любых новых объектов) явно или неявно включает отдельный подпроцесс – выбор из множества доступных только нужных «строительных блоков». Причем выбор здесь – это не столько собственно извлечение конкретного объекта из среды, сколько способ и процесс отождествления объекта либо с образцом (в частности, с аналогом), либо с «нишей», им заполняемой. Однако в большинстве процессов разумной деятельности выбор реализуется не по приведенной «натурной» схеме, а опосредованно – через информационное взаимодействие, где сопоставляются образы: образ объекта-кандидата на отбор с образом целевого объекта (возможно, существующего только гипотетически). Такого рода образами обычно являются описания объектов или процессов (научные, технические публикации, математические модели, проектная документация и т. п.), в той или иной форме представляющие те или иные свойства и поведение объекта с полнотой и точностью, обусловленной характером решаемой практической задачи. Подобный подход позволяет реализовать эффективную (по используемым ресурсам) процедуру: объект с потенциально бесконечным разнообразием свойств и связей заменяется в операции сопоставления компактным хорошо формализованным идентификатором, свойства которого представлены в форме, адекватной операционной среде.

При этом и в человеческом сознании, и в вычислительной среде чаще сравниваются не непосредственно описания, а их «производные» по тем аспектам рассмотрения предметной области, которые отвечают конкретной задаче потребителя – субъекта, формирующего новое знание[1]. Такое положение обусловлено следующим:

[1] Необходимо четко различать описание (зафиксированное на носителе) собственно знания и его внешнее представление (описательный, поисковый образ) как по назначению, так, соответственно, по построению и использованию. Если описание представляет существо объекта «изнутри» (как устроен,

1. С точки зрения общей теории систем [9] знания, как и описания любых объектов, по своей природе системны: состав элементов, их свойства и взаимосвязи для каждого случая (аспекта рассмотрения, разных задач и т. д.) будут определяться своим законом композиции. То есть для отдельной ПрО может быть определено множество систем $S^i = <M^i, A^i, R^i, Z^i>$, где M^i – множество объектов ПрО, A^i – множество характеристических атрибутов, R^i – множество отношений, Z^i – закон композиции, $i = \overline{1,n}$.

2. Любое описание (как информационный объект[2] – материальная форма передачи знаний) является не более чем некоторой моделью рассматриваемой ПрО, отражающей только те её свойства и поведение, которые являются существенными для конкретной задачи и обстоятельств, связанных с созданием, целями и методами использования.

3. Описания имеют семиотическую природу, т.е. представлены знаками, конкретные значения которых определяются некоторой конкретной концептуальной схемой ПрО. В целом построение описания (операционного образа) отвечает схеме, представленной на рисунке.

4. Для того чтобы описание было адекватно воспринято и понято, оно должно быть представлено понятиями той степени общности/детальности и зафиксировано теми терминами, которые будут иметь одинаковые значения как для генератора, так и для приемника сообщения.

Отметим, что успешность восприятия может быть оценена состоянием, когда термины полученного сообщения имеют эквивалентное отображение в понятийно-терминологической системе приемника и образуют логически непротиворечивую структуру. Успешность понимания может быть оценена состоянием, когда с помощью построенной по полученному сообщению понятийной системы можно воспроизвести гипотетический объект, обладающий предполагаемыми свойствами и поведением (в частности, свойствами объекта-оригинала).

Процесс восприятия/понимания по существу является обратным по отношению к представленному на рисунке: производится восстановление объекта по его образу. Очевидно, что в простейшем случае такое восстановление построено по принципу замещения знаков и/или их комбинаций соответствующими «конструкционными» блоками, выбираемыми из хранилища (в частности, памяти субъекта) в соответствии с контекстом, образуемым системой понятий и знаков, а также принятой парадигмой, категориальной системой, конкретными целями и задачами.

Достаточно простая и очевидная схема, приведенная на рисунке, позволяет тем не менее явно выделить принципиальное отличие организации когнитивных процессов в живых системах от искусственных систем сохранения и управления знаниями. Живым системам свойственны целеполагание и целеустремленность. При этом цель, формируемая в среде субъекта после получения внешнего воздействия (сигнала), существенным образом будет определяться контекстом[3], образуемым тесно связанными со средой знаниями субъекта: сигнал не будет воспринят, если у субъекта не возникнет ассоциаций с его личным знанием. То есть данные, представляющие внешнее воздействие, изначально будут связаны с определенным контекстом (хотя, скорее всего, разным даже для разных когнитивных состояний субъекта) и будут обрабатываться совместно, включая, если необходимо, и изменение контекста.

Уместно отметить и особенности, сказывающиеся на условиях формирования контекста в системах сохранения знаний в разные исторические периоды. Для периода индустриального этапа развития общества (собственно, и породившего информационные системы и информационную деятельность как класс) было характерно сравнительно небольшое число глобальных проектов и приоритетных отраслей, причем каждый проект (или отрасль) мог централизованно управляться и реализовываться в рамках одной (чаще всего специально создаваемой) организационной структуры, включающей и целевую подготовку профессиональных кадров. Это обеспечивало контролируемое «позадачное» формирование контекста и в сознании отдельного субъекта – будущего решателя отраслевых задач, и в общественном сознании – общего контекста, который, явно или неявно, определял развитие направления.

как достигаются целевые показатели и т. п., что обеспечивает возможность его воспроизведения, изменения, применения), то идентификатор (поисковый образ) представляет объект «извне» – основными свойствами, целевыми показателями (точнее, наименованиями и значениями), что обеспечивает его узнавание путем соотнесения с соответствующими характеристиками и, таким образом, выделение из множества других.

[2] Знание определяется как редукция информации, представляемой как суперпозиция возможных состояний информационного объекта в различных предметных областях. Информация выполняет коммуникационные функции, обеспечивающие взаимодействия в предметной области, и собственные, реализующие её действенность. С точки зрения механизма взаимодействия информация как структурированный неатомарный объект должна иметь «интерфейсные» элементы, сопряженные с объектами взаимодействия [10].

[3] Используя терминологию автоматизированных ИПС, можно сказать, что контекст является запросом на отбор данных из внешней среды. То есть «информационный ресурс» субъекта формируется «позадачно»: получаемые данные непосредственно связываются с их целевым использованием. Ресурсы же хранилищ обобществленного знания (библиотеки, АИПС), напротив, формируются по принципу накопления документов в основном с целью их последующего использования для построения нового знания, возможно, в областях, далеких от изначальной. Можно сказать, что создание нового знания всегда локализовано и конкретно обусловлено, в то время как его использование всегда вариантно и рассредоточено. При этом для обеспечения широты и многоаспектности использования, в том числе для решения еще не поставленных задач, знание, целостное «исторически» и методологически, публикуется «по частям» (в соответствии с этапами работ, ограничениями, свойственными видам публикаций, правовыми аспектами и т. д.) и приводится в форме с максимальным абстрагированием (что, в принципе, должно увеличивать его универсальность). И, более того, для обеспечения полноты поиска для решения будущих задач справочно-поисковые средства хранилищ (картотеки библиотек, поисковые образы документов в АИПС) представляют эти знания на достаточно общем уровне и в унифицированной форме.

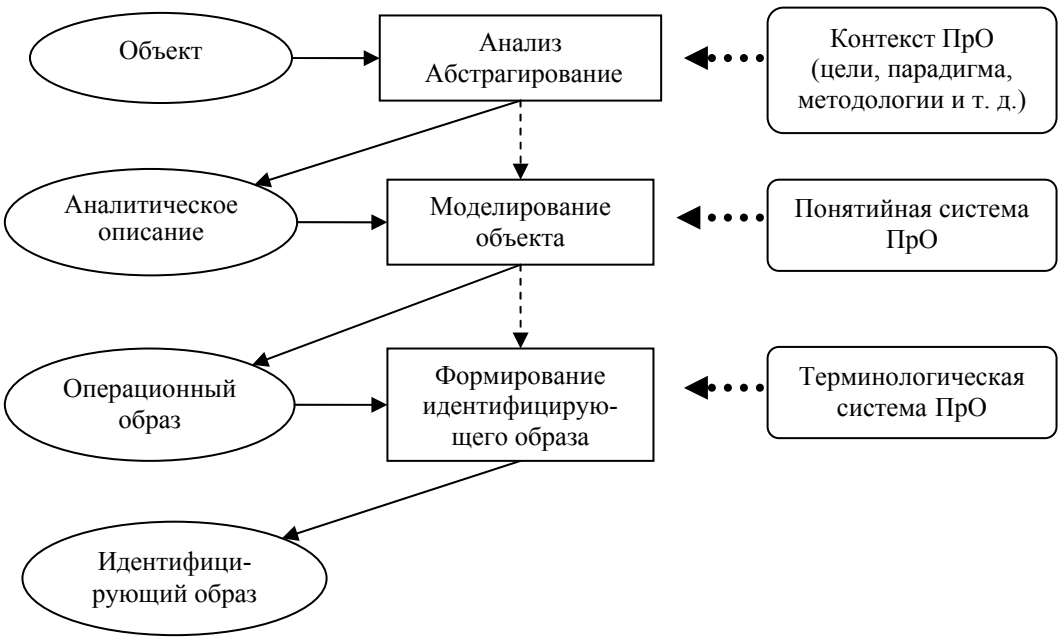

Рис. Построение идентифицирующего образа

Современное существенное увеличение объемов опубликованного знания, диверсификация исследований, преобладающая децентрализация управления привели к необходимости управления собственно контекстом: уже недостаточно перечислить термины темы исследования – необходимо уточнить их значение, недостаточно назвать задачу – необходимо привести цели, условия и парадигму её решения.

Исходя из отмеченных особенностей развития коммуникационных процессов в когнитивных системах, можно сказать, что основой для обеспечения эффективности передачи и расширенного воспроизводства знаний является согласованность и устойчивость используемых понятийно-терминологических систем. В этом смысле учитывающий диалектику процесса познания онтологический подход к формированию образа (дескриптивного описания) знаний представляется конструктивным, поскольку позволяет идентифицировать не только сам объект (информацию, знания, предметы и т. п.) на основе его свойств, но и его «поведение» в концептуальном и историческом пространствах.

Технологически такой подход не является принципиально новым: еще в первых поколениях автоматизированных ИПС для построения поискового образа – идентифицирующего описания объекта (документа, описывающего предмет, процесс и т. д.) – использовались указатели роли и связи, что позволяло в дескриптивной форме отражать существо объекта, в том числе с разных точек зрения. Сформированный образ вполне адекватно отражал и свойства, и связи ПрО, однако использованный контекст описания оставался вне образа. То есть в этом случае не учитываются специфические *особенности и динамика взаимообусловленности предметного, понятийного и знакового пространств*. В современных же условиях, как отмечалось выше, отдельная ПрО может быть представлена разными понятийными и терминологическими множествами (свойство системности) и с разной степенью полноты и детальности (свойство голографичности информации [11]). Отметим, что и операции над такими описаниями должны обеспечить их преобразование и взаимное сопоставление через приведение к целевому контексту, определяемому выбираемыми каждым отдельным субъектом целями и характером решения и соответствующими обстоятельствами конкретной ПрО. Представление образа в вычислительной среде соответствует объектно-ориентированному подходу и будет включать не только данные (свойства), но и методы (поведение) – способы построения, использования и т. д.

Отметим, что термин «онтология» в дальнейшем изложении будет использован в узком смысле поисковых задач (а не задач искусственного интеллекта) – как адаптивный композиционный идентификатор, отражающий основные свойства объекта в сопоставимой форме и не предполагающий построение объектов, обладающих новизной.

ОПРЕДЕЛЕНИЕ ОНТОЛОГИИ ПрО

Рассматривая онтологию ПрО с определенной выше узкой точки зрения (идентификации содержания документа при информационном поиске), будем представлять её как совокупность трех систем: функциональной, понятийной и терминологической.

Функциональная система («рабочий интерфейс» онтологии в деятельности субъекта) представляет объекты ПрО и отношения между ними средствами знакового уровня. Эти отношения имеют функциональную окраску, так как определяют суще-

ственные с точки зрения задач пользователя [12] способы и характер совместного существования и использования объектов.

Однако функциональная система не может рассматриваться вне связи со сложившейся в ПрО *на текущий момент времени* системой понятий, которая обычно фиксируется в форме тезаурусов, рубрикаторов, классификационных схем (логико-семантический базис онтологии). В понятийной системе объектами (также представленными средствами знакового уровня) являются устойчивые понятия ПрО, а набор отношений ограничен родо-видовыми и ассоциативными.

Терминологическая система в онтологии отражает свойства естественного языка (ЕЯ) на уровне знаков – терминов ПрО, которые могут быть связаны отношениями эквивалентности (синонимии) и включения (образования словосочетаний). В качестве термина выступает отдельное слово или словосочетание ЕЯ (а также шифр классификации), которое может быть использовано для описания понятия или объекта в рамках заданной ПрО.

Для определения (установления) взаимосвязи этих систем (а также и их влияния друг на друга в процессе развития) может быть предложена простая операция тождества элементов на знаковом уровне.

Таким образом, некоторая онтология ПрО формально может быть определена как

$$O =< S_f, S_c, S_t, \equiv >,$$

где S_f – функциональная система,

S_c – понятийная система,

S_t – терминологическая система,

$\equiv$ – операция сопоставления[4] элементов различных систем на уровне знаков, обеспечивающая их тождество в функциональной, понятийной и терминологической системах.

Определенная таким образом онтология также обладает свойствами системы (с точки зрения общей теории систем), т.е. для отдельной предметной области (в том числе и отдельного её элемента) может быть построено столько онтологий, сколько может быть аспектов рассмотрения ПрО[5]: $O^i =< S^i_f, S^i_c, S^i_t, \equiv >$. Однако с целью упрощения выкладок, если не будет явной необходимости, мы будем опускать соответствующий индекс.

Отметим, что в качестве функционального (и терминологического) компонента может выступать и система понятий ПрО. В этом случае её понятийной системой будет система метапонятий (это отвечает, например, определению «метаонтологии» [13], где на верхнем уровне есть лишь три категории: «объект», «процесс» и «роль»).

Рассмотрим компоненты, входящие в приведенное выше определение онтологии, более подробно.

Функциональная система определяется как

$$S_f =< M_f, A_f, R_f, Z_f >,$$

где M_f – множество знаковых описаний объектов ПрО,

A_f – множество характеристических атрибутов,

R_f – множество функциональных отношений,

Z_f – закон композиции, в соответствии с которым выбрано конкретное системное основание $\{M_f, A_f, R_f\}$.

Со структурной точки зрения[6] функциональная система может быть представлена помеченным взвешенным направленным мультиграфом

$$MG_f =< V_f, X_f >,$$

где $V_f = \left(v_1^f, v_2^f, \ldots, v_{p_f}^f\right)$ – множество вершин, $p_f = \left|V_f\right|$,

$X_f = \left(x_1^f, x_2^f, \ldots, x_{q_f}^f\right)$ – множество дуг, $q_f = \left|X_f\right|$.

Каждая дуга задается тройкой:

$$x_k^f = \left(v_i^f, v_j^f, w_k^f\right),$$

где v_i^f – вершина начала дуги, v_j^f – вершина завершения дуги, w_k^f – вес дуги (идентификатор функционального отношения).

Направленный мультиграф может иметь дуги $x_k^f = \left(v_i^f, v_j^f, w_k^f\right)$ и $x_l^f = \left(v_i^f, v_j^f, w_l^f\right)$, $w_k^f \neq w_l^f$, но не может при этом содержать дугу $\left(v_j^f, v_i^f, w^f\right)$.

Понятийная система определяется как

$$S_c =< M_c, A_c, R_c, Z_c >,$$

где M_c – множество знаковых описаний понятий ПрО,

A_c – множество характеристических атрибутов знаковых описаний,

R_c – родо-видовые и ассоциативные отношения,

Z_c – закон композиции, в соответствии с которым выбрано конкретное системное основание $\{M_c, A_c, R_c\}$.

Со структурной точки зрения понятийная система представлена помеченным взвешенным направленным графом и (с точки зрения введения операций над онтологиями) описана как

$$G_c =< V_c, X_c >,$$

где $V_c = \left(v_1^c, v_2^c, \ldots, v_{p_c}^c\right)$ – множество вершин, $p_c = \left|V_c\right|$,

$X_c = \left(x_1^c, x_2^c, \ldots, x_{q_c}^c\right)$ – множество дуг, $q_c = \left|X_c\right|$.

Каждая дуга задается тройкой:

$$x_k^c = \left(v_i^c, v_j^c, w_k^c\right),$$

где v_i^c – вершина начала дуги, v_j^c – вершина завершения дуги, w_k^c – вес дуги (идентификатор отношения).

Направленный граф не содержит симметричных дуг, любые две вершины могут быть соединены только одной дугой.

[4] Так как S_f, S_c, S_t являются развивающимися системами (т.е. онтологии могут иметь разные состояния понятийной и знаковой систем в момент сравнения), в операциях над онтологиями они должны быть приведены к общему основанию.

[5] Данное определение расширяет ранее упомянутое в [2].

[6] Закон композиции Z_f и множество характеристических атрибутов A_f представляют интенсиональный способ задания онтологии. В данной статье рассматриваться не будут.

Терминологическая система определяется как

$$S_t =< M_t, A_t, R_t, Z_t >,$$

где M_t – множество терминов ПрО,

A_t – множество характеристических атрибутов терминов,

R_t – отношения эквивалентности и включения,

Z_t – закон композиции, в соответствии с которым выбрано конкретное системное основание $\{M_t, A_t, R_t\}$.

Со структурной точки зрения терминологическая система описывается n-связным графом $G_t = \bigcup_{i=1}^{n} G_i^t$, где каждая компонента связности G_i^t представляет собой полный граф (эквивалентность), дерево (включение) или результат операции объединения полных графов и деревьев (при наличии общих вершин).

ОПЕРАЦИИ НАД ОНТОЛОГИЯМИ

Исследования в области онтологических представлений знаний – интенсивно, хотя и достаточно эклектично развивающееся направление. Это в значительной степени относится и к введению операций над онтологиями. В многочисленных работах предлагаются разнообразные множества операций (обстоятельные обзоры представлены, например, в [14 – 18]), которые достаточно условно можно отнести к следующим классам:

1) аналоги теоретико-множественных операций (пересечение, объединение, разность);

2) операции извлечения и удаления фрагментов онтологий для их использования при создании новых;

3) операции проверки логики онтологий, перевода онтологий на другой формальный язык и т. п.;

4) операции – средства поддержки онтологий в инструментальных системах (создание, сохранение и т. д.), комбинаций информационных объектов (документов, сообщений), представленных онтологиями, на которых субъект может построить новое знание (задачи вывода нового знания на имеющихся онтологиях не рассматриваются).

В качестве основных операций над онтологиями (на структурном уровне) рассмотрим операции из 1-го и 2-го классов – бинарные операции объединения и пересечения и унарные операции построения аспектного представления и масштабирования онтологий, с помощью которых можно, в частности, синтезировать новые онтологии, отражающие предметную область в аспекте, заданном пользователем.

Поскольку операции над онтологиями рассматриваются в рамках одной ПрО, будем считать, что исходные онтологии имеют общие понятийную и терминологическую системы.

