

# Інформаційні технології в медицині

УДК 004.891.3+681.5

К.А. Белецкий, А.И. Поворознюк, О.А. Поворознюк

Национальный технический университет «ХПИ», Харьков

## ПОСТРОЕНИЕ ДЕРЕВА РЕШЕНИЙ ПРИ КОМПЛЕКСНОЙ ОЦЕНКЕ ЛЕЧЕБНО-ДИАГНОСТИЧЕСКОГО ПРОЦЕССА

*В работе рассматривается метод синтеза дерева решений при комплексной оценке рисков на этапах диагностики и медицинской реабилитации в системах информационной поддержки лица принимающего решения. Рассмотрены основные этапы обработки данных. Отмечено, что при синтезе бинарного дерева решений методом иерархической кластеризации диагностируемых состояний в признаковом пространстве, минимизируются только ошибки этапа диагностики. Для минимизации комплексной ошибки обоих этапов в работе предлагается перейти из признакового пространства в пространство фармакологических действий и выполнять иерархическую кластеризацию в указанном пространстве. Для этого разработаны необходимые метрика и критерий кластеризации.*

**Ключевые слова:** диагностика, реабилитация, лицо принимающее решения, дерево решений, отношение правдоподобия, решающее правило, критерий кластеризации, пространство фармакологических действий.

### Введение

**Постановка проблемы.** Процесс реабилитации пациентов состоит из двух взаимосвязанных этапов: диагностики заболеваний и непосредственного лечения выявленных патологий. Между указанными этапами нет четкой границы, потому что после установления диагноза и назначения лечебных процедур необходим мониторинг процесса реабилитации, то есть диагностики текущего состояния пациента с целью оценки эффективности процесса лечения, при необходимости, его коррекции [1, 2].

На сегодняшний день проблема комплексной оценки этапов диагностики и медицинской реабилитации в системах информационной поддержки является актуальной, т.к. несмотря на широкий спектр компьютерных систем, применяемых на этапе диагностики заболевания, на этапе медикаментозной реабилитации поддержка ограничивается медицинскими справочниками фармацевта [3].

**Цель работы.** Разработка метода синтеза дерева решений при комплексной оценке рисков, возникающих на этапах диагностики и медицинской реабилитации.

### Принятие решений в медицинских диагностических системах

Структурная схема системы информационной поддержки лица принимающего решения (ЛПР) изображена на рис. 1.

На первом этапе происходит получение и предварительная обработка набора диагностических признаков  $X_1$  (синдрома), после чего выполняется

постановка развернутого диагноза (преобразование  $X_i \rightarrow D_i$ ). Данная схема изображает систему с верификацией промежуточных данных, при этом ЛПР подтверждает или корректирует полученные результаты (преобразования  $D_i \rightarrow D_i^V$  и  $Y_i \rightarrow Y_i^V$ ).

Для лечения того или иного заболевания необходимо оказание определенных воздействий на организм (во многих областях медицины данные воздействия выполняются медикаментозным путем). Поэтому на следующем этапе выполняется переход от поставленного диагноза к некоторому множеству необходимых фармакологических действий (ФД)  $\{f_i\}_g$ . На основе множества  $\{f_i\}_g$  формируется определенный медикаментозный терапевтический комплекс  $Y_i$ , т.е. набор фармакологических препаратов  $u_i$ , в совокупности охватывающих все соответствующее множество ФД. Для синтеза решающего правила (РП) для преобразования  $X_i \rightarrow D_i$  чаще всего используется вероятностное РП, представляющее собой отношение правдоподобия:

$$\Theta = \prod_i \frac{P(x_i / D_q)}{P(x_i / D_1)}; \quad (1)$$

$$\Theta > A \rightarrow D_q; \Theta < B \rightarrow D_1, \quad (2)$$

где  $A$  и  $B$  – соответственно верхняя и нижняя границы неопределенности, которые необходимы для принятия решения.

