

М.В. БУРЦЕВ, асп. НТУ "ХПИ" (г. Харьков),
А.И. ПОВОРОЗНЮК, канд. техн. наук, проф. НТУ "ХПИ" (г. Харьков)

ПРОГРАММНАЯ РЕАЛИЗАЦИЯ КОМБИНИРОВАННОГО РЕШАЮЩЕГО ПРАВИЛА ДЛЯ ЗАДАЧ МЕДИЦИНСКОЙ ДИАГНОСТИКИ

Рассмотрены современные технологии и средства разработки программного обеспечения, выполнен анализ их достоинств и недостатков. Предложен подход к программной реализации комбинированного решающего правила, включающего вероятностный метод Вальда и описание структуры симптомокомплексов. Илл.: 2. Библиогр.: 11 назв.

Ключевые слова: технологии и средства разработки программного обеспечения, комбинированное решающее правило, метод Вальда, симптомокомплекс.

Постановка проблемы и анализ литературы. Динамичное развитие информационных технологий не только затрагивает совершенствование аппаратной части, но и порождает новые подходы к программированию, а также новую методологию разработки и проектирования программных компонентов и комплексов. Процесс создания современного программного продукта шагнул далеко вперед по сравнению со временами консоли и первых оконных приложений. Разработка всей системы "с нуля" давно сменилась использованием библиотек готовых структур данных и типовых алгоритмов их обработки (Standard Template Library для C++ [1], .NET Framework для языковых средств, входящих в состав Microsoft Visual Studio [2], Java Development Kit [3]). В отношении процесса проектирования программного обеспечения (ПО) также произошли существенные изменения в связи с появлением UML [4], который позволяет абстрагироваться от конкретного языка (в рамках объектно-ориентированной парадигмы) и сосредоточится на создании необходимых для разрабатываемого ПО сущностей. В методологическом плане активно развивается предложенная К. Беком техника экстремального программирования [5]. В связи с многообразием существующих средств разработки ПО, возникает проблема выбора оптимального инструментария и методики для реализации системы поддержки принятия решений в медицине в целом и диагностической подсистемы, основанной на комбинированном решающем правиле, в частности.

Целью статьи является обзор современных средств разработки и проектирования программного обеспечения с целью обоснования методики реализации диагностической подсистемы, основанной на комбинированном решающем правиле, включающем вероятностный метод Вальда и описание структуры симптомокомплексов.

Обзор средств разработки ПО. Сегодня существует большое количество языков программирования, их диалектов, а также сред разработки. Наиболее

привлекательными являются те из них, которые позволяют создавать приложения, являющиеся кроссплатформенными на уровне запуска. Лидерами в данной области являются платформы .NET (Microsoft) и Java (Oracle Corporation (ранее Sun Microsystems)). При этом, последние версии .NET Framework доступны только для ОС Windows (в отличие от Java: Oracle/Sun предоставляет реализацию Java Runtime Environment (JRE) как для Windows, так и для NIX-систем, а Apple – для Mac OS X). Для NIX-систем на момент написания статьи в рамках проекта Mono реализована поддержка .NET Framework 2.0 в то время как готовится к выходу версия 4.0 для Windows.

К тому же, коммерческий характер ПО от Microsoft в ряде случаев делает его менее привлекательным: бюджет отечественных медицинских учреждений в большинстве своем не в состоянии обеспечить требуемую сумму для приобретения всего необходимого лицензионного программного обеспечения. Java лишена этого недостатка, являясь свободно распространяемым инструментом.

Учитывая экономическую составляющую и наличие последних версий сред выполнения для большинства современных платформ (Windows, UNIX, Linux, Mac OS X) в качестве средства разработки выбран Java [6].

В качестве IDE выбран Eclipse [7] – свободная интегрированная среда разработки модульных кроссплатформенных приложений. Развивается и поддерживается Eclipse Foundation. Привлекательность Eclipse Platform состоит в том, что существует множество дополнений, позволяющих снизить временные затраты на разработку ПО. Так, компания Soyatec предоставляет проект eUML2.0, дополняющий Eclipse инструментом UML-моделирования.

Обзор средств проектирования ПО. На стадии проектирования использованы возможности UML – языка графического описания для объектного моделирования в области разработки программного обеспечения. UML является языком широкого профиля, это открытый стандарт, использующий графические обозначения для создания абстрактной модели системы, называемой *UML моделью*. UML был создан для определения, визуализации, проектирования и документирования в основном программных систем. UML не является языком программирования, но на основании разработанной модели возможна кодогенерация.

