

## ФОРМАЛИЗАЦИЯ ДИАГНОСТИЧЕСКИХ ПРИЗНАКОВ В КОМПЬЮТЕРНЫХ СИСТЕМАХ МЕДИЦИНСКОЙ ДИАГНОСТИКИ

к.т.н. А.И. Поворознюк, к.т.н. Н.И. Поворознюк  
(представил д.т.н., проф. В.Д. Дмитриенко)

*Проанализированы диагностические признаки в компьютерных системах медицинской диагностики. Показана разнородность признаков (численные, порядковые, дихотомические, качественные) и их иерархическая структура. Предлагается подход к формализации описания системы признаков и формирования иерархической системы информативных диагностических признаков на основании декомпозиции структур. Рассматривается схема реализации данного подхода.*

Организм человека – сложная, иерархическая, многоуровневая, самоорганизующаяся система, состоящая из взаимосвязанных подсистем различного уровня подчинения [1]. При этом на каждом уровне иерархии регуляторные механизмы соответствующего уровня поддерживают некоторые параметры (физиологические константы), значения которых (установки) меняются системами более высоких уровней или взаимосвязанными системами того же уровня. Сказанное выше говорит о том, что каждая  $i$ -я подсистема организма  $R_i$  (сердечно-сосудистая, дыхания, нервная, иммунная и др.) находится в определенном  $j$ -м состоянии  $S_j^i$  из множества состояний

$$M_i = \{ S_0^i, \dots, S_j^i, \dots, S_n^i \}.$$

Множество состояний  $M^i$  условно разбивается на подмножества:  $M_n^i$  – норма,  $M_{cp}^i$  – пограничное состояние и  $M_n^i$  – патология. Состояние всего организма  $M^{opz}$  определяется состоянием всех его подсистем

$$M^{opz} = \{ M^0, \dots, M^i, \dots, M^n \}.$$

Таким образом, задача диагностики компьютерной медицинской диагностической системы сводится к задаче определения принадлежности текущего состояния организма или отдельной его подсистемы к одному из перечисленных подмножеств ( $M_n^i$ ,  $M_{cp}^i$  или  $M_n^i$ ). Следует отметить, что при конкретной реализации медицинской диагностической системы может диагностироваться не одна, а несколько патологий (и

соответственно несколько пограничных состояний). В этом случае диагностируемая подсистема  $R_i$  имеет внутренние уровни иерархии и подмножества  $M_{zp}^i$  и  $M_n^i$  разбиваются на более мелкие подмножества по числу диагностируемых патологий (заболеваний).

Исходными данными для любой компьютерной медицинской диагностической системы при диагностике  $i$ -й подсистемы организма является принятая в медицинской практике система диагностических признаков  $X^i$ :

$$X^i = (x_0^i, \dots, x_j^i, \dots, x_k^i), \quad (1)$$

которая в какой-то мере отражают текущее  $j$ -е состояние  $i$ -й подсистемы организма  $S_i^j$ , т.е. с позиций теории распознавания образов состояния  $S_i^j$  проецируются на пространство признаков  $X^i$ , которое делится на непересекающиеся подпространства  $X_M^i$  по числу диагностируемых подмножеств.

При проектировании компьютерных медицинских диагностических систем используются исходные данные следующих типов:

1) *физиологические константы организма* – медленно меняющиеся параметры, отражающие определенные установки регуляторных систем, которые в свою очередь классифицируются следующим образом [2]: а) *дихотомические или бинарные* – принимают два значения (0 или 1, “ДА” или “НЕТ”); б) *ранговые или порядковые* – принимают конечное число упорядоченных значений и отражают степень проявления (тяжесть) некоторого процесса. Указанные типы признаков получаются в основном при осмотре пациентов врачом и при субъективной симптоматике путем заполнения различных опросников по различным классам заболеваний; в) *численные* – признаки, получаемые в результате различных инструментальных исследований;

2) *физиологические сигналы* (результаты инструментальных исследований), отражают динамику происходящих в организме процессов;

3) *дву- и трехмерные изображения* – различные виды изображений внутренних органов человека (рентгенограммы, томограммы, скенограммы), характеризующие структуру указанных органов в норме и патологии.

