

УДК 004

ПОСТРОЕНИЕ СИСТЕМ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ В МЕДИЦИНЕ НА ОСНОВЕ ДЕРЕВЬЕВ РЕШЕНИЙ

Жаркова О.С., Шаропин К.А., Сеидова А.С., Берестнева Е.В., Осадчая И.А.
ФГАОУ ВО «Национальный исследовательский Томский политехнический университет»,
Томск, e-mail: osz@tpu.ru

На сегодняшний день по-прежнему актуальной является задача разработки средств выражения медицинской информации и средств для ее обработки и анализа. Данная задача относится к классу трудноформулируемых задач, решение которых связано с применением как формальных, так и эвристических подходов к разработке моделей, адекватно описывающих процессы сбора, обработки, хранения и передачи информации, а также проектированием специальных программных систем. Наиболее эффективным программным средством, наиболее полно решающим данную задачу, являются системы поддержки принятия решений (СППР). В статье рассмотрен один из подходов создания баз знаний для медицинских систем поддержки принятия решений – метод деревьев решений. Представлены примеры использования данного метода для системы поддержки научных исследований бронхиальной астмы.

Ключевые слова: анализ медицинских данных, база знаний, деревья решений, система поддержки принятия решений, медицинская экспертная система

CONSTRUCTION OF DECISION SUPPORT SYSTEMS IN MEDICINE BASED DECISION TREE

Zharkova O.S., Sharopin K.A., Seidova A.S., Berestneva E.V., Osadchaya I.A.
National Research Tomsk Polytechnic University, Tomsk, e-mail: osz@tpu.ru

Today is a very urgent task of development tools of medical information and means of expression, for processing and analysis. This problem belongs to the class of Difficulty problems whose solution involves the use of both formal and heuristic approaches to the development of models that adequately describe the processes of collecting, processing, storing and transmitting information, as well as the design of special software systems. The most effective software tool more fully solve this problem is a decision support system (DSS). The report is considered one of the approaches to the creation of knowledge bases DSS – the method of decision trees. Examples of using this method for a system of scientific support asthma research.

Keywords: analysis of medical data, knowledge base, decision trees, decision support system, medical expert system

Медицинские информационные технологии приобретают все большую актуальность, а программное обеспечение для медицины становится все более востребованным. Под медицинской информационной системой (МИС) понимается комплексная автоматизированная информационная система, в которой объединены электронные медицинские записи о пациентах, данные медицинских исследований, данные мониторинга состояния пациента с медицинских приборов и т.п. [1].

Отличительной особенностью интеллектуальных МИС является наличие базы знаний. База знаний – это особого рода база данных, разработанная для управления знаниями, т.е. сбором, хранением, поиском и выдачей знаний.

Знание – это хорошо структурированные данные, или данные о данных, или метаданные. Для хранения знаний используются базы знаний, которые, в свою очередь, являются основами для любых информационных систем [2, 3].

Существуют три стратегии получения знаний [3]:

- приобретение знаний – это способ автоматизированного наполнения базы зна-

ний посредством диалога эксперта и специальной программы;

- извлечение знаний – процедура взаимодействия инженера по знаниям с источником знаний (экспертом);

- обнаружение знаний.

Извлечение знаний – это процедура взаимодействия эксперта с источником знаний, в результате которой становится явным процесс рассуждений специалистов при принятии решения и структура их представлений о предметной области.

Приобретение знаний – процесс наполнения базы знаний экспертом с использованием специализированных программных средств.

Формирование знаний – процесс анализа данных и выявления скрытых закономерностей с использованием специального математического аппарата и программных средств.

Методы представления знаний

Существуют десятки моделей (или языков) представления знаний для различных предметных областей. Большинство из них может быть сведено к следующим классам [2, 3]:

- семантические сети;
- фреймы;

- формальные логические модели;
- производственные модели.

Термин «семантическая» означает «смысловая», а сама семантика – это наука, устанавливающая отношения между символами объектами, которые они обозначают, т.е. наука, определяющая смысл знаков.

