

УДК 004

ПРИМЕНЕНИЕ МЕТОДОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ В МЕДИЦИНСКИХ ИССЛЕДОВАНИЯХ

Тарасова Л.П., Осадчая И.А.
Научный руководитель: Берестнева О.Г.

Национальный Исследовательский Томский политехнический университет,
634050, Россия, г. Томск, пр. Ленина, 30
E-mail: tarasova.luda31@mail.ru

During research of various methods of mining medical data of patients with asthma have been obtained if – then rules. These rules reveal the regularities characteristic of each type of asthma. Research was conducted on selection of 32 patients on indicators of a physiological and psychological sort.

Key words: bronchial asthma, data mining, if – then rules, rule's probability, significance level.

Ключевые слова: бронхиальная астма, обнаружение данных, «если..., то...» правила, точность правила, статистическая оценка уровня значимости правила.

Введение

Бронхиальная астма является причиной значительных ограничений жизнедеятельности, снижения социальной активности больных, т. е. снижения их качества жизни. Ограничение социальной и физической активности отрицательно сказываются на развитии человека, вызывают существенные трудности у больного [1]. К настоящему времени заболеваемость бронхиальной астмой в мире составляет от 4 до 10 % населения. В России, по разным данным, распространённость среди взрослого населения колеблется от 2,2 до 5–7 %, а в детской популяции этот показатель составляет около 10 %. На развитие болезни влияют не только такие факторы как наследственность, экологические факторы, нервная и иммунная системы, но и возможно ряд других факторов[2].

Постановка задачи

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

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

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 (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).

Для решения поставленной задачи необходимо проанализировать существующие подходы к анализу данных такого типа (т. е. слабо структурированных), выбрать математический метод и способ его компьютерной реализации для построения решающего правила, выявить некоторые закономерности у больных бронхиальной астмой.

Результаты

Поиск логических правил осуществлялся в системе WizWhy, реализующей ограниченный перебор, исключаящий из анализа логические события с низкой частотой. После задания всех параметров и выполнения определенных манипуляций получены результаты:

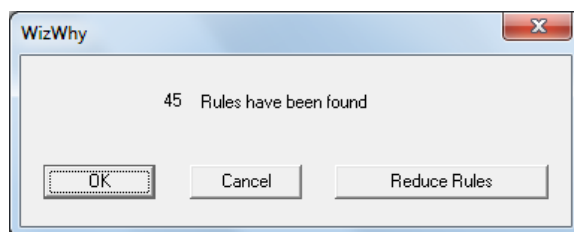


Рис. 1. Окно, отображающее количество найденных правил

Рассмотрим одно из правил полученное при анализе:

If **ADsyst1** is **120.00 ... 135.00** (average = **129.17**)
and **TL1B1H5** is **5.00 ... 6.00** (average = **5.67**)

Then

Diagnosis is BAPI

Rule's probability: **1.000**

The rule exists in **6 records**.

Significance Level: Error probability < 0.01

Positive Examples (records' serial numbers):

1, 2, 3, 4, 6, 7

Высказывание – *Diagnosis is BAPI* – означает, что правило характерно для людей больных бронхиальной астмой. BAPI – вид бронхиальной астмы. *Rule's probability: 1.000* означает, что точность правила в данном случае равна 1.000. Следующая запись *The rule exists in 6 records* характеризует объем множества объектов, для которых справедливо рассматриваемое правило, а другая запись *Significance Level: Error probability < 0.01* касается статистической оценки уровня значимости полученного правила (вероятность ошибки).

Заключение

Эксперимент проводился на выборке из 32 больных по показателям физиологического и психологического рода. Объектом исследования являются данные о больных бронхиальной астмой. Анализ данных показал, что данные слабоструктурированные (имеются как количественные так и качественные показатели), это обусловило выбор продукционной модели в качестве метода представления знаний и методов Data Mining, в частности системы WizWhy, в качестве инструментария обработки.

Список литературы

1. Самарский А.А. Компьютеры, модели, вычислительный эксперимент. – М.: Наука, 1988.
2. Берестнева О.Г., Шаропин К.А., Добрянская Р.Г., Муратова Е.А. Разработка прототипа интеллектуальной системы прогнозирования исхода беременности // Математические методы распознавания образов (ММРО-13): Труды Всероссийской научно-технической конференции – Москва, 10–15 октября 2007. – М.: Физматлит, 2007. – С. 574–577.