

*М.В. БУРЦЕВ*, магистр НТУ "ХПИ" (г. Харьков),  
*А.И. ПОВОРОЗНЮК*, канд. техн. наук, проф. НТУ "ХПИ" (г. Харьков)

## СИНТЕЗ КОМБИНИРОВАННОГО РЕШАЮЩЕГО ПРАВИЛА В ЗАДАЧЕ МЕДИЦИНСКОЙ ДИАГНОСТИКИ

Проведен обзор методов и алгоритмов синтеза решающих правил, применяемых в существующих медицинских системах поддержки принятия решений, выполнен анализ их достоинств и недостатков. Предложено комбинированное решающее правило, включающее вероятностный метод Вальда и описание структуры симптомокомплексов. Ил.: 3. Библиогр.: 14 назв.

**Ключевые слова:** поддержка принятия решений, метод Вальда, симптомокомплекс, комбинированное решающее правило.

**Постановка проблемы и анализ литературы.** Активное развитие информационных технологий ведет к компьютеризации всех областей социальной деятельности человека. Современный вычислитель способен если не полностью заменить специалиста, то значительно упростить его работу. Это особенно важно для областей плохо формализуемых задач, одной из которых является медицинская диагностика.

На сегодняшний день существуют такие медицинские системы поддержки принятия решений: *PUFF* – выполняет диагностику легочных заболеваний; *VM* – осуществляет контроль состояния здоровья пациентов палат интенсивной терапии; *MUCIN* – диагностирует инфекционные заболевания; *AI/COAG* – заболевания крови, а *AI/RHEUM* – ревматические заболевания, [1]; "*Тарвей*" – контролирует состояние сердечно-сосудистой системы (для реанимационных отделений, операционных, палат интенсивной терапии, кабинетов функциональной диагностики); "*Айболит*" – оценивает свойства системы кровообращения по индивидуализированной модели пациента отделения интенсивной терапии [2]; *АСПОН* – автоматизированная система профилактических осмотров населения [3]; *INTERNIST* [4], *MDX2* [5] – интегрированные медицинские системы [6] и другие.

В основу решающих правил (РП), применяемых в этих системах, легли следующие алгоритмы и методы.

1. **Метод Байеса.** Основан на вычислении вероятности возможных диагнозов по вероятностям появления комплексов некоторых признаков и априорным вероятностям самих диагнозов. В расчете используется формула Байеса. Полученное значение вероятности сравнивается с некоторым порогом, в случае превышения которого делается вывод о наличии диагноза.

*Требования:* независимые признаки  $x_j$ .

*Достоинства:* простота реализации.

*Недостатки:* необходим большой объем и репрезентативность обучающей выборки.

2. **Метод Вальда.** Основан на вычислении отношения правдоподобия для

последовательности признаков, и его сравнении со значением области неопределенности, полученным на основании анализа обучающей выборки. Для двух диагнозов  $D_q$  и  $D_w$  и системы признаков  $X$  отношение правдоподобия имеет вид:

$$\Omega = \frac{P(X/D_q)}{P(X/D_w)}, \quad (1)$$

при

$$P(D_q) + P(D_w) = 1, \quad (2)$$

где  $P(X/D_k)$  – условная вероятность наличия системы признаков  $X$  при диагнозе  $D_k$ .

В случае, когда  $X$  рассматривается как система *независимых* признаков вида

$$X = (x_1, x_2, x_3, \dots, x_m), \quad (3)$$

отношение правдоподобия (1) преобразуется в

$$\Omega = \prod_{j=1}^m \frac{P(x_j/D_q)}{P(x_j/D_w)}. \quad (4)$$

Анализ признаков  $x_i$  в (4) осуществляется в порядке убывания их информативности:

$$I_{D_i}(x_1) > I_{D_i}(x_2) > I_{D_i}(x_3) > \dots > I_{D_i}(x_m). \quad (5)$$

Полученное значение  $\Omega$  сравнивается с порогами:

$$A = \frac{1-\beta}{\alpha}; \quad (6)$$

$$B = \frac{\beta}{1-\alpha}, \quad (7)$$

где  $\alpha$  – вероятность ошибки принятия диагноза  $D_w$  при правильном диагнозе  $D_q$ ;  $\beta$  – вероятность ошибочно принятого диагноза  $D_q$  при наличии у больного диагноза  $D_w$ .

