

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Гонсалес, Р. Цифровая обработка изображений в среде MATLAB[Текст] / Р. Гонсалес, Р. Вудс, С. Эдинс. – М.: Техносфера, 2006. – 616 с.
2. Филист С.А., Пихлап С.В., Томакова Р.А. Нечеткие нейросетевые структуры для сегментации изображений глазного дна // Вестник Воронежского государственного технического университета. – Воронеж, 2009. – Т. 5, № 4. – С. 42-45.

Белобров Андрей Петрович

Борисовский Сергей Александрович

Томакова Римма Александровна

ГОУ ВПО «Курский государственный технический университет».

E-mail: SFilist@gmail.com

305040, г. Курск, ул. Челюскинцев 19, к. Б.

Тел.: 84712587098.

Belobrov Andrey Petrovich

Borisovsky Sergey Aleksandrovich

Tomakova Rimma Aleksandrovna

SEI HVT "Kursk state technical university".

E-mail: SFilist@gmail.com

19, Cheljusintsev street, the case B, Kursk, 305040, Russia.

Phone: +74712587098.

УДК 61:681

О.Г. Берестнева, К.А. Шаропин, А.В. Старикова, Л.И. Кабанова

**ТЕХНОЛОГИЯ ФОРМИРОВАНИЯ БАЗ ЗНАНИЙ В МЕДИЦИНСКИХ
ИНФОРМАЦИОННЫХ СИСТЕМАХ***

Рассматриваются вопросы, связанные с технологией формирования баз знаний в медицинских информационных системах, а именно вопрос заполнения базы знаний логическими правилами в виде продукции.

Медицинские информационные системы; методы Data Mining; производственные модели.

O.G. Berestneva, K.A. Sharopin, L.I. Kabanova, A.V. Starikova

**TECHNOLOGY OF KNOWLEDGE BASE DEVELOPMENT FOR MEDICAL
INFORMATION SYSTEMS**

The questions connected with technology of knowledge base development for medical information systems are considered in this article. Filling of knowledge base with logical rules in a form of production model is considered in details.

Medical information systems; methods of Data Mining; production models.

Основным побудительным мотивом работы по внедрению компьютерных технологий в систему здравоохранения является высокая общественная значимость улучшения ситуации в этой сфере, включая повышение качества и скорости лечения, снижение затрат на предоставление услуг и приобретение эффективных средств обеспечения соответствия нормативным документам и прочим требованиям. Врачи консультируют пациентов on-line, диагностическая аппаратура оснащена мощными процессорами, конференции и консилиумы проводятся через интер-

* Работа выполнена при финансовой поддержке РФФИ (проект № 08-06-00313а).

нет. Медицинские информационные технологии приобретают все большую актуальность, а программное обеспечение для медицины становится все более востребованными. Медицинская информационная система (МИС) – комплексная автоматизированная информационная система, в которой объединены электронные медицинские записи о пациентах, данные медицинских исследований в цифровой форме, данные мониторинга состояния пациента с медицинских приборов, средства общения между сотрудниками, финансовая и административная информация, напрямую связанная с медицинской деятельностью (например, стоимость медицинских услуг, расписание приёмов и т. п.) [1].

Отличительной особенностью интеллектуальных МИС является наличие базы знаний. База знаний, БЗ (англ. *Knowledge base, KB*) – это особого рода база данных, разработанная для управления знаниями (метаданными), т.е. сбором, хранением, поиском и выдачей знаний. Под базами знаний понимают совокупность фактов и правил вывода, допускающих логический вывод и осмысленную обработку информации. Существует три стратегии получения знаний:

- 1) приобретение знаний;
- 2) извлечение знаний;
- 3) обнаружение знаний.

Под приобретением (*acquisition*) знаний понимается способ автоматизированного наполнения базы знаний посредством диалога эксперта и специальной программы.

Извлечением (*elicitation*) знаний называют процедуру взаимодействия инженера по знаниям с источником знаний (экспертом, источником знаний и др.).

Термины «обнаружение знаний» (*knowledge discovery*), а также *Data Mining* связывают с созданием компьютерных систем, реализующих методы автоматического обнаружения знаний.

В нашем случае были использованы стратегии *elicitation* и *knowledge discovery*.

