

СИСТЕМА ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЯ В МЕДИЦИНЕ НА ОСНОВЕ СИНТЕЗА СТРУКТУРИРОВАННЫХ МОДЕЛЕЙ ОБЪЕКТОВ ДИАГНОСТИКИ

А. И. ПОВОРОЗНИК

*Национальный
технический
университет
«Харьковский
политехнический
институт»*

e-mail: perederiy@mail.ru

Разработан метод синтеза структурированных моделей объектов диагностики при построении компьютерных систем поддержки принятия решений в медицине на всех этапах преобразования информации. При этом учитывается не только структурный и функциональный базисы, но и неопределенность параметров модели и экспертные оценки.

Ключевые слова: медицинская диагностика, компьютерная система, структурированная модель, неопределенность параметров, экспертные оценки.

Постановка проблемы и анализ литературы

Внедрение новых информационных технологий в медицину обуславливает переход от традиционных медицинских информационно-поисковых систем к интеллектуальным компьютерным системам поддержки принятия решений в медицине (ИКСППМ) с развитым математическим аппаратом и элементами экспертных систем. Разработчики ИКСППМ отмечают следующие актуальные оптимизационные задачи, которые требуют первоочередного решения [1, 2]:

- выделения диагностически-значимых структурных элементов на фоне артефактов и помех (структурная идентификация) при анализе биологических сигналов и медицинских изображений;
- формализация, оценка информативности и формирование минимально необходимого множества информативных признаков;
- синтез решающего правила (РП), согласно которому выполняется процесс диагностики пациентов.

Как отмечается в [1], применение традиционных математических методов (детерминистических, вероятностных, и др.) для решения отмеченных оптимизационных задач в медицине является не совсем корректным, так как применяемые методы рассматривают модель объектов диагностики (ОД) в виде "черного ящика", не учитывают сложности организации диагностируемых подсистем организма, особенностей формирования системы разнотипных диагностических признаков (числовые, ранговые, дихотомические), разнородность методик их получения (анкетирования, осмотр, клинические исследования, анализ биологических сигналов и медицинских изображений), а также сложной итерационной процедуры постановки диагноза (от предварительного до уточненного), поэтому актуальной является задача разработки адекватных моделей ОД и РП на этих моделях.

Цель работы

Целью работы является разработка методов построения интеллектуальных компьютерных системам поддержки принятия решений в медицине на основе единого формализованного подхода синтеза структурированных моделей объектов диагностики на всех этапах преобразования информации.

Синтез структурированных моделей объектов диагностики.

Синтез структурированных моделей ОД основан на концепции эволюционной идентификации структурированных моделей (С-моделей), разработанной в работах Букатовой И. Л. [3], при этом базовым понятием является С-модель, реализующая преобразование $F: X \rightarrow Y$, то есть отображение входного вектора X в выходной Y через набор операторов f из множества F ($f \in F$) в соответствии с некоторой структурой S , которая определяет порядок следования f . Таким образом, С-модель задается графом, вершинами которого являются функциональные элементы f заданного класса моделей, а информационная связь вершин задается структурой S , то есть $C = \{f, S\}$.



Учитывая отмеченные выше оптимизационные задачи при построении ИКСППМ, разнородность диагностических признаков, итерационную процедуру постановки диагноза и необходимость учета экспертных оценок при формировании диагностического вывода, автором в [2] формализованы следующие этапы преобразования информации в ИКСППМ:

- структурная идентификация биологических квазипериодических сигналов (БКС), которые отражают циклические процессы в организме – электрокардиограмма (ЭКГ), реограмма (РГ), и др. и медицинских изображений внутренних органов (МИ) – томограмма, рентгенограмма, УЗИ;
- формализация описания разнородных диагностических признаков;
- синтез иерархических структур диагностируемых состояний S_D и диагностических признаков S_Z ;
- синтез диагностических РП при взаимодействии S_D и S_Z ;
- синтез индивидуальных диагностических прогнозирующих моделей;
- рекомендации по выбору оптимальной методики лечения.