Для формализации операций над онтологиями необходимо определить функцию подобия, вычисляющую меру соответствия при сравнении понятий и связей в исходных онтологиях с целью выделения семантически схожих элементов [17].

Введем функцию подобия $sim(a,b)$, обладающую следующими очевидными свойствами:

1. $sim(a,b) \in [0;1]$;

2. $sim(a,b) = 1 \rightarrow a = b$ (элементы a, b тождественны);

3. $sim(a,b) = 0 \rightarrow a \neq b$ (элементы a, b различны);

4. $sim(a,a) = 1$ (свойство возвратности);

5. $sim(a,b) = sim(b,a)$ (свойство симметричности).

Функция подобия рассчитывается для элементов функциональных систем пары исходных онтологий O_1 и O_2 (в дальнейшем описании индекс i_1 характеризует принадлежность элемента онтологии O_1, а индекс i_2 – онтологии O_2) по следующим правилам.

Для вершин мультиграфов:

1. $sim\left(v_{i_1}^f, v_{i_2}^f\right) = 1$, если $v_{i_1}^f \equiv v_{i_2}^f$.

2. $sim\left(v_{i_1}^f, v_{i_2}^f\right) = F_c\left(w_{k_1}^c, w_{k_2}^c, \ldots, w_{k_n}^c\right)$, если в графе общей понятийной системы онтологий существуют принадлежащие одному маршруту вершины v_k^c и v_l^c, такие, что $v_{i_1}^f \equiv v_k^c$ и $v_{i_2}^f \equiv v_l^c$ (n – длина маршрута, $w_{k_j}^c \left(j = \overline{1,n}\right)$ – вес отдельной дуги).

3. $sim\left(v_{i_1}^f, v_{i_2}^f\right) = f_t(p)$, если в графе общей терминологической системы онтологий существует полный подграф, содержащий вершины v_k^t и v_l^t, такие, что $v_{i_1}^f \equiv v_k^t$ и $v_{i_2}^f \equiv v_l^t$ (p – количество вершин подграфа).

4. $sim\left(v_{i_1}^f, v_{i_2}^f\right) = g_t(q)$, если в графе общей терминологической системы онтологий существует цепь между вершинами v_k^t и v_l^t, такими, что $v_{i_1}^f \equiv v_k^t$ и $v_{i_2}^f \equiv v_l^t$ (q – длина цепи).

5. $sim\left(v_{i_1}^f, v_{i_2}^f\right) = 0$, если не применимо ни одно из правил 1 – 4.

Для дуг мультиграфов:

$$sim\left(x_{i_1}^f, x_{i_2}^f\right) = \begin{cases} 1, x_{i_1}^f \equiv x_{i_2}^f \\ 0, \text{в противном случае} \end{cases}$$

Результат бинарных операций над онтологиями $O_1 = \left\langle S_f^1, S_c, S_t, \equiv \right\rangle$ и $O_2 = \left\langle S_f^2, S_c, S_t, \equiv \right\rangle$, приведенными к одному понятийному и терминологическому основанию, представляет собой онтологию $O_{op} = \left\langle S_f^1 op S_f^2, S_c, S_t, \equiv \right\rangle$, т. е. необходимо формализовать бинарные операции для функциональных систем.

Представление функциональных систем как мультиграфов позволяет свести операции над онтологиями к операциям над мультиграфами MG_f^1 и MG_f^2.

При выполнении операций над мультиграфами будем считать одинаковыми вершинами те, для которых функция подобия отлична от 0, а одинаковыми дугами – те, для которых функция подобия равна 1.

Операция объединения онтологий $\left(O_{\cup} = O_1 \cup O_2\right)$

Формально алгоритм объединения можно представить следующим образом:

1. Вычисляется множество вершин $V_{\cap} = V_f^1 \cap V_f^2$ (для вершин разных мультиграфов, принадлежащих множеству $V_{\cap}$, функция подобия равна 1).

2. Для каждой вершины из множества вершин $\left(V_f^1 \setminus V_{\cap}\right) \cap V_c$ формируется множество маршрутов к вершинам из множества $\left(V_f^2 \setminus V_{\cap}\right) \cap V_c$ в понятийном графе G_c. Если для вершины множество маршрутов не пусто, для каждого из маршрутов вычисляется функция подобия. Две вершины, для которых функция подобия принимает максимальное значение, считаются далее тождественными (в мультиграфах каждая из вершин замещается соответствующим «понятийным маршрутом») и при новом пересечении множеств вершин мультиграфов формируют множество $V_{\cap}^c$ с учетом понятийной системы.

3. Рассматриваются множества вершин $\left(\left(V_f^1 \setminus V_{\cap}\right) \setminus V_{\cap}^c\right) \cap V_t$ и $\left(\left(V_f^2 \setminus V_{\cap}\right) \setminus V_{\cap}^c\right) \cap V_t$. При наличии пары вершин (по одной из каждого множества), входящих в одну компоненту связности графа терминологической системы, происходит замена каждой из вершин соответствующей компонентой (или ее фрагментом, включающим обе вершины) и при новом пересечении множеств вершин мультиграфов множество $V_{\cap}^t$ формируют с учетом терминологической системы.

4. Множество вершин мультиграфа объединения онтологий формируется как результат теоретико-множественных операций:

$$V_f^{\cup} = V_{\cap} \cup V_{\cap}^c \cup V_{\cap}^t \cup$$
$$\cup\left(\left(\left(V_f^1 \setminus V_{\cap}\right) \setminus V_{\cap}^c\right) \setminus V_{\cap}^t\right) \cup$$
$$\cup\left(\left(\left(V_f^2 \setminus V_{\cap}\right) \setminus V_{\cap}^c\right) \setminus V_{\cap}^t\right)$$

5. Инцидентность дуг из множеств X_f^1 и X_f^2 сохраняется в мультиграфе объединения онтологий (с учетом слияния дуг, для которых функция подобия равна 1).

При объединении функциональных систем исходных онтологий по такому алгоритму могут возникнуть противоречия двух типов:

1) В результате операции объединения мультиграф онтологии $O_{\cup}$ может содержать вершины из множеств $V_{\cap}^c$ и $V_{\cap}^t$, являющихся подграфами понятийного и терминологического графов, дуги и ребра которых описывают отношения понятийной и терминологической систем (что противоречит определению мультиграфа функциональной системы). Для разрешения этих противоречий необходима экспертная оценка и замена таких подграфов узлами (и, возможно, дугами), соответствующими функциональной системе.

2) Сохранение в мультиграфе объединения онтологий инцидентности дуг из множеств X_f^1 и X_f^2 может привести к наличию противоречивых дуг, инцидентных одной и той же паре вершин. Такая ситуация также исследуется субъектом – экспертом.

Операция пересечения онтологий $\left(O_1 \cap O_2\right)$

Отличие алгоритма пересечения от алгоритма объединения состоит в способе формирования результирующего мультиграфа:

$$V_f^{\cap} = V_{\cap} \cup V_{\cap}^c \cup V_{\cap}^t,$$

а инцидентность сохраняется в мультиграфе пересечения онтологий только для тех дуг из множеств X_f^1 и X_f^2, для которых функция подобия равна 1.

При пересечении функциональных систем исходных онтологий могут возникнуть противоречия только первого из описанных выше типов.

Построение аспектного представления (проекция онтологии)

Аспект рассмотрения (представления, описания) в свою очередь может быть задан функциональной системой $S_f^i = \left\langle M_f^i, A_f^i, R_f^i, Z_f^i \right\rangle$, тогда операция проекции может быть сведена к операции пересечения исходной $O = \left\langle S_f, S_c, S_t, \equiv \right\rangle$ и аспектной $O_i = \left\langle S_f^i, S_c, S_t, \equiv \right\rangle$ онтологий: $O_{proj} = O \cap O_i$. В результате операции пересечения онтологий на самом деле происходит «обогащение» аспектного описания с учетом понятийной и терминологической систем.

Рассмотрим следующие возможные ситуации:

1. $M_f^i \neq \varnothing, R_f^i = \varnothing$ – аспект задается на уровне объектов – знаковыми описаниями совокупности объектов. Мультиграф аспектной онтологии, участвующий в операции пересечения, при этом представляет собой пустой граф $MG_f^i = \left\langle V_f^i, \varnothing \right\rangle$. В этом случае для множества вершин результирующего мультиграфа V_f^{proj} операция пересечения выполняется по правилам, описанным выше, а множество дуг формируется из дуг множества X_f^i, инцидентных вершинам множества V_f^{proj}.

2. $M_f^i = \varnothing, R_f^i \neq \varnothing$ – аспект задается на функциональном уровне – множеством функциональных отношений. В этом случае множество дуг мультиграфа проекции $X_f^{proj} = X_f \cap X_f^i$, а множество вершин формируется из вершин, инцидентных дугам из множества X_f^{proj}.

3. $M_f^i \neq \varnothing, R_f^i \neq \varnothing$ – аспект задается на смешанном, объектно-функциональном уровне, и для опера-

ции пересечения может быть сформирован мультиграф $MG_f^i = \langle V_f^i, X_f^i \rangle$ с непустыми множествами вершин и дуг.

Масштабирование онтологий

Для описания операций масштабирования (укрупнения или детализации) онтологии определим для исходной онтологии $O = \langle S_f, S_c, S_t, \equiv \rangle$ «онтологию масштабирования» как $O_m = \langle S_f^m, S_c, S_t, \equiv \rangle$, где $S_f^m = \langle M_c, \varnothing, \varnothing, Z_f \rangle$, и сведем операции масштабирования к разрешению противоречий первого типа в онтологии $O \cup O_m$ в соответствии со следующими правилами:

- в случае укрупнения исходной онтологии: в узлах, содержащих дерево понятий, остаются понятия самого верхнего уровня, а в узлах, содержащих подграфы терминологической системы – отдельные слова;
- в случае детализации исходной онтологии: в узлах, содержащих дерево понятий, остаются понятия самого нижнего уровня, а в узлах, содержащих подграфы терминологической системы – словосочетания.

ЗАКЛЮЧЕНИЕ

По определению Гегеля, онтология как наука – это диалектическое учение об абстрактных определениях сущности, рассматривающее категории философских понятий, являющиеся «кирпичиками» формирования картины мира, средством освоения человеком конкретных и абстрактных объектов. Именно с этих позиций в статье предложен (не претендующий на универсальность) подход к определению онтологии в узком смысле и множества операций как инструмента формирования и количественно оцениваемого соотнесения образов объектов предметной области в условиях диалектически взаимосвязанных пространств абстрактных и конкретных объектов ПрО. В задачах информационного поиска такой подход позволит в автоматизированных процессах динамического реформулирования и соотнесения поисковых образов запросов и документов учитывать разнообразие точек зрения человека – определяющего звена в процессах познания – на основе приведения поисковых образов к общему понятийно-терминологическому контексту.

СПИСОК ЛИТЕРАТУРЫ

1. Gruber Th. What is an Ontology [On-line]. – URL: http://www-ksl.stanford.edu/kst/what-is-an-ontology.html
2. Van Heijst G., Schreiber A. T., Wielinga B. J. Using Explicit Ontologies in KBS Development [On-line]. – 1996. – URL: http://ksi.cpsc.ucalgary.ca/KAW/KAW96/borst/node16.html
3. Guarino N. Understanding, Building, and Using Ontologies [On-line]. – URL: http://ksi.cpsc.ucalgary.ca/KAW/KAW96/guarino/guarino.html
4. Genesereth M. R., Nilsson N. J. Logical Foundation of Artificial Intelligence. – Los Altos, California : Morgan Kaufmann, 1987.
5. De Roure D., Jennings N. R., Shadbolt N. R. The Semantic Grid: A Future e-Science Infrastructure [On-line]. – URL: http://www.semanticgrid.org/documents/semgrid-journal/semgrid-journal.pdf
6. Бениаминов Е. М., Болдина Д. М. Система представления знаний Ontolingua – принципы и перспективы // НТИ. Сер. 2. – 1999. – № 10. – С. 26 – 32.
7. Мудрая О. В., Бабич Б. В., Пьяо С. С., Рейсон П., Уилсон Э. Разработка инструментария для семантической разметки текста // Труды международной конференции «Корпусная лингвистика-2006». – СПб. : Изд-во СПбГУ : РХГА, 2006.
8. Uschold M., King M., Moralee S., Zorgios Y. The Enterprise Ontology // The Knowledge Engineering Review. – 1998. – Vol. 13, № l. – P. 31 – 88.
9. Урманцев Ю. А. Общая теория систем: состояние, приложения и перспективы развития // Система, Симметрия, Гармония : сб. – М. : Мысль, 1988. – С. 38 – 124.
10. Максимов Н. В. Информация и знания: природа, концептуальная модель // НТИ. Сер. 2. – 2010. – № 7. – С. 1 – 10.
11. Бониц М. Научное исследование и научная информация. – М. : Наука, 1987. – 156 с.
12. Максимов Н. В., Окропишин А. Е., Окропишина О. В., Передеряев И. И. Использование технологии автоматизированного формирования понятийной структуры предметной области научного исследования в задачах управления научными кадрами // Вестник РГГУ. Серия «Управление». – 2011. – № 4 (66). – С. 175 – 185.
13. Nirenburg S., Raskin V. Ontological Semantics. – Cambridge, 2004.
14. FIPA 98 Specification. Part 12. Ontology Service. – Geneva : Foundation for Intelligent Physical Agents (FIPA), 1998. – Version 1.0.
15. Noy N. F., Musen M. A. SMART: Automated Support for Ontology Merging and Alignment // Proceedings of the Twelfth Workshop on Knowledge Acquisition, Modeling and Management (KAW'99). – Banff, Canada, 1999.
16. Левашова Т. В. Принципы управления онтологиями, используемые в среде интеграции знаний // Труды СПИИ РАН. Вып. 1, т. 2. – СПб. : СПИИ РАН, 2002.
17. Staab S., Studer R., Mitra P., Wiederhold G. Handbook on ontologies. – Berlin : Springer, 2009. – P. 93 – 114.
18. Kaushik S., Farkas C., Wijesekera D., Ammann P. An algebra for composing ontologies // Technical Report / George Mason University. – 2006.

Материал поступил в редакцию 08.02.12.

Сведения об авторах

ГОЛИЦЫНА Ольга Леонидовна – кандидат технических наук, доцент, доцент кафедры системного анализа Национального исследовательского ядерного университета «МИФИ», Москва
E-mail: OLGolitsina@yandex.ru

МАКСИМОВ Николай Вениаминович – доктор технических наук, профессор, профессор кафедры системного анализа Национального исследовательского ядерного университета «МИФИ»
E-mail: nv-maks@yandex.ru

ОКРОПИШИНА Ольга Владимировна – инженер кафедры системного анализа Национального исследовательского ядерного университета «МИФИ»
E-mail: doomguard@yandex.ru

СТРОГОНОВ Владимир Иванович – доктор технических наук, заместитель руководителя Центра перспективных фундаментальных и прикладных исследований ОАО «Научно-исследовательский и проектно-конструкторский институт информатизации, автоматизации и связи на железнодорожном транспорте»
E-mail: Strogonov_vi@mail.ru

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ

УДК [004.89 : 510.64] : 61

А. Ю. Волкова, О. П. Шестерникова

О создании интеллектуальных систем, реализующих ДСМ-метод автоматического порождения гипотез, и результатах их применения для анализа медицинских данных

Рассматриваются две компьютерные интеллектуальные системы типа ДСМ, реализованные в рамках дипломных работ выпускников РГГУ. Подтверждается важная причинно-следственная зависимость, впервые выявленная на онкологических данных в ВИНИТИ РАН, – связь между протеином S100 и продолжительностью жизни больных меланомой.

Ключевые слова: *ДСМ-метод, индукция, аналогия, абдукция, C#, C++, доказательная медицина, РОНЦ им. Н. Н. Блохина, протеин S100, продолжительность жизни, меланома*

ВВЕДЕНИЕ

ДСМ-метод автоматического порождения гипотез (АПГ) был предложен в начале 1980-х гг. Название метода составляют инициалы известного английского философа, логика и экономиста XIX в. Джона Стюарта Милля, который сформулировал методы индуктивного вывода.

ДСМ-метод АПГ – это метод качественного анализа данных. Он является формализованной эвристикой для установления причин наличия или отсутствия изучаемых эффектов в открытых базах структурированных фактов. База фактов «образована фактоподобными высказываниями вида “объект С имеет множество свойств А”, которым приписаны оценки “фактически истинно” (1), “фактически ложно»” (–1), “фактически противоречиво” (0), “неопределенно” (τ)» [1, с. 400].

ДСМ-метод АПГ состоит из трех познавательных процедур: эмпирической индукции (порождения причин эффектов на основе обнаруженных сходств фактов), структурной аналогии (правдоподобных выводов, использующих наличие положительных или отрицательных причин в фактах с неопределенной оценкой, требующей наличия или отсутствия изучаемого эффекта) и абдукции (принятия гипотез посредством объяснения начального состояния базы фактов с помощью (±)-причин) [1]. Как справедливо отмечено в [2], взаимодействие этих процедур «осуществляет согласование идей Д. С. Милля об индукции с абдукцией Ч. С. Пирса, требованием фальсификации порождаемых гипотез К. Р. Поппера и стремлением использовать правдоподобные рассуждения для knowledge discovery согласно Д. Пойа».

ДСМ-метод применим при выполнении следующих условий [1]:

1) данные слабо формализованы, но хорошо структурированы (в частности, для данных должна быть определена операция сходства);

2) в базе данных (базе фактов) для изучаемого эффекта должны содержаться положительные примеры, отрицательные примеры и примеры неопределенности;

3) в базе фактов содержатся зависимости причинно-следственного типа (фрагмент есть причина наличия/отсутствия некоторого множества свойств).

Очевидно, что естественные науки (науки о жизни, и в частности, медицина) удовлетворяют этим условиям. Такие особенности ДСМ-метода АПГ, как способность работать эффективно на малых массивах данных, возможность работы с открытыми массивами (указывая на необходимость расширения базы фактов, если она возникает), позволяют утверждать, что ДСМ-метод АПГ является полезным инструментом интеллектуального анализа данных в науках о жизни. Однако интеллектуальные системы типа ДСМ успешно применялись и в таких предметных областях, как фармакология, техническая диагностика, социология, криминалистика, робототехника [3].

В настоящей статье мы рассмотрим применение ДСМ-систем для задач медицинской диагностики.

За последние годы во Всероссийском институте научной и технической информации РАН (ВИНИТИ РАН) было создано несколько интеллектуальных компьютерных систем типа ДСМ для исследований в следующих областях медицины [1, 4]:

1. Прогнозирование высокопатогенных типов вируса папилломы человека (ВПЧ) по цитологическим результатам исследования мазков (кафедра клиниче-

ской лабораторной диагностики Российской медицинской академии последипломного образования).

2. Диагностика двух заболеваний глаз: дегенеративного ретиношизиса и наследственных витреоретинальных дистрофий (лаборатория клинической физиологии зрения МНИИ глазных болезней им. Гельмгольца).

3. Диагностика системной красной волчанки (Отделение нефрологии Городской клинической больницы им. С. П. Боткина).