Особенностями данного РП является то, что оно может работать только с независимыми признаками и выполняет дифференциацию объектов диа-

гностики (пациентов) всего лишь на 2 класса. Как следствие, применение данного РП требует предварительного синтеза дерева решений, где на каждом

уровне присутствуют два класса  $D_q$  и  $D_l$  [4,5].

Пример такого дерева решений изображен на рис. 2.

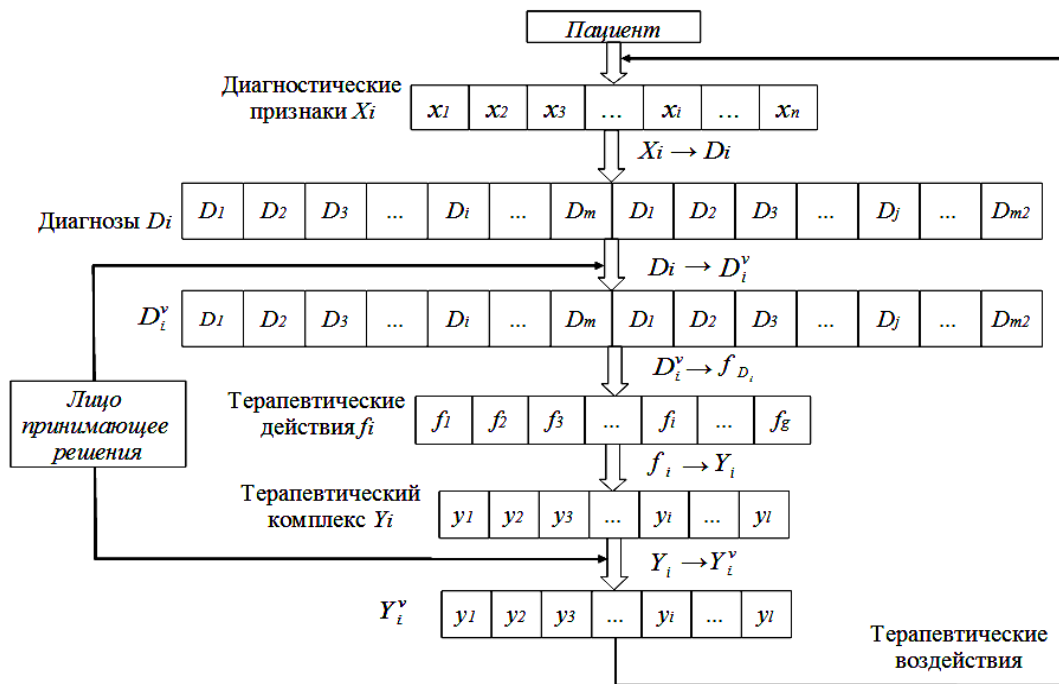


Рис. 1. Информационная поддержка ЛПР

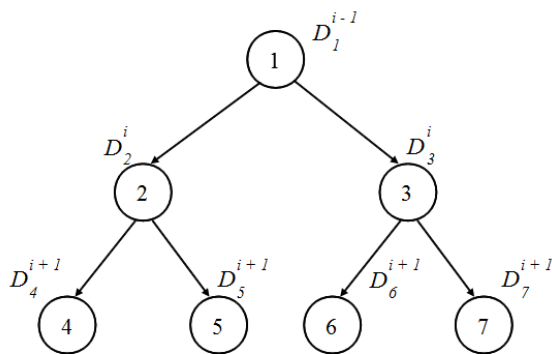


Рис. 2. Дерево решений, построенное классическим способом

Синтез данного дерева решений выполняется линейными методами, например, с использованием последовательной дивизивной (разделительной) кластеризации. Иерархическая кластеризация представляет собой итеративный процесс, в ходе которого крупные кластеры разбиваются на более мелкие, а на следующей итерации полученные кластеры, в свою очередь, также предаются разбиению [5]. Как следствие, в ходе такого процесса образуется бинарное дерево, корнем которого является полное множество диагнозов  $\{D_i\}_n$  в заданной предметной области, в ветвях располагаются кластеры диагнозов, близко расположенных друг к другу, а листьями результирующего дерева, собственно, и являются отдельные диагнозы.

