

ПРИМЕНЕНИЕ МЕТОДОВ DATA MINING ДЛЯ СОЗДАНИЯ МЕДИЦИНСКИХ БАЗ ЗНАНИЙ

А.С. Сеидова, Е.В. Берестнева, И.А. Осадчая

Научный руководитель: доцент, к.т.н. О.В. Марухина

Национальный исследовательский Томский политехнический университет,

Россия, г.Томск, пр. Ленина, 30, 634050

E-mail: aysel4421@mail.ru

APPLICATION OF THE DATA MINING METHODS FOR CREATION OF MEDICAL KNOWLEDGE BASES

A.S. Seidova, E.V. Berestneva, I.A. Osadchaya

Scientific Supervisor: Associate Professor., Ph.D. O.V. Maruhina

Tomsk Polytechnic University, Russia, Tomsk, Lenin str., 30, 634050

E-mail: aysel4421@mail.ru

***Abstract.** Now the problem of development of means of expression and the analysis of medical information is actual. The most effective software solving this problem are the system of support of decision-making (SSDM). In the report one of approaches of creation of knowledge bases for SPPR – a method of trees of decisions is considered.*

Введение

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].

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;
- продукционные модели.

Продукции являются наиболее популярными средствами представления знаний. Продукции, 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 продукции. Если в памяти системы хранится некоторый набор продукций, то они образуют систему продукций.

В Институте кибернетики Томского политехнического университета разработан прототип системы поддержки научных исследований бронхиальной астмы [4-6]. Бронхиальная астма является причиной значительных ограничений жизнедеятельности, снижения социальной активности больных, т.е. снижения их качества жизни. На развитие болезни влияют 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 [4-8].

Математическая постановка задачи

Для нахождения решающего правила посредством определенных математических инструментов, необходимо имеющиеся знания представить и структурировать согласно математическим терминам.

Имеется выборка X из n объектов (больные), характеризующихся m переменными (физиологические, социальные, психологические показатели).

И вектор $Y = \{y_{i1}, \dots, y_{i1} \dots y_{ik}\}$, где y_i – один из возможных диагнозов (BAP1, BASP, BANP, PD), k - количество прогнозируемых классов (диагнозов).

Таким образом, задача заключается в построении решающих правил для выявления закономерностей, т.е. для отнесения i -го объекта ($i = 1 \dots n$) (больного) с определенным признаком j ($j = 1 \dots m$) к одному из имеющихся классов y_i ($i = 1 \dots k$) (диагнозов).

Алгоритмы ограниченного перебора

Алгоритмы ограниченного перебора были предложены в середине 60-х годов М.М. Бонгардом для поиска логических закономерностей в данных.

Эти алгоритмы вычисляют частоты комбинаций простых логических событий в подгруппах данных.

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.

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 фрагмента 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.

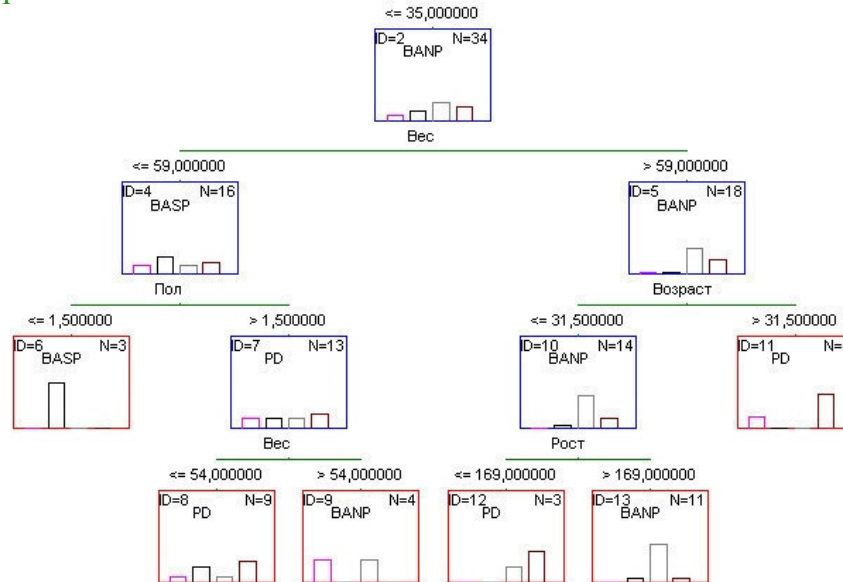


Рис. 1. Фрагмент дерева решений, построенного по данным анамнеза

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.

СПИСОК ЛИТЕРАТУРЫ

1. 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.
2. Берестнева О.Г., Марухина О.В. Базы данных и экспертные системы: Учебное пособие. - Томск: ТПУ, 2010. - 108 с.
3. Загоруйко Н.Г. Вычислительные системы. Экспертные системы и анализ данных: Сборник научных трудов. – Новосибирск, 1991. – 177 с.
4. 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, Дивноморское. - Москва, 2014 - Т. 2 - С. 250-252.
5. Осадчая И.А., Берестнева Е.В. Применение многомерных методов анализа данных в задачах оценки качества жизни // Молодежь и современные информационные технологии: сборник трудов XII Всероссийск. научно-практ. конф. студентов, аспирантов и молодых ученых. - Томск, 2014 - Т. 2 - С. 310-311.
6. Берестнева О.Г., Муратова Е.А. Проблемы унификации данных в научных психологических и медицинских исследованиях // Информатика и системы управления. - 2010. - № 2. - С. 37-40.
7. Марухина О.В., Непотребная А.А. Применение методов Data Mining для анализа медицинских данных // Перспективы развития информационных технологий: труды Всероссийск. молодеж. научно-практ. конф. - Кемерово, 2014 - С. 247-248.
8. 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 Всероссийск. конф. - Иркутск, 2009. - С. 26-35.