В ходе работы над этими системами были выработаны общие принципы построения интеллектуальных ДСМ-систем для решения задач данной предметной области. Результаты проведенных экспериментов показали состоятельность ДСМ-метода автоматического порождения гипотез как нового средства доказательной медицины [5].

Параллельно на кафедре интеллектуальных систем в гуманитарной сфере Института лингвистики Российского государственного гуманитарного университета в рамках выпускных дипломных работ О. П. Шестерниковой [6] и А. Ю. Волковой [7] были созданы две независимые интеллектуальные ДСМ-системы, включающие программные средства, способные помочь в прогнозировании продолжительности жизни больных меланомой кожи. Данные больных были предоставлены Российским онкологическим научным центром РАМН им. Н. Н. Блохина. Задачи, поставленные сотрудниками РОНЦ им. Н. Н. Блохина перед авторами этих систем, включали:

а) определение продолжительности жизни больных меланомой по набору признаков, связанных с социальными факторами, наследственными и врожденными, историей болезни и лабораторными показателями больного;

б) выяснение связи между продолжительностью жизни и значением протеина S100 – биохимическим маркером прогрессирования меланомы кожи.

Анализ данных проводился на трех ДСМ-системах: вначале данные были исследованы с применением системы, созданной в ВИНИТИ РАН, затем с помощью ДСМ-систем О. П. Шестерниковой и А. Ю. Волковой, которые подтвердили результаты, полученные в ВИНИТИ.

РЕЗУЛЬТАТЫ, ПОЛУЧЕННЫЕ В ВИНИТИ РАН

26 марта 2010 г. в Российском государственном гуманитарном университете (РГГУ) на семинаре «Методы искусственного интеллекта в когнитивных исследованиях, в гуманитарных науках и науках о жизни» Е. С. Панкратова и Д. А. Добрынин выступили с докладом о проведенных ими компьютерных экспериментах над клиническими данными больных меланомой. С использованием интеллектуальной ДСМ-системы, программно созданной Д. А. Добрыниным, было установлено, что существует связь между продолжительностью жизни больного и значением протеина S100: значение S100, меньшее 0.12 нг/мл, влечет продолжительность жизни больше 5 лет, а значение S100, большее 0.12 нг/мл, – меньше 5 лет.

В качестве методов были использованы стандартные методы ДСМ-рассуждений: простой метод сходства и запрет на (±)-контрпример.

ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА О. П. ШЕСТЕРНИКОВОЙ

Система О. П. Шестерниковой была создана под руководством В. К. Финна и Е. С. Панкратовой [6]. В ней были реализованы следующие алгоритмы: простой метод сходства и метод с запретом на (±)-контрпример (оба с атомарной правой частью). Для настройки на эксперимент в интерфейс были включены следующие параметры:

- возможность задавать нижний порог для количества объектов, порождающих гипотезу (так называемых «родителей», $n \geq 2$), отдельно для (+)- и (–)-гипотез;
- возможность «фильтровать» гипотезы, т.е. отбирать в ходе порождения гипотез только те, которые удовлетворяют заданным условиям; фильтры в системе были созданы аналогично фильтрам, предложенным в [4]: конъюнктивные (присутствуют все заранее указанные признаки из этой группы) и дизъюнктивные (присутствует хотя бы один из заранее указанных признаков из этой группы) – отдельно для (+)- и (–)-гипотез.

Для оценки работы системы был написан алгоритм «доопределение по одному». Его можно запустить из основной программы как отдельный вид эксперимента.

Система создана на языке C# 2.0 с использованием IDE Microsoft Visual Studio 2005. Одной из целей такого выбора было изучение этого нового языка программирования, а также достоинств и недостатков его применения в так называемом «научном программировании». Из достоинств можно отметить: 1) поддержку платформы .Net Framework, которая предоставляет большой набор готовых библиотек для работы с классами-контейнерами (списки, коллекции, словари и т. д.), xml-документами, потоками, регулярными выражениями; 2) различные стандартные элементы управления, контейнеры и компоненты, которые удобно использовать для быстрого создания графического интерфейса пользователя. Явным недостатком в данном случае является зависимость от платформы (для корректной работы программы необходима версия .Net Framework не ниже 2.0) и операционной системы (Windows XP и выше). Вопрос о преимуществах или недостатках управляемости кода в случае с алгоритмами высокой сложности, каким является и ДСМ-метод, остался открытым.

Идейно система О. П. Шестерниковой создана в концепции объектно-ориентированного программирования. Все понятия предметной области, такие, как «решатель», «эксперимент», «типы данных», представляют собой объекты со своими свойствами, полями, методами.

Система состоит из исполняемого файла, который вызывает основные формы, предоставляющие пользователю интерфейс для работы (редактирование данных, проведение экспериментов и просмотр результа-

тов), и динамической библиотеки, в которую вынесены классы, относящиеся непосредственно к решателю. Таким образом, реализована независимость содержательной части от представления данных.

Данные для системы хранятся в файлах xml. Эти файлы представляют собой сериализованные внутренние объекты системы. Использование такого готового инструмента платформы .Net Framework, как сериализация/десериализация, упрощает работу с данными. Этот же инструмент сериализации/десериализации удобно использовать и при сохранении уже проведенных экспериментов (настроек и результатов), так как понятие «эксперимент» также представляет собой объект и может быть сериализовано.

При проведении экспериментов все примеры делились на положительные и отрицательные в зависимости от продолжительности жизни: примеры с продолжительностью жизни более 5 лет относились к (+)-примерам, а примеры с продолжительностью жизни менее 5 лет – к (–)-примерам. Кроме того, задавались фильтры, содержащие признак «Значение белка S100».

В полученном результате все положительные гипотезы, содержащие признак S100, имеют его значение < 0.12 нг/мл, а отрицательные > 0.12 нг/мл. Поэтому было выдвинуто предположение о зависимости между значением белка S100 и продолжительностью жизни пациента.

ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА А. Ю. ВОЛКОВОЙ

А. Ю. Волковой была создана версия интеллектуальной системы, реализующая различные процедуры ДСМ-метода АПГ, которые применялись к различным предметным областям. В ходе экспериментов эта ДСМ-система была применена к двум предметным областям: фармакологии и медицинской диагностике (диагностика заболеваний глаз – дегенеративного ретиношизиса и наследственных витреоретинальных дистрофий – и прогнозирование продолжительности жизни больных меланомой и оценка прогностического биохимического маркера – протеина S100).

В системе А. Ю. Волковой были реализованы следующие методы: простой метод сходства, запреты на (±)-контрпример, единственность (+)-причины, метод различия в двух вариантах (прямой перевод второго правила индуктивного вывода Д. С. Милля и упрощение метода соединенного сходства-различия), метод соединенного сходства-различия, усеченный метод остатков для случая атомарного эффекта и общий метод остатков для случая неатомарного эффекта [2, 8, 9]. Стратегии ДСМ-рассуждений, содержащие индуктивные методы различия, сходства-различия и остатков, были реализованы впервые. Реализация методов, а также результаты их применения подробно изложены в [10, 11].

ДСМ-решатель был создан на языке программирования C++ в среде Visual Studio 2005 с использованием библиотеки классов MFC (Microsoft Foundation Classes) и библиотеки стандартных шаблонов STL (Standard Template Library), а также интерфейса для доступа и манипулирования внешними данными ADO и языка запросов SQL.

Язык C++, сочетающий свойства как высокоуровневых, так и низкоуровневых языков, был выбран не случайно. Во-первых, при создании ДСМ-решателя необходимо было учесть тот факт, что данные могут представлять собой объекты разной структуры и относиться к разным предметным областям. Эту проблему удалось решить с помощью поддержки в C++ объектно-ориентированного программирования, возможности динамического приведения типов наследуемых классов к базовому и использования виртуальных функций.

Во-вторых, программные реализации процедур ДСМ-рассуждений достаточно ресурсоемки как по времени, так и по памяти. C++ позволяет динамически выделять и освобождать память благодаря наличию в языке указателей (возможность работы на низком уровне с памятью, адресами), что и помогает решить эту проблему при проведении экспериментов.

В-третьих, язык C++ – это один из языков, поддерживаемых платформой .NET Framework, которая обеспечивает совместимость программ, написанных на разных языках программирования.

Так как было необходимо добиться независимости ДСМ-решателя от предметной области, интеллектуальная система включает минимальный пользовательский интерфейс для выбора источника данных и настройки системы на эксперимент: выбор стратегий ДСМ-метода АПГ, задание порога родителей для гипотез, оценка выбранной стратегии процедурой «доопределение по одному».

Для представления медицинских данных и результатов работы ДСМ-системы был разработан документ со специальной структурой в Microsoft Office Excel. В этом документе хранится лишь структурное представление данных, без каких-либо интерфейсных элементов, что позволяет при проведении большого количества экспериментов получать результаты намного быстрее.

Разработанный нами документ Excel состоит из шести листов: «Properties», «Patients», «Hypotheses», «T_Result», «Total» и «Determine_ByOne».

Лист «Properties» содержит информацию о признаках, которые описывают пациентов: номер признака (столбец Id), название (Name), тип (Type) и длина строки (Length) схематичного представления1. Кроме того, дается информация о том, какие признаки включены в фильтры (столбцы для конъюнктивных и дизъюнктивных фильтров). В случае документа для ДСМ-метода с неатомарной правой частью на этом листе также присутствуют столбцы Left (отнести признак в левую часть) и Right (отнести признак в правую часть).

Лист «Patients» хранит данные о пациентах: идентификатор, имя (номер карточки для больных мела-

[1] Например, если признак – «Пол пациента» = мужской/женский , то длина равна 2.

номой), тип значения (+, –, 0 или ?) 2, выбран ли пациент для эксперимента (поле «Included», принимающее значение 0 или 1) и описание структуры (схематичное представление признаков)3.

Из листа «Hypotheses» можно узнать необходимые сведения о полученных гипотезах. У гипотезы есть идентификатор, имя, номер такта ДСМ-рассуждения, на котором она была получена, тип значения (+, – или 0), число родителей, их идентификаторы и тело гипотезы (схематичное представление признаков).

Особенностью структуры листа «Hypotheses» для неатомарного ДСМ-метода является наличие дополнительного поля «method_Got». Это поле заполняется символом «i» (от induction), если гипотеза получена на этапе индукции, или символом «r» (от residues), если гипотеза получена методом остатков. У гипотезы, полученной методом остатков, последний родитель – это идентификатор гипотезы, которая участвовала в её порождении.

Лист «T_Result» хранит информацию о результатах доопределения (τ)-примеров. Для удобства ещё раз выводится идентификатор и имя примера. Каждый (τ)-пример имеет номер такта ДСМ-рассуждения, на котором была последняя попытка доопределения примера, результат доопределения (+, –, 0 или ?), количество доопределивших пример гипотез и список номеров этих гипотез.

Лист «Total» содержит общую информацию об эксперименте: какие стратегии ДСМ-метода АПГ были применены в эксперименте, использовались ли фильтры, количество примеров и полученных гипотез, количество необъясненных гипотезами примеров и др. Если имеются необъясненные примеры, в списке выводятся их идентификаторы (результат проверки аксиомы каузальной полноты).

На листе «Determine_ByOne» представлены результаты процедуры «доопределение по одному». Для каждого примера приводится его идентификатор, имя, тип (+, – или ?) и результат доопределения.

В ходе работы с медицинскими данными стало ясно, что схематичное представление данных в Microsoft Office Excel неудобно для пользователя. Поэтому были также созданы база данных в Microsoft Office Access и интерфейс для визуализации данных с использованием языка программирования Visual Basic for Applications, интерфейса ADO и языка запросов SQL. Важно отметить, что в созданном интерфейсе признаки представлены двумя способами: схематично (в виде строк, состоящих из +, x и t), а также с помощью словесного описания в виде дерева признаков. Подробное описание интерфейса можно найти в [7].

С помощью созданных программных средств были исследованы данные больных меланомой и получены следующие результаты:

1. Оказалось, что метод единственной (+)-причины и метод различия, который является прямым переводом второго правила индуктивного вывода Д. С. Милля, не породили ни одной гипотезы.

2. Метод сходства-различия не подтвердил ни одну положительную гипотезу. Дело в том, что эта стратегия очень чувствительна к данным, так как требует выполнения ряда сильных условий, поэтому полученный результат свидетельствует о том, что в выборке достаточно много субъективных признаков.

3. Стратегия упрощенного метода соединенного сходства-различия с запретом на (±)-контрпример подтвердила гипотезы, полученные стандартной стратегией запрета на (±)-контрпример. Эти гипотезы свидетельствуют о наличии связи между белком S100 и продолжительностью жизни больного меланомой. При этом процедура «доопределение по одному», используемая как критерий оценки подбора параметров и стратегии эксперимента, показала хороший результат. И поскольку упрощенный метод соединенного сходства-различия с запретом на (±)-контрпример – это семантически наиболее сильная стратегия, на исследуемых данных ее стоит признать наилучшей.

Еще один метод, реализованный специально для поставленной задачи, – это усеченный метод остатков. Он заключается в нахождении попарной разности между поступающими на вход гипотезами.

Сначала при помощи программных средств для атомарной версии ДСМ-метода АПГ было проведено три эксперимента.

Первый эксперимент состоял в разделении объектов на (+)- и (–)-примеры в зависимости от продолжительности жизни (продолжительность жизни более 5 лет – это (+)-пример, менее 5 лет – это (–)-пример) и использования фильтров для обязательного наличия в гипотезах белка S100 в левой части. В результате была порождена первая группа гипотез.

Второй эксперимент состоял в разделении объектов на (+)- и (–)-примеры в зависимости от значения белка S100 (значение S100, меньшее 0.12 нг/мл, – это (+)-пример, равное или большее 0.12 нг/мл, – это (–)-пример). В результате была порождена вторая группа гипотез.

Третий эксперимент, как и первый, состоял в разделении объектов на (+)- и (–)-примеры в зависимости от продолжительности жизни, но признак «Уровень S100» в эксперименте не участвовал. В результате была порождена третья группа гипотез.

Затем была найдена разность первой и второй групп гипотез: из (+)-гипотез вычитаем (+)-гипотезы, из (–)-гипотез – (–)-гипотезы. В результате сравнения полученной разности и третьей группы гипотез выяснилось, что пересечения между ними нет. Если бы были найдены равные гипотезы, это означало бы, что S100 не является необходимым признаком. Таким образом, было подтверждено наличие корреляции между уровнем белка S100 и продолжительностью жизни пациента.

Полученные результаты подтвердили, что процедуры ДСМ-рассуждений являются мощным средством доказательной медицины, при этом такие мето-

[2] Знак «?» соответствует тау-примерам.

[3] Схематично один признак представляет собой строку, состоящую из +, x и t: «+» – признак присутствует, «x» – признак отсутствует, «t» – значения нет. Например, если признак – «Пол пациента» = мужской/женский, то у пациента мужского пола схематичное представление признака есть +t.

ды, как упрощенный метод соединенного сходства-различия с запретом на (±)-контрпример и усеченный метод остатков, могут с успехом применяться для решения медицинских задач.

СОЕДИНЕНИЕ ВОЗМОЖНОСТЕЙ ДВУХ СИСТЕМ

Для использования запрограммированных средств вне ДСМ-решателя А. Ю. Волковой была создана динамическая библиотека для анализа медицинских данных, включающая классы представления данных и все необходимые функции для реализации ДСМ-метода, в частности, процедуры простого метода сходства, запрета на контрпример, единственности (+)-причины и метода соединенного сходства-различия.

В ходе работы возникла актуальная сейчас задача взаимодействия управляемого и неуправляемого кода. Созданная библиотека была скомпилирована с поддержкой Common Language Runtime[4], а в ДСМ-систему О. П. Шестерниковой были добавлены элементы интерфейса для обращения к библиотеке.

Нами были написаны дополнительные функции для перевода структуры представления данных в формат данных, необходимый для запуска стратегий ДСМ-метода АПГ, а также функции для обратного перевода уже полученного результата. Следует отметить, что потребовалось перевести не только данные (описания больных), но и информацию о выбранных для эксперимента признаках, их типах и фильтрах.

Созданная динамическая библиотека повысила функциональность существующих ДСМ-систем и может использоваться в дальнейшем.

ЗАКЛЮЧЕНИЕ

Созданные авторами статьи ДСМ-системы имеют важные различия:

- Система О. П. Шестерниковой обладает удобным пользовательским интерфейсом и может применяться непосредственно как инструмент врача; к ней можно подключать дополнительные модули.
- Система А. Ю. Волковой дает возможность исследовать данные разных предметных областей с использованием различных стратегий, но при этом существует четкое разделение: программные средства решателя и средства представления данных.

Проведенное исследование показало, что созданные А. Ю. Волковой и О. П. Шестерниковой интеллектуальные системы типа ДСМ являются важным инструментом для анализа данных, в том числе средством доказательной медицины. В результате экспериментов, проведенных на описанных ДСМ-системах и ДСМ-системе, созданной в ВИНИТИ РАН, были получены одинаковые результаты, выявляющие важную причинно-следственную зависимость: при значении уровня протеина S100 менее 0,12 нг/мл продолжительность жизни больного будет больше 5 лет, при значении уровня S100 более 0,12 нг/мл – меньше 5 лет. Благодаря этому теперь становится возможно формировать группы риска онкологических больных, требующих особого медицинского наблюдения [12, 13].

Возможности описанных выше ДСМ-систем не ограничиваются рассмотренной задачей. И мы надеемся, что полученный нами опыт позволит получить новые результаты при исследовании других задач.

[4] Common Language Runtime (CLR) – «общеязыковая исполняющая среда» – компонент пакета Microsoft .NET Framework, исполняющая программа, написанная на .NET-совместимых языках программирования.

СПИСОК ЛИТЕРАТУРЫ

1. Финн В. К., Блинова В. Г., Панкратова Е. С., Фабрикантова Е. Ф. Интеллектуальные системы для анализа медицинских данных // ДСМ-метод автоматического порождения гипотез: Логические и эпистемологические основания / сост. О. М. Аншаков, Е. Ф. Фабрикантова; под общ. ред. О. М. Аншакова. – М. : Книжный дом «Либроком», 2009. – С. 398 – 428.
2. Финн В. К. Индуктивные методы Д. С. Милля в системах искусственного интеллекта // Искусственный интеллект и принятие решений. – 2010. – № 3, 4.
3. Автоматическое порождение гипотез в интеллектуальных системах / под общ. ред. В. К. Финна. – М. : Книжный дом «Либроком», 2009.
4. Панкратова Е. С., Добрынин Д. А., Цапенко И. В., Зуева М. В., Захарова Г. Ю. Интеллектуальная ДСМ-система для диагностики заболеваний органа зрения // НТИ. Сер. 2. – 2007. – № 3. – С. 14 – 18.
5. Флетчер Р., Флетчер С., Вагнер Э. Клиническая эпидемиология. Основы доказательной медицины. – М. : Медиа Сфера, 2004.
6. Шестерникова О. П. Создание прототипа ДСМ-системы для прогноза продолжительности жизни больных меланомой : дипломная работа / Шестерникова Ольга Павловна; науч. рук. В. К. Финн, Е. С. Панкратова; РГГУ. – М., 2010.
7. Волкова А. Ю. Создание программных средств анализа базы фактов для применения ДСМ-метода автоматического порождения гипотез : дипломная работа / Волкова Анна Юрьевна; науч. рук. В. К. Финн; Ин-т лингвистики Рос. гос. гуманитарного ун-та. – М., 2010.
8. Финн В. К. Индуктивный метод соединенного сходства-различия и процедурная семантика ДСМ-метода // НТИ. Сер. 2. – 2010. – № 4. – С. 1 – 17.
9. Финн В. К. Своевременные замечания о ДСМ-методе автоматического порождения гипотез // НТИ. Сер. 2. – 2009. – № 8. – С. 15 – 26; Finn V. K. Timely Notes about the JSM Method for Automatic Hypothesis Generation // Automatic Documentation and Mathematical Linguistics. – 2009. – Vol. 43, № 5. – P. 257 – 269.
10. Волкова А. Ю. Алгоритмизация процедур ДСМ-метода автоматического порождения гипотез // НТИ. Сер. 2. – 2011. – № 5. – С. 6 – 12;

Algorithmization of Procedures of the JSM Method for Automatic Hypothesis Generation // Automatic Documentation and Mathematical Linguistics. – 2011. – Vol. 45, № 3. – P. 113 – 120.

