

УДК 004.9:616

Л.А. Думанская

С.М. Думанский

ПРИМЕНЕНИЕ ИНФОРМАЦИОННО-АНАЛИТИЧЕСКИХ СИСТЕМ НА ОСНОВЕ OLAP-ТЕХНОЛОГИЙ

Аннотация. В статье описан прототип информационно-аналитической системы (ИАС) на основе OLAP-модели для диагноза заболеваний. Приводятся структурные элементы ИАС и OLAP-куба. Разработан алгоритм распознавания болезненных состояний человека в зависимости от выявляемого набора симптомов и синдромов. Данный прототип ИАС можно использовать для изучения различных предметных областей деятельности человека, где возможно применение задач, относящихся к классу распознавания образов.

Ключевые слова: информационно-аналитическая система, OLAP-кубы, задачи распознавания образов.

Lidiya Dumanskaya

Sergey Dumansky

APPLICATION ANALYTICAL INFORMATION SYSTEMS BASED ON OLAP TECHNOLOGY

Annotation. This article describes a prototype of Information Analytical system (IAS) based on OLAP model for the diagnosis of diseases. It has been shown the structural elements of the IAS and OLAP model. The algorithm elaborated can recognize painful human condition, depending on the set of detectible symptoms and syndromes. This prototype IAS can be used of to explore different domains of human activity, which can be applied to problems related to pattern recognition class.

Keywords: analytical information systems, OLAP cube, pattern recognition.

С развитием вычислительной техники и программного обеспечения широкое применение приобрело исследование предметных областей человека при помощи информационно-аналитических систем (ИАС). В широком понимании ИАС представляет систему, состоящую из структурных элементов, каждый из которых выполняет последовательно следующие функции: сбор должным образом подготовленной информации, очистку, преобразование и передачу информации в хранилища данных, анализ данных при помощи соответствующе подобранных или разработанных алгоритмов, реализующих цель исследования, визуализацию полученных результатов и доставку их конечным пользователям.

Предлагаемый прототип ИАС основывается на алгоритме, который использует правильно разработанную структуру OLAP-куба, позволяющую реализовывать задачи, относящиеся к классу распознавания образов. Задачи такого типа встречаются в различных областях человеческой деятельности: оценка масштабов рудоносности месторождений в геологии, определение нефтеносности пластов при нефтеразведке, диагностике заболеваний в медицине, определение перспективности развития территорий в экономике и т.д. В общем виде задачу распознавания образов можно сформулировать так: имеется некоторая система P , которая в зависимости от значения признаков $\{x_1, x_2, x_3, \dots, x_n\}$, характеризующих ее может принимать различные состояния L . Экспертным путем формируется обучающая матрица $T_{m,n,s}$, состоящая из значений признаков и состояний системы. Строки матрицы $T_{m,n,s} = \|a_{ij}\|$, $1 \leq i \leq m$, $1 \leq j \leq n$ называются эталонами, причем каждый эталон описывает определенное состояние системы P . Матрицу $T_{m,n,s}$ можно поделить на s групп в зависимости от значений состояния системы. В l -ю группу входят те и только те эталоны, которые описывает l -е состояние. Требуется найти решающее правило, которое по определенному набору значений признаков $\{x_1, x_2, x_3, \dots, x_n\}$, описывающих какое-либо состояние системы и необязательно принадлежащих множеству эталонов, определит состояние системы [3]. Существует большое число алгоритмов, позволяющих решать подобного рода задачи, но все они сложны в математическом плане и требуют значи-

тельных временных затрат [2; 3]. Нами же предлагается один из наиболее простых и быстродействующих алгоритмов, время реализации, которого практически не зависит от числа строк обучающей матрицы.

Цель исследования – показать на примере разработанного прототипа ИАС потенциальные возможности использования OLAP-моделей при диагнозе заболеваний. Предметом исследования выбран хронический простатит (ХП), который является одним из наиболее распространенных заболеваний у мужчин независимо от возраста. ХП – это воспаление предстательной железы (ПЖ), обусловленное различными патогенными и этиологическими причинами (внешняя или внутренняя инфекция, ослабление иммунитета, гемодинамические нарушения в ПЖ и т.д.), протекающее в течение длительного периода времени. В зависимости от продолжительности воспалительного процесса, гистологических изменений ПЖ, клинических проявлений различают три стадии хронического простатита [1; 6]. Дифференциация и определение стадий ХП играет важную роль при назначении терапии пациенту, т.е. для каждой стадии заболевания требуется соответствующая схема лечения.