Если  $\Omega > A$ , то делается вывод о наличии диагноза  $D_q$ . Если  $\Omega < B$ , то делается вывод о наличии  $D_w$ . В противном случае выполняется отказ от постановки диагноза.

*Требования:* независимые признаки  $x_j$ .

*Достоинства:* возможность постановки уточняющего диагноза [7].

*Недостатки:* отношение правдоподобия может быть вычислено только для двух состояний, поэтому необходима дихотомизация диагностируемых состояний [8].

3. *Методы распознавание образов.* Основаны на представлении объекта

точкой в пространстве признаков. Данный класс методов включает множество различных алгоритмов, отличающихся вычисляемыми расстояниями (расстояния Эвклида, Минковского, Хэмминга, Камберра, Кендала и др.) [9], а также критериями их оценки (сравнение с прототипом, анализ  $k$  ближайших соседей, алгоритмы голосования и др.). К этому же классу относятся методы, в основе которых лежат нейронные сети.

*Достоинства:* высокая эффективность.

*Недостатки:* сложность реализации; проблема кластеризации объектов в виде компактных, четко различимых множеств.

4. *Детерминистические методы, основанные на применении симптомокомплексов.* Каждый диагноз определяется набором симптомов (диагностических признаков), которые образуют симптомокомплекс данного заболевания (вариантом представления является вектор дихотомических признаков [10]). Информацию о симптомах заболеваний в неформализованном виде можно найти в различных медицинских справочниках [11], сведения которых можно считать экспертной оценкой заболевания, выработанной многими поколениями врачей.

Различают следующие типы симптомов, формирующих симптомокомплексы: *патогномонические* (однозначно свидетельствуют о наличии заданного заболевания), *специфические* (присутствуют при заданном заболевании, но не указывают однозначно на его наличие), *неспецифические* (могут присутствовать при заданном заболевании).

Процесс диагностики представляет собой сравнение вектора признаков диагностируемого объекта с симптомокомплексами известных заболеваний. В качестве критерия сходства в простейшем случае может быть использовано расстояние Хэмминга [12].

*Достоинства:* простота реализации.

*Недостатки:* данный класс методов применим только для дихотомических признаков, для численных – необходимо определение соответствующих областей (например, с помощью лингвистических переменных или разбивки на интервалы [13]); существует проблема составления симптомокомплексов для новых заболеваний.

**Целью статьи** является анализ достоинств и недостатков существующих методов синтеза решающего правила, являющегося ядром диагностической подсистемы, а также разработка комбинированного решающего правила, с целью повышения эффективности процесса постановки диагноза.

**Синтез комбинированного решающего правила.** Анализ достоинств и недостатков существующих алгоритмов показывает, что перспективным является применение комбинированного РП, основанного на объединении вероятностного и детерминистического подходов.

Вероятностный подход, реализуемый методом Вальда, подразумевает вычисление объективных априорных условных вероятностей  $P(x_j / D_k)$  на

основании обучающей выборки. Для этого производится разбивка численных признаков  $x_j$  на непересекающиеся диагностически значимые интервалы  $\tau_i$  ( $i = \overline{1, n}$ ), по которым выполняется построение гистограмм, аппроксимирующих теоретический закон распределения  $f(x_j / D_k)$  анализируемого признака, для двух диагнозов  $D_q$  и  $D_w$  в одной системе координат. Вероятности  $P(x_{ji} / D_k)$ , относящиеся к гистограммам, рассчитываются как

$$P(x_{ji} / D_k) = \frac{N_{x_j D_k}}{N_{D_k}}, \quad (8)$$

где  $N_{x_j D_k}$  – количество объектов обучающей выборки со значением признака  $x_j$  при диагнозе  $D_k$ , попадающих в соответствующий интервал  $\tau_i$  (т.е.  $x_{ji} \in \tau_i$ ),  $N_{D_k}$  – общее число объектов обучающей выборки с диагнозом  $D_k$ .

