

УДК 681.5

А.И. Поворознюк

Национальный технический университет «Харьковский политехнический институт»

РЕКОНФИГУРАЦИЯ СТРУКТУРЫ ДИАГНОСТИЧЕСКИХ ПРИЗНАКОВ ПРИ СИНТЕЗЕ КОМПЬЮТЕРНОГО ДИАГНОЗА

В статье предлагается метод реконфигурации структуры диагностических признаков в разработанном автором методе синтеза уточняющего диагноза, который является модификацией метода последовательного анализа (метода Вальда) и основан на анализе взаимодействия иерархических структур признаков и диагнозов. Синтез и реконфигурация иерархической структуры разнородных диагностических признаков выполняется по критерию автоинформативности (иерархическая кластеризация на основе анализа корреляционных связей признак-признак), а также по критерию диагностической ценности диагностических признаков относительно заданного множества состояний (диагнозов) на каждом уровне взаимодействия иерархических структур признаков и состояний. Применение разработанного метода позволяет выполнять оптимальную схему обследования при постановке уточняющего диагноза в условиях существенной априорной неопределенности.

метод реконфигурации структуры диагностических признаков, синтез уточняющего диагноза

Введение

Постановка проблемы и анализ литературы.

В настоящее время проводятся интенсивные работы по внедрению новых информационных технологий в медицину и построению интеллектуальных компьютерных систем медицинской диагностики (ИКСМД). При этом, в [1, 2] отмечается, что актуальной оптимизационной задачей при синтезе диагностического решающего правила является выбор минимально

необходимого множества информативных признаков и обеспечение оптимальной схемы обследования пациентов. Автором в [3] разработан метод синтеза уточняющего диагноза, который является модификацией метода последовательного анализа (метода Вальда) и основан на анализе взаимодействия иерархических структур диагностических признаков и диагнозов. В [4, 5] разработан метод синтеза иерархических структур признаков S_x^a и диагнозов

S_D^a на основе их автоинформативности (иерархическая кластеризация на основе анализа корреляционных связей) путем реализации преобразований

$$F1: X \rightarrow S_x^a; \quad F2: D \rightarrow S_D^a. \quad (1)$$

И если иерархическая структура диагнозов (бинарное дерево) S_D^a может непосредственно использоваться в диагностической процедуре уточняющего диагноза $S_D \equiv S_D^a$, то иерархическая структура диагностических признаков S_x^a непосредственно использоваться в указанной диагностической процедуре не может по следующим причинам:

- структура S_x^a должна содержать минимально-необходимый объем информативных диагностических признаков для заданной задачи диагностики (заданного множества диагнозов);
- структуры S_x^a и S_D^a получены независимо друг от друга, имеют разное количество уровней иерархии, поэтому структуру S_x^a необходимо согласовать со структурой S_D^a по количеству уровней иерархии;
- должна быть разработана процедура замены кластеров всех уровней иерархии S_x^a интегральными диагностическими признаками.

Исходя из выше изложенного, в работе ставится задача реконфигурации структуры S_x^a , с целью синтеза структуры S_z , которая непосредственно используется в диагностическом решающем правиле, путем выполнения некоторого преобразования

$$F3: S_x^a \rightarrow S_z. \quad (2)$$

Цель статьи – разработка метода реконфигурации структуры диагностических признаков по критерию диагностической ценности признаков относительно заданного множества состояний на каждом уровне взаимодействия иерархических структур признаков и состояний в разработанном автором методе синтеза уточняющего диагноза.

Основной материал

В диагностической процедуре уточняющего диагноза анализируется взаимодействие иерархических структур S_z и S_D , поэтому в работе ставится задача синтеза иерархической структуры информативных признаков S_z , которая учитывает не только автоинформативность системы признаков, но и некоторый "внешний" критерий, отражающий указанное взаимодействие. Так как конечной целью синтеза иерархической структуры информативных признаков является формирование диагнозов пациентов, то в качестве "внешнего" критерия в работе

предлагается использовать информативность признака x_i относительно заданной системы диагнозов D , т.е. количество информации, внесенной в систему после проведения обследования пациента на признак $x_i - I_D(x_i)$ [6].