11. Волкова А. Ю. Анализ данных различных предметных областей с помощью процедур ДСМ-метода автоматического порождения гипотез // НТИ. Сер. 2. – 2011. – № 6. – С. 9 –18; Analyzing the Data of Different Subject Fields Using the Procedures of the JSM Method for Automatic Hypothesis Generation // Automatic Documentation and Mathematical Linguistics. – 2011. – Vol. 45, № 3. – P. 127 – 139.
12. Михайлова И. Н., Панкратова Е. С., Добрынин Д. А., Самойленко И. В., Решетникова В. В., Шелепова В. М., Демидов Л. В., Барышников А. Ю., Финн В. К. О применении интеллектуальной компьютерной системы для анализа клинических данных больных меланомой // Российский биотерапевтический журнал. – 2010. – Т. 9, № 2. – С. 54.
13. Финн В. К. Об определении эмпирических закономерностей посредством ДСМ-метода автоматического порождения гипотез: дополнение к статье В. К. Финна «Индуктивные методы Д. С. Милля в системах искусственного интеллекта» // Искусственный интеллект и принятие решений. – 2010. – № 4. – С. 41 – 48.

Материал поступил в редакцию 18.02.11.

Сведения об авторах

ВОЛКОВА Анна Юрьевна – аспирант Российского государственного гуманитарного университета, Москва

E-mail: anna-.-volkova@mail.ru

ШЕСТЕРНИКОВА Ольга Павловна – аспирант Всероссийского института научной и технической информации РАН, Москва

E-mail: oshesternikova@gmail.com

ИНФОРМАЦИОННЫЙ АНАЛИЗ

УДК 004.94 «7»

И. Н. Карпухин, Ю. П. Кораблин, А. А. Незнанов

Модели времени в имитационном моделировании

Что же такое время? Если никто меня об этом не спрашивает, я знаю, что такое время; если бы я захотел объяснить спрашивающему – нет, не знаю.

Августин Аврелий

Рассмотрены основные подходы к решению классических задач имитационного моделирования и модели времени: дискретно-событийное и непрерывное моделирование, а также моделирование Монте-Карло. Обсуждаются их основные положения, преимущества, недостатки и конкретные реализации. На основе проведённого исследования показано место оригинального программного средства G-IPS Ultimate в ряду других программных продуктов для решения учебных и прикладных задач имитационного моделирования.

Ключевые слова: *имитационное моделирование, время, модель времени, программное средство*

ВВЕДЕНИЕ

Моделирование, т. е. построение и решение моделей, является одним из базовых методов исследования в естественных науках [1, 2]. Имитационным моделированием (*simulation*) неформально принято называть метод исследования, при котором изучаемая система заменяется моделью, с необходимой точностью описывающей реальную систему, после чего над моделью проводятся эксперименты с целью получения новых знаний об этой системе. Имитацией называют процесс проведения эксперимента над моделью с целью предсказать, как поведет себя реальная система, когда такой эксперимент будет поставлен на ней. Заметим, что реальная система на момент имитации может ещё не существовать. Например, в авиации конструкторы строят масштабные модели самолетов, чтобы проверить аэродинамические характеристики их конструкции. Манипулируя конструкцией крыла или формой фюзеляжа на модели, конструктор может определить, насколько эти изменения влияют на поток воздуха, попадающий на соответствующие поверхности. Если модель достаточно «правдиво» (адекватно) представляет реальный самолет, конструктор получает доказательства, что реальная система будет вести себя определенным образом, с учетом привнесенных изменений. Преимущество моделирования заключается в том, что имеется возможность провести несколько экспериментов на модели в различных условиях без необходимости нести расходы по созданию или изменению реальной системы. Гораздо проще вносить изменения в имитационную модель, чем пытаться переработать полноразмерный самолет, а в некоторых случаях эксперимент на реальной системе просто невозможен.

Никакой мыслимый объект или процесс взаимодействия объектов не существует вне времени. Любой, даже самый элементарный процесс, такой, как столкновение элементарных частиц, длится какое-то время. Однако в различных задачах время учитывается по-разному, а стало быть, по-разному отражается на проектируемых моделях и имитации.

Модель, в самом широком смысле, – это любой мысленный или знаковый образ моделируемого объекта (оригинала). К их числу относятся гносеологические образы (воспроизведение, отображение исследуемого объекта или системы объектов в виде научных описаний, теорий, формул и т. п.), схемы, чертежи, графики, планы, карты и т. д. [3]. Построение моделей основано на принципе абстракции. В имитационном моделировании существует и другое определение, согласно которому модель – это статическое представление системы. Имитация добавляет в модель временной аспект, показывая, как будет меняться система с течением времени [4]. Таким образом, имитация может быть определена как изменение состояния модели во времени. Временной фактор настолько важен, что является одним из главных критериев классификации имитационных моделей. Далее мы будем рассматривать компьютерное имитационное моделирование, т. е. исследование поведения математической модели, решаемой с помощью компьютера.

ПОНЯТИЕ ВРЕМЕНИ

На сегодняшний день не существует определения времени, охватывающего все прикладные области. С одной стороны, согласно некоторым определениям время – это одна из осей пространства-времени, согласно другим – фундаментальное понятие человеческого мышления, отображающее изменчивость мира,

процессуальный характер его существования, наличие в мире не только «вещей» (объектов, предметов), но и событий [5]. С другой стороны, время – понятие, позволяющее установить, когда произошло то или иное событие по отношению к другим событиям. Измерение времени подразумевает введение временной шкалы, пользуясь которой можно соотносить эти события[1].

Основной концепцией, используемой сегодня при формализации времени, является ось времени (временная ось или стрела времени – в зависимости от дисциплины), представляющая собой прямую (т. е. математически одномерный объект), протянутую из прошлого в будущее, со шкалой для количественного измерения промежутков времени. Из любых двух несовпадающих точек оси времени одна всегда является будущим относительно другой, что позволяет соотносить совершение событий.

В зависимости от исследуемых процессов рассматривают физическую и логическую ось времени. Физическое время привязано к некоторому устройству, генерирующему физические отметки времени, – *физическим часам*, а логическое время использует только «отношение предшествования» событий на некоторой оси времени, отметки на которой могут не совпадать с физическим временем. Введение логических осей времени связано с локальностью (демонстрируемой теорией относительности) физических часов. Логические оси позволяют рассматривать события в пространственно-протяженных системах. Любая имитация «перемещает» модель по логической оси времени, последовательно совершая действия над ней.

МОДЕЛИ ВРЕМЕНИ

Рассмотрим три типа моделирования, кардинально различными способами учитывающие время:

- дискретно-событийное моделирование,
- непрерывное моделирование,
- моделирование Монте-Карло.

Каждый тип имеет определенный способ добавления аспекта времени к модели. Дискретно-событийное моделирование зависит от наступления определенных событий для перехода модели из одного состояния в другое. Непрерывное моделирование представляет собой изменение состояния в континууме[2] времени независимо от событий, происходящих в системе. Термин «Монте-Карло» применяется к тем моделям, в которых течение времени не является существенным.

Проиллюстрируем различия между первым и вторым типами моделирования на примере светофора и автомобиля. Светофор может находиться в трех состояниях: красный, желтый, зеленый. Они сменяются мгновенно, когда срабатывает внутренний счетчик времени и логическая схема переключает цвет. Автомобиль не может мгновенно изменять скорость своего движения. Он приобретает ускорение, которое определяется как изменение в скорости с течением времени. Это изменение в скорости – непрерывное во времени событие, а не мгновенное изменение [4].

Следует отметить, что в реальном мире все процессы являются непрерывными. Даже в примере со светофором изменение цвета – это также некоторый процесс, который происходит в течение некоторого времени. Однако во многих задачах можно не учитывать внутреннюю структуру процессов малой длительности и считать их мгновенными без потерь смысла и точности результатов эксперимента [6].

Дискретно-событийное моделирование

Формально дискретно-событийное моделирование можно определить как последовательность изменений в модели, вызванных хронологической последовательностью событий, происходящих в системе. Все события считаются происходящими мгновенно и приводят к изменению состояния системы. Состояние системы определяется как совокупность переменных, полностью описывающих систему на уровне абстракции соответствующей модели.

Важнейший компонент дискретно-событийного моделирования – часы. Часы отсчитывают время имитации и могут использоваться для запуска и синхронизации событий в системе (так, в примере со светофором часы могут отслеживать время переключения светофора).

В отличие от реальных систем в имитационных моделях время можно «растягивать», чтобы детально рассмотреть исследуемый процесс, или «ускорять», чтобы значительно быстрее узнать результат эксперимента. И это – одно из ключевых преимуществ имитационного моделирования.

Популярной областью применения дискретно-событийного моделирования является моделирование систем массового обслуживания (СМО), т. е. систем, которые производят обслуживание поступающих в них заявок. Это обслуживание в СМО производится определенными устройствами. Классическая СМО содержит от одного до бесконечного числа устройств. В зависимости от наличия возможности ожидания начала обслуживания СМО подразделяются на три вида:

1) системы с потерями, в которых заявки, не нашедшие в момент поступления ни одного свободного устройства, теряются;

2) системы с ожиданием, в которых имеется накопитель бесконечной ёмкости для буферизации поступивших заявок; при этом ожидающие заявки образуют очередь;

3) системы с накопителем конечной ёмкости (с ожиданием и ограничениями), в которых длина очереди не может превышать ёмкости накопителя; при этом заявка, поступающая в переполненную СМО (отсутствуют свободные места для ожидания), теряется.

Базовыми СМО с очередями являются следующие (рис. 1 – 4):

1. Одна очередь – одно обслуживающее устройство.
2. Одна очередь – много обслуживающих устройств.
3. Много очередей – одно обслуживающее устройство.
4. Много очередей – много обслуживающих устройств.

[1] Энциклопедия Кольера

[2] Континуум (от лат. continuum – непрерывное): в математике – непрерывная совокупность, например, совокупность всех точек отрезка на прямой или всех точек прямой, эквивалентная совокупности всех действительных чисел. (Большой энциклопедический словарь. 2000.)

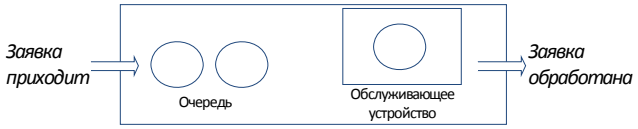

Рис. 1. СМО: одна очередь – одно обслуживающее устройство

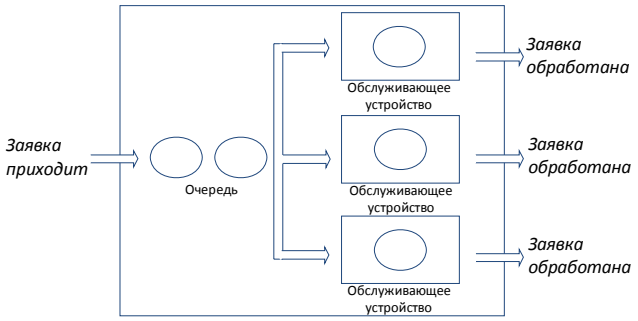

Рис. 2. СМО: одна очередь – много обслуживающих устройств

Преимущество использования моделей СМО заключается в том, что из моделей компонентов некоторой системы можно составить более сложную СМО, описывающую всю систему.

Цифровой компьютер прекрасно подходит для данного типа моделирования, ведь он, по сути, является дискретной машиной [4]. Так, программный пакет *SCADA*, предназначенный для разработки или обеспечения работы в реальном времени систем сбора, обработки, отображения и архивирования информации об объекте мониторинга или управления, может являться частью АСУ ТП, АСКУЭ, системы экологического мониторинга, научного эксперимента, автоматизации здания и т. д. Такого рода системы часто применяют на высокотехнологичном производстве, где присутствует набор датчиков, описывающих состояние системы, и набор устройств, позволяющих изменять это состояние. Пользовательский интерфейс дает возможность диспетчеру в интерактивном режиме отслеживать показатели датчиков реальной системы и получать рекомендации о дальнейших действиях. Датчики опрашиваются в дискретном времени (например, один раз в 5 секунд), и их показания обновляются в диспетчерском интерфейсе.

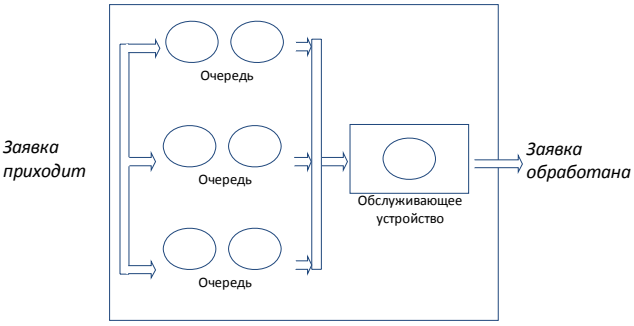

Рис. 3. СМО: много очередей – одно обслуживающее устройство

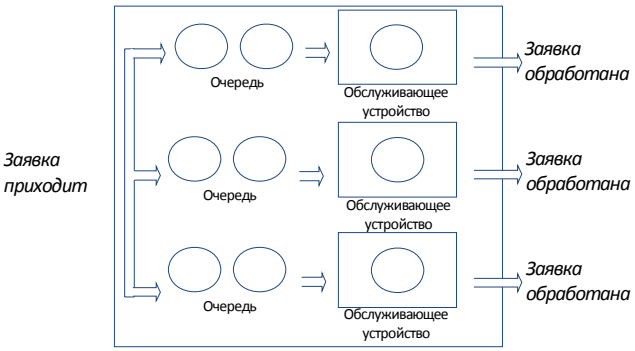

Рис. 4. СМО: много очередей – много обслуживающих устройств

В табл. 1 представлены некоторые широко известные программные продукты, применяемые для дискретно-событийного моделирования [7].

Таблица 1

Программные продукты, применяемые для дискретно-событийного моделирования

Программный продукт	Фирма-разработчик	Адрес в Интернете
Arena	Systems Modeling Corporation	http://www.sm.com/
CSIM18	Mesquite Software, Inc.	http://www.mesquite.com/
Extend	Imagine That, Inc.	http://www.imaginethatinc.com/
GPSS/H	Wolverine Software Corporation	http://www.wolverinesoftware.com/
iGrafx Process	Micrografx, Inc.	http://www.micrografx.com/
Micro Saint	Micro Analysis & Design	http://www.maad.com/
ProcessModel	ProcessModel, Inc.	http://www.processmodel.com/
Promodel/MedModel	PROMODEL Corporation	http://www.promodel.com/
Silk	ThreadTec, Inc.	http://www.threadtec.com/
SIMSCRIPT II.5 and SIMPROCESS	CACI Products Company	http://www.caci.com/
Simul8	Simul8 Corporation	http://www.SIMUL8.com/
Taylor ED	F&H Simulations, Inc.	http://www.taylor-ed.com/
Witness	Lanner Group	http://www.lanner.com/corporate/

Непрерывное моделирование

Как было сказано ранее, дискретно-событийное моделирование основывается на хронологической последовательности событий, каждое из которых приводило к изменению состояния системы. Непрерывное моделирование описывается как система переменных, непрерывно изменяющих свои значения во времени. Рассмотрим такой простой пример, как падающий шарик. Единственная переменная, описывающая данную модель, – это скорость падения. Пока шарик зафиксирован, он имеет нулевую скорость падения, как только мы его выпустили, под действием силы тяжести скорость падения шарика начала возрастать. График зависимости скорости шарика от времени падения представлен на рис. 5. Эта кривая была получена нами на основе моделирования, исходя из предположения, что скорость шарика постоянно растет, поэтому она достаточно точно отражает реальное поведение физической системы.

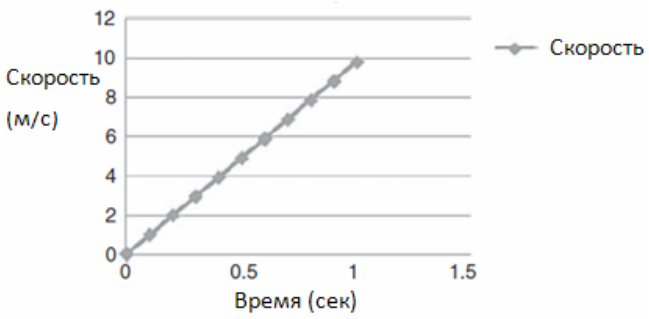

Рис. 5. Зависимость скорости падающего шарика от времени падения

К непрерывному моделированию часто прибегают в случаях исследования физических систем, включающих механические, тепловые или гидравлические компоненты. Но оно может быть использовано и для других целей, например, для исследования распространения заболеваний [4]. График, показывающий количество инфицированных людей в больших популяциях, может быть описан как непрерывная функция от времени, а следовательно, представлен в компьютере (рис. 6).

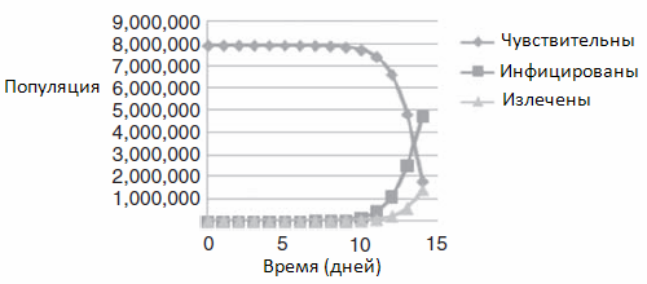

Рис. 6. Распространение заболевания в больших популяциях

Для непрерывного моделирования во многих случаях используют системы дифференциальных уравнений [8], поэтому довольно сложно назвать какие-либо программные продукты, созданные специально для непрерывного моделирования. Обычно для него используют либо универсальные математические пакеты (*Mathcad*, *Mathlab*, *Mathematica* и др.), либо уникальное программное обеспечение, разработанное под конкретную задачу.

