

Технология подготовки и анализа данных для построения шкалы оценки печеночной недостаточности¹

А. Н. Полуянов, С. В. Зыкин

ОФИМ СО РАН, ул. Певцова, д. 13, г. Омск, 644099, Россия.

Andrey.Poluyanov@gmail.com, zykin@ofim.oscsbras.ru

Аннотация. В работе рассматривается технология подготовки и анализа данных для построения шкалы оценки печеночной недостаточности пациентов. Рассмотрены принципы реализации программного обеспечения и некоторые результаты проведенного анализа данных.

Ключевые слова: межмодельные преобразования, гиперкубическое представление данных, шкалы оценки состояния пациентов.

1 Введение

Построение гиперкубического (многомерного) представления данных является основой для интеллектуального анализа данных (Data Mining) [1-6]. В работе предлагается следующая последовательность обработки и преобразования данных:

- Исходные данные должны быть представлены в реляционном нормализованном виде, и к ним обеспечивается доступ по технологии OLTP (online transaction processing).
- Пользовательское представление данных в виде гиперкубов, реализующее технологию OLAP, обеспечивается инструментарием, преобразующим исходные данные в гиперкуб.
- Гиперкубическое представление данных далее используется для выявления закономерностей, предсказания неизвестных значений и т.д.

В работе [7] формулируется следующий тезис: "Прежде чем выполнить любое ... из оптимизирующих преобразований, следует очень тщательно разобраться в том, какой именно потребуется анализ и какие для этого необходимы отчеты и служебная информация". В данной работе предполагается, что основой аналитической работы пользователя является необходимость формирования новых гиперкубов, а не многократное формирование реализации одного и того же гиперкуба. Следовательно, основное внимание необходимо акцентировать на сокращении времени формирования схемы нового гиперкуба.

В работе [8] для автоматизации формирования гиперкуба предлагается использовать последовательность преобразований:

$$RRD \Rightarrow TJ \Rightarrow ST,$$

где **RRD** – реляционное представление данных, **TJ** – таблица соединений, **ST** – гиперкуб "семантическая трансформация". В данном случае **RRD** представление исходной модели данных, **ST** – целевой. Представление **TJ** является промежуточным.

С использованием описанной последовательности преобразований реализованы алгоритмы и программное обеспечение для формирования гиперкубического представления данных.

На основе статистических данных по историям болезни пациентов с печеночной недостаточностью проведена аналитическая обработка данных для обоснования эффективности применения новых методов лечения и реабилитации пациентов с заболеваниями печени.

¹ Работа выполнена по проекту № 2.2 ОМН РАН за 2009 г.

На основании анализа данных построена шкала оценки печеночной недостаточности и реализовано программное обеспечение для хранения данных по пациентам и визуальной интерпретации состояния пациентов.

2 Подготовка и анализ данных

Исходные данные по пациентам содержатся в реляционной базе данных со следующим набором отношений.

Пациенты

ID пациента	ФИО	Год рождения	Дата поступления	№ группы	Адрес
-------------	-----	--------------	------------------	----------	-------

Группы

№ группы	Группа (описание)
1	ПФ
2	ПФ+ГС
3	контрольная

Виды показателей

ID анализа	ID типа	Наименование анализа	Ед. измерения	Комментарий
------------	---------	----------------------	---------------	-------------

Показатели

ID пациента	ID анализа	Дата взятия	Значение анализа	№ диапазона
-------------	------------	-------------	------------------	-------------

Типы значений показателей

ID типа	Тип
1	Числовой
2	Текстовый

Шкала

ID анализа	№ диапазона	Min значение	Max значение	Текстовое значение	Балл
------------	-------------	--------------	--------------	--------------------	------

Баллы

ID пациента	Дата взятия	Баллы (по шкале 0-100)
-------------	-------------	------------------------

В отношении «Пациенты» хранятся данные по пациентам. Отношение «Группы» предназначено для деления пациентов на группы, определяющие вид применяемого метода лечения. Отношение «Виды показателей» содержит описание анализов, применяемых для контроля состояния пациентов. Отношение «Показатели» предназначено для хранения значений показателей анализов по пациентам за определенную дату. Отношение «Шкала» содержит данные для оценки состояния пациентов в баллах (от 0 до 100). Отношение «Баллы» предназначено для хранения расчетных баллов оценки состояния пациентов на определенную дату.