Оценка информативности признаков выполняется на основе анализа отношений признаки – диагнозы. Целью такой оценки является определение диагностической ценности отдельных признаков или системы признаков для конкретных диагнозов или системы диагнозов на каждом уровне взаимодействия иерархических структур признаков/диагнозов. Информативность признаков используется для решения следующих задач:

- исключение из системы неинформативных признаков с целью выделения множества информативных признаков;
- формирования иерархической структуры информативных признаков;
- ранжировка признаков по критерию их информативности для определения порядка их включения в процедуру последовательного анализа.

Учитывая выше изложенное, в работе предлагается следующая процедура реализации преобразования (2)

$$F3: S_x^a \rightarrow S_y \rightarrow S_z. \quad (3)$$

Рассмотрим более подробно реализацию отмеченных в (3) этапов преобразования F2 (реконфигурации структуры S_x^a). Как отмечается в [7], реконфигурация структур может выполняться путем реализации следующих действий:

- 1) замена элементов исходной структуры;
- 2) замена связей между элементами;
- 3) одновременная замена элемента и связей.

Целью преобразования $F4: S_x^a \rightarrow S_y$ является синтез иерархической структуры информативных признаков минимально-необходимого объема относительно заданной системы диагнозов. При реализации преобразования F4 выполняется первый тип отмеченных выше действий – замена элементов, в результате которой графы структур S_x^a и S_y топологически подобны.

В результате преобразования F4 определяется информативность исходного пространства признаков X относительно системы диагнозов $\{D\}$, выполняется замена кластеров нижнего уровня иерархии структуры S_x^a (подмножества коррелированных признаков) наиболее информативным (преобразование элементов S_x^a нижнего уровня) и выполняется замена кластеров остальных уровней S_x^a интегральными признаками. Кроме того, при синтезе S_y , система разнородных элементов S_x^a приводится к еди-

ной шкале – каждый исходный признак x_i представляется непересекающимся множеством диагностически значимых интервалов. Для реализации преобразования F4 выполняется оценка информативности системы диагностических признаков относительно заданной системы диагнозов. Рассмотрим коротко необходимые в дальнейшем выражения отмеченной оценки. Более подробно реализация преобразования F4 рассматривается в [8].

Если система диагнозов образует полную группу несовместных событий (каждому пациенту в обучающей выборке поставлен один диагноз D_i , нет пациентов с несколькими диагнозами), то неопределенность системы диагнозов оценивается с помощью энтропии [6, 8]:

$$H(D) = -\sum_{i=1}^n P(D_i) \cdot \log_2 P(D_i), \quad (4)$$

где $P(D_i)$ – априорная вероятность диагноза D_i .

Выражение (4) является оценкой максимально возможного количества информации, которое может быть внесено системой диагностических признаков, то есть для любой системы диагнозов D и любой системы p диагностических признаков $X = \{x_1, x_2, \dots, x_p\}$ справедливо неравенство

$$I_D(X) \leq H(D), \quad (5)$$

которое превращается в равенство только для системы детерминированных признаков.

Суммарная информативности численного признака x_j относительно всей системы диагнозов D , при условии разбивки его динамического диапазона на m непересекающихся интервалов равна

$$I_D(x_j) = \sum_{i=1}^n \sum_{k=1}^m P(D_i) \cdot P(x_{jk} / D_i) \times \log_2 \frac{P(x_{jk} / D_i)}{P(x_{jk})}. \quad (6)$$

При обследовании пациента по комплексу независимых признаков $X = \{x_1, x_2, \dots, x_p\}$, их диагностическая ценность определяется по

$$I_D(X) = \sum_{j=1}^p I_D(x_j). \quad (7)$$

Целью преобразования F5: $S_y \rightarrow S_z$ является согласование топологий S_z и S_D для реализации рассмотренного в [3] метода уточняющего диагноза и обеспечения оптимального плана диагностического обследования конкретного пациента.

Элементы структуры S_z формируются из элементов структуры S_y с учетом их информационной полноты и диагностической ценности (см. ниже) таким образом, чтобы с их помощью можно было выполнить дифференциальную диагностику на каждом уровне иерархии диагнозов. Таким образом,

каждой паре элементов S_D , имеющей общего родителя, ставится в соответствие один элемент структуры S_z . При реализации преобразования F5 выполняется третий тип преобразования структур – замена элементов и замена связей. Реализация преобразования F5 является предметом исследования данной работы.