Для построения ИАС перечень наименований клинических признаков (показателей) ХП разбиваются на три категории. I категория. Параметры первичного осмотра: возраст, микционные нарушения выражены: сильно, слабо, не выявлены, преобладающий симптомокомплекс: сексуальный, болевой, параметры ПЖ: увеличена, уменьшена, норма. II категория. Показатели уродинамики: СОСМ (средняя объемная скорость мочеиспускания, мл/с), МОСМ (максимальная объемная скорость мочеиспускания, мл/с), Т-МОСМ (время достижения максимальной объемной скорости мочеиспускания, с), Т-50 (время выделения при мочеиспускании первых 50 мл мочи, с), ОСМ-50 (объемная скорость при выделении первых 50 мл мочи, мл/с). III категория. Параметры физикальных исследований: наличие пояснично-крестцового радикулита: да, нет, мочевого пузыря прощупывается: да, нет, выделения из уретры: да, нет, придатки яичек уплотнены: да, нет.

Каждый из показателей, входящих в определенную категорию может иметь по несколько значений. Следует учесть, что в ИАС количество категорий, показателей, входящих в категорию и число их значений возможно при необходимости увеличить. Также значения показателей данной категории в разрезе стадий может пересекаться, что затрудняет отнесение диагноза ХП к той или иной стадии.

По клиническим проявлениям стадии ХП характеризуются следующим образом (см. табл. 1) [6].

Таблица 1

Клинические проявления стадий ХП

1-я стадия	2-я стадия	3-я стадия
ПЖ увеличена в размерах, преобладает отечность; наблюдается боль в покое у 83 % обследованных пациентов; проявления вторичного пояснично-крестцового радикулита у 10,6 % обследованных пациентов; микционные нарушения (дизурия, нарушение процессов мочеиспускания) в большем числе случаев повышены; либидо повышено; ПЖ при пальпации: ощущается болезненность; хроническая почечная недостаточность у 4,3 % обследованных пациентов.	размер ПЖ близок к норме; боль в покое у 34 % обследованных пациентов; проявления вторичного пояснично-крестцового радикулита у 72,4 % обследованных пациентов; микционные нарушения (дизурия, нарушение процессов мочеиспускания) наблюдается у 28,1 % обследованных пациентов; либидо пониженное; ПЖ при пальпации: болезненность ощущается в меньшей степени, чем при 1-й стадии; хроническая почечная недостаточность у 20,5 % обследованных пациентов.	ПЖ уменьшена в размерах; боль в покое менее выражена; проявления вторичного пояснично-крестцового радикулита у 90,5 % обследованных пациентов; микционные нарушения (дизурия, нарушение процессов мочеиспускания) наблюдается у 100 % обследованных пациентов; либидо пониженное; ПЖ при пальпации: болезненность отсутствует; хроническая почечная недостаточность у 81,1 % обследованных пациентов.

Для реализации прототипа ИАС использовалось следующее программное обеспечение: Access 2013 и Excel 2013 как компоненты MS Office 2013; WindowsServer 2012 R2 в качестве операционной системы; SQL Server 2012 Standard Edition с входящими в него службами SSIS (SQL Server Integration Services) и SSAS (SQL Server Analysis Services). Для сбора данных и формирования Базы данных (БД) первичной информации используется СУБД Access 2013 [4]. Для каждой категории наименований клинических показателей составляются справочники нормативно справочной информации (НСИ) в виде соответствующих таблиц, которые используются для ввода информации в учебную матрицу. Форма ввода данных представлена на рисунке 1. Значение каждого из показателей эталонной записи заносится при помощи раскрывающегося списка из справочников НСИ. Заполнив учебную матрицу Tm,n,s необходимым количеством эталонных строк (чем их больше, тем точнее будет прогноз определения стадий ХП), при помощи предварительно подготовленного проекта SSIS переносят таблицы БД Access в формат SQLServer, при этом производится проверка данных на точность, их очистка и преобразование. Следует учесть, что по мере накопления эталонных записей в БД Access (таких баз может быть несколько в зависимости от разбросанности источников сбора), используя SSIS, их можно неоднократно переносить в хранилища данных [7].

