

УДК 681.3

М.В. БУРЦЕВ, НТУ "ХПИ", г. Харьков,
А.И. ПОВОРОВНИЮК, к.т.н., доц. НТУ "ХПИ", г. Харьков

ПОСТРОЕНИЕ ИЕРАРХИЧЕСКОЙ СТРУКТУРЫ ДИАГНОЗОВ ДЛЯ КОМБИНИРОВАННОГО РЕШАЮЩЕГО ПРАВИЛА В КОМПЬЮТЕРНЫХ СИСТЕМАХ МЕДИЦИНСКОЙ ДИАГНОСТИКИ

Предложены варианты построения иерархической структуры диагнозов, используемых комбинированным решающим правилом в процессе постановки уточняемого диагноза в компьютерных системах медицинской диагностики. Выявлены их преимущества и недостатки. Описан способ хранения сформированной структуры в виде знаний системы поддержки принятия решений. Ил.: 4. Библиогр.: 10 назв.

Ключевые слова: иерархическая структура диагнозов, комбинированное решающее правило, уточняемый диагноз, компьютерные системы медицинской диагностики, поддержка принятия решений.

Постановка проблемы и анализ литературы. Несмотря на стремительное развитие информационных технологий, ориентированных на медицину, актуальной оптимизационной задачей остается синтез решающего правила, которое позволит повысить эффективность процесса постановки диагноза. Разработано комбинированное решающее правило (КРП) [1], объединяющее вероятностную составляющую и описание структуры симптомокомплексов с помощью нечеткой логики [2]. В его основе лежит обход дерева диагнозов, в каждом составном узле которого вычисляется отношение правдоподобия. В зависимости от рассчитанного значения принимается решение о наличии соответствующего диагноза (группы диагнозов). Качество работы КРП напрямую зависит от качества предварительно сформированной иерархической структуры диагнозов. Таким образом, актуальной видится проблема организации известных диагнозов таким образом, чтобы эффективность КРП была максимальной.

Необходимая структура иерархии диагнозов имеет вид двоичного дерева [3], каждый простой узел которого содержит конечный диагноз, а составные – две группы диагнозов, относительно которых осуществляется диагностика, а также последовательности наборов диагностически значимых интервалов численных признаков [4]. Интервалы формируются в ходе работы оптимизационной процедуры, выполняющей минимизацию интегральной ошибки аппроксимации теоретического закона распределения гистограммой, и вместе с группами диагнозов участвуют в вычислении априорных условных вероятностей, используемых решающим правилом совместно с функциями принадлежности, заданными экспертами. Сформированная

иерархическая структура является фреймом знаний [5] для системы поддержки принятия решений (СППР) [6], и может быть сохранена в соответствующей таблице базы данных (БД) в виде сериализованного объекта [7].

Целью статьи является описание подходов к формированию иерархической структуры диагнозов для комбинированного решающего правила, выявление их преимуществ и недостатков, а также обоснование способа хранения сформированной структуры.

Структурная декомпозиция (построение иерархии "сверху"). При данном подходе осуществляется последовательное разделение множества состояний (диагнозов) на две группы, до тех пор, пока сформированные группы содержат более одного диагноза (рис. 1). Исходное множество состояний в том числе включает состояние "практически здоров".