Кластеризация всегда выполняется в смысле оптимизации некоторого критерия [6]. В качестве данного критерия чаще всего принимают минимум площади пересечения эллипсоидов рассеивания. Например, для дерева, изображенного на рис. 2 могли бы получиться следующие эллипсоиды рассеивания (рис. 3).

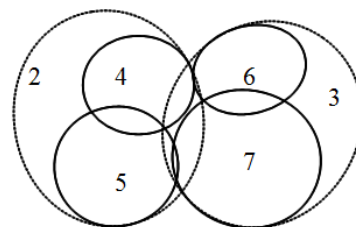


Рис. 3. Пресечения эллипсоидов рассеивания

Более подробный пример иерархической кластеризации и пересечения эллипсоидов рассеивания проиллюстрирован на рис. 4.

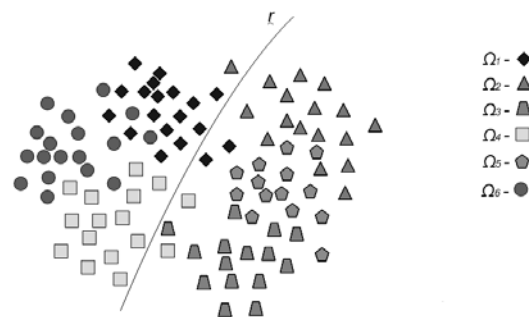


Рис. 4. Более наглядный пример кластеризации

На рис. 4 изображены объекты, принадлежащие 6 классам ( $\Omega_1, \Omega_2, \Omega_3, \Omega_4, \Omega_5, \Omega_6$ ), над которыми производится кластеризация.

Разделение элементов происходит по кривой  $g$  и в результате получаются два кластера  $D_i$  и  $D_j$ , где:

$$D_i = \{\Omega_1, \Omega_4, \Omega_6\};$$

$$D_j = \{\Omega_2, \Omega_3, \Omega_5\}.$$

Классы, принадлежащие одному кластеру, расположены более тесно, и, как следствие, их эллипсоиды рассеивания больше пересекаются между собой, чем с эллипсоидами рассеивания классов из другого кластера. Из-за того, что эллипсоиды рассеивания двух кластеров пересекаются между собой, появляются  $\alpha$ -ошибка (ложная тревога) и  $\beta$ -ошибка (пропуск цели).

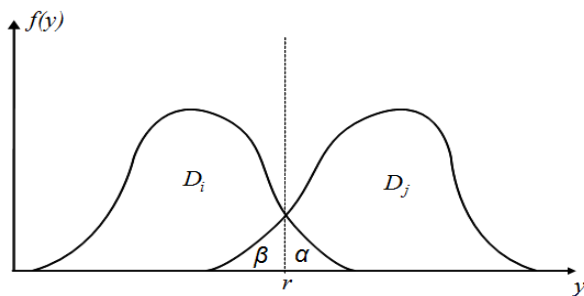


Рис. 5. Ошибки первого и второго рода при пересечении эллипсоидов рассеивания

На рис. 5  $D_i$  и  $D_j$  – кластеры,  $r$  — порог, по которому происходит разделение данных кластеров,  $y$  – некоторая обобщенная координата (ось).  $\alpha$  и  $\beta$  – ошибки первого и второго рода соответственно. Величины ошибок могут быть найдены по следующим формулам [7]:

$$\alpha = \int_r^{\infty} f(y / D_i) dy; \tag{3}$$

$$\beta = \int_{-\infty}^r f(y / D_j) dy. \tag{4}$$

Также при известном распределении можно определить пороговое значение:

$$r = \arg \min_y \{\alpha, \beta\}. \tag{5}$$

Однако, при описанном выше подходе учитывается только расположение эллипсоидов рассеивания в признаковом пространстве, т.е. учитываются ошибки возникающие только на этапе диагностики заболевания. Но ошибки, возникающие в процессе выбора необходимых ФД и последующего назначения терапевтического комплекса, не учитываются. Как было отмечено в цели, в данной статье выполняется синтез дерева диагнозов по критерию минимизации ошибки лечебных процедур.

