

ИНФОРМАЦИОННАЯ ТЕХНОЛОГИЯ АНАЛИЗА РЕЗУЛЬТАТОВ ЛЕЧЕНИЯ ПАЦИЕНТОВ С ЭНДОКРИНОПАТИЯМИ

И.А. Лызин, О.В. Марухина
Томский политехнический университет
i-lyzin@mail.ru

Введение

Ожирение у детей и подростков является актуальной и весьма серьезной проблемой. Распространенность, которой возрастает: в развитых странах мира 25% подростков имеют избыточную массу тела, а 15% страдают ожирением. Избыточный вес и ожирение – это результат формирования аномальных или чрезмерных жировых отложений, которые могут наносить вред здоровью [1].

Целью данной работы была разработка информационной технологии, позволяющей комплексно анализировать результаты лечения детей с применением методов визуализации. В качестве инструмента решения поставленной задачи был выбран язык R, так как R является популярным инструментом для визуализации и анализа данных.

Материалы и методы

Предметом исследования является массив данных предоставленный НИИ курортологии и физиотерапии г. Томска, в котором содержатся группы клинико-лабораторных показателей пациентов. Состояние каждого пациента оценивалось при поступлении на лечение и по окончании курса реабилитации на основании набора лабораторных данных и химического анализа крови. Пациенты – дети и подростки в возрасте от 6 до 17 лет, с избыточной массой тела и ожирением. В состав базы включены данные 276 пациентов, поделенных на 5 групп. Деление пациентов на группы осуществлялось в зависимости от способа лечения

Анализ выбросов – это важный этап в подготовке исходных данных к восстановлению позволяющий избежать отклонения восстановленных данных от разумных значений. Под "выбросом" понимается наблюдение, которое "слишком" велико или "слишком" мало по сравнению с большинством других имеющихся наблюдений. После исключения выбросов был получен пригодный массив данных для дальнейшего анализа.

Поскольку пропуски в данных являются повсеместным явлением, прежде чем перейти к восстановлению пропущенных данных и построению визуальных образов необходимо идентифицировать пропуски. Идентификация недостающих данных является единственным однозначным шагом. Решение о способе устранения пропущенных значений также будет зависеть от оценки того, какие процедуры приведут к самым надежным и точным результатам. Идентификация пропущенных значений обычно легко делается с использованием функций `is.na()` и `complete.cases()`. Анализ пропусков в исходных данных показал, что процент пропусков не превышает 15 в некоторых случаях 20

процентов. Следовательно, можно переходить к процессу восстановления.

Метод множественного восстановления пропущенных данных – это способ заполнения пропусков при помощи повторного моделирования. Множественное восстановление часто применяется для работы с пропущенными данными в сложных ситуациях.

Процесс восстановления пропусков включает в себя несколько этапов:

1. Функция `mice()` обрабатывает исходную таблицу данных с пропущенными значениями, а возвращает объект, содержащий несколько полных наборов данных (пять по умолчанию). Каждый такой полный набор данных получается при восстановлении пропущенных данных исходной таблицы. Все полные наборы данных отличаются друг от друга значениями пропусков, полученных в ходе процедуры Гиббса. Процедура Гиббса предполагает, что значения каждой переменной, содержащей пропуски, предсказываются по значениям остальных переменных. Этот процесс повторяется, пока значения для пропущенных данных не сойдутся: `mice(data, m)`;

2. Затем при помощи функции `with()` применяется статистическая модель (например, линейная): `with(imp, analysis)`;

3. Функция `pool()` производит построение полного набора данных: `pool(fit)` [2].

После восстановления пропусков в исходных наборах был получен полностью укомплектованный набор данных, который можно использовать для моделирования визуальных образов используя лица Чернова. Лица Чернова – это схема визуального представления мультивариативных данных в виде человеческого лица. Каждая часть лица: глаза, нос, рот – представляет собой значение определенной переменной, назначенной для этой части (всего 18) [3]. Основная идея использования лиц заключается в том, что люди легко распознают лица и без труда замечают небольшие изменения, для человека очень естественно смотреть на лица, легко делать сравнения и выявлять отклонения.

Так как количество показателей намного больше чем можно поместить в лица, необходимо подготовить восстановленные данные к построению визуальных образов. Подготовка данных предполагает определения наиболее информативных признаков из всего восстановленного набора данных. Информативность признаков была определена с использованием метода Кульбака. Данный метод предлагает в качестве оценки информативности меру расхождения между двумя классами ко-

торая называется дивергенцией [4]. В данном случае в качестве классов выступают показатели до и после лечения.

Для построения визуальных образов использовались следующие информативные признаки: ТМТ, NO, Гаркави, Каталаза, Лептин, ФНО, АПФ, ИЛ-4, окЛПНП, КК.

Анализируя результаты можно заметить по каким группам лечение было наиболее эффективным. Можно выделить группу 1, по этой группе заметны значительные изменения. Визуальные образы группы 2 и группы 5 практически не изменились, следовательно, можно сделать вывод, что лечение не оказало положительного результата.

| № группы | До лечения | После лечения |
|----------|---|---|
| Группа 1 |  |  |
| Группа 2 |  |  |
| Группа 5 |  |  |

Рис. 1. Визуальные образы

Заключение

В результате выполнения работы, была разработана информационная технология анализа результатов лечения детей с эндокринопатиями.

С использованием разработанной информационной технологии было осуществлено восстановление пропусков в исходных наборах данных и проанализированы результаты лечения пациентов с использованием визуальных образов.

В рамках выполнения работы была построена структурная схема алгоритма информационной технологии и осуществлена её практическая реализация. Алгоритм информационной технологии представлен на рисунке 2.

Разработанная информационная технология имеет широкую сферу применения и может использоваться для комплексного анализа данных в различных медицинских учреждениях.



Рис. 2. Алгоритм ИТ анализа результатов лечения детей

Исследование выполнено при частичной финансовой поддержке РФФИ (проект 18-07-00543-а)

Список использованных источников

1. Всемирная организация здравоохранения / Ожирение и избыточный вес [Электронный ресурс] – URL: <http://www.who.int/mediacentre/factsheets/fs/>.
2. Кабаков Р.И. R в действии. Анализ и визуализация данных в R / Метод множественного восстановления данных. – Москва: Из-во «ДМК Пресс», 2014. – 489 с.
3. Берестнева О.Г. Методы структурного анализа и визуализации экспериментальных данных в социальных и медицинских исследованиях / О.Г. Берестнева, И.А. Осадчая, А.Л. Бурцева. – Томск: Из-во ТПУ, 2014. – 17 с.
4. Гублер Е. В. Вычислительные методы анализа и распознавания патологических процессов. – Москва: Из-во «МЕДИЦИНА», 1978. – 198 с