Во взаимосвязанной многоуровневой системе результаты любых измерений (не говоря об осмотре врача или субъективной симптоматике) можно считать косвенными измерениями, так как с одной стороны один признак  $x_l^i$  (его изменение или проявление) отражает изменение состояний многих подсистем организма  $\{R_0, \dots, R_i, \dots, R_k\}$  (например, такой интегральный признак как температура тела, проявляется при многих патологиях):

$$x_l^i \rightarrow \{S_j^0, \dots, S_j^i, \dots, S_j^k\},$$

а с другой стороны – изменение состояния одной подсистемы регистрируется изменением многих признаков

$$S_j^i \rightarrow \{x_0^i, \dots, x_l^i, \dots, x_k^i\},$$

т.е. отношение между  $S_j^i$  и  $X^i$  имеют тип "многие к многим".

Рассмотренный тип отношений  $S_j^i$  и  $X^i$  относится ко всем перечисленным выше типам исходных данных, но если для первого типа исходных данных (физиологических констант) вектор диагностических признаков (1) является первичным (непосредственно измеряемым), то в случае 2-го и 3-го типов (сигналов и изображений) картина становится намного сложнее, так как диагностические признаки получаются в результате довольно сложных алгоритмов обработки.

При обработке физиологических сигналов во временной области численные значения диагностических признаков получаются в результате следующих этапов обработки сигналов [3]: регистрация и оцифровка сигнала; цифровая фильтрация и предварительная обработка сигнала; выделение диагностических структурных элементов (структурная идентификация сигнала); уточнение параметров структурных элементов; вычисление диагностических признаков, которые являются амплитудно-временными параметрами структурных элементов.

При обработке двумерных изображений (трехмерные изображения получаются путем композиции двумерных срезов) выполняются следующие этапы обработки [4]: регистрация и оцифровка изображения; различные виды двумерной фильтрации с целью снижения уровня шумов и повышения контрастности необходимых структурных элементов изображения; ручная, полуавтоматическая или автоматическая процедура выделения диагностических структурных элементов; вычисление диагностических признаков, которые являются параметрами структурных элементов (количество, геометрические размеры, площадь, периметр и т.д.).

Качество получаемых диагностических признаков вида (1) для 2-го и 3-го типов данных определяется в основном качеством этапа структурной идентификации [3].

Под формализацией системы диагностических признаков [2] подразумевается стандартизация их описания (разработка стандартных форм историй болезней, бланков обследований, анкет, опросников и т.д.) и придание признакам вида, удобного для компьютерной обработки (разработка системы кодировки данных, которая особенно актуальна при обработке дихотомических и порядковых признаков). Однако этого недостаточно для получения качественного компьютерного диагноза, так как представление системы признаков в виде вектора (1) не отражает сложности отношений  $S_j^i$  и  $X^i$ , отдельные признаки  $X_l^i$  взаимосвязанные, зависимые и малоинформативные, система признаков  $X^i$  обычно избыточна.

Из теории идентификации известно, что наилучшие результаты идентификации получаются в том случае, когда структура модели максималь-

ным образом отражает структуру объекта. Применительно к рассматриваемой задаче объектом является сложная иерархическая структура подсистем организма  $R_i$  которая характеризуется системой состояний  $S_i^j$  (структура указанной системы, как правило, нам неизвестна), а моделью является система диагностических признаков  $X^i$ . Поэтому на этапе формализации диагностических признаков предлагается построение формализованного подхода к построению иерархической системы информативных диагностических признаков минимально необходимого объема (рис. 1).