Для этого необходимо рассмотреть, что такое ФД и какая существует связь между ними и диагнозами.

Фармакологическое действие – это влияние активных компонентов лекарственных препаратов на отдельные органы человека и организм в целом.

С каждым диагнозом может быть связано некоторое число необходимых ФД, и для некоторых диагнозов эти множества необходимых ФД могут пересекаться (рис. 6).

ФД представляются в виде бинарных переменных, то есть они могут принимать только два значения 0 и 1, где 0 – это отсутствие необходимости в оказании данного ФД, а 1 – указывает на необходимость присутствия соответствующего ФД.

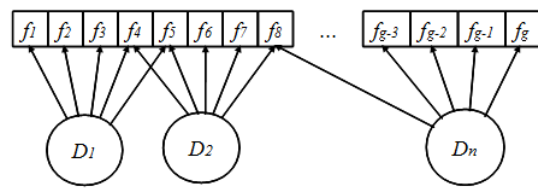


Рис. 6. Связь диагнозов с ФД

Для заданной предметной области, характеризующейся  $n$  диагнозами можно сформировать перечень ФД размерностью  $g$ , и, как следствие, можно построить  $g$ -мерное пространство. Так как ФД являются бинарными величинами, то получено пространство будет иметь форму единичного гиперкуба соответствующей мерности. Тогда в этом пространстве диагнозы будут располагаться в вершинах данного гиперкуба (рис. 7). И в этом пространстве необходимо выполнять иерархическую кластеризацию.

- $D_0 \rightarrow \{\};$
- $D_1 \rightarrow \{f_3\};$
- $D_2 \rightarrow \{f_1, f_2, f_3\};$
- $D_3 \rightarrow \{f_1, f_3\};$

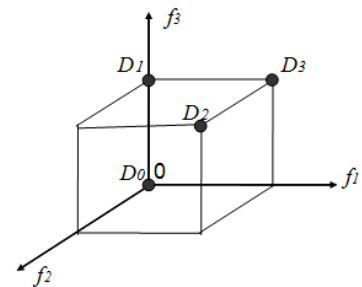


Рис. 7. Трехмерное пространство ФД

Для кластеризации в новом пространстве необходимо выбрать метрику, по которой можно рассчитать критерий. Т.к. ФД являются дихотомическими величинами, то в качестве метрики можно выбрать расстояние Хемминга:

$$r(D_i, D_j) = \sum_{k=1}^g |x_{ki} - x_{kj}|, \tag{6}$$

где  $x_{ki}, x_{kj} \in \{0, 1\}$ ,  $g$  – размерность пространства фармакологических действий.

В случае, если двум диагнозам соответствуют конфликтующие ФД, то выполняется коррекция соответствующего расстояния:

$$r_{ij}' = \omega_{ij} r_{ij}. \quad (7)$$

Таким образом рассматривая диагнозы в пространстве ФД возможно построить полносвязный граф диагнозов. Весами ребер данного графа будут расстояния между диагнозами, полученные с использованием представленной выше метрики (7). А, собственно, кластеризацию можно выполнить любым методом кластерного анализа, причем критерием, по которому будет происходить разделения, будет являться минимизация расстояния между отдельными диагнозами.

### Выводы

В работе предложен метод синтеза дерева решений для дифференциальной диагностики на основе комплексной оценки рисков этапов диагностики и медикаментозного лечения. Предложено выполнять кластеризацию диагнозов в пространстве фармакологических действий. Т.к. фармакологические действия являются дихотомическими величинами, то диагнозы располагаются в вершинах  $n$ -мерного гиперкуба. Для нового пространства были разработаны метрика (основанная на расстоянии Хемминга) и критерий кластеризации.

### Список литературы

1. Поворознюк О.А. Біотехнічна система призначення лікарських препаратів в дерматології : автореф. дис. ... канд. техн. наук : спец. 05.11.17 "Біологічні та медичні прилади і системи" / Поворознюк Оксана Анатоліївна; Харківський нац. університет радіоелектроніки. – Х., 2010. – 23 с.

