

УДК 681.5

А.И. Поворознюк

Национальный технический университет «ХПИ», Харьков

## КОНЦЕПТУАЛЬНАЯ МОДЕЛЬ ОБЪЕКТА ДИАГНОСТИКИ В КОМПЬЮТЕРНЫХ СИСТЕМАХ МЕДИЦИНСКОЙ ДИАГНОСТИКИ

*Для синтеза компьютерного диагноза приводится концептуальная модель объекта диагностики (подсистем организма человека), которая учитывает сложность взаимодействия подсистем друг с другом и окружающей средой. Разработана диагностическая процедура, которая минимизирует риск недостоверных решений, а при их возникновении, позволяет получать достоверные решения на максимально возможном уровне иерархии диагнозов.*

*медицинская диагностика, модель, дерево решений*

### Введение

**Постановка проблемы и анализ литературы.** Синтез решающего правила (РП), согласно которому выполняется процесс диагностики пациентов для заданного множества диагнозов, предполагает разработку концептуальной модели объекта диагностики (ОД) – диагностируемых подсистем организма. Как отмечается в [1, 2], применение традиционных математических методов (детерминистических, вероятностных, логико-лингвистических и др.) для синтеза РП в медицине является не совсем корректным, так как применяемые методы рассматривают модель ОД в виде "черного ящика" и не учитывают сложность и многоуровневый принцип взаимодействия подсистем организма. Кроме того, применяемые методы являются критическими к объему обучающей выборки и имеют ограничения на размерность ОД при их практической реализации.

В [3] выделено 15 уровней взаимодействия подсистем организма друг с другом и внешней среды – от 0-го уровня генетического генератора полевого взаимодействия до 15-го – зоны контакта и интерфейса с природной и антропо-экологической средой. В настоящее время медицина может получить неполные медицинские данные, отражающие динамику взаимодействий уровней 0 – 9. При внутреннем и внешнем взаимодействии происходит обмен веществом, энергией и информацией [3] с целью поддержания необходимого уровня информационно-энергетического гомеостаза. В случае выраженного отклонения от должного гомеостаза, срабатывают сигнальные системы организма и появляются ощущения дискомфорта – субъективные симптомы наличия патологий. Таким образом, диагноз конкретного пациента должен отражать нарушения взаимодействия между различными уровнями подсистем организма при различных состояниях внешней среды.

**Целью статьи** является разработка концептуальной модели ОД, которая учитывает многоуровневый иерархический принцип взаимодействия подсистем организма друг с другом и окружающей сре-

дой и разработка РП на этой модели.

**Разработка концептуальной модели ОД.** В формализованном виде постановка диагноза пациенту с помощью компьютерной системы сводится к задаче определения принадлежности текущего состояния организма или отдельной его подсистемы к одному из формализованных состояний из множества диагнозов  $\{D_i\}$ . При этом анализируется принятая в медицинской практике система диагностических признаков  $X_i$ , которая, в какой-то мере отражает текущее  $j$ -е состояние  $i$ -й подсистемы организма  $S_{ij}$ . Таким образом, внутренние состояния  $S_{ij}$  различных уровней иерархии проецируются на пространство признаков  $X_i$ , а с другой стороны,  $S_{ij}$  проецируются на систему диагнозов  $\{D_i\}$ , при этом задачей диагностики является определение зависимости  $X_i \rightarrow \{D_i\}$ , как показано на рис. 1.

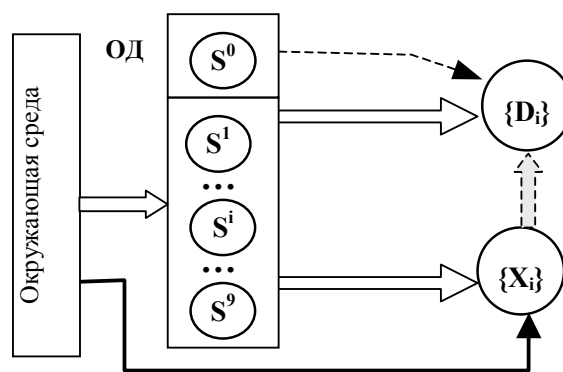


Рис. 1. Схема связей состояний – признаки – диагнозы

Рассмотрим особенности концептуальной модели ОД и характер связей состояний – признаки – диагнозы.