На каждом из отмеченных этапов решается локальная оптимизационная задача, для которой используется свой набор исходных данных, свой критерий оптимальности и алгоритм оптимизационной процедуры, поэтому в [4] автором вводится понятие понятие расширенной C - модели (C' -модель) и процедура реконфигурации модели. C' -модель задается следующим образом

$$C' = \{S, F, E, \varepsilon_\sigma, \varepsilon_\Delta\}, \quad (1)$$

где: $S = \{P, V\}$ – структура модели, которая задана множествами вершин P и дуг V ;

F – функциональный базис модели, причем функции могут приписываться как вершинам f_p , так и дугам f_v ;

E – экспертные оценки, которые при необходимости дополняют F компоненты;

ε_σ – неопределенность параметров F -базиса, которая зависит от статистических свойств обучающей выборки;

ε_Δ – неопределенность параметров F -базиса, которая зависит от точности их задания (шага квантования).

Значения ε_σ и ε_Δ используется не только в качестве оценки параметров модели, но и являются параметрами оптимизационных процедур при синтезе модели. При реализации последовательности этапов преобразования информации, на каждом из которых используется критерий Q_i , выполняется реконфигурация модели $i-1$ этапа путем структурных и функциональных преобразований. Рассмотрим применение C -моделей при реализации отмеченных ранее этапов преобразования информации.

Этап выделение информативных структурных элементов БКС, параметры которых являются диагностическими признаками, является наиболее ответственным и трудоемким этапом их обработки, так как ошибки структурной идентификации (пропуск структурного элемента или ложная идентификация) приводят к грубым ошибкам вычисления диагностических показателей и ошибкам компьютерного диагноза (если не предусмотрены специальные средства выявления указанного типа ошибок, в частности отказ от обработки "подозрительного" периода или подтверждение человека-оператора правильности структурной идентификации).

Сложность идентификации состоит в том, что амплитудные и временные характеристики структурных элементов содержат информацию о работе диагностируемых подсистем организма и подвержены вариабельности в широком диапазоне, кроме того, эти сигналы могут содержать артефакты (резкое изменение некоторых характеристик в отдельных периодах) и снимаются на фоне помех. Традиционные методы обработки БКС во временной области (контурный анализ) являются по сути эвристическими, с трудом адаптируются к другим типам сигналов и не удовлетворяют возрастающим требованиям к качеству структурной идентификации.



Автором разработана формализованная процедура идентификации структурных элементов БКС на основе построения индивидуального пространства параметров для каждого типа структурного элемента с использованием методологии и вычислительных процедур преобразования Хока [4, 5]. При этом строится С'-модель (1) эталона каждого типа структурного элемента, параметрами которой являются:

- S – граф неточечного преобразования Хока фрагмента БКС из временной области $x(t_i)$ в пространство параметров $Y(P)$;
- F – вычислительные процедуры преобразования Хока;
- E – априорная информация о структуре БКС (адаптивная временная маска);
- ε_{Δ} – неопределенность квантования пространства параметров.

Синтез S_D выполняется процедурой иерархической кластеризации $F1: \{D\} \rightarrow S_D$ множества диагностируемых состояний $\{D\}$ в признаковом пространстве $\{X\}$, при этом параметрами С'-модели (1) являются:

- S – граф диагностируемых состояний (бинарное дерево решений);
- F – вычислительные процедуры кластеризации;
- E – экспертная информация о структуре симптомокомплексов.

Синтез S_z выполняется в результате преобразований $F2: \{X\} \rightarrow S_x$, $F3: S_x \rightarrow S_y$ и $F4: S_y \rightarrow S_z$ ($F3$ и $F4$ – реконфигурация структур). Параметры С'-модели (1):

- $S = \{S_x, S_y, S_z\}$ – графы диагностических признаков;
- F – вычислительные процедуры синтеза S;
- E – экспертная информация о структуре симптомокомплексов;
- ε_{Δ} – неопределенность квантования числовых признаков;
- ε_{σ} – неопределенность, обусловленная ограниченностью обучающей выборки.