Синтез иерархической структуры признаков на основе автоинформативности обеспечивает слабую зависимость признаков на всех уровнях иерархии и дает основание для оценки их диагностической ценности $I_D(X)$ по (7). Так как $I_D(X)$ должно удовлетворять неравенству (5), то система диагностических признаков X относительно системы диагнозов D характеризуется коэффициентом информационной полноты $k_{ин}(X, D)$, который определяется по выражению

$$k_{ин}(X, D) = \frac{I_D(X)}{H(D)}. \quad (8)$$

Аналогичным образом определяется коэффициент информационной полноты $k_{ин}(x_j, D)$ каждого признака x_j

$$k_{ин}(x_j, D) = \frac{I_D(x_j)}{H(D)}. \quad (9)$$

Так как процесс измерения диагностических признаков требует определенного медицинского оборудования и ресурсов (временных, материальных, финансовых и т.д.), то с учетом суммарной сложности измерения признака, каждый признак характеризуется коэффициентом диагностической ценности

$$k_{дц}(x_j, D) = \frac{k_{ин}(x_j, D)}{r_t(x_j)}, \quad (10)$$

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

Диагностическая ценность признака, определяемая по (10) повышается у наиболее информативных признаков с наименьшей сложностью их измерения.

С учетом (7), для комплекса независимых признаков справедливы соотношения:

$$k_{ин}(X, D) = \sum_j k_{ин}(x_j, D); \quad (11)$$

$$k_{дц}(X, D) = \sum_j k_{дц}(x_j, D). \quad (12)$$

Следует отметить, что при взаимодействии системы признаков с иерархической системой диагнозов следует различать понятие информативности относительно уровня иерархии (степени детализации) системы диагнозов. Нетрудно заметить, что при повышении степени детализации (движение по иерархической структуре диагнозов вниз) неопреде-

ленность диагнозов, которая определяется по (4) возрастает, поэтому для диагностики каждого k -го уровня иерархии диагнозов система диагностических признаков должна вносить соответствующее количество информации, определяемое по (6), (7).

Иерархическая структура диагнозов S_D представляет собой бинарное дерево, где каждый уровень состоит из i пар элементов (кластеров), имеющих одного родителя $i - 1$ уровня, т.е.

$$D_k^i = \{D_{q,k}^i, D_{q+1,k}^i\}, \quad (13)$$

где верхний индекс i указывает на уровень иерархии структуры S_D ; q – номер элемента SD при последовательной нумерации элементов на i -м уровне иерархии, $q \leq 1, 2^i$; k – номер родителя элемента S_D при последовательной нумерации элементов на $(i - 1)$ -м уровне иерархии S_D .

Отношения элементов S_D между i -м и $(i + 1)$ -м уровнями иерархии (увеличение индекса i указывает на увеличение степени детализации) являются отношениями вхождения. Фиктивный нулевой уровень D^0_1 является множеством всех диагнозов и является родителем одной пары элементов первого уровня $\{D^1_{1,1}, D^1_{2,1}\}$. Каждый элемент первого уровня может быть родителем пары элементов следующего уровня и т.д. Если структура диагнозов S_D содержит p уровней иерархии, то на каждом i -м уровне

не может находиться до 2^i пар элементов, имеющих общих родителей. Таким образом, при реализации диагностической процедуры необходимо на каждом i -м уровне иерархии выполнить дифференциальную диагностику только двух элементов при условии, что на предыдущем этапе определено состояние родителя. Дифференциальная диагностика пары $\{D^i_{q,k}, D^i_{q+1,k}\}$ выполняется с помощью элемента $Z^i_{q,q+1}$, который должен быть достаточно информативен для решения указанной задачи. Исходя из выше изложенного, взаимодействие иерархических структур S_D и S_z представлено на рис. 1.

На рис. 1 показано три типа связей между элементами иерархических структур S_D и S_z :

- отношения вхождения (показана сплошными линиями), которое устанавливается между родителями и потомками структуры диагнозов S_D на смежных уровнях иерархии;

- причинно-следственная связь (показана пунктирными линиями), между элементами структур S_D и S_z i -го уровня иерархии и которая используется для дифференциальной диагностики соответствующей пары диагнозов $\{D^i_{1,k}, D^i_{i+1,k}\}$ с помощью элемента $Z^i_{1,i+1}$;

- последовательность реализации (показана штрих-пунктирными линиями), которая указывает на связь между элементами структуры S_z на смежных