Для анализа эффективности применяемых методов лечения и реабилитации было построено гиперкубическое представление данных следующего вида:

Выбранные отношения: Пациенты, Группы, Виды показателей, Показатели, Типы значений показателей;

Атрибуты множества X: Дата взятия, ФИО пациента;

Атрибуты множества Y: Наименование анализа, № группы;

Атрибуты множества Z: Значение анализа;

Ограничения: F = «Тип = «числовой»», C = {{ Виды показателей, Типы значений показателей }};

Далее данные были проанализированы по непараметрическому критерию Уилкоксона [9]. Получены следующие результаты:

При сравнении групп 1 и 2 с группой 3 (контрольная группа) по значениям параметров Асат, Алат, ЛИИ получено, что количество инверсий параметра в группе пациентов 1 и 2 по сравнению с группой 3 на третий день прохождения лечения лежит в интервале значений с уровнем значимости 5% (вероятность отвергнуть правильную гипотезу). Это позволяет принять гипотезу о том, что значения параметров принадлежат одной генеральной совокупности.

Сравнение тех же больных по тому же параметру на первый день лечения показывает, что эта гипотеза должна быть отвергнута.

Результаты обработки данных по критерию Уилкоксона приведены в таблицах 1 и 2.

Таблица 1. Результаты обработки данных по критерию Уилкоксона (часть 1).

Интервал: 995.3 - 1596.7

№ Дня	Азот	Мочевина	Креатинин	Белок	Биллирубин	ЛИИ
1	1553.0	1639.0	1644.0	753.0	1979.0	2105.0
2	1678.0	1884.0	1682.0	1342.0	1918.0	1162.0
3	1910.0	1916.0	1692.0	1099.0	1941.0	1100.0
4	1662.0	1753.0	1463.0	1379.0	1941.0	1160.0
5	1643.0	1762.0	1451.0	1203.0	1990.0	1420.0
8	1416.0	1625.0	1626.0	1062.0	1948.0	1479.0
10	1435.0	1731.0	1427.0	906.0	1810.0	1543.0

Таблица 2. Результаты обработки данных по критерию Уилкоксона (часть 2).

Интервал: 995.3 - 1596.7

№ Дня	ПТИ	Глюкоза	АСАТ	АЛАТ	Калий	Натрий
1	558.0	1680.0	1774.0	1708.0	1639.0	939.0
2	1208.0	1010.0	892.0	665.0	1224.0	1455.0
3	976.0	1151.0	1105.0	732.0	1059.0	1142.0
4	1311.0	834.0	1035.0	681.0	936.0	1987.0
5	1015.0	999.0	1244.0	858.0	1026.0	1804.0
8	1289.0	1266.0	1573.0	871.0	552.0	1412.0
10	1114.0	1234.0	1661.0	1238.0	1127.0	1420.0

На основании данного исследования можно сделать вывод об эффективности применяемого метода лечения для групп 1 и 2.

Разработанная шкала основана на применяемых в медицине шкалах для оценки состояния больного APACHE 2 и SAPS 2 [10]. По параметрам, отсутствующим в данных шкалах интервальные границы были уточнены при обработке статистических данных с использованием меры Кульбака. Границы интервалов выбирались исходя из максимизации информационной меры Кульбака значимости параметров при сравнении выборок по 1,2 и 3-ей группам.

3 Описание работы программного обеспечения

Для визуализации состояния пациентов и хранения данных по пациентам реализовано программное обеспечение «Шкала оценки печеночной недостаточности», реализованное с помощью пакета разработки программного обеспечения «Delphi 6.0».

Все данные, необходимые для работы программного обеспечения хранятся в базе данных.

Функционально работа с программой разделена на 4 экрана.

Экран "Выбор пациента" (рис. 1) предназначен для осуществления следующих операций:

- ввод в программу данных нового пациента;
- изменение данных о пациенте;
- удаление данных о пациенте;
- выбор пациента из списка пациентов, ранее введенных в программу.