На рис.1 приведены примеры теоретических законов распределения вероятностей  $f(x_j / D_q)$  и  $f(x_j / D_w)$ , а также аппроксимация этих законов распределения гистограммами.

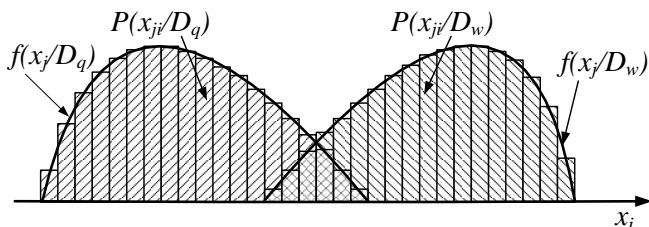


Рис. 1. Теоретические законы распределения условных вероятностей и их аппроксимация гистограммами

Если у пациента измеренный признак  $x_j$  попадает в интервал  $\tau_i$ , то берется соответствующая условная вероятность  $P(x_{ji} / D_k)$ , относящаяся к  $\tau_i$ , и подставляется в (4). Полученное значение сравнивается с порогами (6) и (7), на основании чего делается заключение о наличии диагноза ( $D_q$  или  $D_w$ ), или же выполняется отказ от дальнейшей диагностики, если в РП вошли все признаки  $x_j$ .

Подход, основанный на нечеткой логике [14], является выражением мнения экспертов. Он состоит в вычислении значения некоторой непрерывной функции  $\mu(x_j / D_k)$  – функции принадлежности, полученной исходя из мнения

специалиста, которая выражает степень соответствия значения численного признака  $x_j$  некоторой лингвистической переменной, описывающей рассматриваемый признак (например, "высокая температура" или "повышенное артериальное давление").

На рис. 2 показан внешний вид функций принадлежности  $\mu(x_i / D_q)$  и  $\mu(x_i / D_w)$ . Для данных функций справедливы соотношения:

$$\begin{aligned} \mu(x_i / D_q) &= \bar{\mu}(x_i / D_w) = 1 - \mu(x_i / D_w), \\ \mu(x_i / D_w) &= \bar{\mu}(x_i / D_q) = 1 - \mu(x_i / D_q). \end{aligned} \quad (9)$$

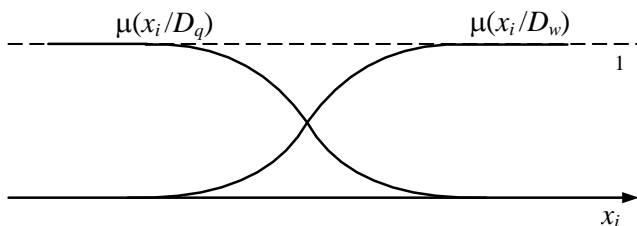


Рис. 2. Функции принадлежности  $\mu(x_i / D_q)$  и  $\mu(x_i / D_w)$

Кроме того, каждому из признаков  $x_j$  выставляется экспертная оценка, выражающая его вес в симптомокомплексе. Эти оценки используются для взвешивания значения  $\mu(x_j / D_k)$  и принимают значения  $e_0, e_1, e_2, e_3$ , где  $e_0$  – вес патогномонических симптомов;  $e_1$  – вес специфических симптомов;  $e_2$  – вес неспецифических симптомов;  $e_3$  – вес показателей, не входящих в симптомокомплекс данного заболевания.