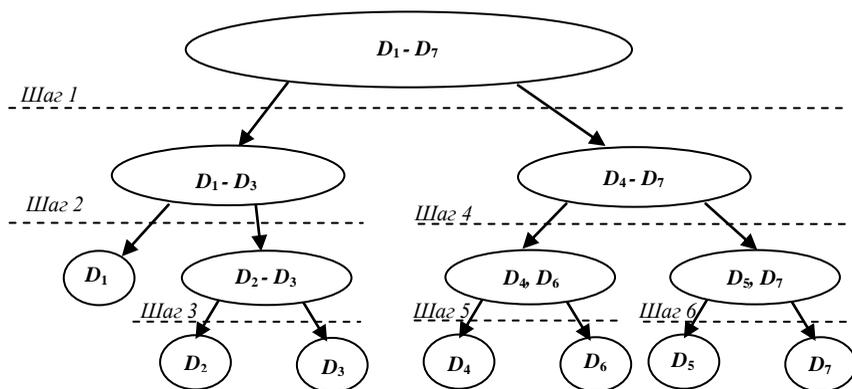


Рис. 1. Пример структурной декомпозиции

Как видно из рис. 1, на первом шаге исходное множество диагнозов $D^0 = \{D_i, i = \overline{1,7}\}$ разбивается на два подмножества $D^1 = \{D_i, i = \overline{1,3}\}$ и $D^2 = \{D_i, i = \overline{4,7}\}$. Затем происходит последовательное разделение подмножеств, сформированных на предыдущем шаге, до тех пор, пока конечные диагнозы не выделяются в отдельные узлы сформированного двоичного дерева.

Формирование подмножеств D^q и D^w (см. рис. 2) на каждом шаге можно выполнять применением генетического алгоритма [8]. Структура гена представляет собой объединение двух частей. Первая содержит

количество диагнозов, включаемых в D^q , вторая – случайным образом сформированную последовательность диагнозов, входящих в разделяемое подмножество. К генам применяются генетические операторы скрещивания (обмен частями между двумя генами), и мутации (качественные и количественные изменения в составах D^q и D^w).

Структура гена и пример его реализации представлены на рис. 2.



Рис. 2. Ген, используемый в структурной декомпозиции:
а) общая структура; б) пример конкретной реализации

Оценка качества гена производится на основании количества правильно классифицированных объектов обучающей выборки с помощью КРП в рамках разделяемого узла.

Данный подход позволяет сократить время получения результата (относительно алгоритма полного перебора), а также задать требуемое качество формируемой структуры. К недостаткам следует отнести существенные вычислительные затраты, связанные с построением диагностически значимых интервалов признаков, используемых КРП, для каждого гена на всех стадиях эволюционного цикла. Таким образом, при программной реализации необходимо обеспечить кэширование формируемых последовательностей интервалов, во избежание повторных вычислений для одних и тех же исходных данных.

Структурная композиция (построение иерархии "снизу"). Построение иерархической структуры начинается с листьев, содержащих конкретные диагнозы (N уровень дерева), при этом состояние "практически здоров" исключается. Для построения $N - 1$ уровня вычисляются вероятности попарных сочетаний диагнозов, представленных на N уровне. Наиболее вероятные пары участвуют в построении $N - 1$ уровня (рис. 3). Аналогичным образом выполняется построение всех вышележащих уровней дерева, путем попарной группировки узлов нижележащего уровня, сочетание которых наиболее вероятно. Таким образом, первый уровень дерева будет включать один

узел, содержащий все известные заболевания. Затем формируется нулевой уровень дерева, также включающий состояние "практически здоров", после чего первый уровень дерева представляет собой разделение состояний "здоров" и "болен". Данный подход позволяет установить множество наиболее вероятных состояний для диагностируемого объекта в том случае, когда в ходе работы решающего правила для некоторого составного узла дерева рассчитанное значение отношения правдоподобия попадает в границы области неопределенности и невозможно установить один конкретный диагноз. Это особенно актуально, если у больного присутствует несколько диагнозов, что само по себе является типичной ситуацией.