Моделирование Монте-Карло

Как было сказано выше, термин «Монте-Карло» применяется к тем моделям, в которых течение времени не является существенным. Например, к сложным финансовым моделям, описанным с помощью электронных таблиц. Если параметрами модели являются такие количественные показатели, как, скажем, процентные ставки или денежные потоки, то задав эти параметры в виде случайных значений с некоторой функцией распределения, на выходе финансовой модели также получим набор случайных переменных. Оценки функции распределения этих переменных можно получить при многократном пересчете электронных таблиц, каждый раз принимая на вход распределения вероятностей, используемые в модели. Популярное программное обеспечение для стохастического моделирования электронных таблиц приведено в табл. 2.

Следует отметить большой потенциал данного подхода, так как аналитики в самых разных отраслях часто представляют данные в виде электронных таблиц [9]. Фактически, моделирование Монте-Карло можно считать частным случаем дискретно-событийного моделирования, в котором результат имитации становится известен после первого же наступившего события (генерации случайных величин).

Таблица 2

Программные продукты, применяемые для моделирования Монте-Карло

Программный продукт	Фирма-разработчик	Адрес в Интернете
@Risk	Palisade	http://www.palisade.com/
Crystal Ball	Oracle	http://www.decisioneering.com/
GPSS/H	Wolverine Software Corporation	http://www.wolverinesoftware.com/
Statistica	StatSoft	http:// www.statsoft.ru
Excel	Microsoft	http://www.microsoft.com

Однако один результат, как правило, не пригоден для статистических оценок, в то время как выборки, полученные при многократных прогонах модели, позволяют получить статистически достоверные данные.

МОДЕЛИРОВАНИЕ В *G-IPS ULTIMATE*

Так как моделирование Монте-Карло можно считать частным случаем дискретно-событийного моделирования с дополнительной статистической обработкой, а задачи непрерывного моделирования слишком предметно-ориентированы, наибольший интерес представляют средства дискретно-событийного моделирования, которые способны охватить задачи из разных прикладных областей. Подобные системы сегодня присутствуют как в промышленности (например, *SCADA*, *GPSS* и т. п.), так и в российской высшей школе [10]. Но в процессе анализа реального применения имитационного моделирования были выявлены не только сильные стороны, но и различные недостатки, начиная с высокой стоимости и заканчивая слишком специфическим способом представления знаний, который достаточно сложно освоить. В результате было решено разработать систему, в которой эти недостатки были бы исключены или минимизированы. В настоящее время такая система разрабатывается на кафедре информационной безопасности и программной инженерии Российского государственного социального университета и называется *G-IPS Ultimate*.

Система *G-IPS Ultimate* в большой степени ориентирована на дискретно-событийное моделирование. В её основе лежит дискретное описание компонентов системы в виде графовых моделей особой топологии и переходов (связей) между этими моделями [11]. Однако она может применяться и для некоторых задач непрерывного моделирования благодаря опции, позволяющей компонентам одной системы действовать абсолютно независимо друг от друга, т.е. параллельно. Кроме того, в *G-IPS Ultimate* реализован метод построения недетерминированных моделей с помощью задания вероятностей для действий, меняющих состояние моделируемой системы. При этом невозможно будет предугадать хронологию событий, воздействующих на систему, и в системе. *G-IPS Ultimate* позволяет также исключить временной фактор в модели, что делает её вполне применимой для решения задач с помощью моделирования Монте-Карло.

Если обратить внимание на то, что большинство реально функционирующих систем можно представить в виде систем массового обслуживания, то благодаря опции параллельного опроса датчиков, параллельного выполнения действий и параллельного прогона моделей можно значительно ускорить процесс имитации и, в частности, принятия решений [12].

СПИСОК ЛИТЕРАТУРЫ

1. Рыжиков Ю. И. Имитационное моделирование. Теория и технологии. – М. : Альтекс, 2004. – 384 с.
2. Шеннон Р. Имитационное моделирование систем – искусство и наука. – М. : Мир, 1978. – 425 с.
3. Ивин А. А., Никифоров А. Л. Словарь по логике. – М. : ВЛАДОС, 1997.
4. Sokolowski J. A., Banks C. M. Principles of modeling and simulation: multidisciplinary approach. – New Jersey : A JOHN WILEY & SONS, 2009.
5. Философия : энциклопедический словарь / под ред. А. А. Ивина. – М. : Гардарики, 2004.
6. Gheorghe L. Continuous/Discrete Co-simulation interfaces from formalization to implementation. – Montreal : Polytechnique de Montreal, 2009.
7. Isken M. Computer simulation in Management Engineering // Management Engineering. – Chicago : Healthcare Information and Management Systems Society, 2001. – P. 179 – 196.
8. Kwan Hee Han. Programmable Logic Controller // Object-Oriented Modeling, Simulation and Automatic Generation of PLC Ladder Logic. – 2010.
9. Томашевский В., Жданова Е. Имитационное моделирование в среде GPSS. – М. : Бестселлер, 2003. – 416 с.
10. Карпухин И. Н., Незнанов А. А. Системы имитационного моделирования учебного назначения в российской высшей школе. – М. : Спутник, 2011. – С. 132 – 142.
11. Карпухин И. Н., Незнанов А. А. Программные средства имитационного моделирования процессов принятия решений реального времени // Интегрированные модели, мягкие вычисления, вероятностные системы и комплексы программ в искусственном интеллекте. Т. 2. – Коломна, 2009. – С. 132 – 140.
12. Рыжиков Ю. И. Имитационное моделирование систем массового обслуживания. – Л. : ВИККИ им. А. Ф. Можайского, 1991. – 111 с.
13. Емельянов А. А., Власова Е. А., Дума Р. В. Имитационное моделирование экономических процессов. – М. : Финансы и статистика, 2002. – 368 с.
14. Павловский Ю. Н., Белотелов Н. В., Бродский Ю. И. Имитационное моделирование. – М. : Академия, 2008. – 240 с.

Материал поступил в редакцию 13.03.12.

Сведения об авторах

КАРПУХИН Илья Николаевич – аспирант кафедры информационной безопасностии программной инженерии Российского государственного социального университета, Москва

E-mail: SurStrat@mail.ru

КОРАБЛИН Юрий Прокофьевич – доктор технических наук, профессор кафедры информационной безопасности и программной инженерии Российского государственного социального университета

E-mail: y.p.k@mail.ru

НЕЗНАНОВ Алексей Андреевич – кандидат технических наук, доцент, заместитель заведующего отделением прикладной математики и информатики Национального исследовательского университета «Высшая школа экономики», Москва

E-mail: aneznanov@hse.ru

УДК 81 : 656.61.017.2.052

В. М. Гриняк, М. В. Трофимов

Оценка характера движения морских судов в лингвистических переменных*

Рассматривается проблема обнаружения маневра объекта в современных информационных системах управления движением на море при обработке навигационных данных двухкоординатной радиолокационной системой кругового обзора. Обсуждается задача оценки интенсивности маневрирования, связанная с выработкой тревожных сигналов и принятием управленческих решений в целях обеспечения безопасности коллективного движения судов. Предлагается модельная интерпретация задачи, основанная на машине нечеткого вывода типа Мамдани и ориентированная на вербальную обобщенную оценку маневренности судна операторами систем управления движением судов и судоводителями.

Ключевые слова: *управление движением судов, обнаружитель маневра, сопровождение объекта, нечеткая система типа Мамдани, лингвистическая оценка траектории движения*

ВВЕДЕНИЕ

Навигационная безопасность движения морских судов является актуальной проблемой эксплуатации водных транспортных путей. В зонах высокой интенсивности движения её решение возложено на особые информационные средства – бортовые и береговые системы управления движением судов (СУДС) [1, 2]. В соответствии с общепринятой концепцией построения таких систем, их задачи реализуются с использованием измерительной информации, доставляемой радарами и/или спутниковыми средствами траекторных измерений – транспондерами.

Сложившаяся практика судовождения, правила которой регламентируют, что управление судном есть исключительное право его капитана, отводит СУДС (как береговым, так и бортовым) роль особого инструмента информирования судоводителя о возможном наступлении опасной ситуации (столкновении) [3]. Генерация тревожного сигнала по какому-либо объекту или их группе служит указанием судоводителю (или оператору СУДС), на основании которого он принимает решение об изменении курса и скорости движения.

Обращение к автоматизированным средствам информационного обеспечения требует предельно формализованных представлений понятия «опасная ситуация», и здесь следует обратиться к опыту практического судовождения, показывающему, что главное условие безопасного движения – стремление не допустить чрезмерного сближения судов. При таком подходе к интерпретации опасности её формальным критерием служит уменьшение расстояния между объектами до некоторой критической величины, определяющей своего рода «зону безопасности» вокруг судна (корабельный домен) [4].

Маневрирующие и не маневрирующие объекты с точки зрения оценки безопасности имеют ряд принципиальных различий [5 – 7]. Во-первых, при внешнем наблюдении полностью достоверный прогноз траектории маневрирующего объекта невозможен. Во-вторых, если исходить из принятого на практике положения, что маневрирование судна, как правило, свидетельствует о попытке судоводителя придать движению безопасный характер и о его контроле над ситуацией, то для маневрирующих объектов вербальный уровень опасности заведомо ниже, чем для не маневрирующих. Это является побудительным мотивом создания таких информационных моделей выработки тревожных сигналов, которые выделяли бы различные уровни опасности ситуации (типа «очень опасная», «опасная», «почти безопасная» и т. п.), учитывая при этом маневренные характеристики траектории движения судна.

Специальные алгоритмы обнаружения маневра нередко реализуются на практике при решении задач сопровождения траектории и наблюдения [8]. Они используются для «переключения» параметров системы на участках маневрирования объекта, характеризующихся несоответствием между моделируемым и реальным движением. Вместе с тем, известные обнаружители маневра предназначены для автоматизированных систем и неудобны для поддержки вербального принятия решений операторами СУДС и судоводителями.

В настоящей работе рассматривается новый подход к обнаружению маневра, связанный с классификацией наблюдаемых объектов по степени интенсивности маневрирования с помощью аппарата нечеткой логики. Получаемые таким образом обобщённые лингвистические представления о характере движения того или иного судна используются при распознавании опасных ситуаций бортовой или береговой СУДС.

* Работа выполнена в рамках Государственного задания высшим учебным заведениям в части проведения научно-исследовательских работ, проект № 7.2104.2011.

ОСНОВНЫЕ МОДЕЛЬНЫЕ ПРЕДСТАВЛЕНИЯ И ПОСТАНОВКА ЗАДАЧИ

Пусть движение объекта описывается следующими уравнениями:

$$\begin{aligned} x(t_{k+1}) &= x(t_k) + v_x(t_k)\tau + q_x(t_k), \\ y(t_{k+1}) &= y(t_k) + v_y(t_k)\tau + q_y(t_k) \end{aligned} \quad (1)$$

Здесь k – идентификатор (порядковый номер) момента времени, $x(t_k)$, $y(t_k)$ – координаты объекта в момент времени t_k; $v_x(t_k)$, $v_y(t_k)$ – компоненты вектора скорости объекта, $q_x(t_k)$, $q_y(t_k)$ – компоненты вектора случайных, немоделируемых параметров движения, $\tau = t_{k+1} - t_k$.

Пусть измеряемыми параметрами являются декартовы координаты объекта. Тогда модель рассматриваемой задачи можно представить следующим дискретным матричным уравнением «состояние-измерение»:

$$\begin{aligned} s(t_{k+1}) &= \Phi s(t_k) + q(t_k), \\ z(t_k) &= Hs(t_k) + r(t_k). \end{aligned} \quad (2)$$

Здесь $s(t_k) = (x(t_k), v_x(t_k), y(t_k), v_y(t_k))^{\mathrm{T}}$ – вектор состояния объекта, включающий его координаты и их производные (T – символ транспонирования), $q(t_k)$ – вектор немоделируемых параметров движения, $z(t_k)$ – вектор измерений, $r(t_k)$ – вектор погрешностей измерений. Имея в виду (1), матричные коэффициенты Ф и H системы уравнений (2) равны соответственно

$$\Phi = \begin{Vmatrix} 1 & \tau & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \tau \\ 0 & 0 & 0 & 1 \end{Vmatrix}, \; H = \begin{Vmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{Vmatrix}.$$

Модель оценивания вектора состояния $s(t_k)$ по измерениям $z(t_k)$ может быть представлена следующим уравнением:

$$\hat{s}(t_{k+1}) = \Phi\hat{s}(t_k) + K(z(t_k) - H\Phi\hat{s}(t_k)) \quad (3)$$

Здесь $\hat{s}(t_k)$ – оценка вектора состояния, K – матричный коэффициент.

Пусть матрица K определяется схемой классического $\alpha - \beta$ алгоритма [9] и имеет вид

$$K = \begin{Vmatrix} \alpha & 0 \\ \beta/\tau & 0 \\ 0 & \alpha \\ 0 & \beta/\tau \end{Vmatrix}, \quad (4)$$

а коэффициенты α и β выбираются по следующему правилу:

$$\alpha_k = \frac{2(2k+1)}{(k+2)(k+1)}, \; \beta_k = \frac{6}{(k+2)(k+1)},$$

где k– порядковый номер момента времени в формуле (3).

Пусть J – число измерений (и, соответственно, итераций), участвующих в оценке вектора состояния $s(t_k)$ итерационной процедурой (3), так что $k = \overline{1,J}$. При увеличении J коэффициенты α и β асимптотически уменьшаются до 0. Поэтому алгоритм (3), реализованный с большим J, будет успешно оценивать координаты и скорости объектов, движущихся прямолинейно и равномерно, а для маневрирующих объектов погрешность оценки вектора состояния будет довольно высокой.

Пусть $\hat{s}_J(t_i)$ – оценка вектора состояния системы в момент времени t_i, полученная итерационным алгоритмом (3) при обработке J последних измерений. Если при этом задача одновременно решается при J, $J-1$, $J-2$, … и, наконец, только при двух измерениях (минимально возможном их количестве), то тогда в момент времени t_i будем иметь кортеж векторов оценки

$$\hat{S}_J(t_i) = \{\hat{s}_2(t_i), \hat{s}_3(t_i), \hat{s}_4(t_i), \ldots, \hat{s}_J(t_i)\}. \quad (5)$$

Введем вектор $\delta z(t_{k+1}) = z(t_{k+1}) - H\hat{s}(t_{k+1})$, характеризующий невязку измерения при оценке вектора состояния уравнением (3). Пусть $\|\delta z_J(t_i)\|$ – евклидова норма вектора невязки $\delta z(t_i)$, полученного в момент времени t_i при реализации итерационного алгоритма (3), обрабатывающего J последних измерений. Тогда при оценке вектора состояния в каждый момент времени t_i, наряду с кортежем векторов оценки (5), будем иметь кортеж норм векторов невязок

$$\begin{aligned} \delta_J(t_i) = \{&\|\delta z_2(t_i)\|, \|\delta z_3(t_i)\|, \\ &\|\delta z_4(t_i)\|, \ldots, \|\delta z_J(t_i)\|\}. \end{aligned} \quad (6)$$

Элементы кортежа (6) являются, по сути, основным информативным признаком, характеризующим качество оценки вектора состояния алгоритмом (3–4) с тем или иным значением J. Задача обнаружения маневра объекта и оценки его интенсивности сводится, таким образом, к анализу свойств кортежа (6).

МЕТОД РЕШЕНИЯ ЗАДАЧИ

Перейдем от (6) к кортежу относительных величин

$$\Delta_J(t_i) = \{L_2(t_i), L_3(t_i), L_4(t_i), \ldots, L_J(t_i)\}, \quad (7)$$

где $L_j(t_i) = \dfrac{\|\delta z_j(t_i)\|}{\sigma}$, σ – величина, характеризующая среднеквадратичное отклонение погрешности измерений $r(t_k)$ в системе (2).

Введем лингвистическую переменную $Q_j(t_i)$, $j = \overline{2,J}$ «Качество оценки вектора состояния алгоритмом (3–4) в момент времени t_i по j последним измерениям» с термами «good» (g, «хорошее») и «bad» (b, «плохое»). Пусть термы имеют следующие функции принадлежности типа «дополнение», определённые на универсальном множестве $u \in [0,3]$:

$$\begin{aligned} \mu_g(\mathrm{u}) &= 1 - \frac{1}{1 + \exp(-\mathrm{a}_1(\mathrm{u} - \mathrm{c}_1))}, \\ \mu_b(\mathrm{u}) &= \frac{1}{1 + \exp(-\mathrm{a}_2(\mathrm{u} - \mathrm{c}_2))}, \end{aligned} \quad (8)$$

где a_1, a_2, c_1, c_2, – настраиваемые параметры.

Введём лингвистическую переменную $P(t_i)$ «Характер движения судна в момент времени t_i » с термами «high-high-maneuverable» (hhm, «очень высокоманев-

ренное»), «high-maneuverable» (hm, «высокоманевренное»), «low-maneuverable» (lm, «низкоманевренное») и «low-low-maneuverable» (llm, «очень низкоманевренное»). Пусть термы имеют следующие функции принадлежности типа «кластер», определённые на универсальном множестве $v \in [2,J]$:

$$\mu_{hhm}(v) = 1 - \frac{1}{1 + \exp(-a_3(v - c_3))},$$

$$\mu_{hm}(v) = \exp\left(-\frac{(v - c_4)^2}{a_4}\right),$$

$$\mu_{lm}(v) = \exp\left(-\frac{(v - c_5)^2}{a_5}\right), \qquad (9)$$

$$\mu_{llm}(v) = \frac{1}{1 + \exp(-a_6(v - c_6))},$$

где a_3–a_6, c_3–c_6, – настраиваемые параметры.

Пусть переменные $Q_j(t_i)$ обрабатываются машиной нечеткого вывода типа Мамдани [10], на вход которой подается кортеж величин (7), а на выходе формируется числовое значение $m(t_i)$ – вещественное число, характеризующее степень интенсивности маневрирования судна. Машина нечеткого вывода работает согласно системе правил, представленной в табл. 1.

Таблица 1

Система правил машины нечеткого вывода типа Мамдани (уровень 1)

№	$Q_2(t_i)$	$Q_2(t_i)$	$Q_4(t_i)$	…	$Q_{J-2}(t_i)$	$Q_{J-1}(t_i)$	$Q_J(t_i)$	$P(t_i)$
1	*g*	*g*	*g*	…	*g*	*g*	*g*	P_1
2	*g*	*g*	*g*	…	*g*	*g*	*b*	P_2
3	*g*	*g*	*g*	…	*g*	*b*	*b*	P_3
…	…	…	…	…	…	…	…	…
$J-1$	*g*	*b*	*b*	…	*b*	*b*	*b*	P_{J-1}
J	*b*	*b*	*b*	…	*b*	*b*	*b*	P_J

Работу нечеткого алгоритма оценки степени интенсивности маневрирования судна можно окончательно представить схемой, показанной на рис. 1. Здесь $L_j(t_i)$ – величины кортежа (7) в момент времени t_i (вход), если $L_j(t_i) > 3$, то вход принимается равным 3; $m(t_i)$ – определённая системой типа Мамдани M_1 (уровень 1) в момент времени t_i степень интенсивности маневрирования судна.