Язык UML является достаточно строгим и мощным средством моделирования, которое может быть эффективно использовано для построения концептуальных, логических и графических *моделей* сложных систем различного целевого назначения. Этот язык вобрал в себя наилучшие качества и опыт методов программной инженерии, которые с успехом использовались на протяжении последних лет при моделировании больших и сложных систем.

Реализация комбинированного решающего правила. При реализации комбинированного решающего правила (РП) [8] необходимо создание классов для реализации сущностей предметной области, а именно:

признак, диагноз, интервал, норма, последовательность интервалов, решающее правило. Эта задача решается с помощью иерархии классов, изображенных на диаграмме (рис. 1).

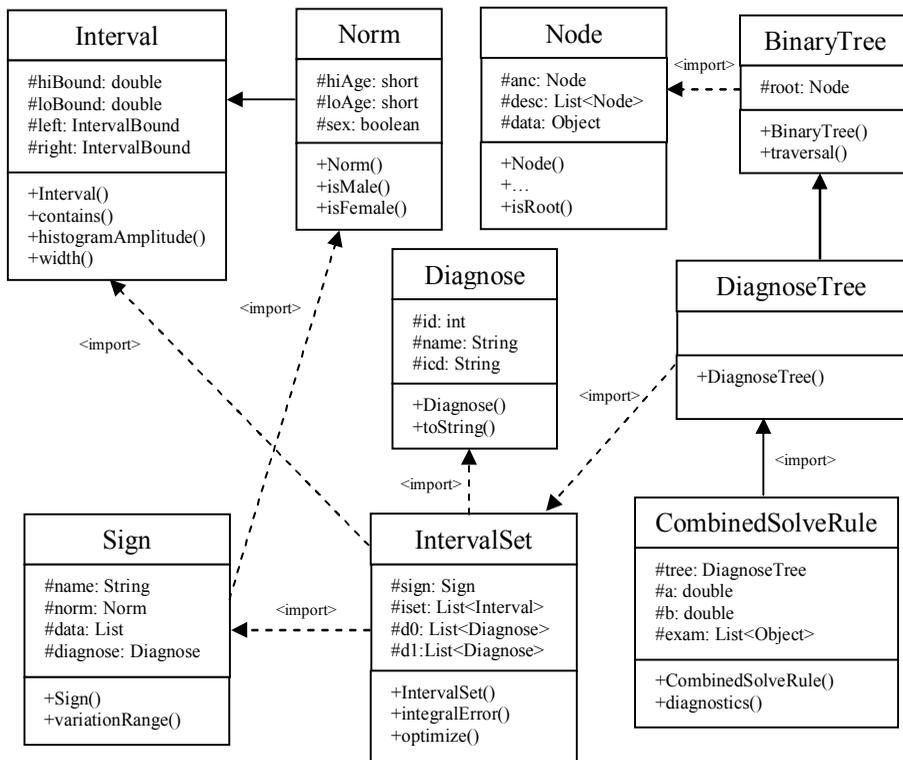


Рис. 1. Диаграмма классов

Основными сущностями являются *признак* и *диагноз*, реализованные классами Sign и Diagnose соответственно. Sign обеспечивает возможность представления в системе выбранного признака с учетом группы диагнозов для подмножества объектов обучающей выборки, хранящихся в базе данных (БД). Остальные сущности импортируют Sign и Diagnose (напрямую, или через другие сущности/классы), осуществляя их обработку: Interval и IntervalSet выполняют разбивку признака на диагностически значимые интервалы [9] с минимизацией интегральной ошибки аппроксимации теоретического закона распределения гистограммой, а CombinedSolveRule – представляет реализацию алгоритма комбинированного решающего правила. CombinedSolveRule

реализует интерфейс `Diagnosics`, который порождает семейство классов, осуществляющих диагностику: возможна реализация детерминистической, информационно-вероятностной логики, или методов распознавания образов, что вместе с подходом позднего связывания, применяемого в Java, позволяет использовать полиморфное создание диагностирующих объектов во время выполнения программы.

Данные обучающей выборки хранятся в MySQL-базе данных [10] (для проектирования использовался `MySQL Workbench`), структура которой приведена на рис. 2.

Таблица `Sign` хранит описание признаков (тип, название или псевдоним) и совместно с `ExaminationContent` используется для построения новых таблиц различных обследований любой структуры. Измеренные значения признаков для объектов диагностики (представленных таблицами `Person` и `Patient`) хранятся в отдельных таблицах, имена которых указаны в поле `Table` таблицы `ExaminationContent`. `SignDiagnose` – представляет структуру симптомокомплексов: поле `SignWeight` указывает на вес признака `idSign` в симптомокомплексе заболевания `idDiagnose`; `FuzzyFunction` – описание лингвистической переменной симптома, а также функций принадлежности [11] на языке `Fuzzy Control Language`, используемый библиотекой `jFuzzyLogic`.