Известно много экспертных систем для постановки медицинских диагнозов. Они построены, главным образом, на основе правил, описывающих сочетания различных симптомов различных заболеваний. С помощью таких правил узнают не только, чем болен пациент, но и как нужно его лечить. Правила помогают выбирать средства медикаментозного воздействия, определять показания – противопоказания, ориентироваться в лечебных процедурах, создавать условия наиболее эффективного лечения, предсказывать исходы назначенного курса лечения и т.п. Технологии *Data Mining* позволяют обнаруживать в медицинских данных шаблоны, составляющие основу указанных правил.

Data Mining – это процесс обнаружения в "сырых" данных ранее неизвестных нетривиальных практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности. Информация, найденная в процессе применения методов *Data Mining*, должна быть нетривиальной и ранее неизвестной. Знания должны описывать новые связи между свойствами, предсказывать значения одних признаков на основе других и т.д. Найденные знания должны быть применимы и на новых данных с некоторой степенью достоверности. Полезность заключается в том, что эти знания могут приносить определенную выгоду при их применении. Знания должны быть в понятном для пользователя виде. Например, проще всего воспринимаются человеком логические конструкции "если ... то ...". Более того, такие правила могут быть использованы в различных СУБД в качестве SQL-запросов. В случае, когда извлеченные знания непрозрачны для пользователя, должны существовать методы постобработки, позволяющие привести их к интерпретируемому виду.

В нашем случае БЗ содержит логические правила в виде продукционных моделей. Продукции являются наиболее популярными средствами представления знаний. В них отсутствуют жесткие ограничения, характерные для логических исчислений, что дает возможность изменять интерпретацию элементов продукции.

В общем виде под продукцией понимается выражение следующего вида:

$$(i); Q; P; A \Rightarrow B; N.$$

Здесь i – имя продукции, с помощью которого данная продукция выделяется из всего множества продукций. Элемент Q характеризует сферу применения продукции. Основным элементом продукции является ее ядро: $A \Rightarrow B$. Интерпретация ядра продукции может быть различной и зависит от того, что стоит слева и справа от знака секвенции \Rightarrow . Обычное прочтение ядра продукции выглядит так: ЕСЛИ A , ТО B , более сложные конструкции ядра допускают в правой части альтернативный выбор, например, ЕСЛИ A , ТО $B1$, ИНАЧЕ $B2$. Секвенция может истолковываться в обычном логическом смысле как знак логического следования B из истинного A (если A не является истинным выражением, то о B ничего сказать нельзя). Возможны и другие интерпретации ядра продукции, например A описывает некоторое условие, необходимое для того, чтобы можно было совершить действие B [1].

Элемент P есть условие применимости ядра продукции. Обычно P представляет собой логическое выражение (как правило, предикат). Когда P принимает значение "истина", ядро продукции активизируется. Если P ложно, то ядро продукции не может быть использовано.

Элемент N описывает постусловия продукции. Они актуализируются только в том случае, если ядро продукции реализовалось.

Если в памяти системы хранится некоторый набор продукций, то они образуют систему продукций. В системе продукций должны быть заданы специальные процедуры управления продукциями, с помощью которых происходит актуализация продукций и выбор для выполнения той или иной продукции из числа актуализированных.

Продукции по сравнению с другими формами представления знаний имеют следующие преимущества [1]:

- ◆ модульность;
- ◆ единообразие структуры (основные компоненты продукционной модели могут применяться для построения интеллектуальных систем с различной проблемной ориентацией);
- ◆ естественность (вывод заключения в продукционной модели во многом аналогичен процессу рассуждений эксперта);
- ◆ гибкость родовидовой иерархии понятий, которая поддерживается только как связь между правилами (изменение правила ведет за собой изменение в иерархии).

Представление знаний с помощью продукций иногда называют «плоским», так как в продукционных моделях отсутствуют средства для установления иерархии правил.

Основные методы, используемые для построения логических правил:

- ◆ деревья решений;
- ◆ метод ограниченного перебора.

Деревья решений – это способ представления правил в иерархической, последовательной структуре, где каждому объекту соответствует единственный узел, дающий решение.

Под правилом понимается логическая конструкция, представленная в виде "если ... то ...". Данный метод реализован в пакете *See 5/C 5.0*. Задача *See 5/C 5.0*

состоит в предсказании диагностического класса какого-либо объекта по значениям его признаков. При этом *See 5/C 5.0* конструирует классификатор в виде дерева решений, которому, в свою очередь, может быть поставлено в соответствие некоторое множество логических правил. В ряде случаев полученное дерево решений может оказаться слишком сложным для восприятия. Например, при построении задач высокой размерности для неоднородных данных дерево нередко получается кустистое и довольно большое. Поэтому, с целью упрощения логического вывода рекомендуется использовать логическую связку "И". Если по смыслу существует логическая связка "ИЛИ", то формируется второе аналогичное правило, содержащее только связки "И".