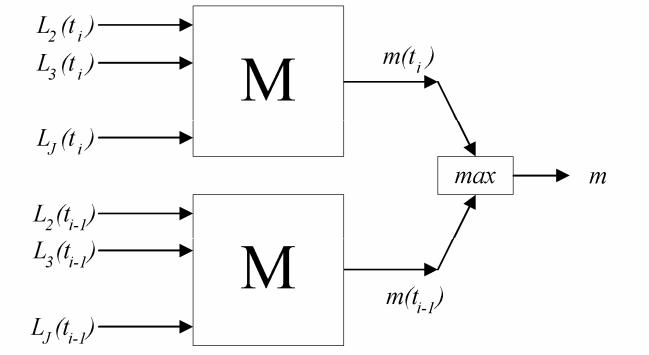

Рис. 1. Схема работы нечеткого алгоритма оценки степени интенсивности маневрирования судна

Из величин $m(t_{i-1})$ и $m(t_i)$ выбирается максимальное значение, которое и принимается за окончательное значение степени интенсивности маневрирования судна. Выбор максимального из двух соседних $m(t_{i-1})$ и $m(t_i)$ необходим для повышения устойчивости работы системы при больших ошибках измерений (для фильтрации случайных выбросов).

Настройка описанной системы состоит в задании максимального количества измерений J, параметров функций принадлежности $a_1,...,a_6$, $c_1,...,c_6$, значений лингвистической переменной P_i и величины σ, характеризующей погрешность измерений.

РЕЗУЛЬТАТЫ ЧИСЛЕННОГО МОДЕЛИРОВАНИЯ

При моделировании задачи было принято, что информационной базой СУДС является двухкоординатный радар кругового обзора (например, типа Raytheon) с периодом обращения 3 с и разрешением по углу и дальности соответственно $\Delta\varphi = 0.03^\circ$ и $\Delta r = 6$ м. Максимальное количество измерений было принято равным $J = 10$. Принятые значения лингвистической переменной P_i приведены в табл. 2.

Таблица 2

Значения лингвистической переменной P_i

i	P_i
1	*llm*
2	*llm*
3	*lm*
4	*lm*
5	*lm*
6	*hm*
7	*hm*
8	*hm*
9	*hhm*
10	*hhm*

Заданные значения параметров функций принадлежности (8) и (9) приведены в табл. 3 (в данном случае параметры задаются экспертом, система не подвергается настройке на обучающей выборке, см. также рис. 2 и 3).

Таблица 3

Значения параметров функций принадлежности

i	a_i	c_i
1	5.0	1.5
2	5.0	1.5
3	6.0	3.0
4	2.0	4.5
5	2.0	7.5
6	6.0	9.0

Если вероятностные характеристики ошибок измерений r_k хорошо известны, то величина σ может зада-

ваться априорно. В тех случаях, когда r_k можно оценить только приблизительно, с точностью до порядка величин, величина σ может быть оценена формулой:

$$\hat{\sigma}_k = \frac{\sum_{i=1}^{k} \left\| \delta z_2(t_i) \right\|}{k},$$

где k – порядковый номер момента времени, прошедшего от начала наблюдения за судном (в настоящей работе рассматривается именно этот случай).

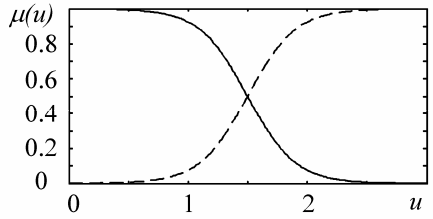

Рис. 2. Функции принадлежности термов «good» (сплошная линия) и «bad» (пунктир)

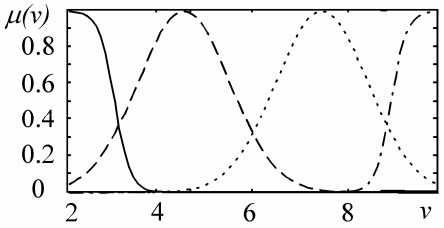

Рис. 3. Функции принадлежности термов «hhm» (сплошная линия), «hm» (пунктир), «lm» (точки) и «llm» (точки и пунктир)

На рис. 4 показана моделируемая траектория движения морского судна. Вначале судно движется прямолинейно и равномерно, а затем совершает манёвр – поворот с радиусом 300 м (такие кинематические свойства вполне характерны для современных судов небольшой размерности).

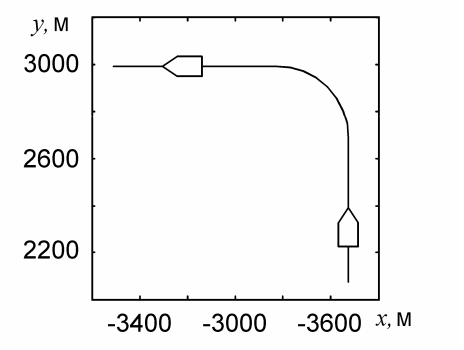

Рис. 4. Траектория движения судна

На рис. 5 показан результат решения задачи оценки степени интенсивности маневрирования для судна, движущегося по изображенной траектории со скоростью 10 м/с (левая колонка рисунков) и 20 м/с (правая колонка рисунков). Здесь t – время, прошедшее от начала работы алгоритма, m – определённое по мере движения судна значение степени интенсивности его маневрирования (рис. 5а и 5б). В данном случае m близко к максимальному значению (около 10) на прямолинейном участке траектории и уменьшается до значений ≈3,5 – 5 при повороте на скорости 10 м/с и до ≈2,5 при повороте на скорости 20 м/с. При этом алгоритм достаточно быстро реагирует на изменение характера движения судна (не более 15 секунд, участок [63, 78] секунд на рис. 5б).

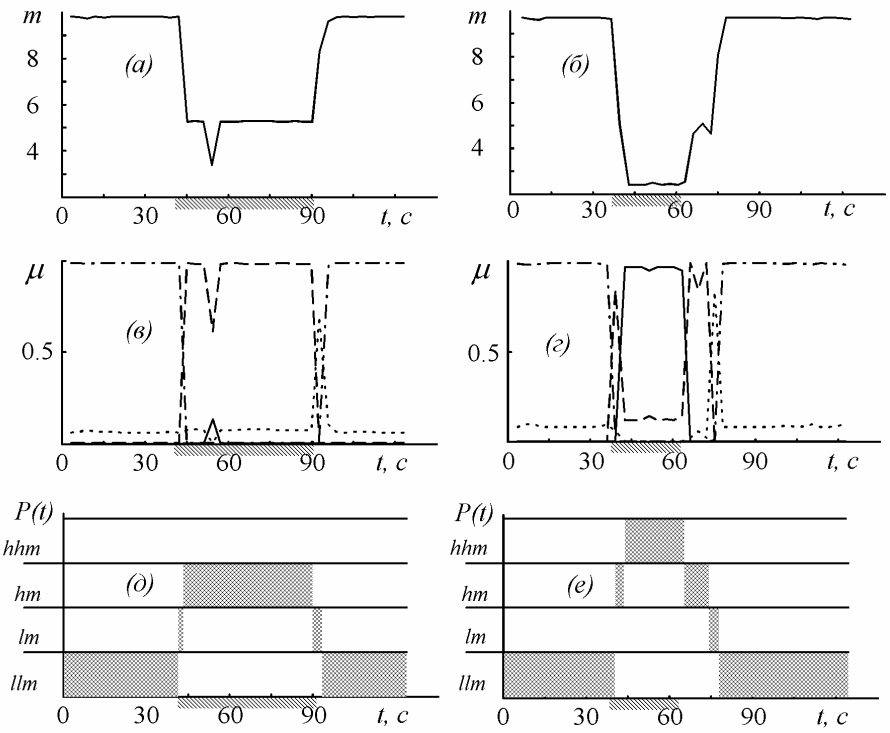

Рис 5. Работа алгоритма нечеткой оценки степени маневрирования судна (штриховкой под осью абсцисс обозначен участок маневрирования судна)

Рис. 5в и 5г показывают значения функций принадлежности термов «hhm» (сплошная линия), «hm» (пунктир), «lm» (точки) и «llm» (точки и пунктир) по мере движения судна. Так, при движении со скоростью 10 м/с на участке [0, 40] секунд движение с наивысшей степенью оценено как «очень низкоманевренное», на участке [40, 43] – как «низкоманевренное», на участке [43, 90] – как «высокоманевренное» и т. д. (рис. 5в). Соответствующие термы с максимальными значениями функции принадлежности приведены на рис. 5д и 5е, которые иллюстрируют обобщённую вербальную лингвистическую интерпретацию траекторных свойств движения в разрезе значений «очень высокоманевренное», «высокоманевренное», «низкоманевренное» и «очень низкоманевренное».

Перечислим основные полученные результаты.

В работе сформулированы методологические основы и дана концепция решения задачи оценки интенсивности маневрирования в лингвистических переменных. Рассмотрена нечеткая модель классификации характера движения морского судна, ориентированная на реализацию в качестве обнаружителя маневра. Предложен нечеткий алгоритм лингвистической оценки степени маневра, основанный на сопровождении траектории судна классическим $\alpha-\beta$ алгоритмом с различной степенью фильтрации и обработки параметров сопровождения машиной нечеткого вывода типа Мамдани. На модельном примере продемонстрирована конструктивность и эффективность предлагаемой методики.

Результаты исследования ориентированы на расширение функциональных возможностей существующих береговых и бортовых систем управления движением судов.

СПИСОК ЛИТЕРАТУРЫ

1. ОАО «Норфес» [Электрон. ресурс]. – URL: http://www.norfes.ru/
2. Группа компаний «ТРАНЗАС» [Электрон. ресурс]. – URL: http://www.transas.ru/
3. Huges T. When is a VTS not a VTS // The J. of Navigation. – 2009. – Vol. 62, № 3. – P. 439 – 442.
4. Pietrzykowski Z., Uriasz J. The Ship Domain – a Criterion of Navigational Safety Assessment in an Open Sea Area // The J. of Navigation. – 2009. – Vol. 62, №1. – P. 93 – 108.
5. Девятисильный А. С., Гриняк В. М. Прогнозирование опасных ситуаций при управлении движением на море // Известия РАН. Теория и системы управления. – 2004. – № 3. – С. 127 – 136.
6. Гриняк В. М., Головченко Б. С., Малько В. Н. Распознавание опасных ситуаций системами управления движением судов // Транспорт: наука, техника, управление. – 2011. – № 8.
7. Гриняк В. М., Дорожко В. М., Лоскутов Н. В., Кириченко О. В. Модели обеспечения безопасности на морских акваториях в условиях высокой интенсивности движения // НТИ. Сер. 2. – 2004. – № 9. – С. 6 – 8.
8. Бакулев П. А., Сычев М. И., Нгуен Чонг Лыу. Многомодельный алгоритм сопровождения траектории маневрирующей цели по данным обзорной РЛС // Радиотехника. – 2004. – №1.
9. Benedict T. R., Bordner G. R. Synthesis of an optimal set of radar track-while-scan smoothing equations // IRE Trans. on AC-1 (July 1962). – 1962. – P. 27 – 32.
10. Круглов В. В., Дли М. И., Голунов Р. Ю. Нечеткая логика и искусственные нейронные сети. – М. : Физматлит, 2001. – 224 с.

Материал поступил в редакцию 07.02.12.

Сведения об авторах

ГРИНЯК Виктор Михайлович – кандидат технических наук, заведующий кафедрой Владивостокского государственного университета экономики и сервиса
E-mail: Viktor.Grinyak@vvsu.ru

ТРОФИМОВ Максим Валерьевич – ассистент Владивостокского государственного университета экономики и сервиса
E-mail: bugzex@yandex.ru

УДК 001.8

В. В. Нешитой

Метод наибольшего правдоподобия, устойчивый метод и энтропия

Исследуется близость двух методов оценивания параметров непрерывных, в том числе ранговых, распределений – метода наибольшего правдоподобия Р. Фишера и устойчивого метода автора. Показывается, что оба метода базируются на одном и том же равенстве $p(x) = \beta z p(z)$, устанавливающем взаимосвязь между обобщенной плотностью p(x) и двухпараметрической плотностью p(z). В методе наибольшего правдоподобия используется логарифм этого равенства. Показывается, что логарифмическая функция правдоподобия, взятая с обратным знаком, представляет собой энтропию распределения. При преобразовании распределений второй системы к форме первой системы энтропия плотности уменьшается, т.е. появляется новая информация.

Ключевые слова: *системы непрерывных распределений, методы оценивания параметров, метод наибольшего правдоподобия, устойчивый метод, логарифмическая функция правдоподобия, критерий согласия, энтропия распределения, извлечение информации из распределения*

МЕТОД НАИБОЛЬШЕГО ПРАВДОПОДОБИЯ

В 1912 г. Р. Фишер предложил для нахождения оценок параметров аппроксимирующих распределений метод наибольшего (максимального) правдоподобия. Суть метода сводится к тому, что наступившие события имели наибольшую вероятность наступить при заданном комплексе условий. Вероятность совместного наступления событий при условии их независимости равна произведению вероятностей наступивших событий. Это произведение называется функцией правдоподобия:

$$L = \prod_{i=1}^{n} f(x_i, \theta_j) \ . \qquad (1)$$

В качестве оценок максимального правдоподобия параметров Θ_j принимаются те их значения, при которых функция правдоподобия имеет максимум. Дифференцируя функцию правдоподобия по параметрам Θ_j и приравнивая частные производные к нулю, получают систему уравнений правдоподобия, решая которую, находят оценки параметров. Однако здесь следует отметить, что в случае распределений с тремя и тем более с четырьмя параметрами получаются весьма сложные уравнения, решение которых сопряжено с большими трудностями.

Ниже будут рассматриваться четырехпараметрические непрерывные распределения автора [1], задающие три системы непрерывных распределений:

$$p(x) = Ne^{k\beta x}(1 - \alpha u e^{\beta x})^{\frac{1}{u}-1};$$

$$p(t) = Nt^{k\beta-1}(1 - \alpha u t^{\beta})^{\frac{1}{u}-1};$$

$$p(y) = \frac{N}{y}(\ln y)^{k\beta-1}\left[1 - \alpha u (\ln y)^{\beta}\right]^{\frac{1}{u}-1}.$$

Приведенные плотности распределения включают как частные случаи широкое разнообразие непрерывных распределений.

ДРУГИЕ ФОРМЫ ЗАДАНИЯ ФУНКЦИИ ПРАВДОПОДОБИЯ

Приведенная форма задания функции правдоподобия (1) не является единственно возможной. Суть метода наибольшего правдоподобия не изменится, если из функции правдоподобия L извлечь корень n-ой степени:

$$L^* = \sqrt[n]{\prod_{i=1}^{n} f(x_i, \theta_j)} = \left[\prod_{i=1}^{n} f(x_i, \theta_j)\right]^{\frac{1}{n}}. \qquad (2)$$

Здесь функция правдоподобия задана средним геометрическим вероятностей n независимых событий. Это вторая возможная форма задания функции правдоподобия.

Третья и четвертая формы получаются путем логарифмирования первых двух форм:

$$\ln L = \sum_{i=1}^{n} \ln f(x_i, \theta_j); \qquad (3)$$

$$\ln L^* = \frac{1}{n}\sum_{i=1}^{n} \ln f(x_i, \theta_j) = \overline{\ln f(x_i, \theta_j)}. \qquad (4)$$

Здесь форма (3) задана суммой логарифмов вероятностей n независимых событий, а форма (4) – средним значением логарифмов вероятностей.

Все четыре формы представления функции правдоподобия являются равноправными в том смысле, что дают одни и те же оценки параметров аппроксимирующего распределения. При этих оценках каждая функция правдоподобия принимает свое максимальное значение.

Рассмотрим пример.

Пусть случайная величина T имеет показательный закон распределения

$$p(t) = \frac{\alpha}{e^{\alpha t}} \ . \qquad (5)$$

Необходимо оценить параметр α по результатам наблюдений $t_1, t_2, ..., t_n$, где n – объем выборки.

Запишем для показательного закона все рассмотренные нами формы задания функции правдоподобия:

1. $L=\prod_{i=1}^{n}p(t_i)=\frac{\alpha}{e^{\alpha t_1}}\cdot\frac{\alpha}{e^{\alpha t_2}}\cdots\frac{\alpha}{e^{\alpha t_n}}=\frac{\alpha^n}{e^{\alpha\sum t_i}}$;
2. $L^*=\left[\prod_{i=1}^{n}p(t_i)\right]^{\frac{1}{n}}=\frac{\alpha}{e^{\alpha(\sum t_i)/n}}=\frac{\alpha}{e^{\alpha\bar{t}}}$;
3. $\ln L=n\ln\alpha-\alpha\sum_{i=1}^{n}t_i$;
4. $\ln L^*=\ln\alpha-\alpha\bar{t}$.

Дифференцируя любую из них по параметру α и приравнивая первую производную к нулю, найдем:

$$\alpha=\frac{1}{\bar{t}}\ . \qquad (6)$$

МОДИФИЦИРОВАННЫЙ МЕТОД НАИБОЛЬШЕГО ПРАВДОПОДОБИЯ

В качестве логарифмической функции правдоподобия для непрерывных распределений в случае генеральной совокупности может быть принято математическое ожидание логарифма плотности распределения [2]:

$$M\left[\ln p(x)\right]=\ln L\ . \qquad (7)$$

Использование этой функции правдоподобия позволяет значительно проще решать такие задачи, как вычисление оценок параметров, вычисление значений функции правдоподобия при заданных значениях параметров, проведение научных исследований. Эта форма задания функции правдоподобия следует из формулы (4), в которой среднее значение логарифмической функции правдоподобия заменено математическим ожиданием.

Используем функцию правдоподобия (7) для нахождения оценки параметра α показательного закона распределения (5). Здесь порядок расчета следующий:

- логарифмируем плотность распределения

$$\ln p(t)=\ln\alpha-\alpha t\ ;$$

- находим математическое ожидание логарифма плотности

$$M\left[\ln p(t)\right]=\ln\alpha-\alpha M(t)\ ;$$

- находим частную производную по параметру α и приравниваем ее к нулю

$$\frac{\partial M\left[\ln p(t)\right]}{\partial\alpha}=\frac{1}{\alpha}-M(t)=0\ ;$$

- из полученного уравнения правдоподобия находим зависимость между параметром α и математическим ожиданием случайной величины T в случае показательного закона распределения

$$\alpha=\frac{1}{M(t)}\ .$$

Последняя формула справедлива для генеральной совокупности. Для перехода к выборочной совокупности заменяем математическое ожидание случайной величины T его оценкой, вычисленной по выборке объемом n:

$$\bar{t}=\frac{1}{n}\sum_{i=1}^{n}t_i\ .$$

Оценка параметра α будет равна $\alpha=1/\bar{t}$, т.е. вычисляется по прежней формуле (6).

