004.9+519.254

Н.В. Максюта, А.И. Поворознюк, С.В. Стешкин

Национальный технический университет "ХПИ", Харьков

ОТБОР ДИАГНОСТИЧЕСКИ ЦЕННЫХ ПОКАЗАТЕЛЕЙ С ИСПОЛЬЗОВАНИЕМ МЕТОДА КОРРЕЛЯЦИОННЫХ ПЛЕЯД

В работе обоснована необходимость формирования информативно полного пространства показателей при проектировании интеллектуальных компьютерных систем медицинской диагностики. Проведен краткий обзор методов отбора диагностически ценных показателей и методов кластеризации. Разработан комплексный подход к процедуре отбора диагностически ценных показателей с использованием метода корреляционных плеяд. Разработана структура подсистемы кластеризации, алгоритм ее функционирования и ее программная реализация.

Ключевые слова: *интеллектуальная компьютерная система медицинской диагностики, отбор, диагностически ценный показатель, кластеризация, метод корреляционных плеяд.*

Введение

Постановка проблемы. Процесс постановки диагноза представляет собой сложный многоуровневый процесс, требующий от врача большого опыта, профессиональных знаний и умений. Врач анализирует огромный спектр показателей, отличающихся друг от друга количественно, качественно и по типу и находящихся в тесной взаимосвязи друг с другом. При этом возможности человеческого мозга по переработке информации ограничены законом "семь плюс минус два", сформулированным специалистами в области инженерной психологии [1, 2]. Это ухудшает качество отбора нужных выводов из увеличивающегося количества фактов – болезней, синдромов, симптомов, тестов и т.д. [3, 4]. Современные интеллектуальные компьютерные системы медицинской диагностики (ИКСМД) позволяют автоматизировать процесс постановки диагноза и выступают в качестве мощного помощника врачу-специалисту, особенно при проведении экспресс-диагностики. Однако включение малоинформативных показателей ухудшает качество решающего правила. Поэтому формирование информативно полного пространства диагностически ценных показателей является актуальной проблемой при проектировании ИКСМД. Необходимость выполнения данного этапа также обусловлена применением в диагностике высокой размерности описания объектов и ограниченностью исходной выборки [4 – 15].

Цель статьи. Разработка комплексного подхода к процедуре отбора диагностически ценных показателей с использованием метода кластеризации.

Анализ литературы. Отбор группы диагностически ценных показателей X'n осуществляется с точки зрения некоторого критерия, в качестве которого чаще всего используется качество распознавания J(X'n). Т.е. выполняется поиск такой группы X'n, для которой J(X'n) имеет наибольшее значение (1):

$$J(X_n^{'}) = \max_{l} J(X_n^{'l}), \qquad (1)$$

где l – номер группы диагностически ценных показателей [8].

При этом выбирается минимальный набор диагностически ценных показателей, т.к. увеличение числа показателей ведет к усложнению анализа и снижению точности оценки параметров связи. Однако полученная группа l должна обеспечить необходимое качество распознавания, поэтому отбор диагностически ценных показателей считается не менее важной задачей по сравнению с синтезом диагностического решающего правила [1, 8 – 12].

На сегодняшний день общеприменимыми являются следующие методы поиска X'n: полный перебор, „k" лучшие признаки, последовательное уменьшение группы признаков, последовательное увеличение группы признаков метод „плюс l минус r", случайный поиск с адаптацией, метод ветвей и границ, а также теоретико-информационный подход, основанный на вычислении условных вероятностей и количества информации [3, 14, 15]. При этом при применении последнего могут использоваться простые бинарные признаки или сложные признаки, представленные в виде простых путем разложения на некоторые диагностические интервалы. [4 – 8, 12 – 15].

Данные методы имеют значительный недостаток, они не учитывают внутреннюю структуру исходных показателей, в связи с чем существует вероятность, что полученная группа X'n будет содержать несколько зависимых показателей вместо одного наиболее ценного. Авторами предлагается решение данной проблемы – выполнять поиск X'n не в исходном пространстве показателей X, а в группах

взаимосвязанных показателей Gi, которые формируются путем кластеризации X. Т.е. на первом этапе выполняется кластеризация на основе некой меры связи (например, коэффициент корреляции), результатом которой будут группы Gi, а на втором этапе – отбор диагностически ценных показателей в каждой такой группе. Это позволит сформировать пространство независимых диагностически ценных показателей и повысить надежность распознавания диагноза.

Общепринимыми методами кластеризации исходного пространства показателей являются: кластерный анализ, расщепление смесей распределений, факторный анализ, главные компоненты, многомерное шкалирование, корреляционные плеяды, экстремальная группировка параметров [4 – 12]. Работа данных методов основана на анализе некоторых мер связи между показателями или объектами, в качестве которых может выступать коэффициент корреляции или расстояние в пространстве признаков. Авторами выбран метод корреляционных плеяд, т.к. среди приведенных методов именно он лучше всех учитывает внутреннюю структуру показателей и позволяет наглядно ее представить. Более подробный сравнительный анализ приведен в [6]. В качестве метода поиска X‘n используется теоретико-информационный подход, выбранный благодаря его теоретической обоснованности и простоте реализации.