Шкала оценки печеночной недостаточности (v 1.0)

Выбор пациента | Ввод данных | График | О программе

Новый пациент | Изменить данные | Удалить данные

Выбор пациента: Петров Петр Петрович, 29.09.1990 г.р.

Данные пациента:

Фамилия: Петров

Имя: Петр

Отчество: Петрович

Адрес: Омск

Дата рождения: 29.09.1990

Дата поступления: 05.10.2001

Сохранить изменения

Рис.1. Экран «Выбор пациента».

После того как пациент выбран, возможен переход на экраны "Ввод данных" (рис. 2) и "График".

На экране "Ввод данных" осуществляется ввод новых данных по пациенту на определенные дату и время по 27 показателям, используемым для оценки состояния пациента. При вводе данных осуществляется расчет баллов по шкале оценки печеночной недостаточности по выбранному пациенту в соответствии с алгоритмами, заложенными в логику работы программы, и баллами по применяемым для оценки параметрам, хранящимся в базе данных. Можно не вводить данные по всем показателям, но чем больше показателей задействовано, тем точнее будет оценка по шкале.

Также на экране "Ввод данных" осуществляется просмотр и редактирование архивных данных по пациенту.

Экран программы "Ввод данных" предназначен для осуществления следующих операций:

- ввод в программу новых данных по пациенту на определенные дату и время по показателям, используемым для оценки состояния пациента;
- расчет и сохранение баллов по шкале оценки печеночной недостаточности;
- просмотр архивных данных пациента;
- редактирование архивных данных по пациенту на определенные дату и время;
- удаление архивных данных по пациенту за определенные дату и время.

Шкала оценки печеночной недостаточности (v 1.0)

Выбор пациента | Ввод данных | График | О программе

Петров Петр Петрович, 29.09.1990 г.р., Дата поступления: 05.10.2001 г.

Ввод новых данных | просмотр архива | Ввод данных за: 05.09.2008 | 10:49 | Сохранить | Сбросить форму

Возраст ☐ нет данных ☐ < 35 ☐ 35 - 59 ☐ 60 - 65 ☒ 66 - 74 ☐ 75 - 79 ☐ >= 80	ЧСС (в минуту) ☐ нет данных ☐ < 40 ☐ 40 - 69 ☐ 70 - 119 ☐ 120 - 159 ☐ >= 160	Систолическое АД ☐ нет данных ☐ < 70 ☒ 70 - 99 ☐ 100 - 199 ☐ >= 200	Температура тела (С) ☐ нет данных ☐ < 39 ☒ >= 39	Диурез (л за 24 часа) ☐ нет данных ☐ < 0,500 ☐ 0,500 - 0,999 ☐ >= 1,000	Белок (г/л) ☐ нет данных ☐ < 50 ☒ 50 - 59 ☐ 60 - 79 ☐ >= 80
Альбумин (г/л) ☐ нет данных ☒ < 30 ☐ 30 - 34 ☐ >= 35	Мочевина (ммоль/л) ☐ нет данных ☐ < 2,8 ☒ 2,8 - 8,3 ☐ >= 8,4	Азот (ммоль/л) ☐ нет данных ☒ < 14,0 ☐ 14,0 - 21,9 ☐ >= 22,0	Креатинин (мкмоль/л) ☐ нет данных ☐ < 26 ☒ 26 - 129 ☐ >= 130	Лейкоциты (1000/л) ☐ нет данных ☒ < 4,0 ☐ 4,0 - 9,9 ☐ >= 10,0	ЛИИ ☐ нет данных ☒ < 1,0 ☐ 1,0 - 3,9 ☐ >= 4,0
Калий (мэкв/л) ☐ нет данных ☒ < 3,0 ☐ 3,0 - 4,9 ☐ >= 5,0	Глюкоза ☐ нет данных ☐ < 3,5 ☒ 3,5 - 6,4 ☐ 6,5 - 10,9 ☐ >= 11,0	АЛАТ ☐ нет данных ☒ < 0,70 ☐ 0,70 - 2,19 ☐ >= 2,20	АСАТ ☐ нет данных ☒ < 0,70 ☐ 0,70 - 1,99 ☐ >= 2,00	Желтуш. прокр. кожи ☐ нет данных ☐ нет ☒ слабое ☐ умеренное ☐ интенсивное	Натрий (мэкв/л) ☐ нет данных ☒ < 135 ☐ 135 - 144 ☐ >= 145
ВЕ (мэкв/л) ☐ нет данных ☒ < 15 ☐ 15 - 19 ☐ >= 20	Билирубин (мг/дл) ☐ нет данных ☒ < 40,0 ☐ 40,0 - 81,9 ☐ >= 82,0	Шкала комы Глазго ☐ нет данных ☐ < 6 ☒ 6 - 8 ☐ 9 - 10 ☐ 11 - 13 ☐ 14 - 15	ПТИ ☐ нет данных ☐ < 60,0 ☒ 60,0 - 72,4 ☐ 72,5 - 84,9 ☐ 85,0 - 109,9 ☐ >= 110,0	Хронические болезни ☐ нет данных ☐ нет болезней ☐ метастазирующая карцинома ☐ онкогенатология ☐ неонкогенная обтурация	
Вид поступления ☐ нет данных ☒ плановая операция ☐ срочная операция ☐ экстренная операция	АРиск ☐ нет данных ☒ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5	Вид операции ☐ нет данных ☐ резекция печени ☐ гемигепатэктомия левосторонняя ☒ гемигепатэктомия правосторонняя ☐ гемигепатэктомия расширенная	Потери желчи ☐ нет данных ☐ < 100 ☐ 100 - 299 ☒ 300 - 499 ☐ 500 - 799 ☐ >= 800		