Данные, полученные из таблиц, представляющих обследования, являются основой для объективной составляющей РП (вероятностного подхода), а его субъективная составляющая основывается на данных из таблицы `SignDiagnose`. Обработкой этих сведений занимается класс `CombinedSolveRule`, осуществляющий последовательный анализ набора признаков, находящихся в узлах двоичного дерева диагнозов, представленного классом `DiagnoseTree`.

Разработанная подсистема постановки диагноза обучалась на данных выборки, представленной таблицей `ClinicalBlood` (клинический анализ крови), содержащей 434 записи, включающей 9 признаков, 10 диагнозов.

Связь уровня данных (БД, рис. 2) и уровня логики системы (Java-классы, рис. 1) осуществляется с помощью технологии объектно-реляционного проецирования. Для этого использована библиотека `Hibernate` [7], предоставляющая лёгкий в использовании каркас для отображения объектно-ориентированной модели данных в традиционные реляционные базы данных. `Hibernate` не только решает задачу связи классов Java с таблицами базы данных (и типов данных Java с типами данных SQL), но также предоставляет средства для автоматической генерации и обновления набора таблиц, построения запросов и обработки полученных данных и может значительно уменьшить время разработки, которое обычно тратится на ручное написание SQL- и JDBC-кода.

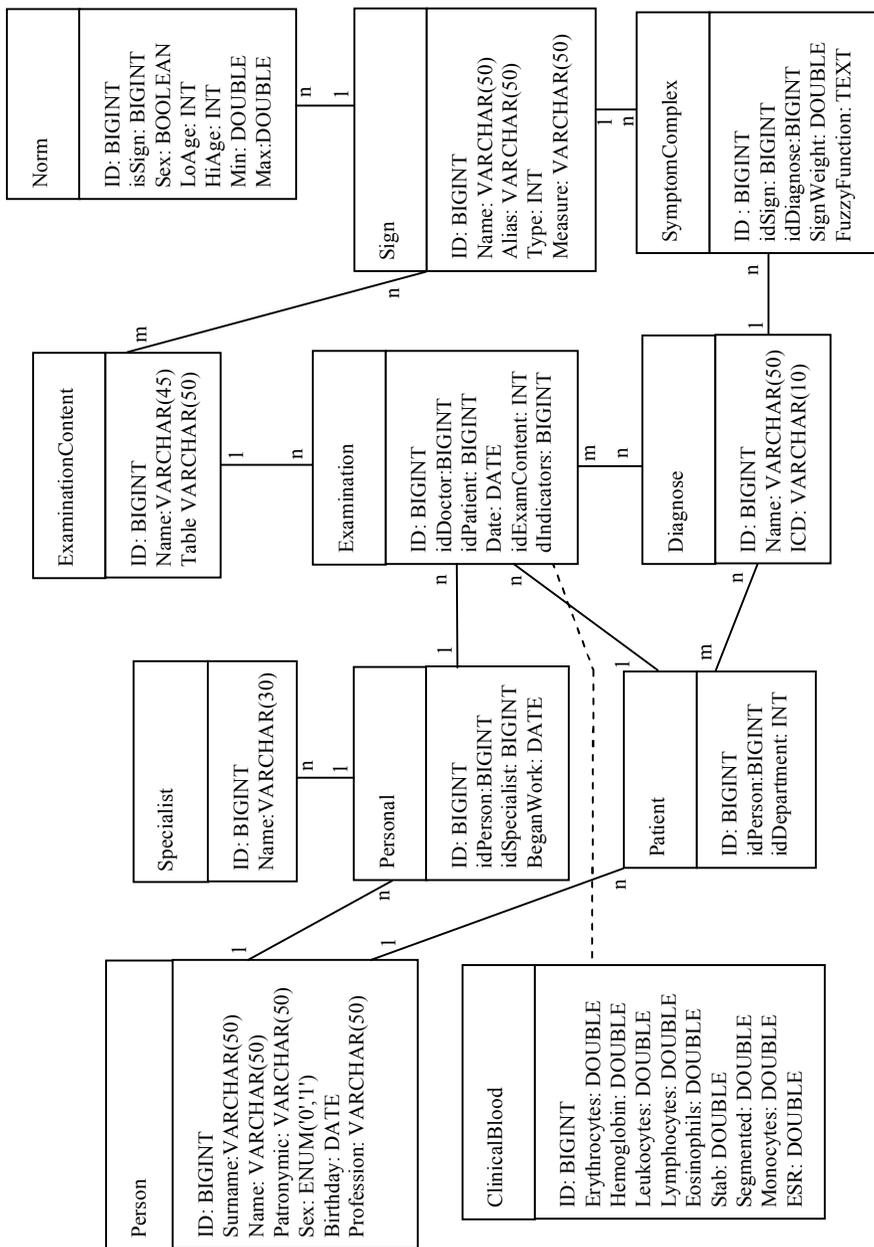


Рис. 2. Структура базы данных

Выводы. Проведен анализ существующих средств разработки и проектирования программного обеспечения. Предложен вариант реализации комбинированного решающего правила, основанного на объединении вероятностного подхода (метода Вальда) и описания симптомокомплексов (нечеткая логика). Приведена диаграмма классов, участвующих в реализации РП, а также структура базы данных, используемой разработанной диагностической подсистемой.

Список литературы: 1. Шилдт Г. Полный справочник по C++: Пер. с англ. / Г. Шилдт. – М.: ООО "И.Д. Вильямс", 2004. – 800 с. 2. Пауэрс Л. Microsoft Visual Studio 2008: Пер. с англ. / Л. Пауэрс, М. Снелл. – СПб.: БХВ-Петербург, 2009. – 1200 с. 3. Шилдт Г. Полный справочник по Java, 7-е издание.: Пер. с англ. / Г. Шилдт. – М.: ООО "И.Д. Вильямс", 2009. – 1040 с. 4. Рамбо Дж. UML 2.0. Объектно-ориентированное моделирование и разработка. 