При этом справедливо неравенство

$$e_0 \geq e_1 \geq e_2 \geq e_3, \quad \sum_{i=0}^3 e_i = 1. \quad (10)$$

Таким образом, учитывая оценки  $e_i$ , функция принадлежности принимает вид

$$\mu_1(x_j / D_k) = e_i \cdot \mu(x_j / D_k). \quad (11)$$

В конечном итоге, разные подходы (вероятностный подход, и подход на основе нечеткой логики) предназначены для решения одной и той же задачи классификации. Будучи полученными из разных предпосылок (статистика и мнение эксперта), они описывают общую проблему с различных позиций. Естественным является предположение, что их совместное использование перспективно, поэтому могут быть предложены следующие варианты их

совместного использования:

1. *Коллектив РП* [9]. Данный метод может быть проиллюстрирован следующим образом (рис. 3).

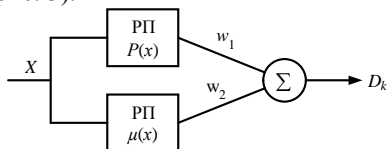


Рис. 3. Структурная схема коллектива решающих правил

Симптомокомплекс  $X$  анализируется с помощью каждого из РП, которым задана степень доверия  $w_i$ . Блок логического вывода формулирует окончательный диагноз  $D_k$  на основании взвешенных результатов работы всех решающих правил. При таком подходе, основанном на нечеткой логике, необходимо выполнение дефазификации.

2. *Взвешивание*. Отношение правдоподобия (4) с учетом (8) и (11) преобразуется к виду:

$$\Omega = \prod_{j=1}^m \frac{P(x_{ji} / D_q) \cdot \mu_1(x_j / D_q)}{P(x_{ji} / D_w) \cdot \mu_1(x_j / D_w)}. \quad (12)$$

Однако, при такой реализации нет возможности установить степень доверия к каждому из подходов отдельно.

3. *Суммирование оценок*. Так как условные вероятности  $P(x_j / D_k)$  и значения функции принадлежности  $\mu(x_j / D_k)$  являются нормированными величинами, то можно выполнить их усреднение с учетом весовых коэффициентов  $k_i$ . Тогда выражение (11) с учетом (8) преобразуется в

$$h(x_j / D_k) = k_1 P(x_{ji} / D_k) + k_2 \mu_1(x_j / D_k), \quad (13)$$

где  $k_i > 0$ ,  $\sum k_i = 1$ ,  $i = \overline{1,2}$ , а отношение правдоподобия приобретает вид

$$\Omega = \prod_{j=1}^m \frac{h(x_j / D_q)}{h(x_j / D_w)}. \quad (14)$$

**Выводы.** Проведен анализ существующих алгоритмов и методов реализации решающих правил, применяемых в медицинских системах поддержки принятия решений, выявлены их сильные и слабые стороны. Предложен вариант комбинированного РП, которое позволяет учесть не только объективные вероятности, но и субъективные оценки экспертов.

**Список литературы:** 1. Джарратано Дж. Экспертные системы: принципы разработки и программирование, 4-е издание. / Дж. Джарратано, Г. Райли. – М.: ООО "И. Д. Вильямс", 2007. – 1152 с. 2. Лицук В.А. Об инфраструктуре информационной поддержки клинической медицины / В.А. Лицук, А.В. Гаврилов, Г.В. Шевченко // Медицинская техника. – 2003. – № 4. – С. 36–42.