Реализация алгоритма, при помощи которого осуществляется анализ данных, происходит в OLAP-кубе. Используемая информация загружается в хранилище данных, состоящее из таблиц двух типов – измерений и фактов. Структурно OLAP-куб формируется по схеме «звезда» или «снежинка». Схема «звезда» заключается в том, что таблицы фактов связаны с таблицами измерений, которые не имеют смежных с собой других таких же таблиц, в то время как схема «снежинка» может состоять из множества связанных между собой цепочек таблиц измерений [5; 7]. В нашем случае выбрана схема «снежинка». К таблицам фактов относятся: Fact_Стадий хронического простатита, Fact_Физикальные исследования, Fact_Исследования уродинамики, Fact_Параметры первичного осмотра. В качестве таблиц измерений выбраны: Справочник стадий хронического простатита, Справочник показателей уродинамики, Справочник физикальных показателей, Справочник параметров первичного осмотра. В таблицах фактов регистрируются события, которые требуют агрегации, а таблицы измерений используются для выполнения «срезов», по которым анализируются данные.

Если OLAP-куб условно представить в виде «шахматки», по строкам которой проставляются таблицы измерений, а по столбцам таблицы фактов, то на пересечении строк и столбцов можно отразить типы связей между таблицами. Существуют следующие типы отношений между таблицами фактов и измерений: нет связи, обычное, факт, ссылочная, многие ко многим, интеллектуальный анализ данных. Искусство проектирования OLAP-кубов как раз и заключается в правильном определении таблиц фактов и измерений и соответствующем распределении типов отношений между ними.

Последняя и наиболее важная функция ИАС – визуализация полученных результатов и доставка их конечным пользователям – осуществляется с помощью компоненты MS Office Excel, которая в данном случае играет роль клиентского приложения. На рисунке 2 показана выходная форма результата, полученного при определении стадии хронического простатита по выбранным клиническим показателям пациента.