Семантическая сеть – это ориентированный граф, вершины которого – понятия, а дуги – отношения между ними.

Проблема поиска решения в базе знаний типа семантической сети сводится к задаче поиска фрагмента сети, соответствующего некоторой подсети, отражающей поставленный запрос к базе. Недостатком этой модели является сложность организации процедуры поиска вывода на семантической сети.

Фрейм – это абстрактный образ для представления некоего стереотипа восприятия. Основным преимуществом фреймов как модели представления знаний является то, что она отражает концептуальную основу организации памяти человека, а также ее гибкость и наглядность.

Традиционно в представлении знаний выделяют формальные логические модели, основанные на классическом исчислении предикатов первого порядка, когда предметная область или задача описывается в виде набора аксиом. Эта логическая модель применима в основном в исследовательских «игрушечных» системах, так как предъявляет очень высокие требования и ограничения предметной области.

Произведения являются наиболее популярными средствами представления знаний. Произведения, с одной стороны, близки к логическим моделям, что позволяет организовывать на них эффективные процедуры вывода, а с другой стороны, более наглядно отражают знания, чем классические логические модели. В них отсутствуют жесткие ограничения, характерные для логических исчислений, что дает возможность изменять интерпретацию элементов произведения.

В общем виде под *произведением* понимается выражение следующего вида:

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

Здесь i – *имя произведения*, с помощью которого данная продукция выделяется из всего множества произведений. В качестве имени может выступать некоторая лексема, отражающая суть данной продукции (например, «покупка книги» или «набор кода замка»), или порядковый номер произведения в их множестве, хранящемся в памяти системы.

Элемент Q характеризует *сферу применения произведения*. Такие сферы легко выделяются в когнитивных структурах человека. Наши знания как бы «разложены по полоч-

кам». На одной «полочке» хранятся знания о том, как надо готовить пищу, на другой – как добраться до работы и т.п. Разделение знаний на отдельные сферы позволяет экономить время на поиск нужных знаний. Такое же разделение на сферы в базе знаний информационной системы целесообразно и при использовании для представления знаний производственных моделей.

Основным элементом произведения является ее *ядро*: $A \Rightarrow B$. Интерпретация ядра произведения может быть различной и зависит от того, что стоит слева и справа от знака секвенции \Rightarrow . Обычное прочтение ядра произведения 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 произведения, например A описывает некоторое условие, необходимое для того, чтобы можно было совершить действие B .

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

Элемент N описывает *постусловия произведения*. Они актуализируются только в том случае, если ядро произведения реализовалось. Постусловия произведения описывают действия и процедуры, которые необходимо выполнить после реализации B . Выполнение N может происходить не сразу после реализации ядра произведения.

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

В Институте кибернетики Томского политехнического университета разработан прототип системы поддержки научных исследований бронхиальной астмы [6–8]. Бронхиальная астма является причиной значительных ограничений жизнедеятельности, снижения социальной активности больных, т.е. снижения их качества жизни. Ограничение социальной и физической активности отрицательно сказывается на развитии человека, вызывают существенные трудности у больного. На развитие болезни

вливают не только такие факторы, как наследственность, профессиональные факторы, экологические факторы, нервная и иммунная системы, но и возможно ряд других

факторов. Для выявления скрытых закономерностей у больных бронхиальной астмой нами были использованы преимущественно продукционные модели [4–9].