3. Ахутин В.М. Формальная модель автоматизированной системы профилактической осмотров населения / В.М. Ахутин, В.В. Шановалов // Медицинская техника. – 2002. – № 1. – С. 3–7.
4. Miller R.A. INTERNIST-1, An Experimental Computer-Based Diagnostic Consultant for General Internal Medicine / R.A. Miller, Jr. H.E. Pople, J.D. Myers // New England Journal of Medicine. – 1989. – Vol. 307. – P. 468–476.
5. Sticklen J. MDX2: An Integrated Medical Diagnostic System / J. Sticklen // PhD thesis, Department of Computer and Information Science, The Ohio State University. – 2000. – P. 746.
6. Генкин А.А. Новая информационная технология анализа медицинских данных (программный комплекс ОМИС) / А.А. Генкин. – СПб.: Политехника, 1999. – 191 с.
7. Поворознюк А.И. Метод постановки уточняющего диагноза в компьютерных системах медицинской диагностики при иерархической структуре диагностических признаков / А.И. Поворознюк // Збірник наукових праць Харківського університету Повітряних Сил. Х.: ХУПС, 2006. – Вип. 3 (9). – С. 125–130.
8. Поворознюк А.И. Синтез иерархической структуры диагностических признаков в компьютерных системах медицинской диагностики / А.И. Поворознюк // Вісник Національного технічного університету "ХПІ". – Харків: НТУ "ХПІ", 2003. – № 7. – Т. 2. – С. 39–44.
9. Дюк В.А. Компьютерная психодиагностика / В.А. Дюк. – СПб.: Братство, 1994. – 364 с.
10. Тимофієва Н.К. Моделирование целевой функции в задаче клинической диагностики на основе теории комбинаторной оптимизации / Н.К. Тимофієва // Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту. Матеріали міжнародної наукової конференції. Євпаторія 18-22 травня 2009 р. – Том 1. – Херсон: ХНТУ, 2009. – С. 219–223.
11. Справочник практического врача / Под ред. А.А. Михайлова, Л.И. Дворецкого. – М.: Новая Волна, 2001. – 528 с.
12. Поворознюк А.И. Компьютерные системы медицинской диагностики: Лабораторный практикум: в 2 ч. – Ч.2. / А.И. Поворознюк, А.Е. Филатова, А.Н. Шеин. – Харьков: НТУ "ХПИ", 2007. – 96 с.
13. Поворознюк А.И. Формирование диагностических интервалов численных признаков при дифференциальной диагностике / А.И. Поворознюк // Вісник Хмельницького національного університету. – Хмельницький: ХНУ. – 2007. – № 3. – Т. 1. – С. 106–109.
14. Рутковская Д. Нейронные сети, генетические алгоритмы и нечеткие системы: / Д. Рутковская, М. Пилиньский, Л. Рутковский. – М.: Горячая линия – Телеком, 2006. – 452 с.

*Статья представлена д.т.н. проф. НТУ "ХПИ" Серковым А.А.*

УДК 681.3

**Синтез комбінованого вирішального правила в задачі медичної діагностики** / Бурцев М.В., Поворознюк А.І. // Вісник НТУ "ХПІ". Тематичний випуск: Інформатика і моделювання. – Харків: НТУ "ХПІ". – 2009. – № 43. – С. 27 – 33.

Проведено огляд методів та алгоритмів синтезу вирішальних правил, що застосовуються в існуючих медичних системах підтримки прийняття рішень, виконано аналіз їхніх переваг і недоліків. Запропоновано комбіноване вирішальне правило, що включає імовірнісний метод Вальда та опис структури симптомного комплексу. Лл.: 3. Бібліогр.: 14 назв.

**Ключові слова:** підтримка прийняття рішень, метод Вальда, симптомний комплекс, комбіноване вирішальне правило.

UDC 681.3

**Synthesis of the combined solving rule in the problem of medical diagnostics** / Burtsev M.V., Povoroznyuk A.I. // Herald of the National Technical University "KhPI". Subject issue: Information Science and Modeling. – Kharkov: NTU "KhPI". – 2009. – №. 43. – P. 27 – 33.

The solving rules methods and synthesis algorithms applied in existing medical systems of decision-making support are reviewed, and analysis of their merits and demerits is made. The combined solving rule including a both Wald's method and the complex of symptoms structure description is offered. Figs.: 3. Refs.: 14 titles.

**Key words:** decision-making support, Wald's method, complex of symptoms, combined solving rule.

*Поступила в редакцію 19.10.2009*