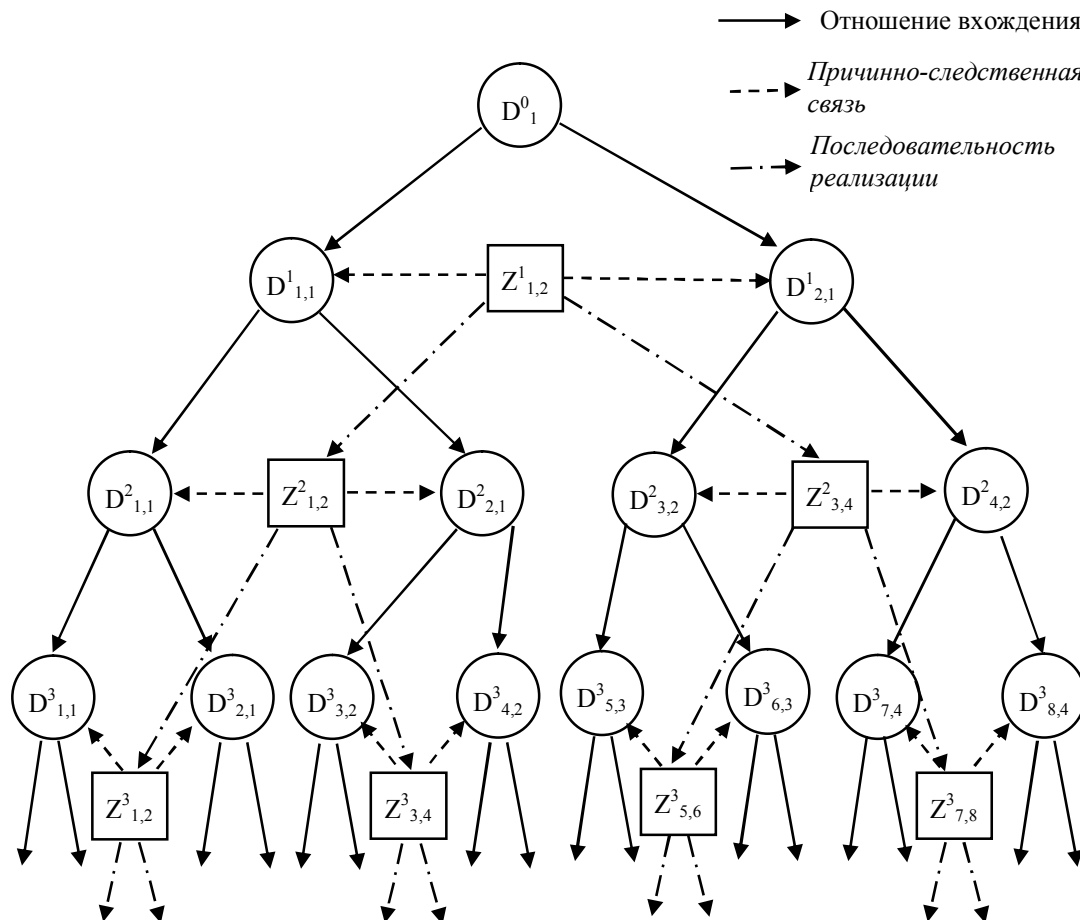


Рис. 1. Типы взаимодействий элементов иерархических структур S_D и S_z

ных уровнях иерархии и используется для формирования оптимального плана обследования конкретного пациента.

Рассмотрим метод формирования элементов структуры S_z , которые должны формироваться из элементов структуры S_y , и удовлетворять следующим требованиям:

– обеспечить требуемое значение коэффициента информационной полноты по (11) для дифференциальной диагностики между элементами S_D на всех уровнях иерархии;

– обеспечить максимальное значение коэффициента диагностической ценности по (12).

Формирование структуры S_z начинается с элемента верхнего уровня, т.е. с элемента $Z^1_{1,2}$, который должен обеспечить дифференциальную диагностику между состояниями $D^1_{1,1}$ и $D^1_{2,1}$.

Вначале рассчитывается энтропия диагнозов $D^1_{1,1}$ и $D^1_{2,1}$ по (4) и необходимая информативность элемента $Z^1_{1,2}$ по (11) при заданном коэффициенте информационной полноты. Следует отметить, что информативность любого элемента структуры S_z не превышает 1 бита и достигает своего максимального значения в случае равновероятных диагнозов в соответствии с (4).

