

ПОСТРОЕНИЕ СППР НА ОСНОВЕ ДЕРЕВЬЕВ РЕШЕНИЙ В МЕДИЦИНСКИХ ИССЛЕДОВАНИЯХ

Сеидова А. С., Берестнева Е. В.

Научный руководитель: Марухина О.В.

Томский политехнический университет

aysel4421@mail.ru

грант РФФИ № 15-0708922, № 14-07-00675

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.

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.

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.

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. Ограничение социальной и физической активности отрицательно сказывается на развитии человека, вызывают существенные трудности у больного. 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 [4-7].

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

$$X = \begin{pmatrix} x_{11} & \dots & x_{1n} \\ \vdots & & \vdots \\ x_{m1} & \dots & x_{mn} \end{pmatrix}$$

(физиологические, социальные, психологические показатели):

где i - номер объекта (больные), j - номер переменной (признак).

И вектор $Y = \{y_1, \dots, y_i, \dots, y_k\}$, где y_i - один из возможных диагнозов (BARI, 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 ($i = 1 \dots k$) (диагноз).

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 [8].

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.

Заключение

Авторы считают, что в данной работе новыми являются следующие положения и результаты:

1. Для формирования медицинских баз знаний наиболее эффективными являются методы интеллектуального анализа данных

2. Продукционные модели (в виде деревьев решений), описывающих выявленные закономерности физиологических и психологических особенностей пациентов с различными формами бронхиальной астмы.

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 использованных источников

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, Дивногорское, 2-9 Сентября 2014. - Москва: Физматлит, 2014 - Т. 2 - С. 250-252.

5. Осадчая И. А. , Берестнева Е. В. Применение многомерных методов анализа данных в задачах оценки качества жизни [Электронный ресурс] // Молодёжь 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.

6. Берестнева О. Г. , Осадчая И. А. , Немеров Е. В. Методы исследования структуры медицинских данных // Вестник науки Сибири. - 2012 - №. 1 (2). - С. 333-338.

7. Берестнева О.Г., Муратова Е.А. Проблемы унификации данных в научных психологических и медицинских исследованиях // Информатика и системы управления, 2010. - № 2 - с. 37-40.

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 Всероссийской конференции - Иркутск, 5-15 июля 2009. - Иркутск: ИСЭМ СО РАН, 2009. - с. 26-35.