Рис. 1. Форма ввода первичной информации

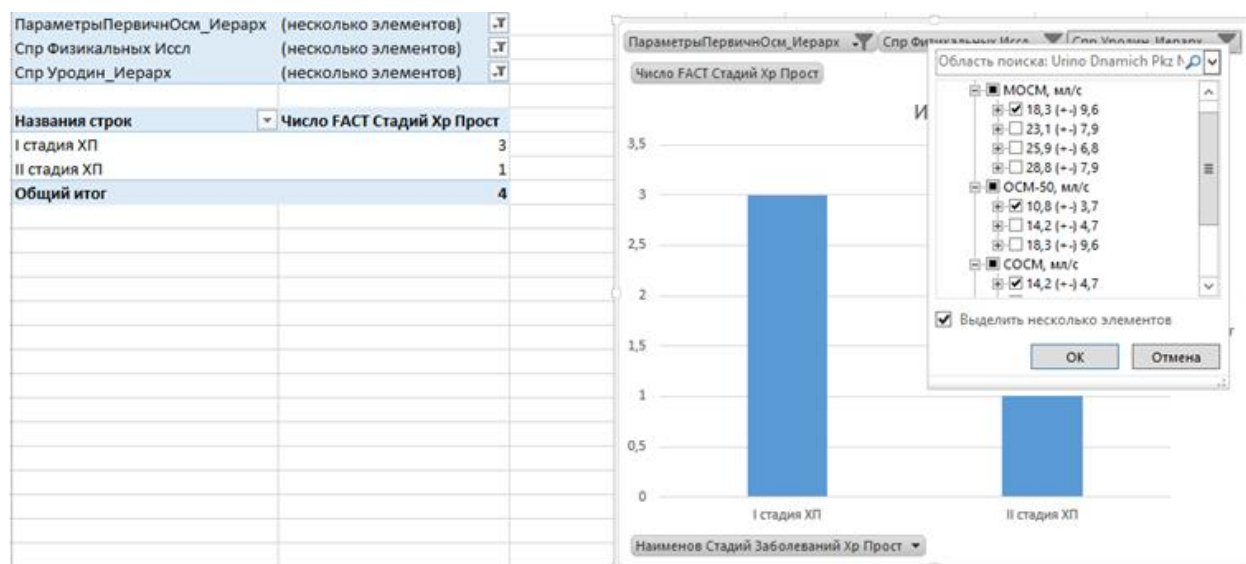


Рис. 2. Выходная форма определения стадии хронического простатита

При помощи строки соединения лист Excel связывают с OLAP-кубом, который удаленно может располагаться в локальной сети организации или в «облаке». При соединении листа Excel с OLAP-кубом на нем высвечиваются иерархические структуры таблиц фактов и измерений, а ниже –

прямоугольник, разбитый на 4 сектора: фильтры, колонны, строки, значения. Перетаскивая реквизиты таблиц в нужные сектора прямоугольника, в результате получают выходные данные по запрашиваемым запросам. Так, например, на рисунке 3 показано распределение значений параметров первичного осмотра пациентов по стадиям хронического простатита. Такие же формы можно получить и для других категорий клинических показателей.

Число ФАКТ	Парам	Первичн	Осмотр	Названия столбцов
Названия строк	I стадия ХП	II стадия ХП	III стадия ХП	Общий итог
Возраст	3	2	3	8
20 - 29	2			2
30 - 39		1		1
40 - 49		1		1
50 - 59			2	2
больше 60			1	1
До 20	1			1
Микционные нарушения выражены	3	2	3	8
Сильно		1	3	4
Умеренно	3	1		4
Параметры ПЖ	2	2	3	7
Норма		1	1	2
Увеличена	2	1		3
Уменьшена			2	2
Преобладающий симптомокомплекс	3	2	3	8
Болевой	3			3
Дизурический			3	3
Сексуальный		2		2
Общий итог	11	8	12	31

Рис. 3. Распределение параметров первичного осмотра пациентов по стадиям хронического простатита

Разработанный прототип ИАС представляет собой комплексную структуру исследования предметной области, начиная от ввода, обработки и визуализации данных. Основным преимуществом системы является то, что она дает конечным пользователям, не владеющим навыками программирования (врачам), возможность использовать такие высокоинтеллектуальные средства анализа информации, как OLAP-кубы. Описанную технологию можно адаптировать не только для определения стадий хронического простатита, но и для диагностики других заболеваний. Также лежащий в основе предлагаемой ИАС алгоритм при желании можно использовать и в других областях человеческой деятельности: геологии, нефтегазовой, экономике, т.е. там, где возможно использование задач, относящихся к классу распознавания образов. Следует особо отметить, что данный прототип ИАС использует клиент-серверный принцип построения, что позволяет обращаться к системе и получать выходные данные неограниченному числу конечных пользователей. Предусмотрена возможность ввода исходных данных из нескольких источников, расположенных удаленно друг от друга.

Библиографический список

1. Господарский, А. С. Простатит. Современный взгляд на лечение и профилактику / А. С. Господарский. – СПб. : ИГ «Весь», 2009. – 128 с.
2. Журавлёв, Ю. И. Распознавание. Классификация. Прогноз. Математические методы и их применение. Вып. 2 / Ю. И. Журавлёв. – М. : Наука, 1989. – 72 с.
3. Константинов, Р. М. О комбинаторно-логическом подходе к задачам прогноза рудоносности. Вып. 31 / Р. М. Константинов, З. Е. Королёва, В. Б. Кудрявцев. – М. : Наука, 1976. – С. 5–35.
4. Мак-Дональд, М. Access 2007 Недостатки руководства / М. Мак-Дональд; пер. с англ. – СПб. : БХВ-Петербург, 2007. – 784 с.

5. Петкович, Д. Microsoft SQL Server 2012. Руководство для начинающих / Д. Петкович. – СПб. : БХВ-Петербург, 2014. – 816 с.
6. Ткачук, В. Н. Хронический простатит / В. Н. Ткачук, А. Г. Горбачёв, Л. И. Агулянский. – Л. : Медицина, 1989. – 208 с.
7. Харинатх, С. Microsoft SQL Server Analysis Services 2008 и MDX для профессионалов / С. Харинатх [и др.]. – М. : И. Д. Вильямс, 2010. – 1072 с.