Рассмотрим пример на вычисление M[lnp(x)]. Логарифмическую функцию правдоподобия можно вычислить двумя способами. Первый, традиционный, способ – это ее вычисление путем интегрирования:

$$M\left[\ln p(x)\right]=\int\left[\ln p(x)\right]p(x)dx\ .$$

Пусть плотность p(x) задается четырехпараметрической формулой

$$p(x)=\frac{\beta(\alpha u)^k\Gamma(\mathrm{k}+1/\mathrm{u})}{\Gamma(\mathrm{k})\Gamma(1/\mathrm{u})}e^{k\beta x}\left(1-\alpha u e^{\beta x}\right)^{\frac{1}{u}-1}\ , \qquad (8)$$

которая относится к первому типу первой системы непрерывных распределений [1].

Тогда логарифмическая функция правдоподобия будет выражаться интегралом

$$M\left[\ln p(x)\right]=\int\left[\begin{array}{l}\ln\dfrac{\beta(\alpha u)^k\Gamma(\mathrm{k}+1/\mathrm{u})}{\Gamma(\mathrm{k})\Gamma(1/\mathrm{u})}+k\beta x+\\ +\left(\dfrac{1}{u}-1\right)\ln\left(1-\alpha u e^{\beta x}\right)\end{array}\right]p(x)dx\ ,$$

где плотность p(x) задается формулой (8).

При втором способе используется метод дифференцирования [2].

Прологарифмируем плотность распределения (8):

$$\ln p(x)=\ln\beta+k\ln\alpha u+\ln\Gamma(\mathrm{k}+1/\mathrm{u})-\ln\Gamma(\mathrm{k})-\ln\Gamma(1/\mathrm{u})+\mathrm{k}\beta\mathrm{x}+(1/\mathrm{u}-1)\ln\left(1-\alpha\mathrm{u}e^{\beta\mathrm{x}}\right).$$

На основании последнего равенства запишем логарифмическую функцию правдоподобия

$$\ln L=M\left[\ln p(x)\right]=\ln\beta+k\ln\alpha u+\ln\Gamma(\mathrm{k}+1/\mathrm{u})-\ln\Gamma(\mathrm{k})-\ln\Gamma(1/\mathrm{u})+k\beta M(x)+(1/\mathrm{u}-1)\mathrm{M}\left[\ln\left(1-\alpha\mathrm{u}e^{\beta\mathrm{x}}\right)\right]. \qquad (9)$$

Возьмем от функции (9) частные производные по параметрам α, β, k, u и приравняем их к нулю. В результате получим систему четырех уравнений правдоподобия с четырьмя неизвестными параметрами:

$$\left\{\begin{array}{l}\dfrac{\partial\ln L}{\partial\alpha}=\dfrac{k}{\alpha}+\left(\dfrac{1}{u}-1\right)M\left(\dfrac{-ue^{\beta x}}{1-\alpha ue^{\beta x}}\right)=0\\ \dfrac{\partial\ln L}{\partial\beta}=\dfrac{1}{\beta}+kM(x)+\left(\dfrac{1}{u}-1\right)M\left(\dfrac{-\alpha ue^{\beta x}x}{1-\alpha ue^{\beta x}}\right)=0\\ \dfrac{\partial\ln L}{\partial k}=\ln\alpha u+\psi\left(k+\dfrac{1}{u}\right)-\psi(k)+\beta M(x)=0\\ \dfrac{\partial\ln L}{\partial u}=\dfrac{k}{u}+\psi\left(k+\dfrac{1}{u}\right)\left(-\dfrac{1}{u^2}\right)-\psi\left(\dfrac{1}{u}\right)\left(-\dfrac{1}{u^2}\right)+\\ +\left(-\dfrac{1}{u^2}\right)M\left[\ln\left(1-\alpha ue^{\beta x}\right)\right]+\left(\dfrac{1}{u}-1\right)M\left(\dfrac{-\alpha ue^{\beta x}}{1-\alpha ue^{\beta x}}\right)=0\ .\end{array}\right. \qquad (10)$$

Оценки четырех параметров находятся путем решения системы уравнений правдоподобия (10).

Полученные уравнения можно несколько упростить. Так, из первого уравнения правдоподобия имеем

$$k=\alpha(1-u)M\left(\frac{e^{\beta x}}{1-\alpha ue^{\beta x}}\right). \qquad (11)$$

Умножив четвертое уравнение правдоподобия на u, получим:

$$k+\psi\left(k+\frac{1}{u}\right)\left(-\frac{1}{u}\right)-\psi\left(\frac{1}{u}\right)\left(-\frac{1}{u}\right)+\left(-\frac{1}{u}\right)M\left[\ln\left(1-\alpha u e^{\beta x}\right)\right]-$$
$$-\alpha(1-u)M\left(\frac{e^{\beta x}}{1-\alpha u e^{\beta x}}\right)=0.$$

С учетом (11) последнее равенство примет вид:

$$\psi\left(k+\frac{1}{u}\right)-\psi\left(\frac{1}{u}\right)+M\left[\ln\left(1-\alpha u e^{\beta x}\right)\right]=0\ . \quad (12)$$

Перепишем систему уравнений правдоподобия для первого типа распределений, заданных плотностью (8):

$$\begin{cases} k-\alpha(1-u)M\left(\dfrac{e^{\beta x}}{1-\alpha u e^{\beta x}}\right)=0 \\ \dfrac{1}{\beta}+kM(x)-\alpha(1-u)M\left(\dfrac{xe^{\beta x}}{1-\alpha u e^{\beta x}}\right)=0 \\ \ln\alpha u+\psi\left(k+\dfrac{1}{u}\right)-\psi(k)+\beta M(x)=0 \\ \psi\left(k+\dfrac{1}{u}\right)-\psi\left(\dfrac{1}{u}\right)+M\left[\ln\left(1-\alpha u e^{\beta x}\right)\right]=0\ . \end{cases} \quad (13)$$

С учетом двух последних уравнений системы (13) логарифмическая функция правдоподобия запишется в окончательном виде [2]:

$$\ln L=M[\ln p(x)]=\ln\beta+\ln\frac{\Gamma(\mathrm{k}+1/u)}{\Gamma(\mathrm{k})\Gamma(1/\mathrm{u})}+$$
$$+k\left[\psi(k)-\psi\left(k+\frac{1}{u}\right)\right]+\left(\frac{1}{u}-1\right)\left[\psi\left(\frac{1}{u}\right)-\psi\left(k+\frac{1}{u}\right)\right]\ . \quad (14)$$

Таким образом, логарифмическая функция правдоподобия получена достаточно простым способом без интегрирования.

Из (14) следует, что логарифмическая функция правдоподобия зависит от трех параметров: β, k, u. Она может быть вычислена по этой формуле для распределений первого типа не только первой системы непрерывных распределений, но и второй и третьей систем, если их привести к форме плотности p(x), т.е. представить в виде tp(t)=f(lnt), yp(y)lny=f(lnlny).

ОЦЕНИВАНИЕ ПАРАМЕТРОВ ПО МЕТОДУ НАИБОЛЬШЕГО ПРАВДОПОДОБИЯ

Традиционно считается, что оценки параметров по методу наибольшего правдоподобия находятся путем решения системы уравнений правдоподобия, при этом число уравнений равно числу параметров.

В нашем примере обобщенное распределение (8) содержит четыре параметра: два параметра формы k, u, масштабный параметр β и параметр сдвига α. Главными здесь являются параметры формы.

Уравнения правдоподобия (13) оказались весьма сложными, причем в каждом из них имеются все четыре параметра. Решить такую систему практически невозможно. Однако, как показали исследования, в этом нет необходимости. Чтобы упростить решение поставленной задачи, ее надо разделить на два этапа.

На первом этапе разрабатываются два критерия (показателя), которые зависят только от двух параметров формы k, u. По этим критериям устанавливается тип аппроксимирующего распределения, а оценки параметров k, u вычисляются либо графическим методом, либо путем решения системы **двух уравнений с двумя неизвестными**. Такая система уравнений решается значительно проще, чем система четырех уравнений с четырьмя неизвестными, заданная, например, формулами (13).

На втором этапе при известных оценках параметров формы по простым формулам вычисляются оценки параметров α, β.

Такой подход позволил автору разработать устойчивый метод вычисления наилучшего аппроксимирующего распределения и нахождения оценок его параметров.

Начнем решать задачу оценивания параметров по методу наибольшего правдоподобия со второго этапа.

Рассмотрим случайную величину

$$Z=\alpha u e^{\beta X}\ . \quad (15)$$

и найдем плотность распределения случайной величины Z по известной формуле *p(z)=p(x)(dx/dz)=p(x)/(dz/dx)*.

Первая производная от z по x равна $dz/dx=\alpha u\beta e^{\beta x}$. Тогда из плотности (8) получим

$$p(z)=\frac{\Gamma(\mathrm{k}+1/\mathrm{u})}{\Gamma(\mathrm{k})\Gamma(1/\mathrm{u})}z^{k-1}(1-z)^{\frac{1}{u}-1}, \quad (16)$$

т.е. имеем бета-распределение.

Приведем плотность (16) к форме плотности (8), т.е. представим ее в виде зависимости *zp(z)=f(lnz)*:

$$zp(z)=\frac{\Gamma(\mathrm{k}+1/\mathrm{u})}{\Gamma(\mathrm{k})\Gamma(1/\mathrm{u})}e^{k\ln z}\left(1-e^{\ln z}\right)^{\frac{1}{u}-1}.$$

Запишем для последней плотности логарифмическую функцию правдоподобия:

$$\ln L_{(z)}=M[\ln zp(z)]=\ln\Gamma\left(\mathrm{k}+\frac{1}{\mathrm{u}}\right)-$$
$$-\ln\Gamma(\mathrm{k})\text{-}\ln\Gamma\left(\frac{1}{\mathrm{u}}\right)+kM(\ln z)+\left(\frac{1}{u}-1\right)M[\ln(1-z)]\ . \quad (17)$$

Найдем уравнения правдоподобия:

$$\begin{cases} \dfrac{\partial\ln L_{(z)}}{\partial k}=\psi\left(k+\dfrac{1}{u}\right)-\psi(k)+M(\ln z)=0 \\ \dfrac{\partial\ln L_{(z)}}{\partial u}=\psi\left(k+\dfrac{1}{u}\right)\left(-\dfrac{1}{u^2}\right)-\psi\left(\dfrac{1}{u}\right)\left(-\dfrac{1}{u^2}\right)+ \\ +\left(-\dfrac{1}{u^2}\right)M[\ln(1-z)]=0. \end{cases}$$

Из первого уравнения правдоподобия имеем:

$$M(\ln z)=\psi(k)-\psi\left(k+\frac{1}{u}\right)\ . \quad (18)$$

Из второго уравнения правдоподобия находим:

$$M[\ln(1-z)]=\psi\left(\frac{1}{u}\right)-\psi\left(k+\frac{1}{u}\right)\ . \quad (19)$$

Логарифмическая функция правдоподобия (17) с учетом формул (18) и (19) перепишется в виде:

$$\ln L_{(z)} = M\left[\ln zp(z)\right] = \ln\frac{\Gamma(k+1/u)}{\Gamma(k)\Gamma(1/u)} + \\ +k\left[\psi(k)-\psi\left(k+\frac{1}{u}\right)\right]+\left(\frac{1}{u}-1\right)\left[\psi\left(\frac{1}{u}\right)-\psi\left(k+\frac{1}{u}\right)\right]. \quad (20)$$

Из сопоставления функций правдоподобия (14) и (20) следует, что первая из них зависит от трех параметров β, k, u, а вторая – только от двух параметров k, u.

На основании формул (14) и (20) можем записать равенство

$$M\left[\ln p(x)\right] = \ln\beta + M\left[\ln zp(z)\right], \quad (21)$$

которое справедливо также для других типов распределений.

Равенство (21) позволяет вычислять оценку наибольшего правдоподобия параметра β при известных оценках двух параметров формы k, u:

$$\beta = e^{M\left[\ln p(x)\right]-M\left[\ln zp(z)\right]} = \frac{e^{M\left[\ln p(x)\right]}}{e^{M\left[\ln zp(z)\right]}} = \frac{\overline{p(x)}_{геом.}}{\overline{\mathrm{zp(z)}}_{геом.}} . \quad (22)$$

Перепишем далее третье уравнение правдоподобия системы (13) с учетом формулы (18):

$$\beta M(x) = M(\ln z) - \ln\alpha u .$$

Отсюда найдем оценку наибольшего правдоподобия произведения αu:

$$\alpha u = e^{M(\ln z)-\beta M(x)} . \quad (23)$$

Входящие в формулы (22), (23) величины *M(lnzp(z))* и *M(lnz)* зависят от двух параметров формы k, u. Они вычисляются по формулам (20) и (18). Другие величины – *M(lnp(x))* и *M(x)* – в общем случае зависят от четырех параметров, но эти величины вычисляются по статистическому распределению.

Переходим к первому этапу решения.

Итак, при известных оценках параметров формы оценки наибольшего правдоподобия параметра β и произведения параметров αu вычисляются по формулам (22) и (23). Остается невыясненным вопрос о вычислении типа аппроксимирующего распределения и оценок параметров k, u по двум показателям, зависящим от этих параметров. Нахождение таких показателей является наиважнейшей задачей любого метода оценивания. Успешное решение этой задачи позволяет отказаться от выдвижения гипотез о виде аппроксимирующего распределения и проверки каждой из них по критериям согласия. **Наличие таких показателей позволяет вычислять наилучшее аппроксимирующее распределение из трех систем непрерывных распределений автора без выдвижения гипотез,** легко решать и другие задачи.

В случае метода наибольшего правдоподобия такие показатели можно получить из равенства (21) в виде центральных моментов второго – четвертого порядков

$$\mu_r = M\left[\ln p(x) - M\left(\ln p(x)\right)\right]^r = \\ = M\left[\ln zp(z) - M\left(\ln zp(z)\right)\right]^r , \quad (24)$$

которые зависят от двух параметров формы k, u [1].

В качестве первого показателя целесообразно принять центральный момент второго порядка. В качестве второго показателя могут быть приняты либо центральный момент третьего порядка, либо разность $\Delta = \ln M[p(x)] - M[\ln p(x)]$. Но решение этой задачи еще требует серьезных исследований.

Если искомые показатели найдены, необходимо при заданных значениях параметров формы построить бинарную сетку (номограмму) зависимости одного показателя от другого. С помощью построенной номограммы можно решать задачу установления типа аппроксимирующего распределения и нахождения в первом приближении оценок двух параметров формы в ручном режиме. Оценки параметров находятся по номограмме при известных значениях двух показателей, вычисленных по статистическому распределению.

В заключение следует отметить, что хорошие показатели должны обеспечить построение номограммы с высокой разрешающей способностью.

УСТОЙЧИВЫЙ МЕТОД

Устойчивым называется метод оценивания параметров, который не чувствителен к выбросам на концах статистического распределения.

Рассмотрим плотности (8) и (16). Случайные величины X и Z связаны соотношением (15)

$$Z = \alpha u e^{\beta X} .$$

Базой устойчивого метода является равенство, устанавливающее взаимосвязь между плотностями *p(x)* и *p(z)*. Найдем его.

Поскольку $p(z) = p(x)(dx/dz)$, $dx/dz = 1/\beta z$, то $p(z) = p(x)/\beta z$, откуда и следует искомое равенство

$$p(x) = \beta z p(z) . \quad (25)$$

Запишем на основе (25) новые равенства [1]

$$M\left[p(x)\right]^r = \beta^r M\left[zp(z)\right]^r \quad (26)$$

или

$$S_r^{(x)} = \beta^r S_r^{(z)} . \quad (27)$$

При известных оценках параметров k, u оценка параметра β устойчивого метода задается равенством (при r=1 в формулах (26), (27))

$$\beta = \frac{M\left[p(x)\right]}{M\left[zp(z)\right]} = \frac{S_1^{(x)}}{S_1^{(z)}} . \quad (28)$$

Здесь математическое ожидание плотности p(x) заменяется средним значением, которое вычисляется по статистическому распределению.

Логарифмируя равенство (25) и переходя к математическим ожиданиям, получим формулу (21), которая является базой метода наибольшего правдоподобия.

Формула (22) дает оценку наибольшего правдоподобия параметра β как **отношение средних геометрических** значений величин *p(x)* и *zp(z),* а формула (28) – как **отношение средних арифметических** тех же величин в случае устойчивого метода. Отсюда следует, что оценки параметра β, вычисленные по обоим методам, должны быть одинаковыми. Оценка

параметра α (или произведения αu) вычисляется в обоих методах по одним и тем же формулам.

Итак, устойчивый метод близок к методу наибольшего правдоподобия, но в то же время он значительно проще последнего. Устойчивый метод обладает тем несомненным преимуществом перед методом наибольшего правдоподобия, что для него разработаны два показателя (асимметрии $B = M[p(x)(x - M(x)]$ и островершинности $H = S_3 / S_1^3$), с помощью которых по заранее построенной бинарной сетке (номограмме) или соответствующей компьютерной программе автора легко вычисляются аппроксимирующие распределения и оценки двух параметров формы k, u [1, 3 – 5]. Для метода наибольшего правдоподобия такие показатели еще предстоит разработать.

С другой стороны, метод наибольшего правдоподобия позволяет легко выражать через параметры распределения математическое ожидание логарифма плотности распределения. Например, для плотности p(x) (см. формулу (8)) величина M[lnp(x)] задается формулой (14). Эту величину можно использовать как естественный критерий близости статистического распределения и вычисленного закона распределения. По найденным оценкам параметров β, k, u следует вычислить теоретическое значение величины M[lnp(x)] и сравнить его с эмпирическим значением $\overline{\ln p(x)}$, рассчитанным непосредственно по статистическому распределению. Оба значения должны практически совпадать.

ФУНКЦИЯ ПРАВДОПОДОБИЯ КАК КРИТЕРИЙ СОГЛАСИЯ

Итак, важнейшим свойством функции правдоподобия является то, что **только при точно установленном аппроксимирующем распределении функция правдоподобия, рассчитанная по оценкам его параметров, будет равна статистической функции правдоподобия, т.е. вычисленной по статистическому распределению.**

При неправильно выбранном теоретическом законе распределения функция правдоподобия, вычисленная по оценкам его параметров, которые в свою очередь вычислены по статистическому распределению, будет **меньше статистической функции правдоподобия, т.е. меньше максимального ее значения.**

К сожалению, метод наибольшего правдоподобия не дает рекомендаций по ***вычислению*** наилучшего аппроксимирующего распределения. Поэтому при использовании этого метода приходится выдвигать различные гипотезы и проверять их с помощью критериев согласия. Поскольку все неподходящие аппроксимирующие распределения будут иметь меньшие значения функции правдоподобия, чем вычисленное непосредственно по статистическому распределению, **в качестве критерия согласия целесообразно использовать степень близости теоретической и статистической функций правдоподобия.**

Рассмотрим пример.

Пусть показательный закон распределения (5) имеет параметр α=1. В этом случае $M(t) = 1$. Тогда логарифмическая функция правдоподобия будет равна

$$\ln L = M\left[\ln p(t)\right] = \ln \alpha - \alpha M(t) = -1 .$$

Вычислим далее для показательного закона распределения некоторые числовые характеристики случайной величины T, необходимые для дальнейших расчетов: начальный момент второго порядка, дисперсию, среднее квадратическое отклонение, математическое ожидание логарифма случайной величины и математическое ожидание случайной величины в степени ½. При α=1 они равны:

$$M(t^2) = 2;\ \ D(t) = 1;\ \ S(t) = 1;$$

$$M(\ln t) = \psi(1) = -0{,}5772164;\ \ M(\sqrt{t}) = \sqrt{\pi}/2 .$$

Теперь используем **метод выдвижения гипотез** для нахождения наилучшего аппроксимирующего распределения по приведенным числовым характеристикам случайной величины.