Отбор диагностически ценных показателей с использованием метода корреляционных плеяд

Суть разработанного подхода заключается в том, что структура объекта представляется в виде графа, вершинами которого являются исходные показатели, а дуги характеризуют взаимосвязь между ними. При этом в качестве веса дуги выступает значение статистической меры связи между исходными показателями (например, значение коэффициента корреляции). Для формализации задачи кластеризации показателей исходный граф разбивается на два и более подграфа, в соответствии с критерием (2), при этом вершины из разных подграфов имеют минимальную корреляционную связь, а вершины внутри каждого из подграфов – максимальную.

$$E = \frac{R_{G_i G_j}}{R_{G_i} + R_{G_j}} \to \min, \qquad (2)$$

где $R_{G_i G_j}$ – корреляция показателей из подграфа G_i с показателями из подграфа G_j (межгрупповая корреляция, -> min);

R_{G_i} – корреляция показателей из подграфа G_i друг с другом (внутригрупповая корреляция, -> max).

Такое представление объекта и группировка вершин используется в методе корреляционных плеяд. А подграфы (плеяды) формируются путем применения итерационной процедуры исключения дуг исходного графа, значение веса которых меньше граничного значения R_0 (задается эвристически). В результате получается структура подграфов, в каждом из которых собраны коррелированные показатели. Далее в каждом из подграфов G_i выбирается один диагностически ценный показатель x_i, информативность которого относительно системы диагнозов $\{D\}_n$ максимальна. Для показателя x_i $I_D(x_i)$ определяется по выражению:

$$I_D(x_i) = \sum_{j=1}^{n}\sum_{k=1}^{m} P(D_j)\cdot P(x_{ik}/D_j)\cdot \log_2 \frac{P(x_{ik}/D_j)}{P(x_{ik})} \qquad (3)$$

где n – количество диагнозов;

m – количество диагностических интервалов (градаций) показателя x_i;

$P(D_j)$ – частота встречаемости диагноза D_j в $\{D\}_n$ (априорная вероятность);

$P(x_{ik}/D_j)$ – вероятность наличия k-го диагностического интервала показателя x_i при диагнозе D_j;

$P(x_{ik})$ – вероятность наличия k-го диагностического интервала показателя x_i в $\{D\}_n$.

Полученный набор диагностически ценных показателей используется для построения диагностических решающих правил.

Для проверки адекватности работы комплексного использования метода кластеризации и метода отбора диагностически ценных показателей авторами разработана структура подсистемы кластеризации (рис. 1), алгоритм ее функционирования и ее программная реализация.

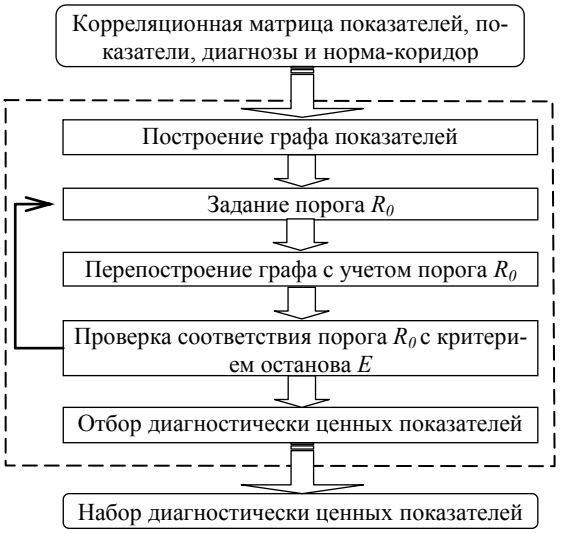

Рис. 1. Структурная схема подсистемы кластеризации диагностических показателей

Исходными данными для подсистемы кластеризации является корреляционная матрица медицинских показателей, представленная в файле Excel

с обязательным обозначением названий показателей. Результатами работы подсистемы являются сформированные плеяды (отображаются в виде подграфов на экране) и набор диагностически ценных показателей.

Выводы

В работе разработан подход к отбору диагностически ценных показателей с использованием метода корреляционных плеяд, а также его программная реализация на языке высокого уровня C++. В настоящее время проводится тестовая проверка работы алгоритма на реальных данных и набор статистики.

Перспективы дальнейших исследований

Перспективы дальнейших исследований состоят в том, чтобы проверить адекватность применения разработанного подхода для синтеза диагностического решающего правила путем анализа надежности распознавания объектов по исходному множеству показателей и по отобранному информативному подмножеству.

Список литературы

1. Miller G. The magical number seven, plus or minus two; some limits on our capacity, for process-ing information / G. Miller, A. Marsh // Psychological Review. – 1956. – Vol.63. – P. 81-87.