Далее рассчитываются коэффициенты диагностической ценности всех элементов структуры S_y относительно пары диагнозов $D^1_{1,1}$ и $D^1_{2,1}$ по (10) и элементы структуры S_y упорядочиваются в соответствии с

$$k_{\text{дц}}(y_i) \geq k_{\text{дц}}(y_j) \geq \dots \geq k_{\text{дц}}(y_q). \quad (14)$$

Для каждого элемента структуры S_y рассчитывается коэффициент информационной полноты относительно пары диагнозов $D^1_{1,1}$ и $D^1_{2,1}$, после чего выполняется последовательная процедура включения элементов y_i из упорядоченной последовательности (14) в элемент $Z^1_{1,2}$ до тех пор, пока не будет достигнут необходимый уровень информационной полноты элемента $Z^1_{1,2}$, определяемого по (9).

Для формирования элементов следующего уровня иерархии $Z^2_{1,2}$ и $Z^2_{3,4}$ из структуры S_y исключаются те элементы, информативность которых исчерпана при формировании $Z^1_{1,2}$, то есть формируются структуры $S_{y^2_{1,2}}$ и $S_{y^2_{3,4}}$ следующим образом.

Для всех элементов S_y , включенных в $Z^1_{1,2}$ формируются вектора пометок, в которых 1 отмечаются те диапазоны, которые подтверждают соответствующий диагноз ($D^1_{1,1}$ и $D^1_{2,1}$), то есть вектора пометок формируются для каждой из структур $S_{y^2_{1,2}}$ и $S_{y^2_{3,4}}$.

Если вектор пометок элемента y_j содержит только одну единицу (такая ситуация всегда возникает у простых дихотомических признаков) то информативность данного признака исчерпана, и он исключается из соответствующей структуры $S_{y^2_{1,2}}$ и $S_{y^2_{3,4}}$.

Оставшиеся элементы $S_{y^2_{1,2}}$ и $S_{y^2_{3,4}}$ используются для формирования соответствующих элементов $Z^2_{1,2}$ и $Z^2_{3,4}$ по приведенной выше методике для дифференциальной диагностики элементов S_D второго уровня иерархии.

Процедура продолжается до достижения последнего уровня иерархии S_D .

Таким образом, каждый элемент структуры S_z является упорядоченным подмножеством элементов структуры S_y и однозначно задает их порядок включения в диагностическую процедуру уточняющего диагноза на каждом этапе реализации.

Выводы

1. Предлагается метод реконфигурации структуры диагностических признаков по критерию информативности в методе постановки уточняющего диагноза в интеллектуальных компьютерных системах медицинской диагностики.

2. Применение метода позволяет реализовать оптимальную схему обследования пациентов и снять ограничения на размерность пространства признаков.

Список литературы

1. Кобринский Б.А. Принципы математико-статистического анализа данных медико-биологических исследований // *Российский вестник перинатологии и педиатрии*. – 1996. – № 4. – С. 60-64.
2. Ахутин В.М., Шаповалов В.В., Иоффе М.О. Оценка качества формализованных медицинских документов // *Медицинская техника*. – М., 2002. – № 2. – С. 27-31.
3. Поворознюк А.И. Метод постановки уточняющего диагноза в компьютерных системах медицинской диагностики при иерархической структуре диагностических признаков // *Зб. наук. пр. Харківського університету Повітряних Сил*. – Х: ХУПС, 2006. – Вып. 3 (9). – С. 125-130.
4. Поворознюк А.И. Синтез иерархической структуры диагностических признаков в компьютерных системах медицинской диагностики // *Вестник НТУ "ХПИ"*. Сб. науч. тр. – Х: НТУ "ХПИ", 2003. – № 7, т. 2. – С. 39-44.
5. Поворознюк А.И. Синтез иерархической структуры решающих правил в компьютерных системах медицинской диагностики // *Моделювання та інформаційні технології*. Зб. наук. пр. Інституту проблем моделювання в енергетиці. – К.: ІПМЕ, 2004. – Вып. 25. – С. 170-174.
6. Постнова Т.Б. Информационно-диагностические системы в медицине. – М.: Наука, 1972. – 376 с.
7. Букатова И.Л. Эвоинформатика: состояние и перспективы // *Эволюционная информатика и моделирование*. – М.: Наука. – 1994. – С. 18-29.
8. Гуторова Т.В., Поворознюк А.И. Оценка информативности диагностических показателей в компьютерных системах медицинской диагностики // *Системы обработки информации*. – Х.: ХВУ, 2004. – Вып. 7 (35). – С. 162-168.

Поступила в редколлегию 31.01.2007

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