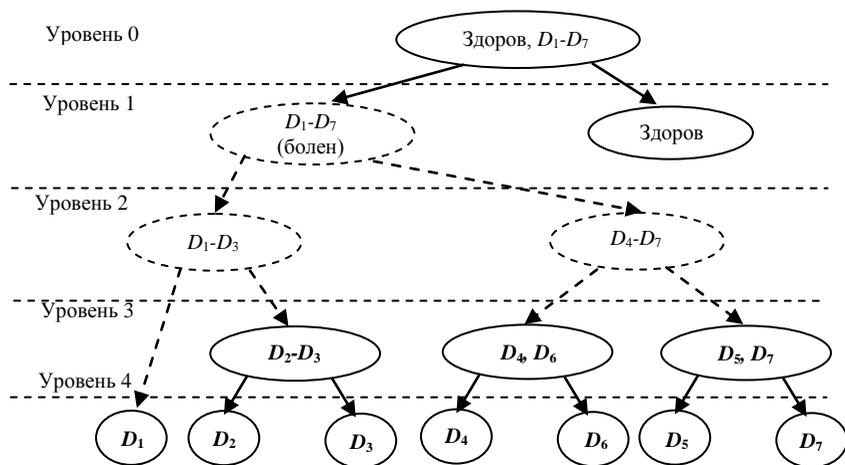


Рис. 3. Пример структурной композиции

Достоинством данного подхода можно считать простоту реализации, а также отсутствие лишних вычислений, связанных с построением интервалов для КРП. Недостатком является риск отнесения объектов, имеющих заболевания, симптомокомплекс которых близок к состоянию "практически здоров", к классу "здоров", либо отказ от диагностирования подобных объектов.

Хранение сформированной иерархической структуры. Полученная иерархическая структура является знаниями для системы поддержки принятия решений. В [9] была рассмотрена программная реализация КРП как части СППР, разрабатываемой на Java платформе. Рассмотренная структура БД, должна быть дополнена таблицей хранения

знаний. Эта таблица связана с таблицей обследований отношением многие-ко-многим, что позволяет использовать одни и те же знания при различных типах обследования (рис. 4).

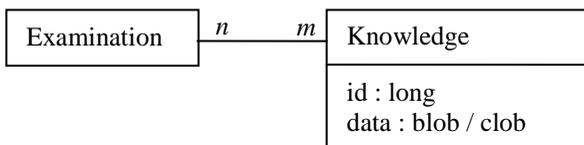


Рис. 4. Дополнение к структуре БД

Поле data таблицы Knowledge предназначено для непосредственного хранения данных. Данные можно сохранить в виде сериализованного объекта, который непосредственно используется КРП. Такими объектами являются экземпляры класса DiagnoseTree [9], сформированные с помощью специальных строителей, содержащие всю информацию, необходимую для комбинированного решающего правила. С помощью стандартных средств они могут быть преобразованы в поток байт и сохранены в виде файлов или записей в таблице БД. При хранении такого представления поле Knowledge.data должно иметь тип blob, что соответствует хранению больших объектов в двоичном представлении. Достоинством такого подхода является то, что необходимые знания после извлечения из базы передаются КРП без предварительной обработки. Недостатком является то, что в случае изменения класса DiagnoseTree, или любой из его зависимостей, десериализация объектов более старых версий будет невозможна.

Альтернативой является хранение XML-описания [10] экземпляров DiagnoseTree. В таком случае в БД сохраняется текстовая информация (Knowledge.data имеет тип clob), содержащая метаданные, а процедуры записи/чтения БД должны сопровождаться маршалингом/демаршалингом, которые можно автоматизировать при помощи существующих библиотек (jaxb, simple-xml).

Выводы. Предложены варианты построения иерархической структуры диагнозов, используемых комбинированным решающим правилом в процессе постановки уточняемого диагноза в компьютерных системах медицинской диагностики. Выявлены их преимущества и недостатки. Описан способ хранения сформированной структуры в виде знаний системы поддержки принятия решений.

Список литературы: 1. Бурцев М.В. Выбор функций принадлежности для описания симптомокомплексов в комбинированном решающем правиле / М.В. Бурцев, А.И. Поворозюк // Вісник НТУ "ХПІ". – Харків: НТУ "ХПІ". – 2010. – № 31. – С. 10 – 15.