2. Дмитриенко В.Д. Многокритериальная оценка лекарственных препаратов / В.Д. Дмитриенко, О.А. Поворознюк // Інформаційні технології та комп'ютерна інженерія. – Вінниця, ВНТУ, 2009. – № 3 – С. 144-148.

3. Компендиум 2007 – лекарственные препараты [Электронный ресурс] / Под ред. В.Н. Коваленко, А.П. Викторова. – Режим доступа к ресурсу: <http://www.compendium.com.ua>.

4. Поворознюк А.І. Інформаційні технології підтримки прийняття рішень в медичній діагностиці на основі синтезу структурованих моделей: автореф. дис. ... доктора техн. наук: спец. 05.13.06 "Інформаційні технології" / Поворознюк Анатолій Іванович. – Х., 2011. – 39 с.

5. Поворознюк А.І. Метод постановки уточнюючого діагноза в комп'ютерних системах медичної діагностики при ієрархічній структурі діагностических ознак / А.І. Поворознюк // Збірник наукових праць Харківського університету Повітряних Сил. – Х: ХУПС, 2006. – Вип. 3 (9). – С. 125-130.

6. Поворознюк А.І. Система підтримки прийняття рішень в медицині на основі структурної ідентифікації об'єктів діагностики / А.І. Поворознюк // Сборник научных трудов СХУЯЕуП. – Севастополь: СХУЯЕуП, 2008. – № 1 (25) – С. 234-245.

7. Айвазян С.А. Прикладная статистика: Классификация и снижение размерности / С.А. Айвазян, В.М. Бухштабер, И.С. Енюков. – М.: Финансы и статистика, 1989. – 607 с.

Поступила в редколлегию 6.03.2012

Рецензент: д-р физ-мат. наук, проф. А.И. Бых, Харьковский национальный университет радиоэлектроники, Харьков.

### ПОБУДОВА ДЕРЕВА РІШЕНЬ ПРИ КОМПЛЕКСНІЙ ОЦІНЦІ ЛІКУВАЛЬНО-ДІАГНОСТИЧНОГО ПРОЦЕСУ

К.А. Білецький, А.І. Поворознюк, О.А. Поворознюк

В роботі розглянуті методи синтезу дерева рішень при комплексній оцінці ризиків на етапах діагностики та медичної реабілітації в системах інформаційної підтримки лиця приймаючого рішення. Розглянуті основні етапи обробки даних. Зазначено, що при синтезі бінарного дерева рішень методом ієрархічної кластеризації станів, що діагностуються, в просторі ознак, мінімізуються лише помилки етапу діагностики. Для мінімізації комплексної помилки обох етапів в роботі пропонується перейти із простору ознак в простір фармакологічних дій і виконувати ієрархічну кластеризацію в указаному просторі. Для цього розроблені необхідні метрика та критерій кластеризації.

**Ключові слова:** діагностика, реабілітація, особа, що приймає рішення; дерево рішень, відношення правдоподібності, вирішуючи правило, критерій кластеризації, простір фармакологічних дій.

### BUILDING OF A DECISION TREE USING COMPLEX EVALUATION OF A CLINICAL PROCESS

K.A. Beletskiy, A.I. Povoroznyuk, O.A. Povoroznyuk

In this work the methods for synthesizing a decision tree using a complex assessment of a clinical process in decision support systems are reviewed. The main data processing stages are also reviewed. It is noted that only the errors of diagnosis stage are minimized, when synthesizing a binary decision tree using hierarchical clustering in symptom's area. It's proposed to make hierarchical clustering in pharmacological area instead of the symptoms area. This is necessary for minimizing a complex error of both the diagnosis stage and the medical rehabilitation stage. Also the necessary metric and the clustering criterion are developed.

**Keywords:** diagnosis, rehabilitation, the person making decisions, the decision tree, likelihood ratio, decision rule, clustering criterion, the pharmacological area.