Данные успешно сохранены

100
47.1 >
0

Рис.2. Экран «Ввод данных».

На экране "График" (рис. 3) осуществляется построение графика на основании всех введенных по пациенту данных. На данном экране можно в динамике отследить изменение состояния пациента.

Экран программы "О программе" содержит краткое описание работы программного обеспечения и ссылку на подробную инструкцию по работе с программой.

Для удобства работы с программой, баллы, характеризующие состояние пациента приводятся к шкале 0 – 100 баллов. Для оценки состояния пациента графическое поле на экране "Ввод данных" и на экране "График" разделено на 3 части:

- зеленая область - от 0 до 30 баллов - удовлетворительное состояние пациента;
- желтая область - от 30 до 65 баллов - состояние средней тяжести;
- красная область - от 65 до 100 баллов - тяжелое состояние пациента.

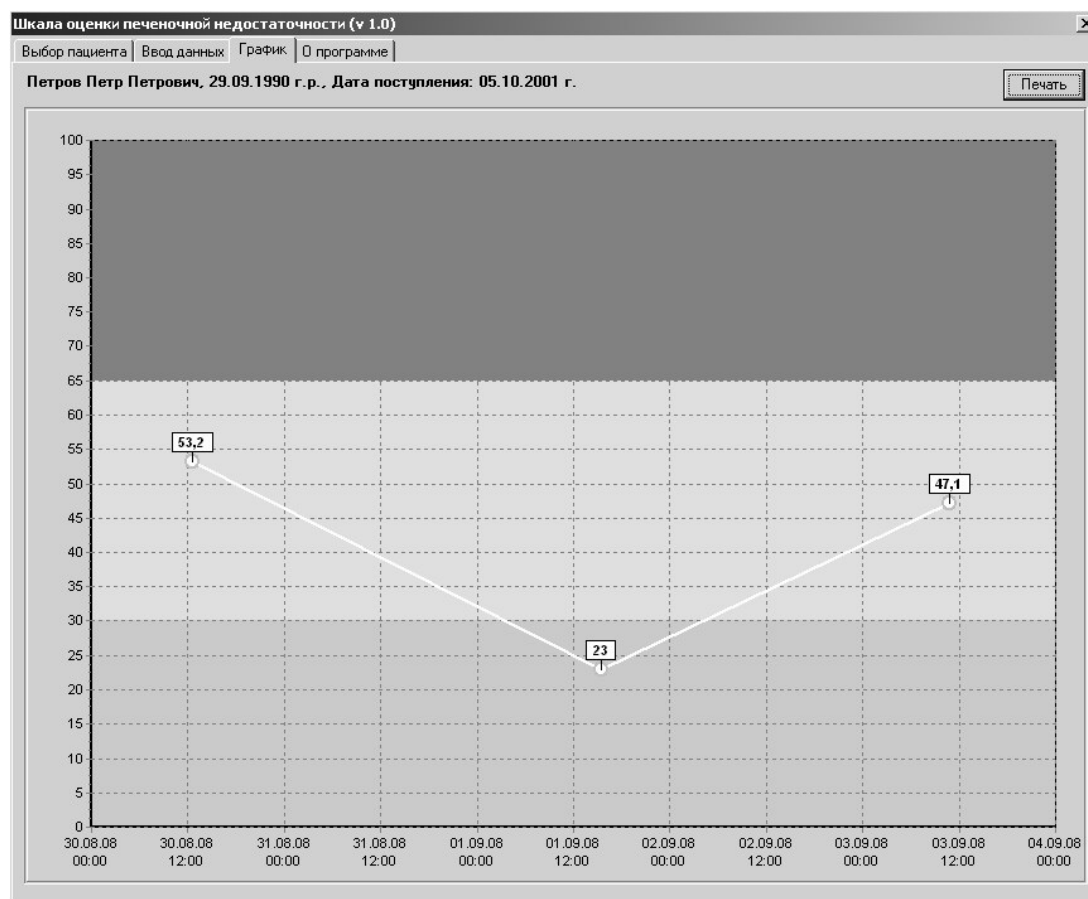


Рис.3. Экран «График».

4 Заключение

Разработанное программное обеспечение «Шкала оценки печеночной недостаточности» позволяет в динамике оценивать состояние пациентов с заболеваниями печени и визуализировать данное состояние, получить заключение об эффективности применяемого метода лечения.

Данное программное обеспечение предполагает возможность дальнейшего развития и модификации шкалы оценочных параметров для расширения области применения программы и приведения шкалы в соответствие потребностям медицинского учреждения, в котором будет применяться программное обеспечение.