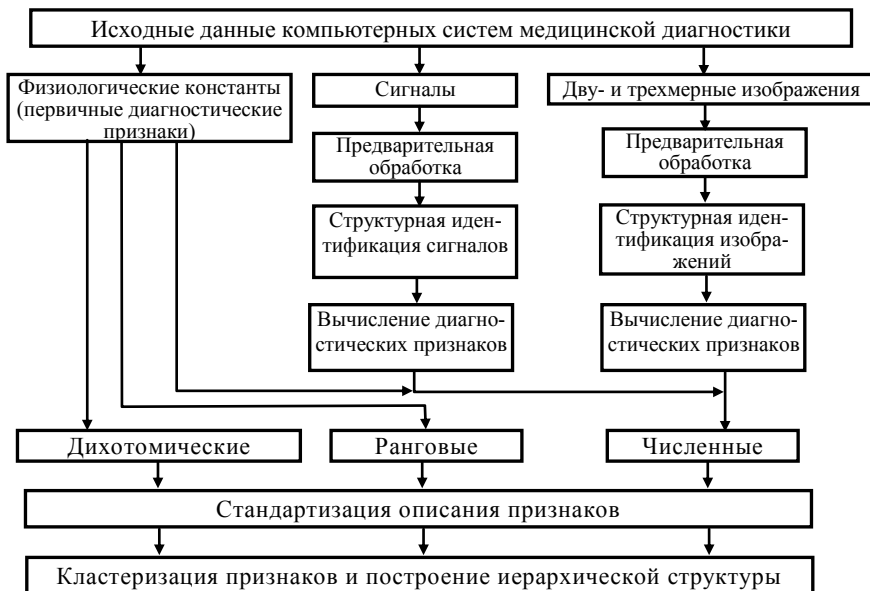


Рис. 1. Структурная схема формализации диагностических признаков

Кроме вышеотмеченных особенностей системы признаков, предпосылкой к такой постановке задачи является тот факт, что в медицинской практике широко используется многоуровневая система диагностических признаков (признаки первого уровня для выявления отклонений от нормы при скрининговых обследованиях, признаки второго уровня для диагностики заболеваний при углубленном обследовании и т.д.).

Задачу построения иерархической системы диагностических признаков  $X^i$  можно свести к потоковой задаче [5]. При этом исходные признаки представляются вершинами полносвязного графа, а дугам такого графа приписываются некоторые численные данные (веса), в качестве которых могут быть использованы статистические меры связи  $K_{ij}$  или расстояния  $D_{ij}$  в пространстве признаков. В качестве статистической меры связи используется

коэффициент парной корреляции для численных признаков, или его аппроксимации (коэффициент сопряженности для дихотомических и коэффициент ранговой корреляции для ранговых признаков). При наличии априорной информации о структуре объекта некоторые вершины могут быть объединены, а дуги исключены. Тогда задача структурной идентификации сводится к задаче группировки вершин графа в некоторые классы (кластеризация вершин) или к эквивалентной ей задаче разреза графа на подграфы. При такой формулировке задача сводится к потоковой [3], для решения которой предлагается адаптация алгоритма “дефекта” [5].

Кластеризация выполняется таким образом, чтобы достичь максимальной связи вершин внутри класса (подграфа) при минимальной связи между классами. Процедура кластеризации является итерационной (после разбивки исходного графа на  $N$  подграфов к каждому из подграфов применяется кластеризация для получения подграфов второго уровня и т.д.), в результате которой получается иерархическая структура подграфов, в каждом из которых собраны коррелированные признаки  $\{x_l^i\}_G$ . Окончательным этапом кластеризации является выделение в группе  $\{x_l^i\}_G$  коррелированных признаков в подграфах последнего уровня иерархии наиболее информативного признака  $\{x_l^i\}_G \rightarrow x_l^i$  (снижение размерности пространства признаков) или замена группы интегральным диагностическим признаком  $\{x_l^i\} \rightarrow y_l^i$  (переход к новому пространству признаков). Полученные в результате кластеризации диагностические признаки используются для построения диагностических решающих правил.

## ЛИТЕРАТУРА

1. *Справочник по функциональной диагностике / Под ред. И.А. Кассирского.* – М.: Медицина, 1970. – 848 с.
2. Дюк В.А. *Компьютерная психодиагностика.* – С.-П.: Братство, 1994. – 364 с.
3. Поворознюк А.И., Филатова А.Е. *Выбор метрики пространства признаков в задаче структурной идентификации квазипериодических сигналов // Вестник ХГПУ.* – Х.: ХГПУ. – 2000. – Вып. 99. – С. 138 – 141.
4. Грузман И.С., Киричук В.С., Косых В.П., Перетягин Г.И., Спектор А.А. *Цифровая обработка изображений в информационных системах: Учебное пособие.* – Новосибирск: Изд-во НГТУ, 2000. – 168 с.
5. Поворознюк А.И. *Структурная декомпозиция биологических объектов на основании потоковых моделей // Вестник ХГПУ.* – Х.: ХГПУ. – 2000. – Вып. 99. – С. 134 – 137.

Поступила 12.09.2002

**ПОВОРОЗНЮК Анатолий Иванович**, канд. техн. наук, доцент, проф. кафедры ВТ и программирования НТУ «ХПИ». В 1977 году окончил ХПИ. Область научных интересов – разработка методов и алгоритмов построения компьютерных систем медицинской диагностики.

**ПОВОРОЗНЮК Назар Иванович**, канд. техн. наук, доцент, доц. кафедры информаци-

*онно-измерительной техники НТУУ «КПИ» (г. Киев). В 1978 году окончил КПИ. Область научных интересов – методы и приборы обработки сигналов.*