На рис. 1 фигурными стрелками отмечено отображение ОД (состояние подсистем уровней 1 – 9) на пространство диагнозов и пространство признаков. Следует отметить, что некоторое формализованное состояние диагноза может соответствовать нарушению гомеостаза на нескольких уровнях взаимодействия, и даже достоверный диагноз не всегда отражает причину такого нарушения (диагноз ставится

на уровне патологии органов и систем – 7 и 8 уровни, а причиной являются более низкие уровни: биомолекулярный – 4-й, клеточный – 5-й, тканевый – 6-й).

Аналогичным образом формируется диагностическое пространство признаков, компоненты которого зависят от методики измерения и используемой медицинской аппаратуры, и соответствуют разным уровням взаимодействия. Кроме того, при постановке диагноза учитывается генетическая предрасположенность пациента к некоторым типам патологий (связь между нулевым уровнем взаимодействия и диагнозом показана на рис. 1 пунктирной линией). Влияние состояния окружающей среды на множество диагностических признаков показано

жирной линией – учет при постановке диагноза метеозависимых факторов.

При постановке уточняющего диагноза [4] следует различать классы предварительных диагнозов разных уровней детализации и классы окончательного диагноза. Показатели качества классификации зависят от расположения эллипсоидов рассеяния объектов, принадлежащим указанным классам. С учетом отмеченных выше особенностей ОД можно констатировать то, что эллипсоиды рассеяния классов окончательного диагноза имеют более тонкую структуру, состоящую из подмножеств внутренних состояний подсистем разных уровней взаимодействия, как показано на рис. 2.

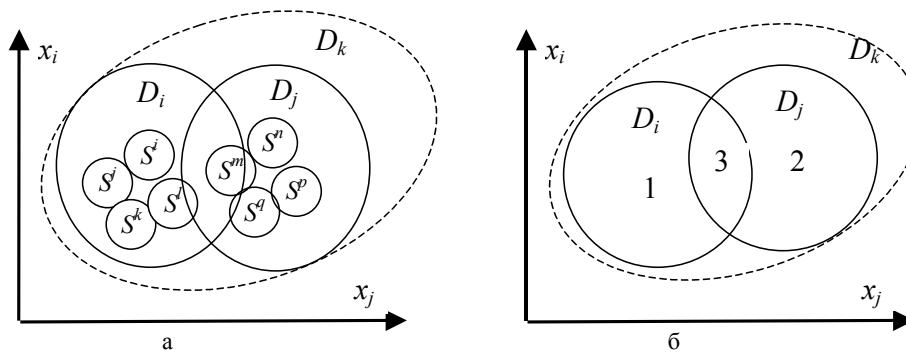


Рис. 2. Иллюстрация расположения эллипсоидов рассеяния классов в пространстве признаков

Рис. 2, а иллюстрирует иерархию внутренних состояний подсистем и диагнозов разных уровней взаимодействия. Диагнозы одного уровня ( $D_i$  и  $D_j$ ) входят в кластер диагноза более высокого уровня  $D_k$ , а каждый из них включает множество внутренних состояний, причем некоторые состояния могут одновременно включаться в несколько диагнозов (в примере состояние  $S^m$  одновременно включается в диагнозы  $D_i$  и  $D_j$ ), что приводит к пересечению эллипсоидов рассеяния. В случае пересекающихся эллипсоидов рассеяния, в каждом эллипсоиде можно условно выделить область достоверного решения (ошибка классификации ниже некоторого порога) и область недостоверного решения. На рис. 2, б цифрами 1, 2 выделены области достоверного решения для состояний  $D_i$  и  $D_j$  соответственно, а цифрой 3 – область недостоверного решения. Различные варианты РП минимизируют область 3 по различным критериям (средний риск, минимаксный, максиминный и т.д. [5]).

В теории равновесия [6] определены понятия сильных (достигающих глобального экстремума) и слабых (достигающих локального экстремума) решений. Применительно к рассматриваемой задаче, определим сильное решение  $Y = f(X)$  таким, при котором достигается глобальный экстремум некоторого критерия  $R_{\{D\}}$  (примем для определенности достижение максимума)

$$R_{\{D\}} = \max_{A_i} R_{\{D\}}(Y^*, X^*), \quad (1)$$

где  $\{D\}$  – множество диагностируемых состояний;

$A_i$  – варианты РП;

$Y^*, X^*$  – точки обучающей выборки.