Рассмотрим технологию построения деревьев решений в задаче прогнозирования исхода беременности.

Исходная выборка содержала 171 объект (беременных женщин), состояние которых описывалось с помощью 13 психофизиологических признаков.

В результате обработки данной выборки с помощью пакета *C5.0* пользователь получает отчет, который содержит следующую информацию.

Отображается построенное дерево решений. Каждая ветка дерева (рис. 1) заканчивается указанием номера класса, к которому она приводит. Сразу за номером следует запись (*n*) или (*n/m*). Например, если первая ветка заканчивается записью *s*, (*10/4*) это означает, что эта ветка описывает класс *s* и сюда попадает 10 объектов, из которых 4 попадает ошибочно.

```

k1 = 5:
...child = 1: k (29/8)
:   child = 0:
:     ...p3 in 2,3,4,5,6,9,10: s (0)
:       p3 = 1: k (3/1)
:       p3 = 7: s (3)
:       p3 = 8: p (1)
    
```

Рис. 1. Одна из ветвей построенного дерева

В следующем разделе отчета приводятся характеристики сконструированного классификатора: дерево решений имеет 37 веток (*size = 37*), а ошибка классификации наблюдается на 42 объектах, что составляет 24,6 %.

В завершающей части отчета дается таблица с детальным разбором результатов классификации. В нашем примере из первого класса *k* (кесарево сечение) правильно классифицируются 56 объектов, а 17 объектов ошибочно относят к классу 2; среди объектов второго класса (срочные роды) 63 диагностируются правильно, а 9 ошибочно «приписывают» исход родов типа кесарево сечение; 10 объектов третьего класса диагностируются верно, 14 относят не к тому классу.

В заключение система *C5.0* выдает сообщение о затраченном на решение времени. В нашем случае оно составило 0,4 с. Здесь надо отметить очень высокую скорость работы алгоритма *C5.0*, позволяющую оперативно обрабатывать высокомерные массивы информации, содержащие тысячи и десятки тысяч записей.

В *C5.0* предусмотрена возможность преобразования дерева решений в набор правил *IF...THEN*. Результаты в виде набора правил являются более простыми и понятными, чем в виде деревьев решений. Каждое правило описывает связь между набором значений признаков и идентификатором класса. Более того, количество правил, сгенерированных из дерева решений, нередко оказывается несколько меньшим, чем число веток на дереве, а результат может оказаться более точным.

Алгоритмы ограниченного перебора были предложены в середине 60-х годов М.М. Бонгардом для поиска логических закономерностей в данных. С тех пор они продемонстрировали свою эффективность при решении множества задач из самых различных областей.

Эти алгоритмы вычисляют частоты комбинаций простых логических событий в подгруппах данных. Примеры простых логических событий: $X = a$; $X < a$; $X > a$; $a < X < b$ и др., где X – какой либо параметр, " a " и " b " – константы. Ограничением служит длина комбинации простых логических событий (у М. Бонгарда она была равна 3). На основании анализа вычисленных частот делается заключение о полезности той или иной комбинации для установления ассоциации в данных, для классификации, прогнозирования и пр.

Данный метод реализован в системе поиска скрытых закономерностей *WizWhy* компании *WizSoft*. Эта система интересна тем, что ее разработчики утверждают, что она способна обнаружить все «если...то...» – правила в данных. Это подтверждение подкрепляется сообщением о весьма большом количестве структур, использующих *WizWhy*. Поиск логических правил осуществляется в системе *WizWhy*, реализующей ограниченный перебор, исключающий из анализа логические события с низкой частотой.

Приведенную выше технологию можно рассмотреть на примере создания информационной системы мониторинга психофизиологического состояния беременных женщин и прогнозирования исходов родов. Вопросы актуальности создания подобной системы и концептуальная схема системы подробно изложены в [2,3]. Ядром системы является подсистема интеллектуального анализа, которая заполняется вышеупомянутыми средствами. Для формирования окончательного варианта БЗ подсистемы были выбраны устойчивые закономерности на основе анализа полученных результатов:

- 1) логических правил, полученных в системе *WizWhy* [4];
- 2) деревьев решений (построенных на базе *See5*) [5];
- 3) закономерностей, сформулированных экспертом (Добрянская Р.Г. [2,3]).