2. Бурцев М.В. Синтез комбинированного решающего правила в задаче медицинской диагностики / М.В. Бурцев, А.И. Поворознюк // Вісник НТУ "ХПІ". – Харків: НТУ "ХПІ". – 2009. – № 43. – С. 27 – 33. 3. Кормен Т. Алгоритмы: построение и анализ / Т. Кормен, Ч. Лейзерсон, Р. Ривест, К. Штайн / Под ред. И.В. Красикова. – 2-е изд. – М.: Вильямс, 2005. – 1296 с. 4. Поворознюк А.И. Формирование диагностических интервалов численных признаков при дифференциальной диагностике / А.И. Поворознюк // Вісник Хмельницького національного університету. – Хмельницький: ХНУ, 2007. – № 3. – Т. 1. – С. 106 – 109. 5. Гаврилова Т. Базы знаний интеллектуальных систем / Т. Гаврилова, В. Хорошевский. – СПб.: Питер, 2001. – 584 с. 6. Джарратано Дж. Экспертные системы: принципы разработки и программирование / Дж. Джарратано, Г. Райли. – М.: ООО "И. Д. Вильямс", 2007. – 1152 с. 7. Шилдт Г. Полный справочник по Java / Г. Шилдт. – М.: ООО "И. Д. Вильямс", 2009. – 1040 с. 8. Рутковская Д. Нейронные сети, генетические алгоритмы и нечеткие системы: Пер. с польск. И.Д. Рудинского / Д. Рутковская, М. Пилиньский, Л. Рутковский. – М.: Горячая линия – Телеком, 2006. – 452 с. 9. Бурцев М.В. Программная реализация комбинированного решающего правила для задач медицинской диагностики / М.В. Бурцев, А.И. Поворознюк // Вісник НТУ "ХПІ". – Харків: НТУ "ХПІ". – 2010. – № 21. – С. 11 – 16. 10. Холзнер С. XML. Энциклопедия, 2-е изд. / С. Холзнер. – СПб.: Питер, 2004. – 1101 с.

Статья представлена д.т.н., проф. НТУ "ХПІ" Ободом І.І.

УДК 681.3

Побудова ієрархічної структури діагнозів для комбінованого вирішального правила в комп'ютерних системах медичної діагностики / Бурцев М.В., Поворознюк А.І. // Вісник НТУ "ХПІ". Тематичний випуск: Інформатика і моделювання. – Харків: НТУ "ХПІ". – 2011. – № 36. – С. 29 – 34.

Запропоновано варіанти побудови ієрархічної структури діагнозів, що використовуються комбінованим вирішальним правилом у процесі встановлення уточнюючого діагнозу в комп'ютерних системах медичної діагностики. Виявлено їх переваги та недоліки. Описано спосіб збереження сформованої структури у вигляді знань системи підтримки прийняття рішень. Лл.: 4. Бібліогр.: 10 назв.

Ключові слова: ієрархічна структура діагнозів, комбіноване вирішальне правило, уточнюючий діагноз, комп'ютерні системи медичної діагностики, підтримка прийняття рішень.

UDC 681.3

Diagnoses hierarchy building for combined decision rule used in the medical diagnostics computer systems / Burtsev M.V., Povoroznyuk A.I. // Herald of the National Technical University "KhPI". Subject issue: Information Science and Modelling. – Kharkov: NTU "KhPI". – 2011. – № 36. – P. 29 – 34.

Diagnoses hierarchy building methods used by combined decision rule during refined diagnosis in the medical diagnostics computer systems were offered. Proposed methods advantages and disadvantages were detected. Formed diagnoses hierarchy storage as support decision-making system knowledge was described. Figs.: 4. Refs.: 10 titles.

Keywords: diagnoses hierarchy, combined decision rule, refined diagnosis, medical diagnostics computer systems, support decision-making.

Поступила в редакцію 15.07.2011