Разработанный программный продукт успешно применяется в МСЧ 10 г. Омска, на что получен акт о внедрении.

Литература

- [1] Harinarayan V., Rajaraman A., Ullman J. D. Implementing Data Cubes Efficiently// SIGMOD Conference. - Montreal, CA. -1996. - P. 205-216.
- [2] Gray J., Chaudhuri S., Bosworth A., etc. Data Cube: A Relational Aggregation Operator Generalizing Group-By, Cross-Tab, and Sub-Totals// Data Mining and Knowledge Discovery. - 1997. - № 1. - P. 29-53.
- [3] Педерсен Т.Б., Йенсен К.С. Технология многомерных баз данных// Открытые системы. - 2002. - № 1. - С. 45-50.

- [4] Parsaye K. Surveying Decision Support: New Realms of Analysis// Database Programming and Design. - 1996. - № 4. - P. 26-33.
- [5] Parsaye K. OLAP and Data Mining: Bridging the Gap// Database Programming and Design. - 1997. - № 2. - P. 30-37.
- [6] Перевозчикова О.Л., Тульчинский В.Г. Структурный анализ и синтез маршрутных схем для задач Macro Mining // Кибернетика и системный анализ. - 2003. - № 1. - С. 82 - 95.
- [7] Armstrong R. Seven Steps to Optimizing Data Warehouse Performance// Computer, V. 34, № 12, 2001. - P. 76-79.
- [8] Зыкин С.В. Формирование гиперкубического представления реляционной базы данных // Программирование. - 2006. - № 6. - С. 348 - 354.
- [9] Бронштейн И.Н., Семендяев К.А. Справочник по математике для инженеров и учащихся вузов. - М.: Наука, 1981.- 720 с.
- [10] Александрович Ю.С., Гордеев В.И. Оценочные и прогностические шкалы в медицине критических состояний. Справочник. – Санкт-Петербург: Сотис, 2007. – 137 с.