В качестве диагностируемых состояний  $\{D\}$  рассматривается тот уровень состояния ОД, который формализован как диагноз или кластер диагнозов различных уровней иерархии диагнозов. Другими словами, при получении сильного решения, система должна установить диагноз с точностью до заданного уровня иерархии диагнозов и обеспечить при этом значение  $R_{\{D\}}$  выше заданного порога  $T$  ( $R_{\{D\}} \geq T$ ).

Слабым решением  $Y'$  является такое решение, при котором  $R_{\{D\}}$  не достигает  $T$ . Если в РП используется итерационная процедура оптимизации  $R_{\{D\}}$  (например, метод Вальда, при котором последовательно увеличивается число анализируемых признаков), то процесс постановки диагноза на  $i$ -м уровне иерархии диагнозов является последовательностью слабых решений, приводящей к сильному решению

$$Y'_{i0} \rightarrow Y'_{i1} \rightarrow \dots \rightarrow Y'_{ij} \rightarrow \dots \rightarrow Y'_{in} = Y_i, \quad (2)$$

где  $Y'_{ij}$  – слабое решение  $i$ -го уровня иерархии диагнозов на  $j$ -м шаге итерационной процедуры;  $Y_i$  – сильное решение  $i$ -го уровня иерархии диагнозов.

Таким образом, следует различать сильные решения на разных уровнях иерархии диагнозов, а процесс постановки уточняющего диагноза является последовательностью сильных решений при повышении уровня иерархии диагнозов

$$Y_0 \rightarrow Y_1 \rightarrow \dots \rightarrow Y_i \rightarrow \dots \rightarrow Y_m, \quad (3)$$

где  $Y_0$  – тривиальное сильное решение нулевого уровня;

$Y_i$  – сильное решение на  $i$ -м шаге уточняющего диагноза (диагностика пациента с точностью до кластера  $i$ -го уровня);

$Y_m$  – сильное решение на последнем  $m$ -м шаге (диагностика пациента с точностью до диагноза).

Однако, с учетом рис. 2, б, на каждом  $i$ -м шаге последовательности (3) для конкретного пациента последовательность слабых решений (2) может привести к сильному решению (точка попадает в область 1 или 2), а может не привести (точка попадает в область 3). В работе ставится задача разработки такой диагностической процедуры, при которой не только минимизируется риск слабых решений, но и при их возникновении, получается сильное решение на максимально возможном уровне иерархии. Такой диагноз системы позволит врачу сделать один из следующих выводов: 1) множество диагностических признаков недостаточно и необходимо провести

дополнительное обследование; 2) патология находится на ранней стадии развития; 3) у пациента наблюдается несколько патологий.

На рис. 2, б показано, что если решение при дифференциальной диагностике диагнозов  $D_i$  и  $D_j$  является слабым (точка принадлежит области 3), то всегда можно найти такой кластер  $D_k = D_i \cup D_j$  для которого решение является сильным. Данное утверждение справедливо для любого уровня иерархии решений, то есть, если на  $i$ -м уровне иерархии для конкретного пациента получено слабое решение, то всегда можно найти кластер сильного решения на  $(i - 1)$ -м уровне. Таким образом, необходимо разработать метод поиска сильного решения  $(i - 1)$ -го уровня при наличии слабого решения  $i$ -го уровня.

**Диагностическая процедура уточняющего диагноза с расщеплением слабых решений.** Рассмотрим часть бинарного дерева иерархии диагнозов, которая представлена на рис. 3.

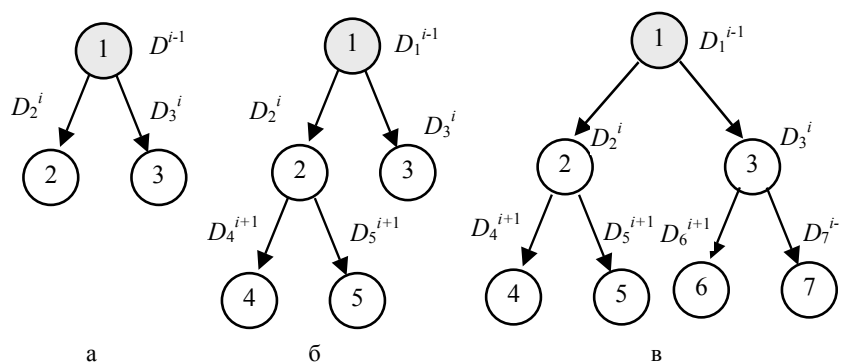


Рис. 3. Примеры иерархии диагнозов (основное дерево решений)

Предположим, что на  $(i - 1)$ -м уровне (вершина 1) получено сильное решение (данное утверждение всегда выполняется для нулевого уровня), но при дальнейшей детализации диагноза (дифференциальная диагностика вершин 2 и 3) решение слабое.