Преобразование F2 удовлетворяет критериям минимума корреляционной связи и для его реализации в [2] предложена процедура иерархической кластеризации, основанная на представлении задачи кластеризации в виде потоковой модели. При этом исходные признаки представляются вершинами полносвязного графа, а дугам такого графа приписываются некоторые численные данные (k_{ij} – коэффициент парной корреляции, хотя возможно применение других статистических мер связи). Тогда задача иерархической кластеризации вершин сводится к последовательной процедуре поиска минимального разреза графа в потоковой задаче с ограничениями, для решения которой предлагается адаптация алгоритма “дефекта” [6]. В потоковых задачах с ограничениями каждая дуга V_{ij} характеризуется следующими параметрами: f_{ij} – поток по дуге; L_{ij} – нижняя пропускная способность дуги; U_{ij} – верхняя пропускная способность дуги; C_{ij} – стоимость прохождения единицы потока по дуге.

Для рассматриваемой задачи кластеризации диагностических признаков $L_{ij} = 0$, $U_{ij} = k_{ij}$. Значение C_{ij} при отсутствии априорной информации принимается равным 1 для всех дуг, а при ее наличии устанавливаются с помощью экспертных оценок, т.е. $C_{ij} = E_{ij}$, $E_{ij} < 1$ для тех вершин, которые соответствуют признакам, входящим в симптомокомплекс одного заболевания. Значения потоков f_{ij} определяются после завершения работы алгоритма “дефекта”, для которого необходимо задать начальную конечную вершины (исток s и сток t).

Задание s и t выполняется на основе оценочной кластеризации вершин по методу корреляционных плеяд, при котором исходное множество вершин N распадается на два непересекающихся подмножества N_s и N_t , после чего в каждом из них определяется наиболее связанная с остальными вершина ($s \in N_s$ и $t \in N_t$), удовлетворяющая условию $N_i = \max_t \sum_j C_{ij}$, причем $s = N_i \forall V_{ij} \in N_s$ и $t = N_i \forall V_{ij} \in N_t$. После определения потоков на всех дугах определяется минимальный разрез графа R, для которого справедливы соотношения:



$$\begin{aligned} f_{ij} &= U_{ij} \quad \forall V_{ij} \in (N_s, N_t), \\ f_{ij} &= 0 \quad \forall V_{ij} \in (N_t, N_s). \end{aligned} \quad (2)$$

Минимальный разрез разделяет все вершины графа на два непересекающихся множества (N_s , включающее начальную вершину s и N_t , включающее конечную вершину t) таким образом, что потоки насыщают все прямые дуги разреза и нулевые на всех обратных дугах разреза (2). В результате иерархической кластеризации синтезируется иерархическая структура S_x , на нижнем уровне иерархии которой могут быть как отдельные диагностические признаки (если они информативны и независимы), так и кластеры коррелированных признаков, а также схема объединения кластеров на остальных уровнях.

Преобразование $F3$ выполняется по критерию минимума интегральной неопределенности ε_k , которая включает ε_Δ и ε_σ и возникает при замене динамического диапазона изменения значений числового признака при разбижке Δ на m неравномерных диагностически-значимых интервалов Δ_k и построения соответствующих гистограмм. В [7] получено выражение оценки ε_k и итерационная процедура формирования Δ_k , при которой минимизируется ε_k .

Целью преобразования $F4$ является согласование топологий S_z и S_D для реализации диагностического РП и обеспечения оптимального плана обследования конкретного пациента. При этом, каждой паре элементов S_D , имеющей общего родителя, ставится в соответствие один элемент структуры S_z . Элементы структуры S_z формируются из элементов структуры S_y с учетом их информационной полноты

$$k_{\text{ин}}(x_j, D) = I_D(x_j) / H(D),$$

где $I_D(x_j)$ – информативность признака x_j относительно состояний D ; $H(D)$ – энтропия D , и диагностической ценности

$$k_{\text{дц}}(x_j, D) = k_{\text{ин}}(x_j, D) / r_t(x_j),$$

где $r_t(x_j)$ – суммарная оценка сложности измерения признака x_j .