Рассмотрим три гипотезы. Запишем закон распределения и его логарифмическую функцию правдоподобия, вычисленную по числовым характеристикам показательного закона.

1. Закон Вейбулла $p(t) = \alpha\beta t^{\beta-1} e^{-\alpha t^{\beta}}$ при β=1/2:

$$p(t) = \frac{\alpha}{2\sqrt{t}\, e^{\alpha\sqrt{t}}};$$

$$M\left[\ln p(t)\right] = -\ln\sqrt{\pi} - 0{,}5M(\ln t) - 1 = -1{,}283757.$$

2. Закон Вейбулла при β=2 (распределение Релея):

$$p(t) = \frac{2\alpha t}{e^{\alpha t^2}};$$

$$M\left[\ln p(t)\right] = \ln 2 - \ln M(t^2) + M(\ln t) - 1 = -1{,}577216.$$

3. Нормальный закон:

$$p(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-a)^2}{2\sigma^2}};$$

$$M\left[\ln p(t)\right] = -\ln\sigma - 0{,}5(\ln 2\pi + 1) = -1{,}418939.$$

Из полученных результатов следует, что наиболее близким к показательному распределению, для которого $M\left[\ln p(t)\right] = -1$, оказался закон Вейбулла с параметром β=1/2. Однако ни одна из выдвинутых гипотез не может быть принята, так как во всех трех случаях логарифмическая функция правдоподобия оказалась значительно меньше –1.

Приведенные расчеты свидетельствуют о неэффективности метода выдвижения гипотез. **Теоретический закон распределения необходимо вычислять по статистическому распределению.** Для решения этой задачи автором разработаны три системы непрерывных четырехпараметрических распределений, методы вычисления типа теоретического распределения и оценок его параметров (универсальный метод моментов и общий устойчивый метод), а также серия компьютерных программ под общим названием SNR (системы непрерывных распределений).

ЭНТРОПИЯ

В качестве меры неопределенности системы принята энтропия. В случае непрерывных распределений она представляет собой математическое ожидание логарифма плотности, взятое с обратным знаком. Другими словами, энтропия – это взятая с обратным знаком логарифмическая функция правдоподобия (7):

$$H_x = -M\left[\ln p(x)\right]. \qquad (29)$$

Рассмотрим единицы измерения энтропии. В приведенной формуле логарифм плотности взят по основанию e=2.71828…. В данном случае в качестве единицы измерения энтропии принят «нат». При основании 10 единица измерения энтропии называется «дит», а при основании 2 – «бит».

С энтропией связано понятие информации. Количество полученной информации о системе уменьшает энтропию системы на это количество информации. Если о системе известно все, то количество информации равно энтропии этой системы, т.е. $I_x = H_x$ [5].

Рассмотрим один способ извлечения информации из ранговых распределений. Пусть ранговое распределение задано плотностью

$$p(t) = \frac{\beta\alpha^k}{\Gamma(\mathrm{k})} t^{k\beta-1} e^{-\alpha t^\beta}, \qquad (30)$$

которая относится ко второму типу второй системы непрерывных распределений автора. Запишем логарифмическую функцию правдоподобия:

$$\ln L = M\left[\ln p(t)\right] = \ln\beta + k\ln\alpha - \ln\Gamma(\mathrm{k}) + (\mathrm{k}\beta\text{-}1)\mathrm{M(lnt)}\text{-}\alpha\mathrm{M}(\mathrm{t}^\beta). \qquad (31)$$

Для того чтобы выразить ее через параметры распределения, найдем уравнения правдоподобия:

$$\frac{\partial \ln L}{\partial \alpha} = \frac{k}{\alpha} - M\left(t^\beta\right) = 0;$$
$$\frac{\partial \ln L}{\partial \beta} = \frac{1}{\beta} + kM(\ln t) - \alpha M\left(t^\beta \ln t\right) = 0;$$
$$\frac{\partial \ln L}{\partial k} = \ln\alpha - \psi(k) + \beta M(\ln t) = 0.$$

Из первого и последнего уравнений правдоподобия имеем:

$$k = \alpha M\left(t^\beta\right), \qquad (32)$$

$$M(\ln t) = \frac{1}{\beta}\left[\psi(k) - \ln\alpha\right]. \qquad (33)$$

Подставляя значения величин k, M(lnt) в формулу (31), найдем:

$$\ln L = M\left[\ln p(t)\right] = \ln\beta - \ln\Gamma(\mathrm{k}) + \mathrm{k}\,\psi(\mathrm{k})\text{-k-M(lnt)}. \qquad (34)$$

Следовательно, энтропия плотности (30) равна

$$H_{p(t)} = -M\left[\ln p(t)\right] = M(\ln t) + \ln\Gamma(\mathrm{k}) + \mathrm{k}\text{-k}\,\psi(\mathrm{k})\text{-ln}\beta. \qquad (35)$$

Пусть параметры рангового распределения (30) равны: α=0,5; β=0,4; k=2. Вычислим его энтропию. Для этого найдем вначале M(lnt) по формуле (33):

$$M(\ln t) = \frac{1}{\beta}\left[\psi(k) - \ln\alpha\right] = \frac{1}{0,4}\left[\psi(2) - \ln 0,5\right] = 2,7898287.$$

Здесь значение пси-функции Ψ(2)=0,4227843 взято из таблицы в [5, с. 53]. Тогда

$$H_{p(t)} = -M\left[\ln p(t)\right] = 2,7898287 + \ln\Gamma(2) + 2\text{-}2\,\psi(2)\text{-ln}0,4 = 4,860551 \quad \text{(нат)}.$$

Преобразуем далее плотность (30) к форме соответствующей плотности p(x) первой системы непрерывных распределений, т.е. представим плотность p(t) в виде tp(t)=f(lnt) (где tp(t)=p(x), lnt=x):

$$tp(t) = \frac{\beta\alpha^k}{\Gamma(\mathrm{k})} e^{k\beta\ln t} e^{-\alpha e^{\beta\ln t}}. \qquad (36)$$

Запишем для нее логарифмическую функцию правдоподобия:

$$\ln L = M\left[\ln tp(t)\right] = \ln\beta + k\ln\alpha - \ln\Gamma(\mathrm{k}) + \mathrm{k}\,\beta\mathrm{M(lnt)}\text{-}\alpha\,\mathrm{M}\left(\mathrm{e}^{\beta\ln\mathrm{t}}\right). \qquad (37)$$

Выраженная через параметры распределения, она равна:

$$\ln L = M\left[\ln tp(t)\right] = \ln\beta - \ln\Gamma(\mathrm{k}) + \mathrm{k}\,\psi(\mathrm{k})\text{ - k}. \qquad (38)$$

Следовательно, энтропия плотности (36) равна:

$$H_{tp(t)} = -M\left[\ln tp(t)\right] = \ln\Gamma(\mathrm{k}) + \mathrm{k}\text{ - k}\,\psi(\mathrm{k})\text{ - ln}\beta. \qquad (39)$$

Сравнивая это равенство с (35), имеем:

$$H_{tp(t)} = H_{p(t)} - M(\ln t), \qquad (40)$$

т.е. энтропия (степень неопределенности) плотности tp(t)=p(lnt)=p(x) оказалась меньше энтропии плотности p(t) на величину M(lnt)=2,7898287 и составила $H_{tp(t)} = 2,0707223$ (нат), т.е. меньше энтропии $H_{p(t)}$ в 2,347 раза.

Таким образом, приведение второй системы непрерывных распределений к форме первой системы уменьшает энтропию второй системы.

Аналогично, приведение третьей системы непрерывных распределений к форме второй (или первой) системы также уменьшает ее энтропию.

Что касается ранговых (убывающих) распределений, то здесь наиболее ярко проявляется уменьшение энтропии, или появление новой информации в виде трех характерных точек: моды и двух точек перегиба, – которые позволяют объективно разделить ранговое распределение на ядро и три зоны рассеяния [3, 4]. Действительно, убывающее ранговое распределение, будучи представленным в виде графика зависимости tp(t)=f(lnt), превращается в одновершинную кривую распределения с тремя характерными точками, которые нельзя обнаружить непосредственно на убывающей кривой.

Два метода оценивания параметров (универсальный метод моментов и общий устойчивый метод) разработаны автором для первой системы непрерывных распределений, заданной плотностью p(x), а другие системы непрерывных распределений при нахождении оценок параметров по этим методам приводятся к первой системе. Это преобразование уменьшает энтропию распределений второй и третьей систем и позволяет находить оценки их параметров методом, пригодным для первой системы непрерывных распределений.

Действительно, метод моментов не может быть применен непосредственно к плотности (30), где значения случайной величины T возводятся в степень β. Но после преобразования той же плотности к виду (36) параметр β уже не является степенью случайной величины T, при этом вычисляются моменты не самой случайной величины T, а ее логарифма. Фактически в этом случае находятся оценки параметров плотности:

$$p(x) = \frac{\beta\alpha^k}{\Gamma(\mathrm{k})} e^{k\beta x} e^{-\alpha e^{\beta x}},$$

где x=lnt, p(x)=tp(t)=p(lnt) [3]. Найденные оценки являются также оценками параметров исходной плотности p(t), которая задана формулой (30).

ЗАКЛЮЧЕНИЕ

В статье рассмотрены четыре формы задания функции правдоподобия. Для проведения научных исследований введена пятая форма как математическое ожидание логарифма плотности распределения.

Дана сравнительная характеристика двух методов оценивания параметров четырехпараметрических непрерывных распределений: метода наибольшего правдоподобия Р. Фишера и устойчивого метода автора. Показано, что оба метода базируются на одном и том же равенстве

$$p(x) = \beta z p(z),$$

где плотность p(x) зависит от четырех параметров, а плотность p(z) – от двух.

В методе наибольшего правдоподобия используется логарифмическая форма приведенного равенства, где в качестве логарифмической функции правдоподобия используется математическое ожидание логарифма плотности:

$$M\left[\ln p(x)\right] = \ln\beta + M\left[\ln zp(z)\right].$$

Устойчивый метод позволяет вычислять закон распределения на базе четырехпараметрических систем непрерывных распределений с помощью показателей асимметрии $B = M[p(x)(x - M(x)]$ и островершинности $H = S_3 / S_1^3$. Оценки последних находятся по статистическому распределению и зависят от двух параметров формы. Тип распределения и оценки параметров формы находятся по заранее построенной бинарной сетке (номограмме) [3, 4] либо рассчитываются по соответствующей компьютерной программе автора. При этом подходящая система непрерывных распределений выбирается в зависимости от свойств случайной величины: это либо первая система, либо вторая и реже – третья система.

В методе наибольшего правдоподобия такие показатели отсутствуют. Поэтому **вид аппроксимирующего распределения подбирается традиционным методом – путем выдвижения гипотез**.

Логарифмическая функция правдоподобия, заданная формулой $\ln L = M\left[\ln p(x)\right]$ и взятая с обратным знаком, представляет собой энтропию. При преобразовании распределений второй системы к форме первой системы энтропия плотности уменьшается, т.е. появляется новая информация.

СПИСОК ЛИТЕРАТУРЫ

1. Нешитой В. В. Элементы теории обобщенных распределений. – Минск : РИВШ, 2009. – 204 с.
2. Нешитой В. В. Исследование статистических закономерностей текста и информационных потоков : дис. … д-ра техн. наук. – Минск, 1987. – 505 с.
3. Нешитой В. В. Законы Ципфа, Бредфорда и универсальные модели // НТИ. Сер. 2. – 2010. – №1. – С. 26 – 33; Neshitoi V. V. Zipf's and Bradford's laws and universal models // Automatic Documentation and Mathematical Linguistics. – 2010. – Vol. 44, №1. – P. 30 – 37.
4. Нешитой В. В. Методы статанализа в библиотечной деятельности: вычисление непрерывных распределений : учеб.-метод. пособие. – Минск : БГУ культуры и искусств, 2010. – 61 с.
5. Вентцель Е. С. Теория вероятностей. – М. : Физматгиз, 1969. – 576 с.

Материал поступил в редакцию 20.10.11.

Сведения об авторе

НЕШИТОЙ Василий Васильевич – доктор технических наук, профессор, заведующий кафедрой информационных ресурсов УО «Белорусский государственный университет культуры и искусств»
E-mail: neshitoi_vv@tut.by

БАЗА ДАННЫХ ВИНИТИ РАН

ВИНИТИ предлагает к использованию через WWW-сервер (**http://www.viniti.ru**) крупнейшую Федеральную базу отечественных и зарубежных публикаций по естественным, точным и техническим наукам. БД ВИНИТИ РАН генерируется с 1981 г., обновляется ежемесячно, пополнение составляет около 1 млн документов в год. БД ВИНИТИ представлена ретроспективными тематическими фрагментами и единой политематической БД (ретроспектива с 2001 г.), объединяющей все тематические фрагменты БД ВИНИТИ.

БД ВИНИТИ РАН в сети INTERNET

Сервер ВИНИТИ – http://www.viniti.ru – обеспечивает on-line доступ к Базе данных ВИНИТИ РАН круглосуточно без выходных.

На основе БД ВИНИТИ РАН предоставляются следующие услуги:

- Диалоговый поиск научно-технической информации **в режиме on-line**;
- **Демо-версия**, позволяющая ознакомиться с основными функциями поисковой системы, составом данных, формами представления документов и получить навыки работы с системой;
- **Поисковые эксперты** ВИНИТИ выполнят тематический поиск по разовым или постоянным запросам, а также окажут **консультационные услуги**.

БД ВИНИТИ РАН на CD-ROM

Любые наборы тематических фрагментов БД ВИНИТИ или их разделов могут быть предоставлены **на CD-ROM в поисковой системе (ИПС) "Сокол"**, обеспечивающей все поисковые функции, доступные в режиме on-line:

- Поиск можно вести в годовом или ретроспективном массиве (за несколько лет сразу) в одном или нескольких тематических фрагментах .
- Поиск по словам и любым словосочетаниям из заглавия, реферата, ключевых слов.
- Использование года, языка, рубрик, шифров тематических разделов БД для уточнения поиска.
- Поиск по словарю, выполняющему функции многоаспектного указателя, в том числе авторского, предметного, источников, индексов МПК, номеров патентных документов и депонированных рукописей и т.д.
- Возможность запоминания запросов для последующего использования и/или редактирования их.
- Чтение документов не только как в РЖ (последовательный просмотр документов одного номера за другим), но и чтение документов нужных тематических фрагментов (разделов) по оглавлению за весь период заказанной ретроспективы.

ИПС "Сокол" является прикладной программой Microsoft Windows.

Любые наборы тематических фрагментов БД ВИНИТИ или их разделов могут быть подготовлены **в коммуникативных форматах** ISO-2709, МЕКОФ, txt на любых видах электронных носителей.

Продукты предоставляются на договорной основе.

Информационная служба БД ВИНИТИ: 125190, Москва, ул. Усиевича 20, ВИНИТИ
Телефон: (499) 155-45-01, 155-45-02, **Факс:** (499) 152-62-31 **e-mail:** csbd@viniti.ru

Москва, ВИНИТИ РАН
28–30 ноября 2012 года

***8-я Международная конференция* «НТИ-2012»,**
посвященная 60-летию ВИНИТИ РАН

«АКТУАЛЬНЫЕ ПРОБЛЕМЫ ИНФОРМАЦИОННОГО ОБЕСПЕЧЕНИЯ НАУКИ, АНАЛИТИЧЕСКОЙ И ИННОВАЦИОННОЙ ДЕЯТЕЛЬНОСТИ»

Для участия **в «НТИ-2012»** приглашаются специалисты в области информационных технологий, телекоммуникаций, создатели и потребители информационных продуктов и услуг, ученые и специалисты РАН, вузовской и отраслевой науки, работники информационных центров и библиотек, служб распространения информационных продуктов и услуг.

Доклады или тезисы докладов направлять в Оргкомитет конференции «НТИ-2012», которые будут опубликованы в специальном сборнике.

Планируется проведение пленарных заседаний, круглых столов и работа по секциям.

Адрес: Россия, 125190, Москва, ул. Усиевича, 20, Всероссийский институт научной и технической информации (ВИНИТИ РАН), ОНИИР, Оргкомитет **«НТИ-2012».**
Тел: (495) 155-44-22, 155-44-29, 152-64-41,
Факс: (495) 152-54-92, 943-00-60
E-mail: conf@viniti.ru , market@viniti.ru
http://www.viniti.ru

Российская академия наук

Федеральное государственное бюджетное учреждение науки
ВСЕРОССИЙСКИЙ ИНСТИТУТ НАУЧНОЙ И ТЕХНИЧЕСКОЙ ИНФОРМАЦИИ РОССИЙСКОЙ АКАДЕМИИ НАУК

предлагает научным работникам, аспирантам и другим специалистам в области естественных, точных и технических наук, желающим быстро и эффективно опубликовать результаты своей научной и научно-производственной деятельности, использовать способ публикации своих работ через *систему депонирования.*

«Депонирование (передача на хранение) – особый метод публикации научных работ (отдельных статей, обзоров, монографий, сборников научных трудов, материалов научных мероприятий – конференций, симпозиумов, съездов, семинаров) узкоспециального профиля, разрешенных в установленном порядке к открытому опубликованию, которые нецелесообразно издавать полиграфическим способом печати, а также работ широкого профиля, срочная информация о которых необходима для утверждения их приоритета. Депонирование предусматривает прием, учет, регистрацию, хранение научных работ и обязательное размещение информации о них в специальных информационных изданиях».

Подготовка и передача на депонирование научных работ происходит в соответствии с «Инструкцией о порядке депонирования научных работ по естественным, техническим, социальным и гуманитарным наукам» (М., 2003).

Результатом депонирования является публикация информации о депонированных научных работах в информационных изданиях ВИНИТИ РАН – реферативном журнале и аннотированном библиографическом указателе «Депонированные научные работы».

В соответствии с “Положением о порядке присуждения ученых степеней”, утвержденном Постановлением Правительства Российской Федерации от 30.01.2002 № 74 (в ред. Постановлений Правительства РФ от 20.04.2006 № 227, от 02.06.2008 № 424, от 20.06.2011 № 475) научные работы, депонированные в организациях государственной системы научно-технической информации, признаны публикациями, учитывающимися при защите кандидатских и докторских диссертаций.

Подать научную работу на депонирование можно, обратившись в Отдел депонирования ВИНИТИ РАН по адресу:

125190, Москва, ул. Усиевича, 20.
ВИНИТИ РАН, Отдел депонирования научных работ.
Тел.: 8 (499) 155-43-28, Факс: 8 (499) 943-00-60.
e-mail: dep@viniti.ru

С инструкцией о порядке депонирования можно ознакомиться на сайте ВИНИТИ РАН: **http://www.viniti.ru**