На рис. 3 показаны возможные варианты потомков анализируемой вершины 1 (кластер  $D^{i-1}$ ). На рис. 3, а анализируемая вершина является родителем (кластером) двух окончательных диагнозов (вершины 2 и 3), на рис. 3, б – одного окончательного (вершина 3) и одного предварительного (вершина 2), а на рис. 3, в – двух предварительных (вершины 2 и 3). Вариант рис. 3, а является тривиальным – окончательным диагнозом является сильное решение указанной вершины (кластер  $D^{i-1}$ ). Проанализируем причину слабых решений при детализации вершины 1 в вариантах рис. 3, б и рис. 3, в, эллипсоиды рассеяния которых показаны на рис. 4 (рис. 4, а соответствует варианту рис. 3, б, а рис. 4, б соответствует варианту рис. 3, в).

Следует отметить, что кластеризация вершин 2 и 3 на рис. 4, а и 4, б проводится наилучшим образом в смысле критерия  $R_{\{D\}} = R_{\{2,3\}}$ , который определяется по (1), т.е. область пересечения вершин 2

и 3  $\Omega_{23}$  является минимальной при всех возможных вариантах разбиения.

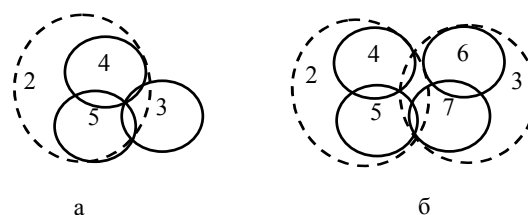


Рис. 4. Расположение эллипсоидов рассеяния иерархических структур

Однако причиной недостоверности решения является попадание точки, которая характеризует исследуемого пациента, в область пересечения вершин 2 и 3, то есть  $Y' \in \Omega_{23}$ . Рассмотрим структуру  $\Omega_{23}$  рис. 4, а. Вершина 2 является кластером вершин 4 и 5, поэтому  $\Omega_{23}$  является объединением  $\Omega_{34}$  и  $\Omega_{35}$ , т.е.  $\Omega_{23} = \Omega_{34} \cup \Omega_{35}$ . Так как  $Y' \in \Omega_{23}$ , то  $\Omega_{23}$  является не пустым множеством, поэтому можно утверждать, что хотя бы одна из областей  $\Omega_{34}$  или  $\Omega_{35}$  является не пустым множеством и предположить, что точка принадлежит одной из них (вероятность пересечения трех вершин  $\Omega_{123}$  и попадание точки в дан-

ную область на порядок ниже и ее можно не учитывать). Величина области пересечения двух эллипсоидов рассеяния в признаковом пространстве определяет суммарную ошибку дифференциальной диагностики между отмеченными вершинами, которая, в свою очередь ухудшает критерий кластеризации между ними. Из сказанного следует, что наиболее вероятной причиной слабого решения рис. 4, а является попадание точки в ту из областей  $\Omega_{34}$  или  $\Omega_{35}$ , которая получена при минимальном критерии кластеризации

$$Y' \in \Omega_{ij} = \operatorname{argmin} \{R_{\{3,4\}}, R_{\{3,5\}}\}. \quad (4)$$

После определения  $\Omega_{ij}$  по (4), начиная от вершины 1, строится альтернативное дерево решений, в котором вершины  $i$  и  $j$  объединяются в один кластер (на рис. 4, а объединяются 3 и 5 вершины). При этом с вероятностью  $P = \Omega_{ij} / \Omega_{23}$  слабое решение при детализации вершины 1 на альтернативные потомки становится сильным, а слабым решением альтернативного дерева становится детализация вершины 2<sub>а</sub>, которая находится на  $(i - 1)$ -м уровне (расщепление слабого решения и его смещение на один уровень вниз) и к ней применяется рассмотренная ранее процедура (анализ вершины 1). Аналогичным образом строится альтернативное дерево для варианта рис. 3, в, но в данном случае выражение (4) записывается в виде:

$$Y' \in \Omega_{ij} = \operatorname{argmin} \{R_{\{4,6\}}, R_{\{4,7\}}, R_{\{5,6\}}, R_{\{5,7\}}\}. \quad (5)$$

Примеры построения альтернативных деревьев решений для вариантов рис. 3, б и 3, в представлены на рис. 5 (рис. 5, а соответствует варианту рис. 3, б, а рис. 5, б соответствует варианту рис. 3, в).