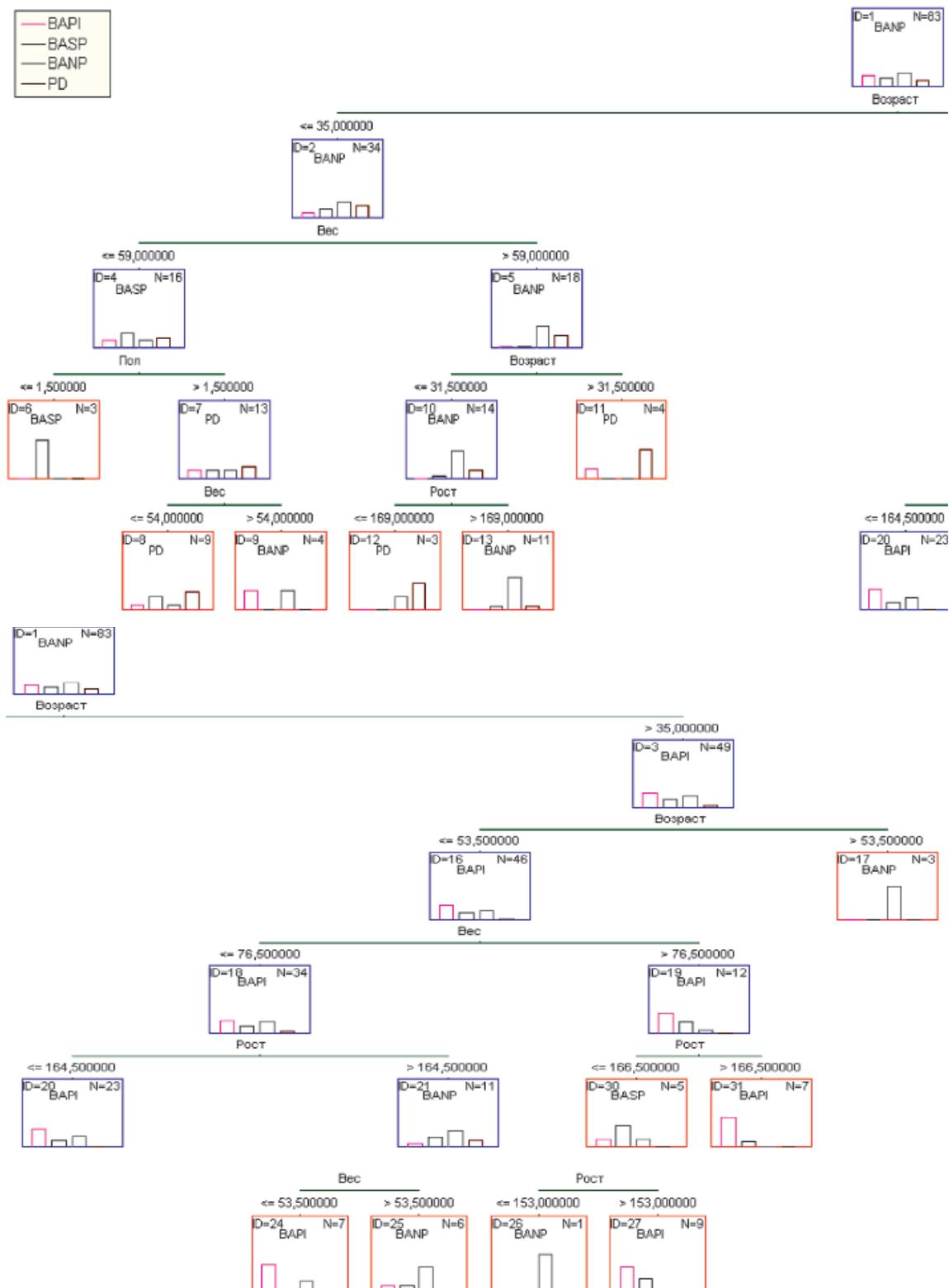


Рис. 1. Дерево решений, построенное по данным анамнеза

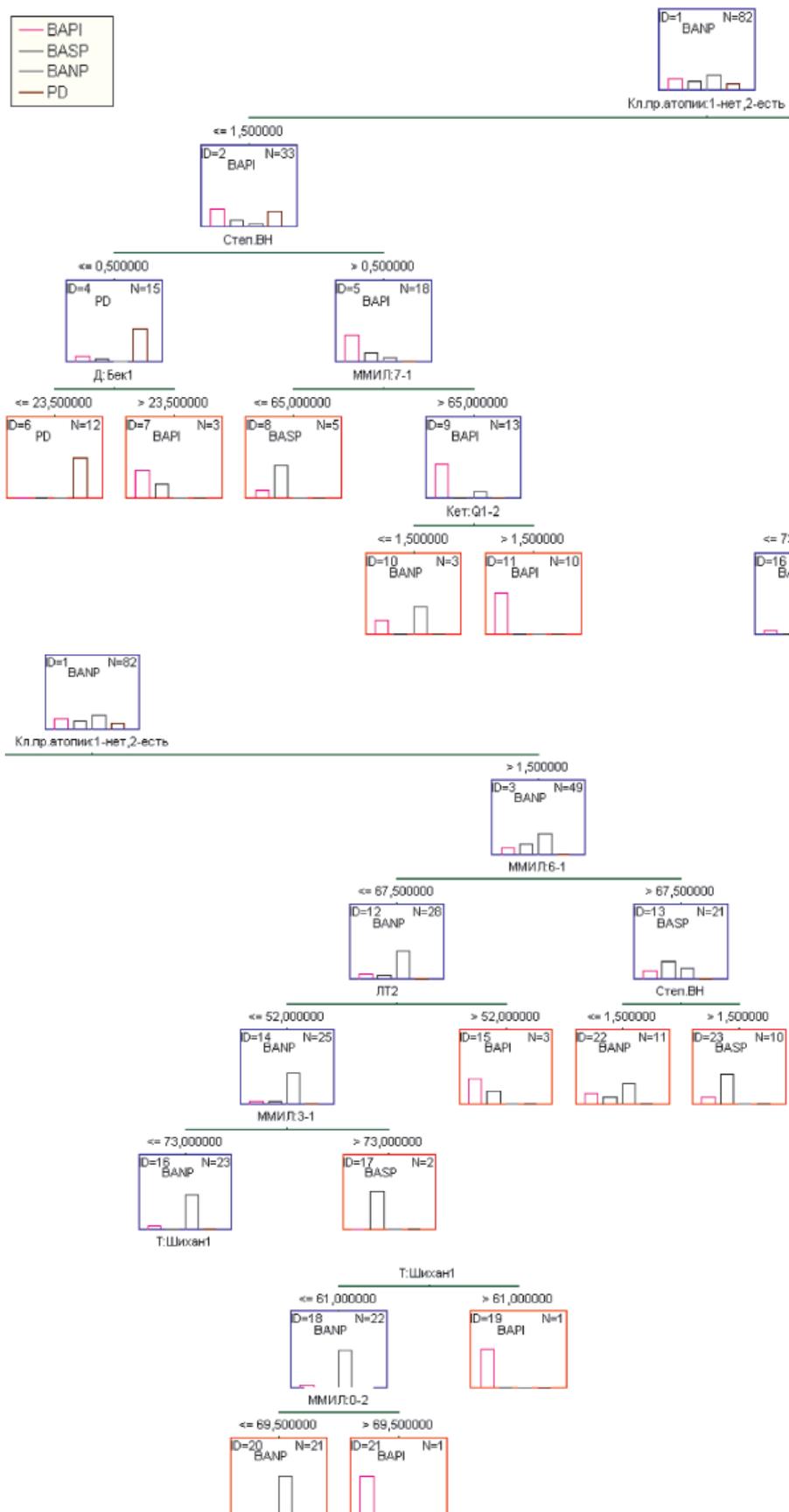


Рис. 2. Дерево решений, построенное по психологическим признакам

Для получения закономерностей в виде продукционных моделей, т.е. поиск в данных «если... то...» правил, чаще всего используется алгоритм ограниченного перебора М.М. Бонгарда и метод деревьев решений [4, 5].

Построение деревьев решений

Дерево решений – это способ представления правил в иерархической, последовательной структуре. Деревья решений способны решать такие задачи, в которых отсутствует априорная информация о виде зависимости между исследуемыми данными.