2. Егоров К.Н Психологические факторы в деятельности врача общей практики/ К.Н. Егоров, В.П. Дуброва // Клиническая медицина. – 2003. – №2. – С. 62-67.

3. Триша Г. Основы доказательной медицины: Пер. с англ. / Г. Триша. – М.: ГЭОТАР-МЕД, 2004. – 240 с.

4. Корбинский Б.А. Принципы математико-статистического анализа данных медико-биологических исследований / Б.А. Корбинский // Российский вестник перинатологии и педиатрии. – 1996. – Вып. 4. – С. 60-64.

5. Айвазян С.А. Прикладная статистика: Классификация и снижение размерности / С.А. Айвазян, В.М. Бухштабер, И.С. Енюков, Л.Д. Мешалкин. – М.: Финансы и статистика, 1989. – 607 с.

6. Максюта Н.В. Алгоритмы и методы снижения пространства диагностических признаков / Н.В. Максюта, А.И. Поворознюк // Вісник НТУ „ХПІ". – Х.: НТУ "ХПІ". – 2005. – №46. – С. 126-131.

7. Брандт З. Анализ данных; статистические и вычислительные методы для научных работников и инженеров / З. Брандт. – М.: Мир: АСТ, 2003. – 686 с.

8. Дюк В.А. Компьютерная психодиагностика / В.А. Дюк. – СПб.: Братство, 1994. – 364 с.

9. Мацуга О.М. Підтримка прийняття рішень під час кластерного аналізу медичних даних / О.М. Мацуга, Т.Г. Ємел'яненко // Актуальні проблеми автоматизації та інформаційних технологій. – Д.: ДНУ, 2008. – Т.12. – С. 28-–36.

10. Емельяненко Т.Г. Принятие решений в системах мониторинга / Т.Г. Емельяненко, А.В. Зберовский, А.Ф. Приставка, Б.Е. Собко. – Д.: ДНУ, 2005. – 224 с.

11. Васильев Д.А. Применение агломеративно-иерархического метода классификации в системах оперативного прогнозирования электрических нагрузок промышленных предприятий / Д.А. Васильев, И.А. Сарафанов // Энергетика. – 2004. – №6. – С. 90-93.

12. Краснопоясовський А.С. Класифікаційний аналіз даних: Навчальний посібник / А.С. Краснопоясовський. – Суми: СумДУ, 2002. – 159 с.

13. Ахутин В.М. Оценка качества формализованных медицинских документов / В.М. Ахутин, В.В. Шаповалов, М.О. Иоффе // Медицинская техника. – 2002. – Вып. 2. – С. 27-31.

14. Ногин В.Д. Принятие решения в многокритериальной среде: количественный подход / В.Д. Ногин. – М.: Физматлит, 2002. – 176 с.

15. Жиглявский А.А. Анализ данных в стоматологии: методология и реализация / А.А. Жиглявский, В.Н. Солнцев // Статистические методы в клинических испытаниях. – СПб.: СПбУ, 1999. – С. 8-56.

Поступила в редколлегию 25.10.2012

Рецензент: д-р техн. наук, проф. В.Д. Дмитриенко, Национальный технический университет «ХПИ», Харьков.

ВІДБІР ДІАГНОСТИЧНО ЦІННИХ ОЗНАК З ВИКОРИСТАННЯМ МЕТОДУ КОРЕЛЯЦІЙНИХ ПЛЕЯД

Н.В. Максюта, А.І. Поворознюк, С.В. Стешкін

В роботі обґрунтована необхідність формування інформативно повного простору показників при проектуванні інтелектуальних комп'ютерних систем медичної діагностики. Проведений короткий огляд методів відбору діагностично цінних ознак та методів кластеризації. Розроблено комплексний підхід до процедури відбору діагностично цінних ознак з використанням методу кореляційних плеяд. Розроблено структуру підсистеми кластеризації, алгоритм її функціонування та її програмна реалізація.

Ключові слова: *інтелектуальна комп'ютерна система медичної діагностики, відбір, діагностично цінна ознака, кластеризація, метод кореляційних плеяд.*

DIAGNOSTIC SELECTION OF INDICATORS USING METHOD CORRELATION PLEIADES

N.V. Maksyuta, A.I. Povoroznyuk, S.V. Steshkin

We justify the need for a complete space informative indicators for the design of intelligent computer systems for medical diagnostics. The brief overview of the methods of selection of diagnostically valuable indicators and methods of clustering. Developed a comprehensive approach to the selection procedure diagnostically valuable indicators using the method of correlation pleiades. The structure subsystem clustering algorithm for its operation and its software implementation.

Keywords: *Intelligent computer system of medical diagnosis, selection, diagnostically valuable indicator of the clustering, the method of correlation pleiades. selection, diagnostically valuable indicator of the clustering, the method of correlation pleiades.*