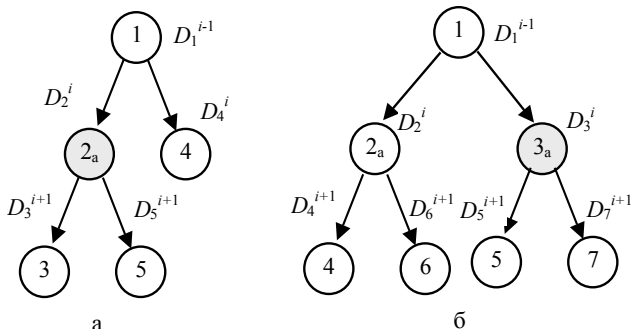


Рис. 5. Примеры альтернативных деревьев решений

Сравнительный анализ рис. 3 и рис. 5 показывает, что основное и альтернативное деревья решений отличаются на  $i$ -м уровне (добавляется два альтернативных потомка), остальные уровни совпадают. Оценим повышение трудоемкости реализации РП при использовании альтернативного дерева решений. Если диагностика выполняется на  $N$  состояний ( $N$  окончательных диагнозов), то основное бинарное дерево решений содержит  $N - 1$  вершин  $V_i$ , которые соответствуют предварительным диагнозам всех уровней иерархии.

Для каждой вершины основного дерева решений  $V_i$ , формируется матрица условных вероятностей

информативных признаков относительно диагностируемых потомков. В каждой вершине  $V_i$  для расщепления слабого решения используются альтернативные потомки, для дифференциальной диагностики которых необходимо дополнительно сформировать матрицу условных вероятностей. Исключив тривиальные случаи рис. 3, а, число которых равно  $N/2$ , окончательно получим число дополнительных матриц условных вероятностей  $N_d = N/2 - 1$ .

Если в анализируемой вершине выполняется дифференциальная диагностика основных потомков с суммарной ошибкой классификации  $\gamma = \alpha + \beta$ , то вероятность сильных решений  $P_0 = 1 - \gamma$ , а вероятность слабых решений  $P_1 = \gamma$ . Таким образом, существенная доля пациентов диагностируется при анализе основного дерева решений, и только незначительная часть – при переходе на альтернативное дерево решений.

## Выводы

Разработана концептуальная модель ОД, которая учитывает многоуровневый иерархический принцип взаимодействия подсистем организма друг с другом и окружающей средой. Проанализированы причины появления слабых решений при постановке уточняющего диагноза и разработано РП, которое в случае появления слабого решения осуществляет переход на альтернативное дерево решений, в котором выполняется расщепление слабого решения и его смещение вниз на один уровень иерархии диагнозов.

## Список литературы

1. Ахутин В.М., Шаповалов В.В., Иоффе М.О. Оценка качества формализованных медицинских документов // *Медицинская техника*. – М.: 2002. – № 2. – С. 27-31.
2. Поворознюк А.И. Формализация этапов проектирования интеллектуальных компьютерных систем медицинской диагностики // *Электронное моделирование*. – 2006. – Т. 28, № 1. – С. 85-97.
3. Весненко А.И., Попов А.А., Проненко М.И. Типология структуры развернутого клинического диагноза в современных медицинских информационных системах и технологиях // *Кибернетика и системный анализ*. – 2002. – № 6. – С. 143-154.
4. Поворознюк А.И. Метод постановки уточняющего диагноза в компьютерных системах медицинской диагностики при иерархической структуре диагностических признаков // *Збірник наукових праць Харківського університету Повітряних Сил*. – Х: ХУПС, 2006. – Вып. 3(9). – С. 125-130.
5. Айвазян С.А., Бухитабер В.М., Енюков И.С. Прикладная статистика: Классификация и снижение размерности. – М.: Финансы и статистика, 1989. – 607 с.
6. Смоляков Э.Р. Равновесные модели при несовпадающих интересах участников. – М.: Наука, 1986. – 223 с.

Поступила в редколлегию 12.12.2007

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