В настоящее время система внедрена в опытную эксплуатацию в одной из женских консультаций г. Томска.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Дюк В., Эмануэль В., Информационные технологии в медико-биологических исследованиях. – СПб.: Питер, 2003. – 528 с.
2. Добрянская Р.Г., Евтушенко И.Д., Берестнева О.Г. Опыт использования информационной системы для мониторинга психофизиологического состояния беременных женщин и выявление групп социально-психологического риска // Мать и дитя: Материалы IX Всероссийского научного форума. – Москва, 2-5 октября 2007. – М., 2007. – С. 70-71.
3. Берестнева О.Г., Шаропин К.А., Добрянская Р.Г., Муратова Е.А. Разработка прототипа интеллектуальной системы прогнозирования исхода беременности // Математические методы распознавания образов (ММО-13): Труды Всероссийской научно-технической конференции - Москва, 10-15 октября 2007. – М.: Физматлит, 2007. – С. 574-577.
4. Старикова А.В. База данных информационной системы мониторинга психофизиологического состояния беременных женщин // Технологии Microsoft в теории и практике программирования: Материалы конференции студентов, аспирантов и молодых ученых – Новосибирск, 1-2 марта 2008. – Новосибирск: НГУ, 2008. – С. 95.
5. Берестнева О.Г., Добрянская Р.Г., Марухина О.В., Шаропин К.А., Муратова Е.А. Формирование базы знаний для экспертной системы прогнозирования исхода беременности // Труды Международных научно-технических конференций «Интеллектуальные системы» (AIS-07) и «Интеллектуальные САПР» (CAD-2007). – Дивноморское, 3-10 сентября 2007. – М.: Физматлит, 2007. – С. 424-429.

Берестнева Ольга Григорьевна

Национальный исследовательский Томский политехнический университет.

E-mail: ogb@rambler.ru.

634004, г. Томск, ул. Советская, 84, к. 109.

Тел.: 83822426100.

Шаропин Константин Александрович

E-mail: kashar@mail.ru.

Старикова Анастасия Викторовна

E-mail: astarikova@yandex.ru.

Кабанова Людмила Игоревна

E-mail: am@am.tpu.ru.

Berestneva Olga Grigorevna

National Research Tomsk Polytechnic University.

E-mail: ogb@rambler.ru.

84/109, Sovetskaya street, Tomsk, 634004, Russia.

Phone: +73822426100.

Sharopin Konstantin Aleksandrovich

E-mail: kashar@mail.ru.

Starikova Anastasia Viktorovna

E-mail: astarikova@yandex.ru.

Kabanova Ludmila Igorevna

E-mail: am@am.tpu.ru.

УДК 616.28-008.1-053.2-073.97

Г.Ш. Гафиятуллина, Е.В. Трофимова

**ПРИНЦИПЫ АНАЛИЗА ФУНКЦИОНАЛЬНЫХ ПАРАМЕТРОВ ПАМЯТИ,
ВНИМАНИЯ И МЫШЛЕНИЯ У СЛАБОСЛЫШАЩИХ ДЕТЕЙ**

Изучены психофизиологические особенности детей с нейросенсорной тугоухостью. Показано, что способы восприятия, формирующиеся через овладение словесными обобщениями, запаздывают в развитии.

Нейросенсорная тугоухость; вызванные потенциалы; вербальный и невербальный анализ и синтез.

G.Sh. Gafiyatullina, E.V. Trofimova

**ANALYSIS PRINCIPLES OF MEMORY, ATTENTION AND THINKING
PARAMETERS IN CHILDREN WITH NEUROSENSORY BRADYACUASIA**

The psychophysiological features in children with neurosensory bradyacuasia were investigated. It was detected that the ways of perception formed through mastering by verbal generalisations are late in development.

Neurosensory bradyacuasia; verbal nonverbal analysis and synthesis.

Влияние биологических и социальных факторов на познавательное развитие детей с нейросенсорной тугоухостью различного генеза приобретает особое значение в условиях их социально-психологической адаптации к школьному обучению. При поступлении в образовательное учреждение центральным блоком мыслительной деятельности ребенка становится умственная работа, однако у слабо-