2-е издание. / Дж. Рамбо, М. Блаха. – СПб.: Питер, 2007. – 544 с. 5. Бек К. Экстремальное программирование: разработка через тестирование. / К. Бек. – СПб.: Питер, 2003. – 224 с. 6. Эккель Б. Философия Java, 4-е издание: Пер. с англ. / Б. Эккель. – СПб.: Питер, 2009. – 640 с. 7. Хемраджани А. Гибкая разработка приложений на Java с помощью Spring, Hibernate и Eclipse: Пер. с англ. / А. Хемраджани – М.: ООО "И.Д. Вильямс", 2008. – 352 с. 8. Бурцев М.В. Синтез комбинированного решающего правила в задаче медицинской диагностики / М.В. Бурцев, А.И. Поворозник // Вісник Національного технічного університету "Харківський політехнічний інститут". Збірник наукових праць. Тематичний випуск: Інформатика і моделювання. – Харків: НТУ "ХПІ". – 2009. – № 43. – С. 27 – 33. 9. Поворозник А.И. Формирование диагностических интервалов численных признаков при дифференциальной диагностике / А.И. Поворозник // Вісник Хмельницького національного університету. – Хмельницький: ХНУ, 2007. – Т. 1. – № 3. – С. 106–109. 10. Кузнецов М.В. Самоучитель MySQL 5 / М.В. Кузнецов, И.В. Симдянов. – СПб.: БХВ-Петербург, 2007. – 560 с. 11. 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.

Статья представлена д.т.н. проф. НТУ "ХПИ" Серковым А.А.

УДК 681.3

Програма реалізація комбінованого вирішального правила для задач медичної діагностики / Бурцев М.В., Поворозник А.І. // Вісник НТУ "ХПІ". Тематичний випуск: Інформатика і моделювання. – Харків: НТУ "ХПІ". – 2010. – № 21. – С. 12 – 17.

Розглянуто сучасні технології та засоби розробки програмного забезпечення, виконаний аналіз їх переваг та недоліків. Запропоновано підхід до програмної реалізації комбінованого вирішального правила, що включає імовірнісний метод Вальда та опис структури симптомокомплексів. Іл.: 2. Бібліогр.: 11 назв.

Ключові слова: технології та засоби розробки програмного забезпечення, комбіноване вирішальне правило, метод Вальда, симптомокомплекс.

UDC 681.3

Software implementation of combined solving rule for medical diagnostics tasks / Burtsev M.V., Povoroznik A.I. // Herald of the National Technical University "KhPI". Subject issue: Information Science and Modelling. – Kharkov: NTU "KhPI". – 2010. – № 21. – P. 12 – 17.

The modern technology and software development tools are reviewed, and analysis of their merits and demerits is made. The software implementation of the combined solving rule, including a probabilistic Wald's method and complex of symptoms structure description is offered.

Key words: technology and software development technology, combined solving rule, Wald's method, complex of symptoms.

Поступила в редакцію 10.04.2010