В работе реализуется метод синтеза уточняющего диагноза [8], который является модификацией метода последовательного анализа (метода Вальда) и основан на анализе взаимодействия иерархических структур диагностических признаков S_z и диагнозов S_D . На каждом i -м этапе последовательного анализа, при дифференциальной диагностике между двумя диагнозами D_q и D_l , вычисляются отношения правдоподобия

$$\Theta = \prod_i \frac{E_{ikq} P(x_{ik} / D_q)}{E_{ikl} P(x_{ik} / D_l)}, \quad (3)$$

которое сравнивается с порогами

$$\Theta > A, \quad \Theta < B, \quad (4)$$

где A и B – верхняя и нижняя границы неопределенности, необходимые для принятия решения.

При выполнении одного из условий (4) принимается решение о диагнозе D_q или D_l соответственно и выполняется переход на более низкий уровень иерархии диагнозов с целью уточнения диагноза. При невыполнении обеих неравенств (4) добавляется следующий $i + 1$ признак и процедура повторяется. Для реализации комбинированного решающего правила, в выражении (3) учитывается неопределенность оценок условных вероятностей, кроме того, условные вероятности взвешиваются системой экспертных оценок о структуре симптомокомплексов E_{ikq} и E_{ikl} , методика расчетов которых приводится в [9].

Разработка структурной схемы СППРМ.

Для программной реализации рассмотренных ранее математических методов структурной идентификации ОД, СППРМ должна работать в двух режимах: режим обу-

чения и рабочий режим постановки диагноза. В режиме обучения выполняется формирование БЗ, в процессе которого выполняется синтез и настройка параметров структурированных моделей структурных элементов БКС и медицинских изображений, диагностических признаков и диагностируемых состояний, комбинированного РП. В режиме постановки диагноза выполняется анализ биологической информации, полученной от пациента и формирование диагностического заключения. На рис. 1 показана структурная схема СППРМ, в которой сплошными стрелками показаны информационные потоки в режиме обучения, а пунктирными – в режиме постановки диагноза.

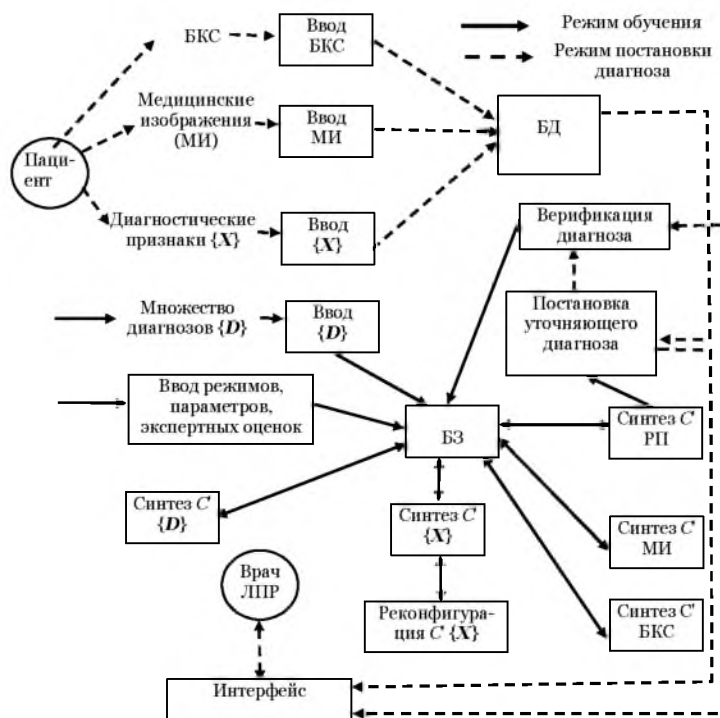


Рис. 1. Структурная схема СППРМ

Назначение модулей следующее:

- БД – база данных, которая хранит результаты обследований пациентов и верификацию диагностического заключения. Следует отметить, что при анализе БКС и медицинских изображений, в БД хранятся как исходные сигналы, так и результаты их структурной идентификации и сформированное множество вторичных диагностических признаков;
- БЗ – база знаний, которая хранит параметры отмеченных ранее структурированных моделей ОД.
- модуль ввода перечня диагностических признаков, диагностируемых состояний и экспертных оценок структур симптомокомплексов;
- модуль ввода значений первичных диагностических признаков;
- модули ввода БКС и медицинских изображений;
- модуль синтеза структурированных моделей структурных элементов БКС и медицинских изображений;
- модуль синтеза структурированной модели (иерархическая кластеризация) диагностируемых состояний;
- модуль синтеза и реконфигурации структурированной модели диагностических признаков;
- модуль реализации комбинированного РП;
- модуль формирования выходных экранных форм и документов, а также реализации интерактивного интерфейса с ЛПР.



Выполнен анализ средств для создания программного обеспечения, в результате которого обоснован выбор интерактивной среды программирования Inprise (Borland) Delphi на базе языка программирования высокого уровня Pascal, который характеризуется высокой точностью вычислений при работе с вещественными типами данных. Отмечена особенность языка явилась определяющей при значительном объеме вычислений, который выполняет СППРМ (особенно в режиме обучения). На этапе исследований и тестовых проверок использовался пакет "СТАТИСТИКА".

Апробация метода и обсуждение результатов.

Для апробация метода синтеза S_z и реализации РП, взята выборка из 434 пациентов, у которых диагностировалось 10 диагнозов (в скобках указано количество пациентов): D_1 – здоров (243), D_2 – анемия неуточненная (29), D_3 – кардиомиопатия (10), D_4 – хроническая ишемическая болезнь сердца неуточненная (17), D_5 – аллергический геморрагический васкулит (24), D_6 – хронический холецистит (21), D_7 – пневмония неуточненная (16), D_8 – инфекционный мононуклеоз неуточненный (28), D_9 – хронический лимфоидный лейкоз (26), D_{10} – почечная недостаточность неуточненная (20). Анализировались 9 показателей клинического анализа крови: Eritr – эритроциты, Hb – гемоглобин, Leic – лейкоциты, Limf – лимфоциты, Eozin – эозинофилы, Pal – нейтрофилы палочкоядерные, Seg – нейтрофилы сегментоядерные, Mon – моноциты, Сое – скорость оседания эритроцитов. Полученные показатели качества классификации приведены в табл. 1.

Таблица 1

Показатели качества классификации

	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	D_{10}	Среднее
Ψ %	97,53	75,86	40,00	29,41	91,67	85,71	62,50	85,71	88,46	95,0	88,48
Ψ' %	98,5	67	62	69	98,4	97,3	79	98,4	93	95	92,8
M_x	2,2	6,68	7,82	6,42	5,39	3,67	4,58	5,39	5,75	3,73	3,61

В табл. 1 в первой строке указана достоверность классификации Ψ стандартным методом дискриминантного анализа в пространстве 9 исходных признаков, а во второй Ψ' – предлагаемым методом уточняющего диагноза с использованием оптимальной схемы обследования пациентов. В третьей строке M_x – среднее число анализируемых признаков у одного пациента. Последний столбец таблицы рассчитан по формуле среднего для неравномерных групп

$$\bar{x} = \sum_i x_i \frac{N_i}{N},$$

где x_i – показатель качества; N_i – количество объектов в i -й группе; N – суммарный объем выборки.