Иерархическое строение дерева классификации – одно из наиболее важных его свойств.

В настоящее время на рынке программных продуктов имеется достаточно большой выбор инструментария для компьютерной реализации метода деревьев решений. В нашем исследовании был использован пакет STATISTICA – система для статистического анализа данных, включающая широкий набор аналитических процедур и методов.

Пример построения дерева решений представлен на рис. 1. Для примера определим зависимость между видами бронхиальной астмы и такими полями, как «пол», «возраст», «вес», «рост».

При построении дерева, представленного на рис. 1, использовался критерий останова $N=9$, т.е. если количество объектов в данном правиле меньше либо равно 9, то выборка прекращается. Если же $N > 9$, то дерево продолжает ветвиться.

Дерево на рис. 2 выявляет закономерности вида бронхиальной астмы и психологического состояния пациента.

Заключение

Полученные результаты будут использованы для расширения имеющейся базы знаний системы поддержки научных исследований бронхиальной астмы, а также при создании прототипа виртуального центра оценки и мониторинга состояния детей

с наиболее распространенными неинфекционными заболеваниями.

Исследование выполнено при частичной финансовой поддержке РФФИ в рамках выполнения научных проектов № 15-07-08922, № 14-07-00675, № 14-06-00026.

Список литературы

1. Берестнева О.Г., Немеров Е.В., Языков К.Г., Фокин В.А., Карпенко П.В., Бурцева А.Л. Проблемы формирования базы знаний психогенных форм бронхиальной астмы // Конгресс по интеллектуальным системам и информационным технологиям (IS-IT'14): труды конгресса, Дивногорское, 2–9 Сентября 2014. – М.: Физматлит, 2014. – Т. 2. – С. 250–252.
2. Берестнева О.Г., Осадчая И.А., Немеров Е.В. Методы исследования структуры медицинских данных // Вестник науки Сибири. – 2012 – № 1 (2). – С. 333–338.
3. Берестнева О.Г., Марухина О.В. Базы данных и экспертные системы: учебное пособие. – Томск: ТПУ, 2010. – 108 с.
4. Берестнева О.Г., Муратова Е.А. Выявление скрытых закономерностей в социально-психологических исследованиях // Известия Томского политехнического университета. – 2003. – т. 306, № 5. – С. 86–91.
5. Берестнева О.Г., Муратова Е.А. Проблемы унификации данных в научных психологических и медицинских исследованиях // Информатика и системы управления. – 2010. – № 2. – С. 37–40.
6. Берестнева О.Г., Уразаев А.М., Марухина О.В. Технологии выявления скрытых закономерностей на основе интеллектуального анализа данных // Информационные и математические технологии в науке и управлении: труды XIV Байкальской Всероссийской конференции – Иркутск, 5–15 июля 2009. – Иркутск: ИСЭМ СО РАН, 2009. – С. 26–35.
7. Загоруйко Н.Г. Вычислительные системы. Экспертные системы и анализ данных: Сборник научных трудов. – Новосибирск, 1991. – 177 с.
8. Марухина О.В., Мокина Е.Е., Берестнева О.Г. Применение методов DATA MINING для выявления скрытых закономерностей в задачах анализа медицинских данных // Фундаментальные исследования. – 2015. – № 4. – С. 28–31.
9. Осадчая И.А., Берестнева Е.В. Применение многомерных методов анализа данных в задачах оценки качества жизни // Молодежь и современные информационные технологии: сборник трудов XII Всероссийской научно-практической конференции студентов, аспирантов и молодых ученых: в 2 т., Томск, 12–14 ноября 2014. – Томск: ТПУ, 2014 – Т. 2 – С. 310–311.
10. Старикова А.В., Берестнева О.Г., Шевелев Г.Е., Шаропин К.А., Кабанова Л.И. Создание подсистемы принятия решений в медицинских информационных системах // Известия Томского политехнического университета. – 2010. – Т. 317, № 5. – С. 194–197.