Из табл. 1 следует, что в предлагаемом методе среднее число анализируемых признаков равно 3,61, что в 2,5 раза меньше исходных 9 признаков. При применении предлагаемого метода средняя достоверность диагноза повысилась с 88,48% до 92,8%, при этом из общего числа отрицательных ответов (31 ответ из 434) получена дифференциация 25 отрицательных ответов о невозможности постановки диагноза на предварительные диагнозы разного уровня. Кроме того, существенно повысилась достоверность тех диагнозов, у которых в обучающей выборке имеется недостаточно объектов (D_3, D_4, D_7, \dots) за счет незначительного снижения достоверности диагноза с наибольшей выборкой – D_1 . Это свидетельствует об эффективности применения предлагаемого метода на обучающих выборках с существенной неравномерностью количества объектов в анализируемых классах.

Выводы

Разработаны методы построения ИКСППМ на основе синтеза структурированных моделей ОД, которые учитывают не только структурный и функциональные



базисы, но и неопределенность параметров, а также позволяют учитывать экспертные оценки. Показано применение метода структурной идентификации на всех этапах преобразования информации. Разработанные методы позволяют снять ограничения на размерность пространства диагностических признаков, повысить достоверность компьютерного диагноза и адаптироваться к конкретным объектам медицинской диагностики.

Литература

1. Ахутин В.М. Оценка качества формализованных медицинских документов [Текст] / В.М. Ахутин, В.В. Шаповалов, М.О. Иоффе // Медицинская техника М.: 2002. № 2. С. 27 – 31.
2. Поворознюк А.И. Формализация этапов проектирования интеллектуальных компьютерных систем медицинской диагностики [Текст] / А.И. Поворознюк // Электронное моделирование К.: ИПМЕ, 2006. – Т. 28. №1 С. 85 – 97.
3. Букатова И.Л. Эвоинформатика: теория и практика эволюционного моделирования. [Текст] / И.Л. Букатова, Ю.И. Михасев, А.М. Шаров – М.: Наука, 1990.– 212с.
4. Поворознюк А.И. Система поддержки принятия решений в медицине на основе структурной идентификации объектов диагностики [Текст] / А.И. Поворознюк //Сборник научных трудов СНУЯЕиП. Севастополь: СНУЯЕиП, 2008. № 1 (25) С. 234 – 245.
5. Поворознюк А.И. Применение преобразования Хока для структурной идентификации физиологических сигналов // Моделювання та інформаційні технології. Збірник наукових праць інституту проблем моделювання в енергетиці. . К.: ИПМЕ, 2003. – Вип. 22, с.143 – 149.
6. Филлипс Д. Методы анализа сетей / Пер. с англ. [Текст] / Д. Филлипс, А. Гарсиа-Диас – М: Мир, 1984. – 648 с.
7. Поворознюк А.И. Формирование диагностических интервалов численных признаков при дифференциальной диагностике// Вісник Хмельницького національного університету. – Хмельницький: ХНУ, 2007.-№3, Т.1. С.106 – 109.
8. Поворознюк А.И. Метод постановки уточняющего диагноза в компьютерных системах медицинской диагностики при иерархической структуре диагностических признаков [Текст] / А.И. Поворознюк // Збірник наукових праць Харківського університету Повітряних Сил. Х: ХУПС, 2006. – Вип. 3(9). С. 125 – 130.
9. Поворознюк А.И. Интеллектуальная система поддержки принятия решений в медицине [Текст] / А.И. Поворознюк // Материалы XIII международной конференции с автоматического управления (Автоматика 2006), г.Винница 25-28 сентября 2006г. – Винница: УНИ-ВЕРСУМ-Винница, 2007. С. 364 – 368.

DECISION SUPPORT SYSTEM IN MEDICINE ON BASE OF THE SYNTHESSES OF THE OUTLINE MODELS OBJECT DIAGNOSTICES

A. I. POVOROZNYUK

*National technical university
"Kharkov polytechnic institute",
Kharkov, Ukraine*

e-mail: perederiy@mail.ru

The Designed method of the syntheses of the outline models object diagnosticses at building computer decision support system in medicine on all stage of the transformation to information. Is it Here-with taken into account not only structured and functional bases, but also uncertainty parameter to models and expert estimations.

Key words: medical diagnostics, computer system, outline model, uncertainty parameter, expert estimations.